Stochastic Games on a Product State Space

János Flesch
Department of Quantitative Economics, Maastricht University, 6200 MD Maastricht, The Netherlands, J.Flesch@ke.unimaas.nl

Gijs Schoenmakers, Koos Vrieze
Department of Mathematics, Maastricht University, 6200 MD Maastricht, The Netherlands
GM.Schoenmakers@micc.unimaas.nl, kvrieze@ictvrieze.nl

We examine so-called product-games with an aperiodic transition structure, with respect to the average reward, for which we present an approach based on communicating states. For the general $n$-player case, we establish the existence of $0$-optimal strategies. In addition, for the special case of two-player zero-sum games of this type, we show that both players have stationary 0-optimal strategies.

Key words: noncooperative games; stochastic games; Markov decision problems; equilibria
MSC2000 subject classification: Primary: 91A15
ORMS subject classification: Primary: stochastic games
History: Received March 16, 2007; revised May 7, 2007 and November 10, 2007.

1. Introduction.

Stochastic games and product-games. An $n$-player stochastic game is given by (1) a set of players $N = \{1, \ldots, n\}$, (2) a nonempty and finite set of states $S$, (3) for each state $s \in S$, a nonempty and finite set of actions $A_i^s$ for each player $i$, (4) for each state $s \in S$ and each joint action $a_s = \bigtimes_{i \in N} A_i^s$, a payoff $r_s(a_s) \in \mathbb{R}$ to each player $i$, (5) for each state $s \in S$ and each joint action $a_s = \bigtimes_{i \in N} A_i^s$, a transition probability distribution $p_{s|a_s} = (p_{s|a_s}(i))_{i \in S}$.

The game is to be played at stages in $\mathbb{N}$ in the following way. Play starts at stage 1 in an initial state, say in state $s_1 \in S$. In $s_i$, each player $i \in N$ has to choose an action $a_i^s$ from his action set $A_i^s$. These choices have to be made independently. The chosen joint action $a_s = (a_1^s, \ldots, a_n^s)$ induces an immediate payoff $r_s(a_s)$ to each player $i$. Next, play moves to a new state according to the transition probability distribution $p_{s|a_s}$, say to state $s_2 \in S$. At stage 2, a new action $a_2^s \in A_2^s$ has to be chosen by each player $i$ in state $s_2$. Then, given action combination $a_s = (a_1^s, \ldots, a_n^s)$, player $i$ receives payoff $r_s(a_s)$ and the play moves to some state $s_3$ according to the transition probability distribution $p_{s_3|a_2^s}$, and so on. We assume complete information (i.e., the players know all the data of the stochastic game), full monitoring (i.e., the players observe the present state and the actions chosen by all the players), and perfect recall (i.e., the players remember all previous states and actions).

A Markov transition structure $\Gamma^i$ for player $i \in N$ is given by (1) a nonempty and finite state space $S_i^i$; (2) a nonempty and finite action set $A_i^s$ for each state $s^i \in S_i^i$; (3) a transition probability distribution $p_{s^i|a_i^s}$ over the state space $S^i$ for each state $s^i \in S_i^i$ and for each action $a_i^s \in A_i^s$. Note that, if we also assigned a payoff in every state to every action, then we would obtain the well-known model of Markov decision problems for player $i$.

We will now consider a special type of $n$-player stochastic games, called product-games, in which the transition structure is derived by taking the product of $n$ Markov transition structures. A product-game $G^i$, associated to the Markov transition structures $\Gamma^i, \Gamma^2, \ldots, \Gamma^n$, is an $n$-player stochastic game for which (1) the set of players is $N = \{1, \ldots, n\}$; (2) the state space is $S = S^1 \times \cdots \times S^n$; (3) the action set for each player $i \in N$ in each state $s = (s^1, \ldots, s^n) \in S$ is $A_i^s = A_i^{s^i}$; (4) the transition probability distribution $p_{s|a_s}$, for each state $s = (s^1, \ldots, s^n) \in S$ and for each joint action $a_s = (a_1^s, \ldots, a_n^s) \in \bigtimes_{i \in N} A_i^s$, is

$$p_{s|a_s}(\bar{s}) = \prod_{i \in N} p_{s_i|a_{s_i}}(\bar{s_i})$$

for state $\bar{s} = (s^1, \ldots, s^n) \in S$. Note that there is no condition imposed on the payoff structure.

