OR Spektrum (1993) 15:9–15

OR Spektrum

Stationary ε-optimal strategies in stochastic games

Frank Thuijsman and Koos Vrieze

Department of Mathematics, University of Limburg, P.O. Box 616, 6200 MD Maastricht, The Netherlands

Received May 27, 1991 / Accepted in revised form October 6, 1992

Summary. We deal with stochastic games with finite state and action spaces for which we examine players' possibilities for playing limiting average (ε-)optimal by means of stationary strategies ($\varepsilon > 0$). It is well-known that stationary limiting average ε-optimal strategies need not exist for all initial states, hence we focus on those particular initial states for which the limiting average value is either maximal or minimal over the set of states.

Zusammenfassung. Für stochastische Zwei-Personen-Null-Summen-Spiele mit endlichen Zustands- und Aktions-Räumen untersuchen wir die Möglichkeiten eines Spielers, um mit Hilfe von stationären Strategien ε-optimal zu spielen bezüglich des Durchschnittsgewinn-Kriteriums ($\varepsilon > 0$). Es ist bekannt, daß solches im allgemeinen nicht für alle Zustände möglich ist. Deshalb konzentrieren wir unsere Untersuchung auf die Zustände für welche der Durchschnittsgewinn-Wert entweder maximal oder minimal ist.

Key words: Stochastic game, limiting average rewards, stationary optimal strategies

Schlüsselwörter: Stochastisches Spiel, Durchschnittsgewinn Auszahlungen, stationäre optimale Strategien

1. Introduction

Two-person stochastic games are non-cooperative games of the following type: Let A_1, A_2, ..., A_z be a finite collection of finite matrices, where each entry (i,j) of matrix A_s consists of a number $r(s,i,j) \in \mathbb{R}$ and a probability vector $p(s,i,j) = (p(1|s,i,j),\ p(2|s,i,j),\ \ldots,\ p(z|s,i,j))$. Such a game can start in any "state" of $S = \{1, 2, \ldots, z\}$ and is to be played as follows: if at stage $n \in \{1, 2, 3 \ldots\}$ play is in state $s \in S$, then independently from each other, player 1 has to choose a row i of A_s and player 2 has to choose a column j of A_s; having done so player 1 receives $r(s,i,j)$ from player 2 and with probability $p(t|s,i,j)$ play moves to state $t \in S$, where actions have to be chosen at stage $n+1$. Proceeding this way there is at each stage n a payoff R_n by player 2 to player 1. In literature two major ways of evaluating the sequence R_1, $R_2, \ldots$ as a single "reward" to player 1 are the β-discounted reward, $\beta \in (0,1)$) and the limiting average reward. The assumption normally is that both players use the same evaluation where player 1 wishes to maximize his reward, while player 2 wishes to minimize the same. In order to achieve those goals the players make use of strategies, plans to play the game. They may use all information they have. At each stage of play each player knows all the entries of the matrices as well as the history of play, i.e. the sequence of past states visited and the pure actions chosen in those states. The players are allowed to randomize over their (pure) actions, so in general a strategy will assign a mixed action to each triple (state, stage, history). Stationary strategies are strategies where a player neither uses information about stage nor about history. So a stationary strategy for player 1 can be seen as $x = (x_1, x_2, \ldots, x_z)$, where x_s is a mixed action for player 1 to be used whenever play is in state s. Stationary strategies for player 2 will be denoted y. For general strategies we write π for player 1, σ for player 2. For $\beta \in (0,1)$ the β-discounted reward to player 1 for initial stage $s \in S$ and strategies (π, σ) is given by $\gamma^\beta(s,\pi,\sigma) = E_{s\pi\sigma}[(1-\beta)\sum_{n=1}^{\infty}\beta^{n-1}R_n]$, while the limiting average reward is given by $\gamma(s,\pi,\sigma) = E_{s\pi\sigma}[\liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}R_n]$. Here $E_{s\pi\sigma}$ denotes expectation with respect to s, π and σ.

From the seminal paper on stochastic games by Shapley [10] it directly follows, using Blackwell [2], that for all $\beta \in (0,1)$ there exist $v^\beta \in \mathbb{R}^z$ and stationary strategies x^β for player 1 and y^β for player 2 such that for all π and σ (using vector notation):

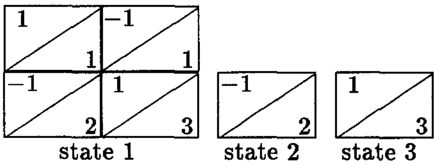
$$\gamma^\beta(x^\beta, \sigma) \geq v^\beta \geq \gamma^\beta(\pi, y^\beta).$$

The vector v^β is called the β-discounted value of the game and strategy $x^\beta(y^\beta)$ is a (stationary) β-discounted optimal strategy for player 1 (2). Also, a stationary strategy $\bar{x}(\bar{y})$ is

β-discounted optimal if and only if for each $s \in S$ the mixed action $\bar{x}_s(\bar{y}_s)$ is optimal for player 1 (2) in the matrix game $[(1-\beta)r(s,i,j)+\beta\sum_t p(t|s,i,j)v_t^\beta]_{i,j}$, while this matrix game has value v_s^β.

