# Stochastic games with additive transitions

Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 179 (2007) 483–497
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Decision Support

# Stochastic games with additive transitions

J. Flesch, F. Thuijsman *, O.J. Vrieze

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

Received 23 March 2005; accepted 29 March 2006
Available online 24 May 2006

**Abstract**

We deal with $n$-player AT stochastic games, where AT stands for additive transitions. These are stochastic games in which the transition probability vector $p_s(a_s)$, for action combination $a_s = (a_s^1, \ldots, a_s^n)$ in state $s$, can be decomposed into player-dependent components as:

$$p_s(a_s) = \sum_{i=1}^{n} \lambda_s^i \cdot p_s^i(a_s^i),$$

where $\lambda_s^i \in [0,1]$ for all players $i$, and $\sum_{i=1}^n \lambda_s^i = 1$, and where $p_s^i(a_s^i)$ is a probability distribution on the finite set of states $S$. Here, $\lambda_s^i$ reflects the influence of player $i$ on the transitions in state $s$. As such the class of AT stochastic games covers several other well-known classes such as perfect information stochastic games, stochastic games with switching control, and so-called ARAT stochastic games.

With respect to the average reward it is not clear whether $\varepsilon$-equilibria always exist in general $n$-player stochastic games. For the class of $n$-player AT games we establish the existence of 0-equilibria, although the strategies involved may be history dependent. In addition we have the following results for the two-player case: (1) for zero-sum AT games, stationary 0-optimal strategies always exist; (2) for two-player general-sum AT absorbing games, there always exist stationary $\varepsilon$-equilibria, for all $\varepsilon > 0$.

Several examples are provided to clarify the issues and to demonstrate the sharpness of the results.


*Keywords:* Non-cooperative games; Multi-stage; Nash-equilibrium

## 1. Introduction

An $n$-player stochastic game $\Gamma$ can be described by (1) a set of players $N = \{1, \ldots, n\}$, (2) a nonempty and finite set of states $S$, (3) for each state $s$, a nonempty and finite set of actions $A_s^i$ for each player $i$, (4) for each state $s$ and each joint action $a_s \in \times_{i \in N} A_s^i$, a payoff $r_s^i(a_s) \in \mathbb{R}$ to each player $i$, (5) for each state $s$ and each joint action $a_s \in \times_{i \in N} A_s^i$, a transition probability vector $p_s(a_s) = (p_s(t|a_s))_{t \in S}$.

* Corresponding author. Fax: +31 31 43 3883489.
*E-mail address:* frank@math.unimaas.nl (F. Thuijsman).

0377-2217/$ - see front matter 
doi:10.1016/j.ejor.2006.03.031

The game is to be played at stages in $\mathbb{N}$ 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, each player $i$ is to choose an action $a^i_{s^1} \in A^i_{s^1}$. These choices induce an immediate payoff $r^i_{s^1}((a^j_{s^1})_{j\in N})$ to each player $i$. Next, the play moves to a new state according to the probability vector $p_{s^1}((a^j_{s^1})_{j\in N})$, say to state $s^2$. At stage 2 a new action $a^i_{s^2} \in A^i_{s^2}$ is to be chosen by each player $i$ in state $s^2$. Then player $i$ receives payoff $r^i_{s^2}((a^j_{s^2})_{j\in N})$ and the play moves to some state $s^3$ according to the probability vector $p_{s^2}((a^j_{s^2})_{j\in N})$, and so on.

A mixed action $x^i_s$ for player $i$ in state $s$ is a probability distribution on $A^i_s$. The set of mixed actions for player $i$ in state $s$ is denoted by $X^i_s$. A strategy $\pi^i$ for player $i$ is a decision rule that prescribes a mixed action $\pi^i_s(h) \in X^i_s$ in the present state $s$ depending on the past history $h$ of the play. We use the notation $\Pi^i$ for the set of history dependent strategies for player $i$. A strategy $\pi^i$ for player $i$ is called pure if $\pi^i$ prescribes, for each state and any past history, one specific action to be played with probability 1.

If the mixed actions prescribed by a strategy only depend on the present stage and state then the strategy is called Markov, while if they only depend on the present state then the strategy is called stationary. Thus, for player $i$, the Markov strategy space is $F^i := \times_{k\in\mathbb{N},s\in S} X^i_s$, while the stationary strategy space is $X^i := \times_{s\in S} X^i_s$. We will use the notations $f^i$ for Markov strategies and $x^i$ for stationary strategies for player $i$, while $f^i_s(k)$ and $x^i_s$ refer to the corresponding mixed actions for player $i$ in state $s$ at stage $k$. Note that the set of pure stationary strategies for player $i$ is $A^i = \times_{s\in S} A^i_s$.

We will often deal with quantities which depend on the player and the state. If $z^i_s \in \mathbb{R}$ for all $i \in N$, $s \in S$, then $z^i$ denotes the column-vector $(z^i_s)_{s\in S}$, $z_s$ denotes the row-vector $(z^i_s)_{i\in N}$, while $z$ denotes the matrix $(z^i_s)_{s\in S,i\in N}$. Similarly, if $Z^i_s$ are sets for all $i \in N$, $s \in S$, then let $Z^i := \times_{s\in S} Z^i_s$, $Z_s := \times_{i\in N} Z^i_s$, $Z := \times_{s\in S,i\in N} Z^i_s$.

A joint strategy $\pi = (\pi^i)_{i\in N}$ with an initial state $s \in S$ determines a stochastic process on the payoffs. The sequences of payoffs are evaluated by the average reward and by the $\beta$-discounted reward, $\beta \in (0,1)$, which are given for player $i$ by

$$\gamma^i_s(\pi) := \liminf_{K\to\infty} \mathbb{E}_{s\pi}\left(\frac{1}{K}\sum_{k=1}^{K} R^i_k\right) = \liminf_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{s\pi}(R^i_k),$$

$$\gamma^i_{\beta s}(\pi) := \mathbb{E}_{s\pi}\left((1-\beta)\sum_{k=1}^{\infty} \beta^{k-1} R^i_k\right),$$

where $R^i_k$ is the random variable for the payoff for player $i$ at stage $k$, and where $\mathbb{E}_{s\pi}$ stands for expectation with respect to the initial state $s$ and the joint strategy $\pi$.

A joint stationary strategy $x \in X$ determines a Markov-chain with transition matrix $P(x)$ on $S$, where entry $(s,t)$ of $P(x)$ gives the transition probability $p_s(t|x_s)$ for moving from state $s$ to state $t$ when $x_s$ is played in state $s$.

With respect to this Markov-chain, we can speak of transient and recurrent states. A state is called recurrent if, when starting there, it will be visited infinitely often with probability 1; otherwise the state is called transient. We can group the recurrent states into minimal closed sets, into so-called ergodic sets. An ergodic set is a collection $E$ of recurrent states with the property that, when starting in one of the states in $E$, all states in $E$ will be visited and the play will remain in $E$ forever with probability 1. Let

$$Q(x) := \lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} P^k(x); \tag{1}$$

the limit is known to exist (cf. Doob (1953, Theorem 2.1, p. 175)). Entry $(s,t)$ of the stochastic matrix $Q(x)$, denoted by $q_s(t|x)$, is the expected frequency of stages for which the process is in state $t$ when starting in $s$. The matrix $Q(x)$ has the well known properties (cf. Doob, 1953) that

$$Q(x) = Q(x)P(x) = P(x)Q(x) = Q^2(x). \tag{2}$$

For $x_s \in X_s$ let $r^i_s(x_s)$ denote the expected immediate payoff for player $i$ in state $s$ if the joint mixed action $x_s$ is played. By definition, for the average reward we have

$$\gamma(x) = Q(x)r(x), \tag{3}$$

hence by (2) we also obtain

$$\gamma(x) = P(x)\gamma(x), \tag{4}$$

$$\gamma(x) = Q(x)r(x) = Q^2(x)r(x) = Q(x)\gamma(x). \tag{5}$$

For $i \in N$, let $N^{-i} = N - \{i\}$ denote the set of opponents of player $i$, and let

$$X^{-i} := \times_{j\in N^{-i}} X^j, \quad F^{-i} := \times_{j\in N^{-i}} F^j, \quad \Pi^{-i} := \times_{j\in N^{-i}} \Pi^j$$

denote the sets of (different types of) joint strategies of the opponents of player $i$.