Observe that (1) the action space of player $i$ only depends on the $i$th coordinate of the state, (2) the $i$th coordinate of the transitions from any state $s$ only depend on the $i$th coordinate $s^i$ of the state and on the action $a_i^s$ chosen by player $i$; i.e. for any $\bar{s} \in S^i$ we have

$$p_{s|a_s}(S^1, \ldots, S^{i-1}, \bar{s}, S^{i+1}, \ldots, S^n) = p_{s_i|a_{s_i}}(\bar{s}).$$
Therefore, as far as the actions and the transitions are concerned, player $i$ can play on the $i$th coordinate of the game $G$ without the interference of the other players. As a consequence, play of the product-game $G$ can be viewed as simultaneous play of the $n$ Markov transition structures $\Gamma^1, \ldots, \Gamma^n$, which are linked by payoff functions $r^1, \ldots, r^n$ that may depend on all $n$ current states as well as on all $n$ actions chosen by the players.

Product-games have been introduced in Altman et al. [1], although in a somewhat different fashion. They only examine two-player games in which the sum of the payoffs is always equal to zero (zero-sum games), and drop the assumption of full monitoring by letting each player only observe his own coordinate of the present state and only the action chosen by himself. As a result, both players have to make choices without noticing anything about the other player’s behavior. They show that a linear programming formulation is sufficient to solve these games; i.e., to find the value and stationary optimal strategies (cf., the definitions below).

Note that the class of product-games, as defined in our paper, differs essentially from other known classes of $n$-player stochastic games. Stochastic games with a single controller; i.e., when one player controls the transitions, however, fall into the class of product-games. Indeed, a stochastic game that is controlled by player $i$ can be seen as a product-game in which $S^j$ is a singleton for all players $j \neq i$. Finally, we wish to mention the class of stochastic games with additive transitions (AT-games, cf. Flesch et al. [8]); i.e., when the transitions are additively decomposable into player-dependent components, in contrast with a product decomposition. Not surprisingly, the structure of product-games and AT-games differ essentially, and product-games require new ideas and an entirely different approach.

From now on, we will consequently use the upper-index for the player and the lower-index for the state. Whenever one of them is omitted, we will then mean a vector in the case of quantities and a product in the case of ideas and an entirely different approach.

Strategies. A mixed action $x^i_s$ for player $i$ in state $s \in S$ 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 mixed action is called completely mixed if it assigns a positive probability to each available action. A (history dependent) strategy $\pi^i$ for player $i$ is a decision rule that prescribes a mixed action $\pi^i(h) \in X^i_{s^h}$ in the present state $s$ depending on the past history $h$ of play (i.e., the sequence of all past states and all past actions chosen by the players). We use the notation $H^i$ for the set of strategies for player $i$. A strategy $\pi^i$ for player $i$ is called pure if $\pi^i$ prescribes, for every state and every possible history, one specific action to be played with probability 1. Given a strategy $\pi^i$ for player $i$ and a history $h$, the strategy $\pi^i$ conditional on $h$, denoted by $\pi^i[h]$, is the strategy which prescribes a mixed action $\pi^i[h](s^h)$ in any present state $s$ for any history $s^h$ as if $h$ had happened before $s^h$; i.e., $\pi^i[h](s^h) = \pi^i(h \oplus s^h)$, where $h \oplus s^h$ is the history consisting of $h$ concatenated by $s^h$. In fact, $\pi^i[h]$ is just the continuation strategy of $\pi^i$ after history $h$.