Mertens and Neyman [7] have shown that also the limiting average value v always exists and that $v=\lim_{\beta\uparrow 1} v^\beta$. Nevertheless, neither limiting average optimal strategies nor history independent limiting average ε-optimal strategies need exist. The classic example for this phenomenon is "the big match" of Blackwell and Ferguson [3], which can be given as:

Example 1.1

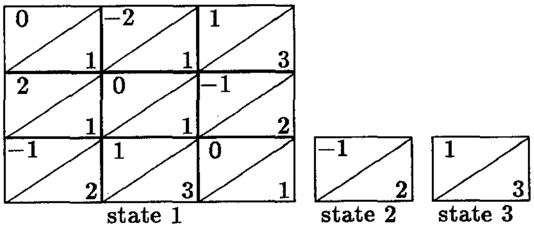


The numbers in the left-upper corners are the payoffs to player 1; the numbers in the right-lower corners are the states to which play will move (with probability 1) once the entry is chosen.

For this game, for initial state 1, the unique stationary β-discounted optimal strategy for player 1 is $(1/(2-\beta), (1-\beta)/(2-\beta))$ and $v_1^\beta=0$. The limiting average value of this game for state 1 is also equal to 0 but player 1 has only history dependent limiting average ε-optimal strategies $(\varepsilon>0)$. Player 2 on the other hand has a stationary limiting average optimal strategy: he can play (1/2, 1/2) in state 1 at all stages.

Following example shows that there is not necessarily for each initial state a stationary limiting average optimal strategy for at least one of the players; here both players need history dependent strategies for limiting average ε-optimal play for initial state 1:

Example 1.2



Here the unique stationary β-discounted optimal strategy for player 1, as for player 2, is $(1/(4-2\beta), 1/(4-2\beta), (2-2\beta)/(4-2\beta))$ and $v^\beta=0$. Each player faces a "big match like" situation for initial state 1.

In this paper we discuss player 1's possibilities for limiting average (ε-)optimal play by means of a stationary strategy, but obviously similar results can be derived for player 2. We focus our attention on initial states in $S^{\max}:=\{s\in S: v_s=\max_t v_t\}$ and initial states in $S^{\min}:=\{s\in S: v_s=\min_t v_t\}$. If player 1 has a stationary limiting average (ε-)optimal strategy for initial state s, then we call s an *(ε-)easy initial state for player* 1 $(\varepsilon>0)$. In Sect. 2 we show that there are always easy initial states for player 1 in $S^{\max}$, while there need not be easy initial states for player 1 in $S^{\min}$. Some of these results can also be found in Thuijsman and Vrieze [11], where the emphasis is mainly on general-sum stochastic games. In Sect. 3 we show that all states in $S^{\min}$ are ε-easy initial states for player 1, while there may be states in $S^{\max}$ that are not ε-easy for player 1. However a sufficient condition is presented for which all initial states in $S^{\max}$ are ε-easy for player 1. In Sect. 4 we show that the techniques we use allow for new and simple proofs for the existence of limiting average value and stationary limiting average (ε-)optimal strategies in several special classes of stochastic games: unichain stochastic games; stochastic games for which $\lim_{\beta\uparrow 1} v^\beta$ does not depend on the initial state; stochastic games with state independent transitions.

In several of our proofs we will use the argument that: if player 1 plays a stationary strategy, then player 2 has a pure stationary strategy as a (β-discounted or limiting average) best reply (cf. Hordijk et al. [6]). We close this section by introducing some notations for a pair of stationary strategies (x,y).

The payoff vector $r(x,y)$ is the vector $(r(1,x,y), r(2,x,y), \ldots, r(z,x,y))$ with $r(s,x,y):=\sum_i\sum_j x_s(i)\, r(s,i,j)y_s(j)$.

The transition matrix $P(x,y)$ is the $z\times z$-matrix of which entry (s,t) is given by $p(t|s,x,y):=\sum_i\sum_j x_s(i)\, p(t|s,i,j)y_s(j)$.

The matrix $Q(x,y)$ is defined by $Q(x,y)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}P^n(x,y)$, where $P^n(x,y)$ denotes the n-fold product of $P(x,y)$ with itself. Using these notations the β-discounted reward and the limiting average reward can be given respectively as:

$$\gamma^\beta(x,y)=(1-\beta)(I-\beta P(x,y))^{-1}r(x,y) \text{ and}$$

$$\gamma(x,y)=Q(x,y)r(x,y).$$

In this paper we also use the following convention. Let $(\beta_n)_{n\in\mathbb{N}}$ be a sequence of discount factors converging to 1 and let $(x^{\beta_n})_{n\in\mathbb{N}}$ be a sequence of stationary β_n-discounted optimal strategies converging to x^1. We write $\lim_{\beta\uparrow 1}x^\beta=x^1$ instead of $\lim_{n\to\infty}x^{\beta_n}=x^1$. We will do similarly for other limits, assuming each time that we are dealing with a converging sequence. By taking subsequences all limits can be assumed to exist. If we write "$x^\beta\ldots$ for all β close to 1", then this should be interpreted as "$x^{\beta_n}\ldots$ for some sequence $(\beta_n)_{n\in\mathbb{N}}$ converging to 1 and n sufficiently large".

Since the number of pure stationary strategies is finite and since for each n there is a pure stationary limiting average best reply against x^{β_n}, we conclude that there is a pure stationary limiting average best reply against x^{β_n} for all n sufficiently large.

2. Easy initial states

Theorem 2.1. *For player 1 there are easy initial states in* $S^{\max}$.

This theorem was first proved by Tijs and Vrieze [12], based on a result of Bewley and Kohlberg [1] which says that for all β sufficiently close to 1, the β-discounted value v^β as well as a pair of stationary β-discounted optimal strategies can be expanded in Puiseux series in fractional powers of $(1-\beta)$. Here we present an elementary proof.