It is well known (cf. Hordijk et al. (1983)) that, against a fixed joint stationary strategy $x^{-i} \in X^{-i}$, there always exists a pure stationary best reply $a^i \in A^i$ of player $i$, i.e.

$$\gamma^i(a^i, x^{-i}) \geqslant \gamma^i(\pi^i, x^{-i}) \quad \forall \pi^i \in \Pi^i. \tag{6}$$

For $i \in N$, $s \in S$, $\beta \in (0,1)$, let

$$v_s^i := \inf_{\pi^{-i}\in\Pi^{-i}} \sup_{\pi^i\in\Pi^i} \gamma_s^i(\pi^i, \pi^{-i})$$

$$v_{\beta s}^i := \inf_{\pi^{-i}\in\Pi^{-i}} \sup_{\pi^i\in\Pi^i} \gamma_{\beta s}^i(\pi^i, \pi^{-i}).$$

Here $v_s^i$ and $v_{\beta s}^i$ are called the average and the $\beta$-discounted minmax values for player $i$ in state $s$, respectively. Intuitively, these are the highest average and $\beta$-discounted rewards that player $i$ can defend against any strategies of the other players if the initial state is $s$. Neyman (1986) showed that, for any $i \in N$ and $\beta \in (0,1)$, there exists an $x^{-i} \in X^{-i}$ satisfying

$$\gamma_\beta^i(\pi^i, x^{-i}) \leqslant v_\beta^i \quad \forall \pi^i \in \Pi^i \tag{7}$$

and

$$v^i = \lim_{\beta\uparrow 1} v_\beta^i. \tag{8}$$

It is clear from the definition of $v_s^i$ and (4) that

$$v_s^i = \min_{x_s^{-i}\in X_s^{-i}} \max_{x_s^i\in X_s^i} \sum_{t\in S} p_s(t|x_s^i, x_s^{-i}) v_t^i. \tag{9}$$

A joint strategy $\pi = (\pi^i)_{i\in N}$ is called an $\varepsilon$-equilibrium, $\varepsilon \geqslant 0$, with respect to the average reward, if

$$\gamma_s^i(\sigma^i, \pi^{-i}) \leqslant \gamma_s^i(\pi) + \varepsilon \quad \forall \sigma^i \in \Pi^i, \ \forall i \in N, \ \forall s \in S,$$

which means that for every initial state $s \in S$, no player can gain more than $\varepsilon$ by a unilateral deviation. Equivalently, strategy $\pi^i$ is an $\varepsilon$-best reply for each player $i$ against $\pi^{-i}$. The definition of $\beta$-discounted equilibria is similar. For simplicity, 0-equilibria are also called equilibria.

It is clear from the definitions of the minmax values $v$ and $v_\beta$ that if $\pi$ is an $\varepsilon$-equilibrium then $\gamma^i(\pi) \geqslant v^i - \varepsilon$ for each player $i$; while if $\pi$ is a $\beta$-discounted $\varepsilon$-equilibrium then $\gamma_\beta^i(\pi) \geqslant v_\beta^i - \varepsilon$ for each player $i$.

Fink (1964) and Takahashi (1964) showed that $\beta$-discounted equilibria always exist in terms of stationary strategies. The structure of average equilibria is however substantially more complex and the question of existence of average $\varepsilon$-equilibria, for all $\varepsilon > 0$, has not yet been answered. The famous game introduced by Gillette (1957), the Big Match, which was solved by Blackwell and Ferguson (1968), and the game in Sorin (1986) demonstrate that, in general, average 0-equilibria do not exist and history dependent strategies are indispensable for establishing average $\varepsilon$-equilibria. The general existence of average $\varepsilon$-equilibria for two-player stochastic games was finally shown by Vieille (2000a,b).

In the development of stochastic games, a special role has been played by the class of zero-sum stochastic games, which are two-player stochastic games for which $r_s^2(a_s) = -r_s^1(a_s)$, for each state $s$ and for each joint action $a_s$. In these games the two players have completely opposite interests. Mertens and Neyman (1981) showed that for such games $v^2 = -v^1$. Here $v := v^1$ is called the value of the game. They also showed that, if instead of using liminf one uses limsup in the definition of the average reward, one would find precisely the same value $v$. Thus, in a zero-sum game, player 1 wants to maximize his own reward, while at the same

time player 2 tries to minimize player 1's reward. For simplicity, let $\gamma = \gamma^1$. A strategy $\pi^1$ for player 1 is called $\varepsilon$-optimal if $\gamma(\pi^1, \pi^2) \geqslant v - \varepsilon$ for any strategy $\pi^2$ of player 2; while a strategy $\pi^2$ for player 2 is called $\varepsilon$-optimal if $\gamma(\pi^1, \pi^2) \leqslant v + \varepsilon$ for any strategy $\pi^1$ of player 1. Mertens and Neyman (1981) proved that both players have $\varepsilon$-optimal strategies for any $\varepsilon > 0$; even though history dependent strategies are necessary for $\varepsilon$-optimality.

From now on when we speak of rewards, minmax values, or equilibria, we will always have the average reward in mind, unless mentioned otherwise.

A stochastic game is said to have an additive transition structure if, for any state $s \in S$ and any joint action $a_s \in A_s$, the transition probabilities can be additively decomposed as

$$p_s(a_s) = \sum_{i \in N} \lambda_s^i p_s^i(a_s^i),$$

where $\lambda_s^i \in [0,1]$ for all $i \in N$, $\sum_{i \in N} \lambda_s^i = 1$, and $p_s^i(a_s^i)$ is a probability distribution on $S$. Here, the component $p_s^i(a_s^i)$ only depends on the action $a_s^i$ of player $i$ in state $s$, so $\lambda_s^i$ reflects the influence of player $i$ on the transitions in state $s$.

Stochastic games with an additive transition structure shall be called AT stochastic games for short.

The class of AT stochastic games includes several important classes, such as stochastic games with switching control (namely when in each state $s$, one player controls the transitions: $\lambda_s^i = 1$ for some $i \in N$), or stochastic games with ARAT structure (namely when besides having additive transitions, the payoffs are also additively decomposable). Note that the class of switching control games further contains the well-known classes of single controller stochastic games (when there is a player $i$ for whom $\lambda_s^i = 1$ for all $s \in S$) and perfect information stochastic games (when in any state $s \in S$, there is at most one player who has more than one action).

In this paper we generalize the results achieved for these subclasses by Liggett and Lippmann (1969), Filar (1981), Raghavan et al. (1985), Thuijsman and Raghavan (1997), and Evangelista et al. (1996), by showing, using a different approach, for AT stochastic games: (i) the existence of 0-equilibria in terms of history dependent strategies; (ii) in zero-sum AT games, the existence of stationary 0-optimal strategies; (iii) in two-player absorbing AT games, the existence of stationary $\varepsilon$-equilibria for all $\varepsilon > 0$. An absorbing game is a stochastic game with the property that all states but one are absorbing, i.e. once play gets there, it will stay there forever.

We remark that the results (ii) and (iii) are based exclusively on stationary strategies. Therefore, these solutions are subgame perfect. We can not strengthen result (i) to the existence of subgame perfect $\varepsilon$-equilibria. At this moment hardly anything is known about existence of subgame perfect equilibria for stochastic games. We emphasize, once again, that the general existence of $\varepsilon$-equilibria has only been shown for two-player stochastic games using the idea of threats, which are not necessarily subgame perfect. Our main result (i) solves the fundamental existence problem for the particular class of $n$-player AT stochastic games.

The outline of the paper is as follows: In Section 2 we provide some preliminary results; in Section 3 we exhibit result (ii) on zero-sum AT stochastic games; Section 4 is devoted to result (iii) on two-player AT absorbing games; and in Section 5 we prove our main result (i) on is on general $n$-player AT stochastic games. We provide several examples to illustrate the issues and to demonstrate the sharpness of the results.

## 2. Preliminaries

The following lemma exhibits an important relationship between the average and discounted rewards for stationary strategies.

**Lemma 1.** *Let $x \in X$. Suppose that $E \subset S$ is an ergodic set with respect to $x$. Let $s \in E$ and $\beta \in (0,1)$. Then*

$$\min_{t \in E} \gamma_{\beta t}(x) \leqslant \gamma_s(x) \leqslant \max_{t \in E} \gamma_{\beta t}(x).$$

**Proof.** By the definition of the $\beta$-discounted reward $\gamma_\beta$, we have

$$\gamma_\beta(x) = (1-\beta) r(x) + \beta P(x) \gamma_\beta(x).$$

In view of (2), multiplying this equality by $Q(x)$ yields

$$Q(x)\gamma_\beta(x) = Q(x)r(x),$$

hence by (3)

$$\gamma(x) = Q(x)\gamma_\beta(x).$$

Since $s \in E$, the closedness of $E$ for $x$ implies that if $q_s(t|x) > 0$ then $t \in E$. Therefore,

$$\gamma_s(x) = \sum_{t\in E} q_s(t|x)\gamma_{\beta t}(x).$$

Now from

$$q_s(t|x) \geqslant 0 \quad \forall t \in E, \quad \text{and} \quad \sum_{t\in E} q_s(t|x) = 1,$$

the result immediately follows. □

**Lemma 2.** *Let $\phi_s \in \mathbb{R}$ for all $s \in S$, and $\phi := (\phi_s)_{s\in S}$. Let $x \in X$ be such that*

$$P(x)\phi \geqslant \phi.$$

*Suppose $E$ is an ergodic set with respect to $x$. Then we necessarily have $\phi_s = \phi_t$ for all $s, t \in E$. Moreover, if it also holds that $\gamma_s(x) \geqslant \phi_s$ for all recurrent states $s$ then $\gamma(x) \geqslant \phi$.*

**Proof.** Let $\overline{E} := \{s \in E | \phi_s = \max_{t\in E}\phi_t\}$ and $s \in \overline{E}$. By the closedness of $E$ for $x$ we obtain

$$\phi_s \leqslant \sum_{t\in S} p_s(t|x_s)\phi_t = \sum_{t\in E} p_s(t|x_s)\phi_t \leqslant \phi_s.$$

The above inequalities imply that from state $s$, transition can only occur to states in $\overline{E}$ with respect to $x$. So, the set $\overline{E}$ is a closed set of states for $x$. Since $E$ is an ergodic set for $x$, we must have $\overline{E} = E$. Therefore $\phi_s = \phi_t$ for all $s, t \in E$.