If the mixed actions prescribed by a strategy only depend on the present state, then the strategy is called stationary. Thus, the stationary strategy space for player $i$ is $X^i = \times_{s \in S} X^i_s$. We use the notation $x^i$ for stationary strategies for player $i$, while $x^i_s$ refers to the corresponding mixed action for player $i$ in state $s$. Note that the set of pure stationary strategies for player $i$ is simply $A^i = \times_{s \in S} A^i_s$.

A joint stationary strategy $x = (x^i)_{i \in N}$ induces a Markov-chain on the state space $S$ with transition matrix $P(x)$, where entry $(s, \tilde{s})$ of $P(x)$ gives the transition probability $p_{s\tilde{s}}(x)$ for moving from state $s$ to state $\tilde{s}$ when the joint mixed action $x^i_s$ 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, play will eventually return to that state with probability 1; otherwise the state is called transient. If play is in a recurrent state, then this state will be visited infinitely often with probability 1, while transient states can only be visited finitely many times, with probability 1. We can group the recurrent states into minimal closed sets, so-called ergodic sets. An ergodic set is a collection $F$ of recurrent states with the property that, when starting in any of the states in $F$, all states in $F$ will be visited infinitely often and the play will remain in $F$ forever with probability 1.

Let
\[
Q(x) := \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} P^m(x);
\]
the limit is known to exist (cf., Doob [4], Theorem 2.1, p. 175). Entry $(s, \tilde{s})$ of the stochastic matrix $Q(x)$, denoted by $q_{s\tilde{s}}(x)$, is the expected frequency of stages for which the process is in state $\tilde{s}$ when starting in $s$. The matrix $Q(x)$ has the well known properties (cf., Doob [4]) that
\[
Q(x) = Q(x)P(x) = P(x)Q(x) = Q^2(x).
\]
Rewards. For a joint strategy \( \pi = (\pi^i)_{i \in N} \) and initial state \( s \in S \), the sequences of payoffs are evaluated by the (expected) average reward, which is given for player \( i \) by

\[
\gamma_i(\pi) := \liminf_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} R^i_m \quad = \liminf_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}_s (R^i_m),
\]

where \( R^i_m \) is the random variable for the payoff for player \( i \) at stage \( m \), and where \( \mathbb{E}_s \) stands for expectation with respect to the initial state \( s \) and the joint strategy \( \pi \). We wish to remark that all our further results hold for the limsup as well.

With regard to a joint stationary strategy \( x = (x^i)_{i \in N} \), we obtain more explicit formulas for the average reward. Let \( r^i(x, s) \) denote the expected immediate payoff for player \( i \) in state \( s \) if the joint mixed action \( x \) is played. By definition, for the average reward of every player \( i \) we have

\[
\gamma_i(x) = Q(x) r^i(x); \quad (3)
\]

hence by (2) we also obtain

\[
\gamma_i(x) = P(x) \gamma_i(x), \quad (4)
\]

\[
\gamma_i(x) = Q(x) r^i(x) = Q(x) r^i(x) = Q(x) \gamma_i(x). \quad (5)
\]

Every player \( i \) has a pure stationary best reply against any fixed joint stationary strategy of his opponents (cf., Blackwell [2], Hordijk et al. [11]); i.e., for any \( x^i \in X^i \) there exists an \( x^i \in X^i \) such that \( \gamma_i(x^i, x^{-i}) \geq \gamma_i(x^i, x^{-i}) \) for all initial states \( s \in S \) and for all strategies \( \pi^i \in \Pi^i \).

For any player \( i \in N \) and initial state \( s \in S \), let

\[
v^i_s := \inf_{\pi^i \in \Pi^i} \sup_{\pi^{-i} \in \Pi^{-i}} \gamma_i(\pi^i, \pi^{-i}). \quad (6)
\]