Proof of Theorem 2.1. Let x^β be a stationary β-discounted optimal strategy, for $\beta<1$, and let $x^1=\lim_{\beta\uparrow 1}x^\beta$. Also, let $\bar{y}$ be a stationary limiting average best reply against x^1.

Define $Z^\beta=(1-\beta)(I-\beta P(x^\beta,\bar{y}))^{-1}$ and let $Z=\lim_{\beta\uparrow 1}Z^\beta$.Then for all β the matrix Z^β is nonnegative and has rowsums all equal to 1; hence such is Z. Then $\gamma^\beta(x^\beta,\bar{y})=Z^\beta r(x^\beta,\bar{y})\geq v^\beta$ for all β, and $\lim_{\beta\uparrow 1}\gamma^\beta(x^\beta,\bar{y})=Zr(x^1,\bar{y})\geq\lim_{\beta\uparrow 1}v^\beta=v$.

Now observe that $ZP(x^1,\bar{y})=Z$, because for all β we have $Z^\beta(I-\beta P(x^\beta,\bar{y}))=(1-\beta)I$.

Hence for each $s\in S$, row Z_s of Z is a stationary distribution of the Markov chain related to $P(x^1,\bar{y})$. Let $S^1,S^2,\ldots,S^H$ be the ergodic sets for $P(x^1,\bar{y})$ and let $q^1,q^2,\ldots,q^H$ be the unique stationary distributions on those ergodic sets. Then for each s we have that Z_s is a convex combination of $q^1,q^2,\ldots,q^H$ and therefore there are $\mu_s^1,\mu_s^2,\ldots,\mu_s^H\geq 0$ with $\sum_{h=1}^H\mu_s^h=1$ such that $Z_s=\sum_{h=1}^H\mu_s^h q^h$.

Altogether we have that for each s:

$$v_s=\lim_{\beta\uparrow 1}v_s^\beta\leq\lim_{\beta\uparrow 1}\gamma^\beta(s,x^\beta,\bar{y})$$

$$=\lim_{\beta\uparrow 1}Z_s^\beta r(x^\beta,\bar{y})=Z_s r(x^1,\bar{y})$$

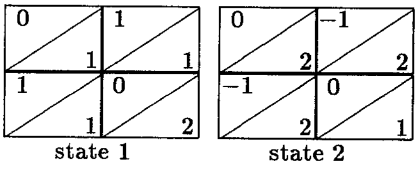
$$=\sum_{h=1}^H\mu_s^h q^h r(x^1,\bar{y})=\sum_{h=1}^H\mu_s^h\gamma^h(x^1,\bar{y}),$$

where $\gamma^h(x^1,\bar{y})$ denotes the limiting average reward ($\in\mathbb{R}$) for each of the states in S^h.

The above equation implies the existence of an ergodic set S^{h*} with $\gamma^{h*}(x^1,\bar{y})\geq\max_t v_t$. Because $\bar{y}$ is defined as a limiting average best reply against x^1, this means that for each $s\in S^{h*}$ and for any strategy τ : $\gamma(s,x^1,\tau)\geq\gamma(s,x^1,\bar{y})\geq v_s=\max_t v_t$. □

In fact Tijs and Vrieze [12] show that all states s with $v_s^\beta=\max_t v_t^\beta$ for all β close to 1, are easy initial states for player 1. The following example shows that this condition can not be weakened to $v_s=\max_t v_t$.

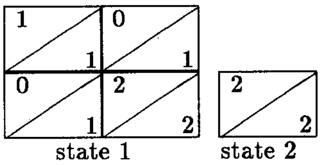
Example 2.2



In this example $v=(0,0)$ while $v_1^\beta=(-1+\beta+\sqrt{1-\beta^2})/2\beta>0$ and $v_2^\beta=-v_1^\beta<0$ for all $\beta\in(0,1)$. State 1 is easy for player 1 while state 2 is not. The unique stationary β-discounted optimal strategy for player 1, as for player 2, is to play $((1+\beta-\sqrt{1-\beta^2})/2\beta,\ (-1+\beta+\sqrt{1-\beta^2})/2\beta)$ in both states.

In the following example player 1 has no easy initial states in $S^{\min}$.

Example 2.3



Here we have $v=(1,2)$ but state 1 is not easy for player 1. The unique stationary β-discounted optimal strategy for player 1, as for player 2, is to play $((3-\sqrt{9-8\beta})/2\beta,\ (-3+2\beta+\sqrt{9-8\beta})/2\beta)$ in state 1.

In both examples of this section player 1 has a stationary limiting average ε-optimal strategy for both initial states in each of the games. Hence, in each of these games all states are ε-easy for player 1.

3. ε-Easy initial states

In this section let $v^{\min}=\min_s v_s$ and let $v^{\max}=\max_s v_s$.

Theorem 3.1. *All states in* $S^{\min}$ *are* ε*-easy initial states for player 1, for all* $\varepsilon>0$.

Proof. Let $\varepsilon>0$ and let $\bar{y}$ be a pure stationary limiting average best reply to x^β for all β close to 1, where the x^β's are stationary β-discounted optimal strategies. Because of the β-discounted optimality of x^β we have:

$$v^\beta\leq(1-\beta)r(x^\beta,\bar{y})+\beta P(x^\beta,\bar{y})v^\beta\text{ for all }\beta.$$

Multiplying this inequality by $Q(x^\beta,\bar{y})$ gives us

$$Q(x^\beta,\bar{y})v^\beta\leq(1-\beta)Q(x^\beta,\bar{y})r(x^\beta,\bar{y})+\beta Q(x^\beta,\bar{y})v^\beta$$

for all β.