Assume further that $\gamma_s(x) \geqslant \phi_s$ for all recurrent states $s$. Then

$$\phi \leqslant Q(x)\phi \leqslant Q(x)\gamma(x) = \gamma(x),$$

where the first inequality follows from $\phi \leqslant P(x)\phi$, the second inequality from the fact that entry $(t, s)$ of $Q(x)$ is only non-zero for recurrent states $s$, and the final equality from (5). □

The next technical lemma on the transition structure of AT games shall be used in the proofs.

**Lemma 3.** *Let $G$ be an arbitrary two-player AT game. Let $s \in S$ and $S_* \subset S$. Suppose that, in state $s$, for every action $a_s^1 \in A_s^1$ there exists an action $a_s^2 \in A_s^2$ such that moving to $S_*$ has probability 1: $p_s(S_*|a_s^1, a_s^2) = 1$. Then for any $a_s^2$ we have either $p_s(S_*|a_s^1, a_s^2) = 1$ for all $a_s^1$ or $p_s(S_*|a_s^1, a_s^2) < 1$ for all $a_s^1$.*

**Proof.** Suppose by way of contradiction that $p_s(S_*|a_s^1, a_s^2) = 1$ and $p_s(S_*|b_s^1, a_s^2) < 1$. Clearly, we must have $\lambda_s^1 > 0$ (or equivalently $\lambda_s^2 < 1$), which implies $p_s^1(S_*|a_s^1) = 1$. If $\lambda_s^1 < 1$ (equivalently $\lambda_s^2 > 0$) then $p_s^2(S_*|a_s^2) = 1$ and therefore $p_s^1(S_*|b_s^1) < 1$; while $\lambda_s^1 = 1$ also yields $p_s^1(S_*|b_s^1) < 1$. Hence $p_s^1(S_*|b_s^1) < 1$, which in combination with $\lambda_s^1 > 0$ yields $p_s(S_*|b_s^1, b_s^2) < 1$ for all $b_s^2$, contradicting the assumption of the lemma. □

**Lemma 4.** *Let $\alpha^i \in \mathbb{R}$ for each player $i$ and let $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$ be a monotone decreasing sequence of reals converging to 0. For each $m$ let $x_m$ be a joint stationary strategy. Assume that for all strategies $x_m$ the carrier is the same, i.e. the set of actions which have positive weight for $x$. Suppose that $E$ is an ergodic set with respect to $x_m$ and $\gamma_s^i(x_m) \geqslant \alpha^i - \varepsilon_m$ for all $s \in E$ and for each player $i$. Then there exists a pure joint strategy $\pi$ such that $\pi$ only uses actions within the carrier of $x_m$ and such that $\gamma_s^i(\pi) \geqslant \alpha^i$ for all $s \in E$ and for each player $i$. Moreover, at any point of time, after any history $h$, the continuation strategy $\pi|h$ also yields at least $\alpha^i$ for any present state $s \in E$ and for each player $i$.*

**Proof.** We only show the statement for player $i$. For each $m \in \mathbb{N}$, by Lemma 6 in Dutta (1995), there exists a joint pure strategy $\pi_m$ which only uses actions in the carrier of $x_m$ and for which $|\gamma_s^i(\pi_m) - \gamma_s^i(x_m)| \leqslant \frac{1}{2}\varepsilon_m$ for all $s \in E$ and for each player $i$. Let $\overline{K}_m \in \mathbb{N}$ be such that

$$\frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{s\pi_m}(R_k^i) \geqslant \gamma_s^i(\pi_m) - \frac{1}{2}\varepsilon_m \quad \forall K \geqslant \overline{K}_m, \ \forall s \in E, \ \forall i, \tag{10}$$

where $R_k^i$ denotes the random variable for the payoff to player $i$ at stage $k$. Then

$$\frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{s\pi_m}(R_k^i) \geqslant \gamma_s^i(x_m) - \varepsilon_m \geqslant \alpha^i - 2\varepsilon_m \quad \forall K \geqslant \overline{K}_m, \ \forall s \in E, \ \forall i. \tag{11}$$

Define

$$r^i := \min\left\{\alpha^i - 2\varepsilon_1, \min_{a_s \in A_s, s \in S} r_s^i(a_s)\right\}.$$

Given $\overline{K}_m$, $m \in \mathbb{N}$, choose an arbitrary $K_1 \geqslant \overline{K}_1$ and choose $K_m \geqslant \overline{K}_m$, $m \geqslant 2$, inductively so that

$$\frac{\sum_{k=1}^{m} K_k \cdot (\alpha^i - 2\varepsilon_k) + \overline{K}_{m+1} \cdot r^i}{\sum_{k=1}^{m} K_k + \overline{K}_{m+1}} \geqslant \alpha^i - 2\varepsilon_{m-1} \quad \forall m \geqslant 2, \ \forall i. \tag{12}$$

By the definition of $r^i$, inequality (12) implies

$$\frac{\sum_{k=1}^{m} K_k \cdot (\alpha^i - 2\varepsilon_k)}{\sum_{k=1}^{m} K_k} \geqslant \alpha^i - 2\varepsilon_{m-1} \quad \forall m \geqslant 2, \ \forall i. \tag{13}$$

Now we define a pure joint strategy $\pi$ as playing $\pi_1$ for the first block of $K_1$ stages, $\pi_2$ for the next block of $K_2$ stages, etcetera. We only need to show that $\gamma_s^i(\pi) \geqslant \alpha^i$ for all $s \in E$ and for any player $i$. Since, at each point in time, the strategy only uses the history of the current block, all continuation strategies $\pi|h$ will also yield at least $\alpha^i$. Suppose we are at the $T$th stage of block $m+1$. Then, in any block $k$, where $k \leqslant m$, player $i$ has received a total expected payoff of at least $K_k \cdot (\alpha^i - 2\varepsilon_k)$. In block $m+1$, if $T < \overline{K}_{m+1}$, then player $i$ received at least $T \cdot r^i$. If, on the other hand, $T \geqslant \overline{K}_{m+1}$, then player $i$ received at least $T \cdot (\alpha^i - 2\varepsilon_{m+1})$. So player $i$'s expected average payoff up to the $T$th stage of block $m+1$ is in the former case at least

$$\frac{\sum_{k=1}^{m} K_k \cdot (\alpha^i - 2\varepsilon_k) + T \cdot r^i}{\sum_{k=1}^{m} K_k + T} \geqslant \frac{\sum_{k=1}^{m} K_k \cdot (\alpha^i - 2\varepsilon_k) + \overline{K}_{m+1} \cdot r^i}{\sum_{k=1}^{m} K_k + \overline{K}_{m+1}} \geqslant \alpha^i - 2\varepsilon_{m-1};$$

while in the latter case it is at least

$$\frac{\sum_{k=1}^{m} K_k \cdot (\alpha^i - 2\varepsilon_k) + T \cdot (\alpha^i - 2\varepsilon_{m+1})}{\sum_{k=1}^{m} K_k + T} \geqslant \frac{\sum_{k=1}^{m} K_k \cdot (\alpha^i - 2\varepsilon_k)}{\sum_{k=1}^{m} K_k} \geqslant \alpha^i - 2\varepsilon_{m-1}.$$

So player $i$'s expected average payoff up to any stage of block $m+1$ is at least $\alpha^i - 2\varepsilon_{m-1}$, which implies $\gamma_s^i(\pi) \geqslant \alpha^i$ for all $s \in E$ and for any player $i$. □

## 3. Zero-sum AT games

Take an arbitrary zero-sum AT game and let $v = v^1 = -v^2$ denote the value. It follows from Eq. (9) and from the additive transition structure of the game that

$$v_s = \min_{x_s^2 \in X_s^2} \max_{x_s^1 \in X_s^1} \sum_{t \in S} p_s(t|x_s^1, x_s^2) v_t = \lambda_s^1 \cdot \max_{a_s^1 \in A_s^1} \sum_{t \in S} p_s^1(t|a_s^1) v_t + \lambda_s^2 \cdot \min_{a_s^2 \in A_s^2} \sum_{t \in S} p_s^2(t|a_s^2) v_t.$$

If $\lambda_s^1 > 0$, then let $\overline{A}_s^1$ be the set of actions of player 1 in state $s$ that maximize the expression $\sum_{t \in S} p_s^1(t|\cdot) v_t$, while if $\lambda_s^1 = 0$ (meaning that in state $s$ player 1 has no influence on the transitions), then we define $\overline{A}_s^1 = A_s^1$. Consequently, $\overline{A}_s^1$ is the set of those actions $a_s^1$ for which

$$v_s \leqslant \sum_{t \in S} p_s(t|a_s^1, x_s^2) v_t \quad \forall x_s^2 \in X_s^2.$$