Here \( v^i_s \) is called the min max level for player \( i \) in state \( s \). Intuitively, this is the highest reward that player \( i \) can defend against any strategies of the other players if the initial state is \( s \). It is thus also the lowest reward that the other players can inflict on player \( i \). Note that, against different joint strategies of players \(-i\), player \( i \) may have to use different strategies to defend his min max level. It is known that the min max level of any player \( i \) satisfies

\[
v^i_s = \min_{x^{-i} \in X^{-i}} \max_{x^{-i} \in X^{-i}} \sum_{t \in T} p_s(x^{-i}(t)) v^i_t, \quad (7)
\]

which is an easy consequence of the definition of \( v^i_s \) and Equality (4). Furthermore, by Thuijsman and Vrieze [23] (their proof is given for only two players but directly extends to the \( n \)-player case in combination with Neyman [14], who showed that the min max levels equal the limit of the discounted min max levels in \( n \)-player stochastic games), there always exists an initial state \( s \) in the set \( \{ t \in S \mid v^i_t = \min_{x^{-i} \in X^{-i}} v^i_s \} \) for which players \(-i\) have a joint stationary strategy \( x^{-i} \) such that \( \gamma_i(\pi^i, x^{-i}) \leq v^i_s \) for all strategies \( \pi^i \) and player \( i \). In other words, the infimum in expression (6) is attained for state \( s \) at stationary strategies.

Equilibria. A joint strategy \( \pi = (\pi^i)_{i \in N} \) is called a (Nash) \( \varepsilon \)-equilibrium for initial state \( s \in S \), for some \( \varepsilon \geq 0 \), if

\[
\gamma_i(\pi^i, \pi^{-i}) \leq \gamma_i(\pi) + \varepsilon \quad \forall \pi^i \in \Pi^i, \forall i \in N,
\]

which means that no player can gain more than \( \varepsilon \) by a unilateral deviation. If \( \pi \) is an \( \varepsilon \)-equilibrium for all initial states, then we call \( \pi \) an \( \varepsilon \)-equilibrium. It is clear from the definition of the min max level \( v^i \) that if \( \pi \) is an \( \varepsilon \)-equilibrium, then \( \gamma_i(\pi) \geq v^i - \varepsilon \) for each player \( i \) and each initial state \( s \in S \).

Regarding general stochastic games, the famous game called the big match, which was introduced by Gillette [10] and solved by Blackwell and Ferguson [3], and the game in Sorin [21] demonstrated that 0-equilibria do not necessarily exist with respect to the average reward. They made it clear, moreover, that history-dependent strategies are indispensable for establishing \( \varepsilon \)-equilibria, for \( \varepsilon > 0 \).

For two-player stochastic games, Vieille [24, 25] managed to establish the existence of \( \varepsilon \)-equilibria, for all \( \varepsilon > 0 \). However, little is known about \( n \)-player stochastic games, and it is unresolved whether they always have \( \varepsilon \)-equilibria, for all \( \varepsilon > 0 \). This is probably the most challenging open problem in the field of stochastic games.

For the class of \( n \)-player aperiodic product-games, we will answer this question in the affirmative by proving the existence of 0-equilibria (cf., Main Theorem 1). Here aperiodicity refers to an aperiodic transition structure and will be given a precise definition later.
**Zero-sum games and optimality.** 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 \( r_i^j(a_i) = -r_i^j(a_i) \) (meaning that the sum of the payoffs is zero), for each state \( s \) and for each joint action \( a_i \). In these games the two players have completely opposite interests. Mertens and Neyman [13] showed that for such games \( v^i = -v^1 \).

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 \( v \). We also refer to Maitra and Sudderth [12] for the existence of the value in a very general setup.

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 \( y^SLgamma \) in the definition of the average reward, one would find precisely the same value \( v^Ei \) or \( y^SLpi \).

For the class of aperiodic zero-sum product-games, we will provide a proof that both players have stationary 0-optimal strategies (cf., Main Theorem 2). In addition, we analyse the structure of the value of these games.

The structure of the article. In §2, we will discuss preliminary concepts and results. Then, in §3, we will present and prove the two main theorems, and provide illustrative examples. Finally, §4 concludes with a short discussion on the case of periodic product-games.

2. Preliminary concepts and results. Some of the content of this section is very similar to the decomposition presented in Ross and Varadarajan [15] for Markov decision problems (i.e., stochastic games with only one player).