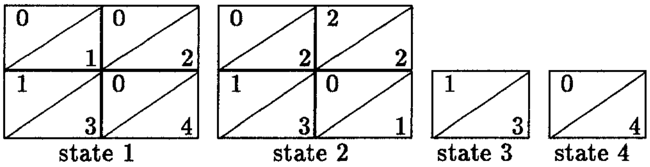
Hence we have that for β sufficiently close to 1:

$$\gamma(x^\beta,\bar{y})=Q(x^\beta,\bar{y})r(x^\beta,\bar{y})\geq Q(x^\beta,\bar{y})v^\beta\geq(v^{\min}-\varepsilon)1_z.$$

Because $\bar{y}$ is a limiting average best reply to x^β, the stationary strategy x^β, with β close to 1, is limiting average ε-optimal on $S^{\min}$. □

One could hope that similarly all states in $S^{\max}$ are ε-easy for player 1. However the next example illustrates that within $S^{\max}$ there may be states that are neither easy nor ε-easy for player 1.

Example 3.2

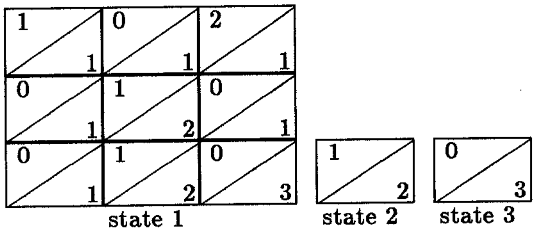


In this example $v=(1,1,1,0)$. It is not hard to verify that for any stationary strategy x player 2 has a best reply y with $\gamma(1,x,y)=\gamma(2,x,y)=0$, since player 2 can either steer the process to state 4, or he can make it remain in state 1 or 2 with payoff 0.

To see that $v=(1,1,1,0)$ examine the stationary strategy $\hat{x}^\beta$ given by the mixed action $((1-\sqrt{1-\beta})/\beta,$ $(-1+\beta+\sqrt{1-\beta})/\beta)$ for states 1 and 2, and find that: $1=\lim_{\beta\uparrow 1}\gamma^\beta(s,\hat{x}^\beta,y)\leq\lim_{\beta\uparrow 1}v_s^\beta=v_s\leq 1$ for $s=1,2$ and any pure stationary β-discounted best reply y. Here the last inequality holds because player 2 can play his first action in each state, thus guaranteeing payoffs of at most 1. We do not claim that $\hat{x}^\beta$ is β-discounted optimal, we just use it to provide a lower bound for v^β.

In some cases however one may find that, even though there are stationary limiting average ε-optimal strategies for player 1 ($\varepsilon>0$), neither x^1 nor x^β, with β close to 1, are limiting average ε-optimal in $S^{\max}$. We examine one such example and then we present a sufficient condition for all states in $S^{\max}$ to be ε-easy for player 1.

Example 3.3



For this game $v^\beta=((1-\sqrt{1-\beta})/\beta,1,0)$ for β close to 1 and a stationary β-discounted optimal strategy for player 1 is for instance $x^\beta=((1-\sqrt{1-\beta})/\beta,(-2+2\beta+\sqrt{1-\beta})/\beta,(1-\beta)/\beta)$. Recall that there is a pure stationary β-discounted best reply against x^β. For stationary strategies $y^1=(1,0,0)$, $y^2=(0,1,0)$ and $y^3=(0,0,1)$, one can verify that $\gamma^\beta(1,x^\beta,y^1)=\gamma^\beta(1,x^\beta,y^2)=\gamma^\beta(1,x^\beta,y^3)$ $=(1-\sqrt{1-\beta})/\beta=v_1^\beta$. It is straightforward to see that $\gamma(1,x^1,y^2)=0$ and $\gamma(1,x^\beta,y^3)=0$ for all β close to 1. Against x^β the strategy y^3 is a pure limiting average best reply, giving limiting average reward 0 to player 1 by a transition to state 3. This transition is clearly caused by $x_3^\beta\neq 0$. Thus, we have found that neither x^1 nor x^β, with β close to 1, are limiting average ε-optimal, since $v_1=\lim_{\beta\uparrow 1}v_1^\beta=1$.

Nevertheless, we would like to be able to derive a stationary limiting average ε-optimal strategy from any arbitrary converging sequence of stationary discounted optimal strategies. Thus we should find a way to deal with sequences x^β, like in this example. We examine x^β more closely by expanding it as a Puiseux series (cf. Bewley and Kohlberg [1]):

$$x^\beta=\begin{pmatrix}1\\0\\0\end{pmatrix}+\begin{pmatrix}-1\\1\\0\end{pmatrix}(1-\beta)^{1/2}$$

$$+\begin{pmatrix}1\\-2\\1\end{pmatrix}(1-\beta)+\begin{pmatrix}-1\\1\\0\end{pmatrix}(1-\beta)^{3/2}+\ldots$$

Notice that $x_3^\beta\neq 0$ is caused by the vector corresponding to $(1-\beta)$. If we let $\bar{x}^\beta$, with β near 1, consist of the first terms of this expression, up to and excluding $(1-\beta)$, then we get a stationary strategy for player 1: $\bar{x}^\beta=(1-\sqrt{1-\beta},\sqrt{1-\beta},0)$. Not only do we now have a strategy which avoids the third action, but also $\gamma(1,\bar{x}^\beta,y^1)=1-\sqrt{1-\beta}$, $\gamma(1,\bar{x}^\beta,y^2)=1$ and $\gamma(1,\bar{x}^\beta,y^3)=2-2\sqrt{1-\beta}$ which implies that $\bar{x}^\beta$ is limiting average ε-optimal for β close to 1.