Let $\overline{A}_s^2$ be defined similarly, where in state $s$ player 2 is minimizing the expression $\sum_{t\in S} p_s^2(t|\cdot)v_t$. Therefore, $\overline{A}_s^2$ is the set of those actions $a_s^2$ for which

$$v_s \geqslant \sum_{t\in S} p_s(t|x_s^1, a_s^2)v_t \quad \forall x_s^1 \in X_s^1.$$

The main idea we shall use in the analysis is that of the restricted game $\bar{G}$ derived from $G$ by restricting the players to actions in $\overline{A}_s^1$ and $\overline{A}_s^2$ in all states $s$. Then $\overline{G}$ is an AT stochastic game as well. Obviously, in $\overline{G}$ mixed actions only use actions that are still available. Thus we define $\overline{X}_s^1$ and $\overline{X}_s^2$ as the sets of mixed actions on $\overline{A}_s^1$ and $\overline{A}_s^2$ respectively. We shall denote the value of $\overline{G}$ by $\bar{v}$. In a natural way, $\overline{X}_s^1$ and $\overline{X}_s^2$ can be seen as subsets of $X_s^1$ and $X_s^2$ respectively. Observe now that if $x_s^1 \in X_s^1$ then

$$x_s^1 \in \overline{X}_s^1 \Longleftrightarrow v_s \leqslant \sum_{t\in S} p_s(t|x_s^1, x_s^2)v_t \quad \forall x_s^2 \in X_s^2$$

and if $x_s^1 \in \overline{X}_s^1$ and $x_s^2 \in X_s^2$ then

$$x_s^2 \in \overline{X}_s^2 \Longleftrightarrow v_s = \sum_{t\in S} p_s(t|x_s^1, x_s^2)v_t. \tag{14}$$

**Lemma 5.** *Let $G$ be an arbitrary zero-sum AT game and let $\overline{G}$ be the corresponding restricted game. Then $v = \bar{v}$.*

**Proof.** Suppose by way of contradiction that $\bar{v}_s < v_s$ for some state $s$ (the arguments are similar for the case when $\bar{v}_s > v_s$). Let $d_1 = v_s - \bar{v}_s$, and let

$$d_2 = \min_{a_t^1 \in A_t^1 - \overline{A}_t^1, a_t^2 \in \overline{A}_t^2, t\in S} \left[ v_t - \sum_{w\in S} p_t(w|a_t^1, a_t^2)v_w \right];$$

the minimized expression is in fact independent of the choice of $a_t^2 \in \overline{A}_t^2$. Here $d_2$ is the minimal decrease in the expectation of the value after transition if player 1 chooses an action outside $\overline{A}_t^1$ in some state $t$, given player 2 plays an action in $\overline{A}_t^2$. Notice that by the assumption $\bar{v}_s < v_s$, there must be a state $t$ such that $A_t^1 - \overline{A}_t^1 \neq \emptyset$. So, we minimize sover a non-empty set. Because of the definition of the sets $\overline{A}_t^1$, we have $d_2 > 0$.

Now let $\bar{\pi}^2$ denote a $\frac{d_1}{2}$-optimal strategy for player 2 in $\overline{G}$ and $\sigma^2$ a $\frac{d_2}{2}$-optimal strategy for player 2 in $G$. Consider the strategy $\pi^2$ for player 2 in $G$ which prescribes to play as follows: play $\bar{\pi}^2$ as long as player 1 chooses actions in the sets $\overline{A}_t^1$, $t \in S$, and as soon as player 1 takes an action outside, start playing $\sigma^2$.

Take an arbitrary $\varepsilon$-best reply $\pi^1$ to $\pi^2$ for player 1 in G for initial state $s$. Note with respect to $(\pi^1, \pi^2)$ and initial state $s$ that as long as player 1 chooses actions in the sets $\overline{A}_t^1$, player 2 is also using only actions in the sets $\overline{A}_t^2$, so the value $v$ does not change in expectation. Notice also that if player 1 ever chooses an action outside $\overline{A}_t^1$ in some state $t$, then the value $v$ drops at least by $d_2$ in expectation and afterwards player 2 plays a $\frac{d_2}{2}$-optimal strategy, so player 1's reward cannot be more than $v_t - \frac{d_2}{2}$. This means that the probability of ever choosing an action outside the sets $\overline{A}_t^1$ is close to zero (if $\varepsilon$ is small). But then player 1 is facing strategy $\bar{\pi}^2$ in the game $\overline{G}$ for the whole play with probability almost 1, and in that case his reward is at most $\bar{v}_s + \frac{d_1}{2} = v_s - \frac{d_1}{2}$, which contradicts the definition of the value $v_s$. □

The following lemma exhibits the advantage of $\overline{G}$.

**Lemma 6.** *If a stationary strategy $\bar{x}^1$ is optimal in $\overline{G}$, then $\bar{x}^1$ is optimal in $G$ as well.*

**Proof.** Let $x^2$ be a stationary best reply to $\bar{x}^1$ in $G$. Then we have to show that $\gamma(\bar{x}^1, x^2) \geqslant v$.

For this purpose consider an arbitrary ergodic set $E$ for $(\bar{x}^1, x^2)$. Since $\bar{x}^1 \in \overline{X}^1$, we have $v \leqslant P(\bar{x}^1, x^2)v$. Hence, by Lemma 2 (with $\phi = v$) we have that the value $v$ must be constant on $E$. This also means by (14) that $x_s^2 \in \overline{X}_s^2$ for all $s \in E$. Because $\bar{x}^1$ is optimal in $\overline{G}$, Lemma 5 yields $\gamma_s(\bar{x}^1, x^2) \geqslant \bar{v}_s = v_s$ for all $s \in E$. By using Lemma 2 again, we obtain $\gamma(\bar{x}^1, x^2) \geqslant v$. Thus, $\bar{x}^1$ is optimal in $G$. □

**Theorem 7.** *In every zero-sum AT game, both players have a stationary optimal strategy.*

**Proof.** We will only prove it for player 1; for player 2 the proof is similar. In view of the previous lemma it is sufficient to show the existence of a stationary optimal strategy in $\overline{G}$. So we may forget about the original game $G$ and only consider $\overline{G}$ from now on.

It is well-known that in any zero-sum stochastic game there is a stationary strategy $\bar{x}^1 \in \overline{X}^1$ such that $\bar{x}^1$ is optimal for at least one initial state with maximal value (cf. Tijs and Vrieze, 1986 or Thuijsman and Vrieze 1991). Take such a strategy $\bar{x}^1$ and let $\overline{E}$ be the set of all states with maximal value for which this particular strategy $\bar{x}^1$ is optimal. Clearly, if play starts in $\overline{E}$ and player 1 plays $\bar{x}^1$, then play will remain in $\overline{E}$, irrespective of player 2's strategy.

If $\overline{E} = S$, then we are done. Otherwise, let $E_1$ be the set of states $s$ in $S - \overline{E}$ for which player 1 has an action $\bar{a}^1_s \in \overline{A}^1_s$ such that transistion occurs to $\overline{E}$ with positive probability, irrespective of the action of player 2:

$$p_s(\overline{E}|\bar{a}^1_s, \bar{a}^2_s) > 0 \quad \text{for all } \bar{a}^2_s \in \overline{A}^2_s.$$

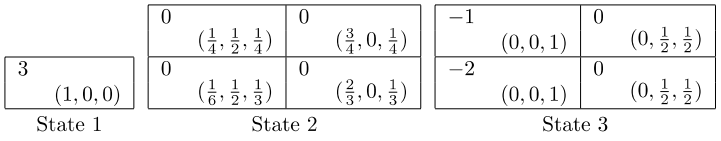
Further, let $E_2$ be the set of states $s$ in $S - (\overline{E} \cup E_1)$ for which player 1 has an action $\bar{a}^1_s \in \overline{A}^1_s$ such that $p_s(\overline{E} \cup E_1|\bar{a}^1_s, \bar{a}^2_s) > 0$ for all $\bar{a}^2_s \in \overline{A}^2_s$. This way we proceed by considering sets $E_n$, defined as the set of states $s$ in $S - (\overline{E} \cup E_1 \cup \ldots \cup E_{n-1})$ for which player 1 has an action $\bar{a}^1_s \in \overline{A}^1_s$ such that $p_s(\overline{E} \cup E_1 \cup \ldots \cup E_{n-1}|\bar{a}^1_s, \bar{a}^2_s) > 0$ for all $\bar{a}^2_s \in \overline{A}^2_s$, until either $\overline{E} \cup E_1 \cup \ldots \cup E_n = S$ or $E_n = \emptyset$.

*Case 1*: $\overline{E} \cup E_1 \cup \ldots \cup E_n = S$. Consider the strategy $\tilde{x}^1 \in \overline{X}^1$ for player 1 which prescribes to play $\bar{x}^1$ in $\overline{E}$ and $\bar{a}^1_s$ in each $s \in E_1 \cup \ldots \cup E_n$. Then, notice that from any initial state, irrespective of the strategy of player 2, play eventually moves to set $\overline{E}$ and from that stage on, play remains in $\overline{E}$ forever. As the value $\bar{v}$ is maximal in $\overline{E}$, it easily follows that $\bar{v}$ is a constant over the whole state space $S$. Since $\bar{x}^1$ is optimal for initial states in $\overline{E}$ in $\overline{G}$, we deduce that $\tilde{x}^1$ must be optimal in $\overline{G}$ for all initial states.