Classification of states. As we know, in a product-game, each player \( i \) has a Markov transition structure \( \Gamma^i \) of his own. We start by analysing each \( \Gamma^i \) separately. We would like to emphasise that such a separate analysis is only possible because, in a product-game, each player affects the transitions of his own states only. We distinguish between two basic types of states in the state space \( S^i \) of \( \Gamma^i \). A state \( s^i \in S^i \) belongs to type 1 if it is transient for each stationary strategy \( s^i \) of player \( i \). Otherwise, \( s^i \) belongs to type 2, in which case player \( i \) has a stationary strategy for which \( s^i \) is recurrent.

Maximal communicating sets. Two states \( s^i_1 \) and \( s^i_2 \) of type 2 are said to communicate with each other, if there exists a stationary strategy \( s^i \) of player \( i \) such that \( s^i_1 \) and \( s^i_2 \) belong to the same ergodic set. We note that communication between states have been used extensively in the literature of stochastic games (cf., Vieille [24, 25], Solan and Vieille [19], Solan [18]).

This relationship of communication is an equivalence relation on the set of states of type 2. As such, it induces equivalence classes, which for obvious reasons are called maximal communicating sets. Therefore, every maximal communicating set \( E^i \) has the properties that (1) player \( i \) can go from any state in \( E^i \) to any other state in \( E^i \), possibly in a number of moves, without leaving \( E^i \) with probability 1 and (2) if player \( i \) decides to leave \( E^i \), the probability that he ever comes back to \( E^i \) is strictly less than 1, regardless of his strategy. The latter observation further implies that (3) the total number of times during the whole play that player \( i \) switches from a maximal communicating set to another one is finite with probability 1, regardless of the initial state and player \( i \)’s strategy; (4) there is always at least one among the maximal communicating sets which player \( i \) is unable to leave; i.e., there are no transitions to states outside; (5) for any strategy of player \( i \), regardless of the initial state, player \( i \) eventually settles, with probability 1, in one of his maximal communicating sets \( E^i \); i.e., after finitely many stages, player \( i \) remains forever in \( E^i \) (it is possible that player \( i \) would be able to leave \( E^i \) with a different strategy).

Let \( E^i_k \), where \( k^i \in K^i \), denote the maximal communicating sets for player \( i \). Further, let \( K := \times^n_{i=1} K^i \). For any \( k = (k^1, \ldots, k^n) \in K \), the product \( E^i_k := \times^n_{i=1} E^i_k \) is called a joint maximal communicating set.

In every state \( s^i \) of the communicating set \( E^i_k \), for every \( k^i \in K^i \), let \( A^i_j \) denote the set of those actions \( a^i_j \in A^i_j \) which keep play in \( E^i_k \) with probability 1. The sets \( A^i_j \) are clearly nonempty. For every state \( s = (s^1, \ldots, s^n) \in S \), we also let \( A^i_j := A^i_j \).
Aperiodicity. A maximal communicating set $E^i_t$ of player $i$ is called aperiodic, if the Markov chain associated to a stationary strategy $x'$ of player $i$ that only uses completely mixed actions on $A^i_t$, for all $s' \in E^i_t$, is aperiodic on $E^i_t$. (Obviously, this definition is independent of the particular choice of $x'$.) The importance of the notion of aperiodicity lies in the fact that for such a completely mixed stationary strategy, after sufficiently many stages, the system can be in any state $s' \in E^i_t$, with positive probability, regardless of the initial state in $E^i_t$.

We will call a product-game aperiodic if all maximal communicating sets, for all players, are aperiodic. From now on, we will only consider aperiodic product-games, with the exception of §4.

Restricted games. Take an aperiodic product-game and some $k = (k_1, \ldots, k^n) \in K$. By restricting the state space to $E_k \subseteq S$, and the action set of each player $i$ in any state $s \in E_k$ to $A^i_k$, we obtain a restricted game $\tilde{G}_k$.

Obviously, $\tilde{G}_k$ is an aperiodic product-game itself.