We generalize the observation of Example 3.3 to derive a sufficient condition for all states in $S^{\max}$ to be ε-easy for player 1. First we recall that by Bewley and Kohlberg [1] we may assume that there are $N\in\mathbb{N}$, $x_0\in\times_{s=1}^{z}\Delta^{m_s}$, x_1, x_2, $\ldots\in\times_{s=1}^{z}\mathbb{R}^{m_s}$, such that:

$$x^\beta=\sum_{n=0}^{\infty}x_n(1-\beta)^{n/N}$$

is a stationary β-discounted optimal strategy for all β close to 1. Here m_s is the number of rows of A_s and Δ^{m_s} is the simplex of mixed actions over those rows.

Because x^β is a stationary strategy for β close to 1, we have $x^1=x_0$; $\sum_i x_{ns}(i)=0$ for all $n\geq 1$ and for all s; if for $s\in S$ and $n\geq 1$ one has $x_{0s}(i)=x_{1s}(i)=\ldots=x_{n-1s}(i)=0$, then $x_{ns}(i)\geq 0$.

We now define a strategy for player 1 that will turn out to be limiting average ε-optimal when a certain condition, to be specified later, is fulfilled. We also distinguish some sets of states within $S^{\max}$.

Definition 3.4. *Let $S^*=\{s\in S^{\max}$: for player 1 strategy x^1 is limiting average optimal for initial state $s\}$. Define $\bar{x}^\beta=\sum_{n=0}^{N-1}x_n(1-\beta)^{n/N}$ and $\underline{x}^\beta=\sum_{n=N+1}^{\infty}x_n(1-\beta)^{n/N}$, for β close to 1. Define $\tilde{x}_s^\beta=x_s^1$ for $s\in S^*$ and $\tilde{x}_s^\beta=\bar{x}_s^\beta$ otherwise, for β close to 1, and let $\tilde{y}$ be a pure stationary limiting average best reply to $\tilde{x}^\beta$ for all β close to 1.*

Define $S^{**}=S^*\cup\{E\subset S^{\max}$: E ergodic with respect to $(\tilde{x}^\beta,\tilde{y})\}$. Define $A=S^{\max}\setminus S^{**}$.

Here the existence of $\tilde{y}$ is again due to the finite number of pure stationary strategies. Recall that we are dealing with a countable sequence of strategies $\tilde{x}^\beta$, against each of which a pure stationary best reply exists. Hence we can take a subsequence of $\tilde{x}^{\beta}$'s, all of which have the same pure limiting average best reply $\tilde{y}$.

Lemma 3.5. *The stationary strategy $\tilde{x}^{\beta}$ is limiting average ε-optimal for all initial states in S^{**}, for β sufficiently close to 1.*

Proof. First of all notice that $S^{**} \neq \emptyset$ by the proof of Theorem 2.1 and second that by definition $\tilde{x}^{\beta}$ is limiting average optimal for initial states in S^*. Moreover on S^* we have $\tilde{x}^{\beta} = x^1$ and if player 1 uses x^1, then play will remain in S^* with probability 1 regardless of player 2's strategy. So let $E \subset S^{\max} \setminus S^*$ be ergodic with respect to $(\tilde{x}^{\beta}, \tilde{y})$. We show that $\gamma(s, \tilde{x}^{\beta}, \tilde{y}) \geq v^{\max} - \varepsilon$ for β close to 1.

For β close to 1 we have $v^{\beta} \leq (1-\beta) r(x^{\beta}, \tilde{y}) + \beta P(x^{\beta}, \tilde{y}) v^{\beta}$ and hence

$$\begin{aligned} v^{\beta} \leq{} & (1-\beta) r(\bar{x}^{\beta}, \tilde{y}) + \beta P(\bar{x}^{\beta}, \tilde{y}) v^{\beta} \\ & + (1-\beta)^2 r(x_N, \tilde{y}) + \beta(1-\beta) P(x_N, \tilde{y}) v^{\beta} \\ & + (1-\beta) r(\underline{x}^{\beta}, \tilde{y}) + \beta P(\underline{x}^{\beta}, \tilde{y}) v^{\beta}. \end{aligned}$$

Let Q_E^{β} denote the restriction of $Q(\bar{x}^{\beta}, \tilde{y})$ to rows corresponding with states in E. Multiplying above inequality by Q_E^{β} yields, after division by $1-\beta$ (recall that $Q_E^{\beta} P(\bar{x}^{\beta}, \tilde{y}) = Q_E^{\beta}$):

$$\begin{aligned} Q_E^{\beta} v^{\beta} \leq{} & Q_E^{\beta} r(\bar{x}^{\beta}, \tilde{y}) \\ & + (1-\beta) Q_E^{\beta} r(x_N, \tilde{y}) + \beta Q_E^{\beta} P(x_N, \tilde{y}) v^{\beta} \\ & + Q_E^{\beta} r(\underline{x}^{\beta}, \tilde{y}) + \beta(1-\beta)^{-1} Q_E^{\beta} P(\underline{x}^{\beta}, \tilde{y}) v^{\beta}. \end{aligned}$$