*Case 2*: $\overline{E} \cup E_1 \cup \ldots \cup E_n \neq S$ and $E_n = \emptyset$. Then $S_* := S - (\overline{E} \cup E_1 \cup \ldots \cup E_{n-1})$ is a non-empty set of states. By Lemma 3 for any $s \in S_*$, player 2 has a set $\overline{A}^2_{*s} \subset \overline{A}^2_s$ of actions $\bar{a}^2_s$ such that $p_s(S_*|a^1_s, \bar{a}^2_s) = 1$ for all $a^1_s \in \overline{A}^1_s$. However, because of this property, if player 2 only uses actions in $\overline{A}^2_{*s}$ then play will always remain in $S_*$ and therefore we might as well consider the restricted stochastic game $G_*$ with state space $S_*$, where player 1 chooses actions from $\overline{A}^1_s$ and player 2 is restricted to $\overline{A}^2_{*s}$. Again, $G_*$ is an AT game. Moreover, since only player 2's action sets have been restricted, we have $v_{*s} \geqslant \bar{v}_s$ for all $s \in S_*$. Also $|S_*| < |S|$ and therefore, by an induction argument on the number of states, we can assume player 1 to have a stationary optimal strategy $x^1_*$ for the game $G_*$.

Now, consider the strategy $\tilde{x}^1 \in \overline{X}^1$ for player 1 which prescribes to play $\bar{x}^1$ in $\overline{E}$ and $\bar{a}^1_s$ in each $s \in E_1 \cup \ldots \cup E_{n-1}$ and $x^1_*$ in $S_*$. We will now show that $\tilde{x}^1$ is optimal for player 1 (in $\overline{G}$). Take a stationary best reply $\bar{x}^2 \in \overline{X}^2$ for player 2 against $\tilde{x}^1$. Suppose $W$ is an arbitrary ergodic set for $(\tilde{x}^1, \bar{x}^2)$. Then, by the definition of $\tilde{x}^1$, either $W \subset \overline{E}$ or $W \subset S_*$. Notice that, by the definition of $\tilde{x}^1$, if $W \subset \overline{E}$ then $\gamma_s(\tilde{x}^1, \bar{x}^2) = \gamma_s(\bar{x}^1, \bar{x}^2) \geqslant \bar{v}_s$ for all $s \in W$; while if $W \subset S_*$ then by Lemma 3, the strategy $\bar{x}^2_s$ can only use actions in $\overline{A}^2_{*s}$ for all $s \in W$, hence $\gamma_s(\tilde{x}^1, \bar{x}^2) \geqslant v_{*s} \geqslant \bar{v}_s$. So in both cases, $\gamma_s(\tilde{x}^1, \bar{x}^2) \geqslant \bar{v}_s$ for all $s \in W$. Thus, $\gamma_s(\tilde{x}^1, \bar{x}^2) \geqslant \bar{v}_s$ for all recurrent states $s$. As $\tilde{x}^1 \in \overline{X}^1$ we have by Lemma 5 that $\bar{v} \leqslant P(\tilde{x}^1, \bar{x}^2)\bar{v}$. Hence, Lemma 2 (with $\phi = \bar{v}$) yields $\gamma(\tilde{x}^1, \bar{x}^2) \geqslant \bar{v}$. As $\bar{x}^2$ was a best reply to $\tilde{x}^1$, the strategy $\tilde{x}^1$ is optimal (in $\overline{G}$) indeed. □

**Example**




In this game representation the entries in the upper-left corners are the payoffs to player 1 who chooses rows, while the entries in the lower right corners are the transition probability vectors. This game is a zero-sum AT game where the transition probabilities for the respective states can be decomposed as:

$$p_1^1(1) = p_1^2(1) = (1,0,0), \quad \lambda_1^1 = \lambda_1^2 = \frac{1}{2};$$
$$p_2^1(1) = \left(\frac{1}{2},0,\frac{1}{2}\right), \quad p_2^1(2) = \left(\frac{1}{3},0,\frac{2}{3}\right), \quad p_2^2(1) = (0,1,0), \quad p_2^2(2) = (1,0,0), \quad \lambda_2^1 = \lambda_2^2 = \frac{1}{2};$$
$$p_3^1(1) = (0,0,1), \quad p_3^1(2) = (0,0,1), \quad p_3^2(1) = (0,0,1), \quad p_3^2(2) = (0,1,0), \quad \lambda_3^1 = \lambda_3^2 = \frac{1}{2}.$$

For this game we find that $v = \bar{v} = (3,1,-1)$, where the game $\overline{G}$ is the game derived by restricting player 1 to use only action 1 in state 2, while player 2 should only use action 1 in any state. The set $\overline{E}$ consists of state 1, the set $E_1$ consists of state 2, and the set $S_*$ is the singleton state 3. The stationary optimal strategy for player 1 is given by $((1),(1,0),(1,0))$.

**Remark 8.** We wish to remark that the existence of stationary $\varepsilon$-optimal strategies and Markov 0-optimal strategies for zero-sum AT games follows from Flesch et al. (1998) (see theorem 1 and the first concluding remark there), but Theorem 7 is stronger, as it concerns stationary 0-optimality.

The AT structure of the transitions implies that, for each state $s$ and actions $a_s^1, b_s^1 \in A_s^1$ the following holds: if

$$\sum_{t\in S} p_s(t|a_s^1, a_s^2)v_t > \sum_{t\in S} p_s(t|b_s^1, a_s^2)v_t$$

for some action $a_s^2 \in A_s^2$, then

$$\sum_{t\in S} p_s(t|a_s^1, b_s^2)v_t > \sum_{t\in S} p_s(t|b_s^1, b_s^2)v_t$$

for all $b_s^2 \in A_s^2$; and similarly for player 2. In other words, the AT transition structure induces a complete ordering on $A_s^1$ and $A_s^2$, with $\overline{A}_s^1$ and $\overline{A}_s^2$ as the sets of "best" actions. In fact the assumption of having such a complete ordering would already be sufficient for the previously mentioned alternative proof based on Flesch et al. (1998).

For the zero-sum case we have seen that player 1 has a strategy which guarantees that he receives at least the value. In other words, he has a strategy which guarantees that player 2 cannot get any reward better than the value. For the $n$-player case we obtain the following result along similar lines.

**Lemma 9.** *For each player $i$ there exists a joint strategy $\sigma^{-i}$ such that $\gamma_s^i(\pi^i, \sigma^{-i}) \leqslant v_s^i$ for each initial state $s$.*

## 4. Two-player absorbing AT games

An absorbing game is a stochastic game in which all states but one are absorbing, i.e. once play reaches an absorbing state, it will stay there forever. Therefore, an absorbing state corresponds to a repeated game. Clearly there are equilibria for each absorbing state and, by taking one for each of them, we can assume, without loss of generality, that the players have only one action in each absorbing state.

Suppose that the initial state is state 1 and it is the non-absorbing one. By the structure of the game, strategies are completely determined by giving the choices for state 1.

**Theorem 10.** *In any two-player absorbing AT stochastic game stationary $\varepsilon$-equilibria exist for all $\varepsilon > 0$.*

**Proof.** Let $G$ be a two-player absorbing AT stochastic game.

Suppose first that $\lambda_1^1, \lambda_1^2 \in (0,1)$. We partition the action sets of the players into an absorbing and a non-absorbing part by defining

$$A_1^{1*} = \{a_1^1 \in A_1^1 | p_1^1(1|a_1^1) < 1\}$$
$$A_1^{1\diamond} = \{a_1^1 \in A_1^1 | p_1^1(1|a_1^1) = 1\}.$$

Let $A_1^{2*}$ and $A_1^{2\diamond}$ be defined analogously for the action set $A_s^2$ of player 2. We distinguish two cases.

*Case 1*: $[A_1^{1\diamond} = \phi$ or $A_1^{2\diamond} = \phi]$. In this case all action combinations are absorbing. Therefore, if $\beta_1, \beta_2, \ldots$ is a sequence of discount factors converging to 1, and if $(x^1_{\beta_1}, x^2_{\beta_1}), (x^1_{\beta_2}, x^2_{\beta_2}), \ldots$ is a sequence of stationary $\beta_m$-discounted equilibria converging to $(x^1, x^2)$, then the latter pair is an average equilibrium since for any arbitrary stationary strategy $y^1$ for player 1 we would have:

$$\gamma^1(y^1, x^2) = \lim_{m\to\infty} \gamma^1_{\beta_m}(y^1, x^2_{\beta_m}) \leqslant \lim_{m\to\infty} \gamma^1_{\beta_m}(x^1_{\beta_m}, x^2_{\beta_m}) = \gamma^1(x^1, x^2)$$

while a similar statement applies for strategies of player 2. For the equality signs we refer to Lemma 4(a) in Vrieze and Thuijsman (1989). (By taking subsequences a limit point $(x^1, x^2)$ of a converging sequence of stationary $\beta_n$-discounted equilibria can always be assumed to exist.)