These restricted games play a key role in the analysis of product-games because of the following observation. As is pointed out above, for any initial state and strategy of the players, each player $i$ eventually settles in one of his maximal communicating sets $E^i$, with probability 1. Hence, with probability 1, player $i$ will eventually settle in a restricted game $\tilde{G}_k$. The study of these restricted games is therefore of great importance.

For a restricted game $\tilde{G}_k$, let $v^i_k$ denote the min max level of player $i$ in $\tilde{G}_k$ for initial state $s \in E_k$. If, for some player $i$, the inequality $v^i_k \geq v^i_t$ holds for all initial states $s \in E_k$, then we call $\tilde{G}_k$ satisfactory to player $i$. Otherwise, $\tilde{G}_k$ is called unsatisfactory to player $i$. In words, if $\tilde{G}_k$ is satisfactory to player $i$, then player $i$ weakly prefers $\tilde{G}_k$ to $G$, as far as his min max level is concerned on $E_k$. Let $K^*$ denote the set of indices $k \in K$ such that $\tilde{G}_k$ is satisfactory to all players. Moreover, let $K^*[i]$ denote the set of indices $k \in K$ such that $\tilde{G}_k$ is unsatisfactory to player $i$ but satisfactory to all players $j \in \{1, \ldots, n-i\}$. Obviously, $K^*[1], \ldots, K^*[n]$, forms a partition of $K$.

Example 1. As an illustration, consider the product-game with two players given in Figure 1. This is a game with six states. In each state, the actions of player 1 are represented by the rows, and the actions of player 2 by the columns. Therefore, each cell of the state space corresponds to a pair of actions. In each cell, the two payoffs to the respective players are given in the upper-left corner, while the next state is indicated in the bottom-right corner. In this game all the transitions are pure; i.e., each transition probability distribution assigns probability 1 to the respective players are given in the upper-left corner, while the next state is indicated in the bottom-right corner.

A maximal communicating set $E^i_t$ is aperiodic if the Markov chain associated to a stationary strategy $x'$ of player $i$ that only uses completely mixed actions on $A^i_t$, for all $s' \in E^i_t$, is aperiodic on $E^i_t$. (Obviously, this definition is independent of the particular choice of $x'$.) The importance of the notion of aperiodicity lies in the fact that for such a completely mixed stationary strategy, after sufficiently many stages, the system can be in any state $s' \in E^i_t$, with positive probability, regardless of the initial state in $E^i_t$.

We will call a product-game aperiodic if all maximal communicating sets, for all players, are aperiodic. From now on, we will only consider aperiodic product-games, with the exception of §4.

Restricted games. Take an aperiodic product-game and some $k = (k_1, \ldots, k^n) \in K$. By restricting the state space to $E_k \subseteq S$, and the action set of each player $i$ in any state $s \in E_k$ to $A^i_k$, we obtain a restricted game $\tilde{G}_k$.

Obviously, $\tilde{G}_k$ is an aperiodic product-game itself.

These restricted games play a key role in the analysis of product-games because of the following observation. As is pointed out above, for any initial state and strategy of the players, each player $i$ eventually settles in one of his maximal communicating sets $E^i$, with probability 1. Hence, with probability 1, player $i$ will eventually settle in a restricted game $\tilde{G}_k$. The study of these restricted games is therefore of great importance.

For a restricted game $\tilde{G}_k$, let $v^i_k$ denote the min max level of player $i$ in $\tilde{G}_k$ for initial state $s \in E_k$. If, for some player $i$, the inequality $v^i_k \geq v^i_t$ holds for all initial states $s \in E_k$, then we call $\tilde{G}_k$ satisfactory to player $i$. Otherwise, $\tilde{G}_k$ is called unsatisfactory to player $i$. In words, if $\tilde{G}_k$ is satisfactory to player $i$, then player $i$ weakly prefers $\tilde{G}_k$ to $G$, as far as his min max level is concerned on $E_k$. Let $K^*$ denote the set of indices $k \in K$ such that $\tilde{G}_k$ is satisfactory to all players. Moreover, let $K^*[i]$ denote the set of indices $k \in K$ such that $\tilde{G}_k$ is unsatisfactory to player $i$ but satisfactory to all players $j \in \{1, \ldots, n-i\}$. Obviously, $K^*[1], \ldots, K^*[n]$, forms a partition of $K$.