It is straightforward to verify that:

$$\lim_{\beta \uparrow 1} (1-\beta) Q_E^{\beta} r(x_N, \tilde{y}) = 0,$$

$$\lim_{\beta \uparrow 1} Q_E^{\beta} r(\underline{x}^{\beta}, \tilde{y}) = 0,$$

$$\lim_{\beta \uparrow 1} \beta(1-\beta)^{-1} Q_E^{\beta} P(\underline{x}^{\beta}, \tilde{y}) v^{\beta} = 0.$$

We show that we also have $\lim_{\beta \uparrow 1} \beta Q_E^{\beta} P(x_N, \tilde{y}) v^{\beta} \leq 0$:

Take $\delta > 0$. For $s \in E$ we have

$$\textstyle\sum_t p(t|s, x_{Ns}, \tilde{y}) v_t^{\beta} = \sum_i \sum_t x_{Ns}(i) p(t|s, i, \tilde{y}) v_t^{\beta}.$$

For β close to 1 we have:

$$\textstyle\sum_t p(t|s, i, \tilde{y}) v_t^{\beta} \leq v^{\max} + \delta \text{ for all } i.$$

For i with $x_{Ns}(i) < 0$ we have $\bar{x}_s^{\beta}(i) > 0$ and hence, since $s \in E \subset S^{\max}$ and E ergodic w.r.t. $(\bar{x}^{\beta}, \tilde{y})$, for such i and for β close to 1:

$$\sum_{t \in S} p(t|s, i, \tilde{y}) v_t^{\beta} = \sum_{t \in E} p(t|s, i, \tilde{y}) v_t^{\beta} \geq v^{\max} - \delta.$$

Combining these inequalities gives:

$$\begin{aligned} \sum_i \sum_t x_{Ns}(i) p(t|s, i, \tilde{y}) v_t^{\beta} \leq{} & \sum_{i, x_{Ns}(i) < 0} x_{Ns}(i)(v^{\max} - \delta) \\ & + \sum_{i, x_{Ns}(i) \geq 0} x_{Ns}(i)(v^{\max} + \delta) \\ ={} & 2\delta \sum_{i, x_{Ns}(i) \geq 0} x_{Ns}(i). \end{aligned}$$

Since $\delta > 0$ is arbitrary, we conclude that $\lim_{\beta \uparrow 1} P(x_N, \tilde{y}) v_t^{\beta} \leq 0$ and hence that $\lim_{\beta \uparrow 1} \beta Q_E^{\beta} P(x_N, \tilde{y}) v_t^{\beta} \leq 0$.

Altogether we have:

$$\begin{aligned} v^{\max} 1_E &= \lim_{\beta \uparrow 1} Q_E^{\beta} v^{\beta} \leq \lim_{\beta \uparrow 1} Q_E^{\beta} r(\bar{x}^{\beta}, \tilde{y}) \\ &= \lim_{\beta \uparrow 1} \gamma(\bar{x}^{\beta}, \tilde{y})_E = \lim_{\beta \uparrow 1} \gamma(\tilde{x}^{\beta}, \tilde{y})_E. \end{aligned}$$

Because $\tilde{y}$ is a limiting average best reply to $\tilde{x}^{\beta}$ for β sufficiently close to 1, this shows that $\tilde{x}^{\beta}$ is limiting average ε-optimal on E for β close to 1. □

In the above lemma we have seen that $\tilde{x}^{\beta}$ is limiting average ε-optimal for all initial states in $S^{**} \subset S^{\max}$. However, it may well be that $S^{**} \neq S^{\max}$ and that the states in $A = S^{\max} \setminus S^{**}$ are not ε-easy for player 1. Such is illustrated by Example 3.2 where $A = \{1, 2\}$, $S^{**} = S^* = \{3\}$ and $S^{\max} = \{1, 2, 3\}$. Nevertheless, in Theorem 3.6 below, we present a sufficient condition for $\tilde{x}^{\beta}$ to be limiting average ε-optimal for initial states in A as well. Obviously, this condition is not fulfilled for Example 3.2.

Below we use the notation v_A^{β}, $r(x^{\beta}, \tilde{y})_A$, 1_A, etc. for the restrictions of v^{β}, $r(x^{\beta}, \tilde{y})$, $1 = (1, 1, 1, \ldots, 1)$, etc. to coordinates in A. Let v_{Ac}^{β} (etc.) denote restriction to coordinates in $A^c = S \setminus A$. Also let $P(x^{\beta}, \tilde{y})^A$, $P(x^{\beta}, \tilde{y})^{Ac}$ and $P(x^{\beta}, \tilde{y})_A$ respectively denote restriction to rows and columns in A, rows in A and columns in A^c, rows in A.