*Case 2*: $[A_1^{1\diamond} \neq \phi$ and $A_1^{2\diamond} \neq \phi]$. Observe that game entries in $A_1^{1\diamond} \times A_1^{2\diamond}$ are non-absorbing, while all other game entries are absorbing by the action of at least one player. Let $(x_1^{1\diamond}, x_1^{2\diamond})$ be an equilibrium in the one-shot game on $A_1^{1\diamond} \times A_1^{2\diamond}$. Note that for all actions $a_1^1, b_1^1 \in A_1^{1\diamond}$ and any action $a_1^2 \in A_1^{2*}$ we have

$$p_1(a_1^1, a_1^2) = p_1(b_1^1, a_1^2).$$

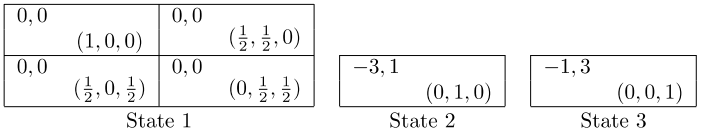
Hence, if $x_1^2$ only uses actions from $A_1^{2*}$ then for any $\eta \in [0,1]$

$$\gamma^1(x_1^{1\diamond}, \eta x_1^{2\diamond} + (1-\eta)x_1^2) \geqslant \gamma^1(x_1^1, \eta x_1^{2\diamond} + (1-\eta)x_1^2)$$

for all $x_1^1$ which only uses actions from $A_1^{1\diamond}$. So, against any stationary strategy of player 2 which only uses $x_1^{2\diamond}$ and actions from $A_1^{2*}$, player 1 cannot do any better in $A_1^{1\diamond}$ than to use $x_1^{1\diamond}$. Obviously, a similar property holds with exchanged roles of the players. Therefore, we may restrict the action spaces of the players and define a related absorbing AT game $G^\diamond$ in which the action set for player 1 is $\{x_1^{1\diamond}\} \cup A_1^{1*}$ and that for player 2 is $\{x_1^{2\diamond}\} \cup A_1^{2*}$, and where the payoffs and transitions are corresponding straightforwardly to the structure in the original game $G$. Notice that $G^\diamond$ has only one non-absorbing entry. Suppose that the payoffs of this entry are $(\alpha^1, \alpha^2)$. Then by subtracting $\alpha^1$ from all payoffs for player 1, and $\alpha^2$ from all payoffs for player 2, we obtain an absorbing game where the payoffs in the only non-absorbing entry are 0, while strategically nothing changes. For all joint stationary strategies the average rewards in this game are the same as those in the related game with all state 1 payoffs equal to 0. Then the game is a recursive repeated game with absorbing states, for which it is shown in Flesch et al. (1996) that stationary $\varepsilon$-equilibria exist. In a natural way, this $\varepsilon$-equilibrium in $G^\diamond$ induces a stationary $\varepsilon$-equilibrium in the original game $G$.

Suppose now that $\lambda_1^1 = 1$ and $\lambda_1^2 = 0$ (if $\lambda_1^1 = 0$ and $\lambda_1^2 = 1$ then the proof is similar). Since player 1 fully controls the transitions, we may redefine $p_1^2$ by $p_1^2(1|a_1^2) = 1$ for all actions $a_1^2 \in A_1^2$. Then $A_1^{2\diamond} = A_1^2$, and the same line of proof as above remains applicable. □

**Example**




We show that there are no stationary 0-equilibria for initial state 1 in this game, which illustrates that the above theorem is sharp. First notice that this is a stochastic game with AT structure, in which for state 1:

$$p_1^1(1) = p_1^2(1) = (1,0,0), \quad p_1^1(2) = (0,0,1), \quad p_1^2(2) = (0,1,0), \quad \lambda_1^1 = \lambda_1^2 = \frac{1}{2}.$$

Suppose by way of contradiction that $(x^1, x^2)$ is a stationary 0-equilibrium. If $x_1^2 = (1,0)$, then $x_1^1 = (1,0)$ is player 1's unique best reply; but then $x^2$ is no best reply to $x^1$ since by playing Right player 2 could get 1 instead of 0. On the other hand, if $x_1^2 \neq (1,0)$ then $x_1^1 = (0,1)$ is player 1's unique best reply; but then $x^2$ is no best reply to $x^1$ since by playing Left exclusively player 2 could get 3.

Note however, that $((0,1),(1-\varepsilon,\varepsilon))$ represents a stationary $\varepsilon$-equilibrium for small $\varepsilon > 0$.
This game can also be represented as

| | |
|---|---|
| $0,0$ | $-3,1$ $\frac{1}{2}*$ |
| $-1,3$ $\frac{1}{2}*$ | $-2,2$ $1*$ |

Now we use a slightly different notation for the transitions. By choosing entry Top-Left we remain with probability 1 in the non-absorbing initial state 1 with direct payoff 0; by choosing entry Top-Right play moves with probability $\frac{1}{2}$ to a $1\times 1$ absorbing state in which the payoffs are $-3$ and 1 for players 1 and 2 respectively and with probability $\frac{1}{2}$ play remains in the initial state with direct payoff 0; by choosing entry Bottom-Left play moves with probability $\frac{1}{2}$ to a $1\times 1$ absorbing state in which the payoffs are $-1$ and 3 for players 1 and 2 respectively and with probability $\frac{1}{2}$ play remains in the initial state with direct payoff 0; by choosing entry Bottom-Right play moves with probability 1 to a $1\times 1$ absorbing state in which the payoffs are $-2$ and 2 for players 1 and 2 respectively, which is equivalent to moving to either state 2 or state 3 with probability $\frac{1}{2}$ in the original game. So the asterisks correspond to the absorbing entries.

**Example.** We now consider the following AT-game:

Near

| | Left | Right |
|---|---|---|
| Top | $0,0,0$ | $0,1,3$ $\frac{1}{3}*$ |
| Bottom | $1,3,0$ $\frac{1}{3}*$ | $\frac{1}{2},2,\frac{3}{2}$ $\frac{2}{3}*$ |

Far

| | |
|---|---|
| $3,0,1$ $\frac{1}{3}*$ | $\frac{3}{2},\frac{1}{2},2$ $\frac{2}{3}*$ |
| $2,\frac{3}{2},\frac{1}{2}$ $\frac{2}{3}*$ | $\frac{4}{3},\frac{4}{3},\frac{4}{3}$ $1*$ |

Note that this game is similar to the game in Flesch et al. (1997). This is a three-player absorbing AT game, where an asterisk in any particular entry denotes a transition to an absorbing state with the same payoff as in this particular entry. There is only one entry for which play will remain in the non-trivial initial state. One should picture the game as a $2\times 2\times 2$ cube, where the layers belonging to the actions of player 3 (Near and Far) are represented separately. As before, player 1 chooses Top or Bottom and player 2 chooses Left or Right.

Note that this is an AT game, in which for state 1 (the non-absorbing state) we have $\lambda_1^1 = \lambda_1^2 = \lambda_1^3 = \frac{1}{3}$, and regarding $p_1^1, p_1^2, p_1^3$: each of the actions Top, Left, Near leads to state 1, while actions Bottom, Right, Far lead to absorption with payoffs $(1,3,0)$, $(0,1,3)$, and $(3,0,1)$ respectively. Note that all entries but entry (Top, Left, Near) are absorbing, so the play absorbs as soon as one of the players chooses his second action. Besides, the payoff and the transition structure is cyclically symmetric, namely it holds for any entry $(i_1,i_2,i_3)\in\{1,2\}^3$ that

$$r^1(i_1,i_2,i_3) = r^2(i_3,i_1,i_2) = r^3(i_2,i_3,i_1)$$
$$p(i_1,i_2,i_3) = p(i_3,i_1,i_2) = p(i_2,i_3,i_1).$$

Similarly to the game in Flesch et al. (1997), an example of a Markov equilibrium for this game is $(f,g,h)$, where $f$ is defined by: at stages $1,7,13,19,\ldots$ play Bottom with probability 1, at stages $2,8,14,20,\ldots$ play Bottom with probability $\frac{3}{4}$, and at all other stages play Top with probability 1. Similarly, $g$ is defined by: at stages $3,9,15,21,\ldots$ play Right with probability 1, at stages $4,10,16,22,\ldots$ play Right with probability $\frac{3}{4}$, and at all other stages play Left with probability 1. Likewise, $h$ is defined by: at stages $5,11,17,23,\ldots$ play Far with probability 1, at stages $6,12,18,24,\ldots$ play Right with probability $\frac{3}{4}$, and at all other stages play Near with probability 1. The average reward corresponding to this equilibrium is $(1,2,1)$.