Example 1. As an illustration, consider the product-game with two players given in Figure 1. This is a game with six states. In each state, the actions of player 1 are represented by the rows, and the actions of player 2 by the columns. Therefore, each cell of the state space corresponds to a pair of actions. In each cell, the two payoffs to the respective players are given in the upper-left corner, while the next state is indicated in the bottom-right corner. In this game all the transitions are pure; i.e., each transition probability distribution assigns probability 1 to a certain state.

The underlying Markov transition structure for player 1 is given by state space $S^1 = \{1, 2, 3\}$, action sets $A^1_1 = A^2_1 = \{1, 2\}$, $A^1_2 = \{1\}$, and transitions

\[
p^1_{11} = (1, 0, 0), \quad p^1_{12} = (0, 1, 0), \quad p^1_{21} = (1, 0, 0), \quad p^1_{22} = (0, 0, 1), \quad p^1_{31} = (0, 0, 1).
\]

Therefore, in state 1, player 1 can either stay or leave for state 2, from state 2 he can either go to state 1 or to state 3, while state 3 is absorbing. Regarding the classification of the states in $S^1$, both $E^1_1 := \{1, 2\}$ and $E^1_2 := \{3\}$ are maximal communicating sets, with index-set $K^1 = \{I, I\}$. Moreover, both $E^1_1$ and $E^1_2$ are aperiodic. As for the actions which keep play in these maximal communicating sets, we obtain $\tilde{A}^1_1 = \{1, 2\}$, $\tilde{A}^1_2 = \{1\}$, $\tilde{A}^2_1 = \{1\}$.

The underlying Markov transition structure for player 2 is given by state space $S^2 = \{1, 2\}$, action sets $A^2_1 = \{1, 2\}$, $A^2_2 = \{1\}$, and transitions $p^2_{11} = (0, 1), \quad p^2_{12} = (0, 1), \quad p^2_{21} = (0, 1)$. Further, both $E^2_1 := \{1\}$ and $E^2_2 := \{2\}$ are aperiodic maximal communicating sets, with $K^2 = \{I, I\}$ and $\tilde{A}^2_1 = \{1\}$.

As all maximal communicating sets are aperiodic, we may conclude that the game is aperiodic as well. Finally, we have $K = \{I, I\}$, which yields four joint maximal communicating sets and four corresponding restricted games. For example, $E_{(I, I)} = \{1, 2\} \times \{1\}$, and the corresponding restricted game $\tilde{G}_{(I, I)}$ consists of cells $(1, 1)$ and $(2, 1)$ in state $(1, 1)$ and cell $(1, 1)$ in state $(2, 1)$.

Later we will examine which restricted games are satisfactory to the players.

3. The main results and the proof.

3.1. The main theorems. For the class of aperiodic product-games, we present the following result concerning existence of equilibria.

Main Theorem 1. There exists a 0-equilibrium in every aperiodic n-player product-game.

The idea of the construction of an equilibrium $\eta$, for an aperiodic product-game $G$, is as follows. The equilibrium $\eta$ will prescribe to follow a joint strategy $\pi$, unless some player $i$ deviates from $\pi^i$ by playing an action outside the support of $\pi^i$ (i.e., an action on which $\pi^i$ puts probability zero). If player $i$ deviates in such a way, then from the next state, say state $s$, players $-i$ switch to a joint stationary strategy $y^{-i}$ and push down
player $i$'s reward to his min max level $v'_i$. In fact, $y^{-i}$ acts as a threat strategy, whose task is to force player $i$ to follow the prescriptions of $\pi^i$. Punishment with $y^{-i}$ will be shown to be severe enough. Finally, our construction will guarantee that no deviation inside the support of $\pi^i$ (such deviations are hard to detect) is profitable for any player $i$.