Theorem 3.6. *If* $\lim_{\beta \uparrow 1} (1-\beta)(I^A - \beta P(\tilde{x}^{\beta}, \tilde{y})^A)^{-1} = 0$, *then $\tilde{x}^{\beta}$ is limiting average ε-optimal for all initial states in $S^{\max}$, for β close to 1.*

Proof. For initial states in S^{**} the ε-optimality follows from Lemma 3.5, so we only have to consider initial states in A. As above we start with $v^{\beta} \leq (1-\beta) r(x^{\beta}, \tilde{y}) + \beta P(x^{\beta}, \tilde{y}) v^{\beta}$, which implies:

$$\begin{aligned} v_A^{\beta} \leq{} & (1-\beta) r(x^{\beta}, \tilde{y})_A + \beta P(\bar{x}^{\beta}, \tilde{y})^A v_A^{\beta} + \beta P(\bar{x}^{\beta}, \tilde{y})^{Ac} v_{Ac}^{\beta} \\ & + \beta(1-\beta) P(x_N, \tilde{y})_A v^{\beta} + \beta P(\underline{x}^{\beta}, \tilde{y})_A v^{\beta}. \end{aligned}$$

Subtracting $\beta P(\bar{x}^{\beta}, \tilde{y})^A v_A^{\beta}$ on both sides, multiplying by $(I^A - \beta P(\bar{x}^{\beta}, \tilde{y})^A)^{-1}$ and taking limits gives:

$$\begin{aligned} v^{\max} 1_A = \lim_{\beta \uparrow 1} v_A^{\beta} \leq{} & \lim_{\beta \uparrow 1} \beta(I^A - \beta P(\bar{x}^{\beta}, \tilde{y})^A)^{-1} P(\bar{x}^{\beta}, \tilde{y})^{Ac} v_{Ac}^{\beta} \\ & + \lim_{\beta \uparrow 1} (1-\beta)(I^A - \beta P(\bar{x}^{\beta}, \tilde{y})^A)^{-1} [r(x^{\beta}, \tilde{y})_A \\ & + \beta P(x_N, \tilde{y})_A v^{\beta} + \beta(1-\beta)^{-1} P(\underline{x}^{\beta}, \tilde{y})_A v^{\beta}]. \end{aligned}$$

Note that each term within the square brackets is bounded uniformly in β. By the condition of the theorem we obtain:

$$\begin{aligned} v^{\max}1_A &\le \lim_{\beta\uparrow 1} \beta(I^A - \beta P(\bar{x}^\beta,\tilde{y})^A)^{-1}P(\bar{x}^\beta,\tilde{y})^{Ac}v^\beta_{Ac} \\ &= \lim_{\beta\uparrow 1} (I^A - \beta P(\bar{x}^\beta,\tilde{y})^A)^{-1}P(\bar{x}^\beta,\tilde{y})^{Ac}v_{Ac}. \end{aligned}$$

Now we use that for any square stochastic matrix P for which $(I-P)^{-1}$ exists, one has:

$$(I-\beta P)^{-1} = (I-P)^{-1} - (1-\beta)(I-\beta P)^{-1}P(I-P)^{-1}.$$

This can be verified by (left-)multiplying both sides with $(I-\beta P)$.

Applying this relation for $P=P(\bar{x}^\beta,\tilde{y})^A$, using that $P(\bar{x}^\beta,\tilde{y})^A(I^A-P(\tilde{x}^\beta,\tilde{y})^A)^{-1}P(\bar{x}^\beta,\tilde{y})^{Ac}$ is bounded and using the assumption that $\lim_{\beta\uparrow 1}(1-\beta)(I^A-\beta P(\bar{x}^\beta,\tilde{y})^A)^{-1}=0$, yields:

$$\begin{aligned} v^{\max}1_A &\le \lim_{\beta\uparrow 1} (I^A - P(\bar{x}^\beta,\tilde{y})^A)^{-1}P(\bar{x}^\beta,\tilde{y})^{Ac}v_{Ac} \\ &= \lim_{\beta\uparrow 1} (I^A - P(\tilde{x}^\beta,\tilde{y})^A)^{-1}P(\tilde{x}^\beta,\tilde{y})^{Ac}v_{Ac}. \end{aligned}$$

Since $v_{Ac}\le v^{\max}1_{Ac}$, the inequality sign in the above equation can be replaced by an equality sign. Next observe that entry (s,t) of $(I^A-P(\tilde{x}^\beta,\tilde{y})^A)^{-1}\,P(\tilde{x}^\beta,\tilde{y})^{Ac}$ denotes the total probability of ever entering A^c at state t, when originally starting in $s\in A$. Hence the probability of entering S^{**}, when starting in A, is close to 1 for β sufficiently near 1. Thus we have $\gamma(s,\tilde{x}^\beta,\tilde{y})\ge v^{\max}-\varepsilon$ for $s\in A$ and β close to 1. □

Remark 3.7. The condition of Theorem 3.6 not only means that starting in A one has to leave A with probability 1, but that one has to leave A sufficiently fast (cf. Example 3.2).

Observe that each entry of $(I^A-\beta P(\tilde{x}^\beta,\tilde{y})^A)^{-1}$ can be written as $\sum_{n=m}^{\infty} c_n(1-\beta)^{n/N}$, with $m\in\mathbb{Z}$ and $c_n\in\mathbb{R}$. The condition of Theorem 3.6 holds if and only if $m>-N$ for each entry. This occurs for instance if all states in A are transient with respect to $(x^1,\tilde{y})$, since in that case $\lim_{\beta\uparrow 1}(I^A-\beta P(x^\beta,\tilde{y})^A)^{-1}$ exists.

If $P(\tilde{x}^\beta,\ \tilde{y})^{Ac}_s\ne 0$ for each $s\in A$, then the condition of Theorem 3.6 also holds, because then each entry of $P^n(\tilde{x}^\beta,\tilde{y})^A$ is at most $(1-k(1-\beta)^{1/N})^n$ for some constant $k\in\mathbb{R}$.

4. Special classes of stochastic games

A unichain stochastic game is a stochastic game with the property that for every pair of stationary strategies there is precisely one ergodic set of states.