However, there are no stationary $\varepsilon$-equilibria for small $\varepsilon > 0$ in this game. First we will argue that there are no stationary 0-equilibria. Suppose by way of contradiction that $(x,y,z)$ is a stationary equilibrium, where

$x, y, z$ are the probabilities on actions Bottom, Right and Far respectively. First we prove that $0 < x, y, z < 1$. If $x = 0$ then, because of a best reply argument, $y > 0$ (and we would have $y = 1$ if $z > 0$). However, if $y > 0$, then $z = 0$, which contradicts $x = 0$. On the other hand $x = 1$ would imply $y = 0$, hence $z = 1$, which contradicts $x = 1$. So $0 < x < 1$, and by symmetry we also have $0 < y, z < 1$. Then

$$\gamma^1(0,y,z) = \frac{3 \cdot \frac{1}{3}(1-y)z + \frac{3}{2} \cdot \frac{2}{3}yz}{\frac{1}{3}y(1-z) + \frac{1}{3}(1-y)z + \frac{2}{3}yz} \quad \text{and}$$

$$\gamma^1(1,y,z) = \frac{1 \cdot \frac{1}{3}(1-y)(1-z) + \frac{1}{2} \cdot \frac{2}{3}y(1-z) + 2 \cdot \frac{2}{3}(1-y)z + \frac{4}{3} \cdot yz}{\frac{1}{3}(1-y)(1-z) + \frac{2}{3}y(1-z) + \frac{2}{3}(1-y)z + yz}.$$

Since $0 < x < 1$ we must have $\gamma^1(0,y,z) = \gamma^1(1,y,z)$, from which we find

$$\frac{3z}{y+z} = \frac{1+3z}{1+y+z},$$

which implies that $y = 2z > z$. By symmetry $z > x$ and $x > y$. Hence $y > z > x > y$ and therefore there are no stationary 0-equilibria. The proof that there are no stationary $\varepsilon$-equilibria is similar to the Proof of Lemma 3.2 in Flesch et al. (1997) for the related game.

## 5. *n*-Player AT games

In this section we shall establish the existence of history-dependent 0-equilibria for all $n$-player AT games. It follows from Eq. (9) and from the additive transition structure of the game that

$$v_s^i = \min_{x_s^{-i} \in X_s^{-i}} \max_{x_s^i \in X_s^i} \sum_{t \in S} p_s(t|x_s^i, x_s^{-i}) v_t^i = \lambda_s^i \cdot \max_{a_s^i \in A_s^i} \sum_{t \in S} p_s^i(t|a_s^i) v_t^i + \sum_{j \neq i} \lambda_s^j \cdot \min_{a_s^j \in A_s^j} \sum_{t \in S} p_s^j(t|a_s^j) v_t^i.$$

We now introduce some notations similar to the 2-player zero-sum case. If $\lambda_s^i > 0$ then let $\overline{A}_s^i$ be the set of actions of player $i$ in state $s$ that maximize the expression $\sum_{t \in S} p_s^i(t|\cdot) v_t^i$; while if $\lambda_s^i = 0$ (meaning that player $i$ has no influence on the transitions in state $s$) then we define $\overline{A}_s^i = A_s^i$. Consequently, $\overline{A}_s^i$ is the set of those actions $a_s^i$ for which

$$v_s^i \leqslant \sum_{t \in S} p_s(t|a_s^i, x_s^{-i}) v_t^i \quad \forall x_s^{-i} \in X_s^{-i}.$$

The main idea we shall use in the analysis is that of the restricted game $\overline{G}$ derived from $G$ by restricting each player $i$ to actions in $\overline{A}_s^i$ in all states $s$. Then $\overline{G}$ is an AT stochastic game as well. Obviously, in $\overline{G}$ players can only randomize on the remaining actions. Thus we define $\overline{X}_s^i$ as the set of mixed actions on $\overline{A}_s^i$. In a natural way $\overline{X}_s^i$ can be seen as a subset of $X_s^i$. We shall denote the minmax value of $\overline{G}$ by $\bar{v}$.

**Lemma 11.** *For each player $i$ and each state $s$, let $z_s^i$ be a completely mixed action for player $i$ on $\overline{A}_s^i$, i.e. $z_s^i(a_s^i) > 0$ for all $a_s^i \in \overline{A}_s^i$. Suppose $E$ is an ergodic set with respect to the joint stationary strategy $z$. Then for any player $i$*

$$\bar{v}_s^i \geqslant v_s^i$$

*for all $s \in E$.*

**Proof.** Take an arbitrary player $i$. Notice that as $v^i \leqslant P(z)v^i$, Lemma 2 (with $\phi = v^i$ for any player $i$) yields $v_s^i = v_t^i$ for all $s, t \in E$. Note that, in any state $s \in E$, for any joint action $a_s^{-i} \in \overline{A}_s^{-i}$, if player $i$ chooses any $a_s^i \in \overline{A}_s^i$ then play remains in $E$, so

$$\sum_{t \in S} p_s(t|a_s^i, a_s^{-i}) v_t^i = v_s^i.$$

Hence, by the definition of $\overline{A}_s^i$, for any $s \in E$, we have

$$\sum_{t \in S} p_s(t|b_s^i, a_s^{-i}) v_t^i < v_s^i$$

for all actions $b_s^i \in A_s^i - \overline{A}_s^i$, so such actions $b_s^i$ outside $\overline{A}_s^i$ lead to a decrease in the expectation of $v^i$ after transition. Thus

$$d_2 := \min_{a_s^i \in A_s^i - \overline{A}_s^i,\ a_s^{-i} \in \overline{A}_s^{-i}, s \in E} \left[ v_s^i - \sum_{t \in S} p_s(t|a_s^i, a_s^{-i}) v_t^i \right] > 0.$$

Suppose, by way of contradiction, that $\bar{v}_s^i < v_s^i$, for some state $s \in E$. Let $d_1 := v_s^i - \bar{v}_s^i$. Now let $\bar{\pi}^{-i}$ be a joint strategy in $\overline{G}$ for the opponents of $i$, against which player $i$ can get at most $\bar{v}^i + \frac{d_1}{2}$ in $\overline{G}$. Similarly $\sigma^{-i}$ is a joint strategy in $G$ for the opponents of $i$, against which player $i$ can get at most $v^i + \frac{d_2}{2}$ in $G$. Consider the strategy $\pi^{-i}$ for the opponents of player $i$ in $G$ which prescribes to play as follows: play $\bar{\pi}^{-i}$ as long as player $i$ chooses actions in the sets $\overline{A}_t^i$, $t \in S$, and as soon as player $i$ takes an action outside, start playing $\sigma^{-i}$.

Take an arbitrary $\varepsilon$-best reply $\pi^i$ to $\pi^{-i}$ for player $i$ in $G$ for initial state $s$. With respect to $\pi$ and initial state $s$ we obtain a contradiction with the definition of $v_s^i$ similarly to the last part of the Proof of Lemma 5. □

**Theorem 12.** *There exists a* 0*-equilibrium in every n-player AT stochastic game.*

**Proof.** Let $z_s^i$ be a completely mixed action for player $i$ on $\overline{A}_s^i$ and let $W_1, \ldots, W_K$ be the ergodic sets with respect to the joint stationary strategy $z$. Notice that as $v \leqslant P(z)v$, Lemma 2 (with $\phi = v^i$ for any player $i$) yields $v_s = v_t =: \alpha_k$ for all $s, t \in W_k$, and for all $k$.

Let $\bar{x}_\beta$ be a stationary $\beta$-discounted equilibrium in $\overline{G}$, for all $\beta \in (0,1)$. By the finiteness of the state and action spaces there is a set $D \subset (0,1)$ such that 1 is a limit point of $D$ and for all $\beta \in D$ the carrier of $\bar{x}_\beta$ is the same. Clearly, for any $k$, there must be an ergodic set $E_k$ for $\bar{x}_\beta$ so that $E_k \subset W_k$. Let $\varepsilon > 0$. Then for $\beta \in D$ close to 1 it follows from Lemma 1, from the fact that $\bar{x}_\beta$ is a $\beta$-discounted equilibrium, from equality (8), from Lemma 11 (as $E_k \subset W_k$) and the previous observation that the minmax value is a constant on $E_k \subset W_k$:

$$\gamma_s^i(\bar{x}_\beta) \geqslant \min_{t \in E_k} \gamma_{\beta t}^i(\bar{x}_\beta) \geqslant \min_{t \in E_k} \bar{v}_{\beta t}^i \geqslant \min_{t \in E_k} \bar{v}_t^i - \varepsilon \geqslant \min_{t \in E_k} v_t^i - \varepsilon = v_s^i - \varepsilon \tag{15}$$

for any state $s \in E_k$ and for any player $i$. This means that, for $\beta \in D$ sufficiently close to 1, the joint stationary strategy $\bar{x}_\beta$ is individually rational on $E_k$ (up to $\varepsilon$).

Let $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$ be a monotone decreasing sequence of reals converging to 0 and let $\beta_1, \beta_2, \beta_3, \ldots \in D$ be a monotone increasing sequence of discount factors converging to 1, which by (15) can be taken such that for each $m$ we have $\gamma_s^i(\bar{x}_{\beta_m}) \geqslant v_s^i - \varepsilon_m$ for all $s \in E_k$, for each ergodic set $E_k$ and for each player $i$. Then, by Lemma 4 (as the minmax value is constant on $E_k$) there exists a pure joint strategy $\pi$ such that $\pi$ only uses actions that have positive weight for $\bar{x}_{\beta_m}$ and such that $\gamma_s^i(\pi) \geqslant v_s^i$, as well as $\gamma_s^i(\pi|h) \geqslant v_s^i$, for all $s \in E_k$, for all $k$, for any history $h$, and for each player $i$.

By the definition of $z$, one can select a joint pure stationary strategy $\bar{a} \in \overline{A}$ such that, with respect to $\bar{a}$, play eventually reaches $E_1 \cup \ldots \cup E_K$ from any initial state. Since $\bar{a} \in \overline{A}$ we have that $P(\bar{a})v^i \geqslant v^i$ for each player $i$. Let $\pi^*$ be the joint pure strategy defined by playing $\bar{a}$ in any state outside $E_1 \cup \ldots \cup E_K$ and by switching to $\pi$ as soon as one of the sets $E_1, \ldots, E_K$ is entered. By construction $\gamma_s^i(\pi^*) \geqslant v_s^i$ (and $\gamma_s^i(\pi^*|h) \geqslant v_s^i$) for any initial state $s \in S$ and for each player $i$. Now $\pi^*$ is a pure strategy which tells each player exactly which action to play at any state and stage. Therefore any deviation by a player can be detected immediately. Because of Lemma 9 any deviation by player $i$ can be punished by his opponents jointly playing $\sigma^{-i}$. Let $\pi^*(\sigma)$ denote the joint strategy defined by playing $\pi^*$ as long as no one deviates, and by switching to $\sigma^{-i}$ as soon as any player $i$ deviates. (In case more than one player would deviate at the same time, take the smallest index $i$. For verification of the equilibrium property only unilateral deviations need to be considered and simultaneous deviations play no role at all.)

Notice that $\gamma_s^i(\pi^*(\sigma)) = \gamma_s^i(\pi^*) \geqslant v_s^i$ for all $i$ and $s$ (and similarly $\gamma_s^i(\pi^*(\sigma)|h) = \gamma_s^i(\pi^*|h) \geqslant v_s^i$ for any history $h$ that may occur under $\pi^*$). Now $\pi^*(\sigma)$ is a 0-equilibrium because of the following arguments. Suppose that the first deviation occurs by player $i$ deviating in state $s$ at stage $k$ after history $h_k$ by playing action $d_s^i$, where all players were supposed to play $a_s \in \overline{A}_s$ according to $\pi^*$. Then, player $i$'s opponents will start playing their punishment strategies at stage $k+1$. Therefore, from this very stage player $i$'s reward will be kept down to $v_t^i$, where $t$ is the state at stage $k+1$. Thus, player $i$ will receive at most

$$\sum_{t\in S} p_s(t|d_s^i, a_s^{-i})v_t^i \leqslant \sum_{t\in S} p_s(t|a_s^i, a_s^{-i})v_t^i \leqslant \sum_{t\in S} p_s(t|a_s^i, a_s^{-i}) \cdot \gamma_t^i(\pi^*(\sigma)|h_k \oplus (s, a_s)) = \gamma_s^i(\pi^*(\sigma)|h_k).$$

Here $h_k \oplus (s, a_s)$ denotes the concatenation of the history $h_k$, state $s$ and actions $a_s$. The final inequality implies that the deviation reward is at most the continuation reward for the original strategies. □

**Remark 13.** Notice that the AT structure of the transitions was only used to achieve that, for each player $i$, state $s$ and actions $a_s^i, b_s^i \in A_s^i$ of player $i$ in state $s$ the following holds: if

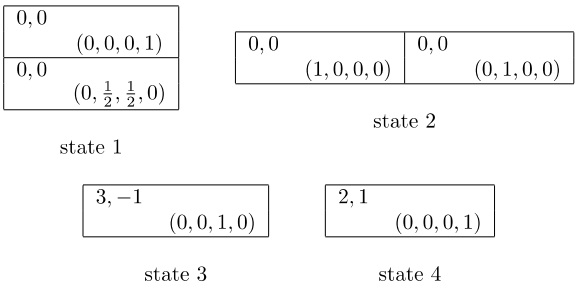
$$\sum_{t\in S} p_s(t|a_s^i, a_s^{-i})v_t^i > \sum_{t\in S} p_s(t|b_s^i, a_s^{-i})v_t^i$$

for some joint action $a_s^{-i} \in A_s^{-i}$, then

$$\sum_{t\in S} p_s(t|a_s^i, b_s^{-i})v_t^i > \sum_{t\in S} p_s(t|b_s^i, b_s^{-i})v_t^i$$

for all $b_s^{-i} \in A_s^{-i}$. In other words, the AT transition structure induces a complete ordering on $A_s^i$, with $\overline{A}_s^i$ as the set of "best" actions.

We now consider the following two-player AT game taken from Flesch et al. (1996):




This game is a two-player perfect information game, for which there is no stationary $\varepsilon$-equilibrium for small $\varepsilon > 0$. One can prove this as follows. Suppose player 2 puts positive weight on Left in state 2, then player 1's only stationary $\varepsilon$-best replies are those that put weight at most $\frac{\varepsilon}{2-\varepsilon}$ on Top in state 1; against any of these strategies, player 2's only stationary $\varepsilon$-best replies are those that put weight 0 on Left in state 2. So there is no stationary $\varepsilon$-equilibrium where player 2 puts positive weight on Left in state 2. But neither is there a stationary $\varepsilon$-equilibrium where player 2 puts weight 0 on Left in state 2, since then player 1 should put at most weight $2\varepsilon$ on Bottom in state 1, which would in turn contradict player 2's putting weight 0 on Left.

Notice that we obtain an equilibrium by letting the players play as follows: player 1 plays Top in state 1 as long as player 2 has never played Left and plays Bottom otherwise; player 2 plays Right in state 2. Another equilibrium is: player 1 plays Top in state 1; player 2 plays Left in state 2 as long as player 1 has never played Bottom and plays Right otherwise.

We remark that in Thuijsman and Raghavan (1997) existence of average 0-equilibria is shown for arbitrary $n$-player games with perfect information.

## Acknowledgements

We would like to thank two anonymous referees for their constructive remarks, by which the presentation has improved a lot.

## References

Blackwell, D., Ferguson, T.S., 1968. The big match. Annals of Mathematical Statistics 39, 159–163.
Doob, J.L., 1953. Stochastic Processes. Wiley, New York.
Dutta, P.K., 1995. A folk theorem for stochastic games. Journal of Economic Theory 66, 1–32.

Evangelista, F., Raghavan, T.E.S., Vrieze, O.J., 1996. Repeated ARAT games. In: Ferguson, T.S., Shapley, L.S., MacQueen, J.B. (Eds.), Statistics, Probability and Game Theory, pp. 13–28.

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 34, 503–515.

Fink, A.M., 1964. Equilibrium in a stochastic *n*-person game. Journal of Science of Hiroshima University 28, 89–93, Series A-I.

Flesch, J., Thuijsman, F., Vrieze, O.J., 1996. Recursive repeated games with absorbing states. Mathematics of Operations Research 21, 1016–1022.

Flesch, J., Thuijsman, F., Vrieze, O.J., 1997. Cyclic Markov equilibria in stochastic games. International Journal of Game Theory 26, 303–314.

Flesch, J., Thuijsman, F., Vrieze, O.J., 1998. Simplifying optimal strategies in stochastic games. SIAM Journal on Control and Optimization 36 (4), 1331–1347.

Gillette, D., 1957. Stochastic games with zero stop probabilities. In: Dresher, M., Tucker, A.W., Wolfe, P. (Eds.), Contributions to the Theory of Games III, Annals of Mathematical Studies, 39. Princeton University Press, pp. 179–187.

Hordijk, A., Vrieze, O.J., Wanrooij, G.L., 1983. Semi-Markov strategies in stochastic games. International Journal of Game Theory 12, 81–89.

Liggett, T.M., Lippmann, S.A., 1969. Stochastic games with perfect information and time average payoff. SIAM Review 11, 604–607.

Mertens, J.F., Neyman, A., 1981. Stochastic games. International Journal of Game Theory 10, 53–66.

Neyman, A., 1986. Lecture Notes of a Course in "Stochastic Games". The Hebrew University, Jerusalem, Israel.

Raghavan, T.E.S., Tijs, S.H., Vrieze, O.J., 1985. On stochastic games with additive reward and transition structure. Journal of Optimization Theory and Applications 47, 451–464.

Sorin, S., 1986. Asymptotic properties of a non-zerosum game. International Journal of Game Theory 15, 101–107.

Takahashi, M., 1964. Equilibrium points of stochastic noncooperative n-person games. Journal of Science of Hiroshima University 28, 95–99, Series A-I.

Thuijsman, F., Raghavan, T.E.S., 1997. Perfect information stochastic games and related classes. International Journal of Game Theory 26, 403–408.

Thuijsman, F., Vrieze, O.J., 1991. Easy initial states in stochastic games. In: Raghavan, T.E.S., Ferguson, T.S., Vrieze, O.J., Parthasarathy, T. (Eds.), Stochastic Games and Related Topics. Kluwer, Dordrecht, pp. 85–100.

Tijs, S.H., Vrieze, O.J., 1986. On the existence of easy initial states for undiscounted stochastic games. Mathematics of Operations Research 11, 506–513.

Vieille, N., 2000a. 2-person stochastic games I: A reduction. Israel Journal of Mathematics 119, 55–91.

Vieille, N., 2000b. 2-person stochastic games II: The case of recursive games. Israel Journal of Mathematics 119, 93–126.

Vrieze, O.J., Thuijsman, F., 1989. On equilibria in repeated games with absorbing states. International Journal of Game Theory 18, 293–310.