Theorem 4.1. *Let x^β be a stationary β-discounted optimal strategy, $\beta\in(0,1)$, in a unichain stochastic game. Then the unichain stochastic game has limiting average value $v=\lim_{\beta\uparrow 1}v^\beta$ and x^1 is limiting average optimal for player 1.*

Proof. Using a method similar to that of the proof for Theorem 2.1 one can show that for each stationary strategy y of player 2 it holds that: $\gamma(x^1,y)=\lim_{\beta\uparrow 1}\gamma^\beta(x^\beta,y)\ge\lim_{\beta\uparrow 1}v^\beta$. Because there is a stationary limiting average best reply to x^1, the limiting average value is at least $\lim_{\beta\uparrow 1}v^\beta$. A player 2 version gives that the limiting average value is at most $\lim_{\beta\uparrow 1}v^\beta$. □

Other proofs for the above theorem have been given by Gillette [4], Hoffman and Karp [5] and Rogers [9].

Bewley and Kohlberg [1] and Vrieze [13] give proofs for the following theorem, which is in fact a corollary of our Theorem 3.1.

Theorem 4.2. *If for a zero-sum stochastic game $\lim_{\beta\uparrow 1}v^\beta$ is independent of the initial state, then $\lim_{\beta\uparrow 1}v^\beta$ equals the limiting average value and a stationary β-discounted optimal strategy x^β is limiting average ε-optimal for player 1 for β close to 1.*

A stochastic game with state independent transitions (SIT) is a stochastic game for which all game matrices have equal size and for which furthermore $p(s,i,j)=p(t,i,j)$ for all $s,t\in S$ and all actions i,j.

Theorem 4.3. *For a* SIT *stochastic game the limiting average value is independent of the initial state, equals* $\lim_{\beta\uparrow 1}v^\beta$ *and both players have stationary limiting average optimal strategies.*

Proof. Let again x^β be a stationary β-discounted optimal strategy for player 1, $\beta\in(0,1)$. Let $x^1=\lim_{\beta\uparrow 1}x^\beta$ and let $\bar{y}$ be a stationary limiting average best reply to x^1. By the proof of Theorem 2.1 there is a non-empty set $S^*\subset S^{\max}$ such that for all $s\in S^*:\gamma(s,x^1,\bar{y})\ge\max_t(\lim_{\beta\uparrow 1}v^\beta_t)$. Then $p(t|s,x^1,y)=0$ for all $s\in S^*$, $t\in S\setminus S^*$ and all y.

Take $s^*\in S^*$ and define $x^*_s=x^1_s$ for $s\in S^*$ and $x^*_s=x^1_{s^*}$ for $s\in S\setminus S^*$. Now $\gamma(s,x^*,y)=\gamma(s^*,x^1,y)\ge\gamma(s^*,x^1,\bar{y})\ge\max_t(\lim_{\beta\uparrow 1}v^\beta_t)$ for all $s\in S$ and for all y. So for each initial state the limiting average value is at least $\max_t(\lim_{\beta\uparrow 1}v^\beta_t)$. A player 2 version would show that the limiting average value is at most $\min_t(\lim_{\beta\uparrow 1}v^\beta_t)$ for each initial state s. So $v_s=\lim_{\beta\uparrow 1}v^\beta_t$ for all s, $t\in S$. □

Parthasarathy et al. [8] derived a similar result for SER-SIT stochastic games, i.e. SIT stochastic games which have the additional property of separable rewards (SER): $r(s,i,j)=c(s)+a(i,j)$ for all s,i,j.

References

1. Bewley T, Kohlberg E (1976) The asymptotic theory of stochastic games. Math Oper Res 1:197–208
2. Blackwell D (1962) Discrete dynamic programming. Ann Math Statist 33:719–726
3. Blackwell D, Ferguson TS (1968) The big match. Ann Math Statist 39:159–163
4. Gillette D (1957) Stochastic games with zero stop probabilities.

In: Dresher M,Tucker AW, Wolfe P (eds) Contributions to the theory of games III (Ann Math Studies, vol 39). Princeton University Press, Princeton, pp 179–187

5. Hoffman AJ, Karp RM (1966) On non-terminating stochastic games. Manag Sci 25:352–362
6. Hordijk A, Vrieze OJ, Wanrooij GL (1983) Semi-Markov strategies in stochastic games. Int J Game Theory 12:81–89
7. Mertens JF, Neyman A (1981) Stochastic games. Int J Game Theory 10:53–66
8. Parthasarathy T, Tijs SH,Vrieze OJ (1984) Stochastic games with state independent transitions and separable rewards. In: Hammer G, Pallaschke D (eds) Selected topics in operations research and mathematical economics. Springer, Berlin Heidelberg New York, pp 262–271
9. Rogers PD (1969) Non-zerosum stochastic games. PhD thesis, report ORC 69-8, Operations Research Center, University of California, Berkeley
10. Shapley LS (1953) Stochastic games. Proc Natl Acad Sci USA 39:1095–1100
11. Thuijsman F, Vrieze OJ (1991) Easy initial states in stochastic games. In: Raghavan TES et al. (eds) Stochastic games and related topics. Kluwer Academic Publishers, Dordrecht, pp 85–100
12. Tijs SH, Vrieze OJ (1986) On the existence of easy initial states for undiscounted stochastic games. Math Oper Res 11:506–513
13. Vrieze OJ (1987) Stochastic games with finite state and action spaces. CWI-tract 33, Centre of Mathematics and Computer Science, Amsterdam