Now we briefly describe the construction of $\pi$, which shows a number of similarities with the construction in Vieille [24, 25]. The joint strategy $\pi$ prescribes to play in the following way:

1. When entering some $E_k$, with $k \in K^*$ (i.e., the restricted game $\tilde{G}_k$ is satisfactory to all players): In this case, $\pi$ will prescribe to stay on $E_k$ and play a certain equilibrium in $\tilde{G}_k$. Here, the players collect “high” payoffs (cf., solvable sets in Vieille [24, 25].)

2. When entering some $E_k$, with $k \in K[i]$ (i.e., the restricted game $\tilde{G}_k$ is unsatisfactory to player $i$): In this case, $\pi$ will prescribe player $i$ to exit $E_k$, (and thereby to leave $E_k$), while all other players patiently wait for player $i$’s exit. It will be taken care of that no player’s min max level drops in expectation. Payoffs in $E_k$ are disregarded (cf., controlled sets in Vieille [24, 25]).

3. Outside all joint maximal communicating sets: In this case, $\pi$ will let the players play for their future perspectives. Payoffs in these states are disregarded.

Note that, according to $\pi$, play will surely settle in a restricted game belonging to case (1).

In addition, for the special case of two-player zero-sum product-games, we show the existence of stationary solutions.

**Main Theorem 2.** In two-player aperiodic zero-sum product-games, both players have a stationary $0$-optimal strategy.

**3.2. The proofs of Main Theorems 1 and 2.** In this section, we provide a proof for Main Theorems 1 and 2. We will focus on Main Theorem 1, as Main Theorem 2 will follow (cf., the end of §3.2.3) along the way without major additional difficulties. Since an equilibrium reward for player $i$, from any initial state $s$, is at least his min max level $v'_i$, we start by studying the min max levels of product-games.

**3.2.1. The structure of the min max levels of the restricted games.**

**Lemma 3.1.** Let $G$ be an arbitrary aperiodic product-game and consider the restricted game $\tilde{G}_k$, for any $k = (k^1, \ldots, k^n) \in K$, and an arbitrary player $i$. Then, the min max level $\tilde{v}_{k,i}^i$ of any player $i$ in $\tilde{G}_k$ is constant; i.e., $\tilde{v}_{k,s}^i = \tilde{v}_{k,t}^i$ (=: $\tilde{v}_k^i$) for all states $s, t \in E_k$. Moreover, in $\tilde{G}_k$, players $-i$ have a joint stationary strategy $x_i$ which guarantees that player $i$’s reward from any initial state $s \in E_k$ is at most his min max level $\tilde{v}_{k,s}^i$; i.e., for all strategies $\pi^i$ for player $i$ in $\tilde{G}_k$ we have

$$\tilde{\gamma}^i(\pi^i, x_i) \leq \tilde{v}_{k,s}^i,$$

where $\tilde{\gamma}$ denotes the average reward for the game $\tilde{G}_k$.
Proof. Consider such a restricted game $\tilde{G}_k$ and a player $i$. Let $\alpha^i := \min_{\pi \in \mathcal{E}_k} \tilde{v}^i_k$. As is mentioned in the introduction, by applying Thuilisman and Vrieze [23] together with Neyman [14] for the game $\tilde{G}_k$, there exists a state $s' \in \{t \in E_k | \tilde{v}^i_k = \alpha^i\}$ for which players $-i$ have a joint stationary strategy $x^{-i}$ such that for all strategies $\pi^i$ for player $i$ in $\tilde{G}_k$, we have

$$\tilde{v}^i_k(\pi^i, x^{-i}) \leq \tilde{v}^i_k = \alpha^i.$$  

Let $Z$ denote the set of all those states $s \in \{t \in E_k | \tilde{v}^i_k = \alpha^i\}$ for which this $x^{-i}$ satisfies, for all strategies $\pi^i$ for player $i$ in $\tilde{G}_k$, that

$$\tilde{v}^i_k(\pi^i, x^{-i}) \leq \alpha^i.$$  

Let $x^i$ be a completely mixed stationary strategy in $\tilde{G}_k$ for player $i$. Consider the joint stationary strategy $(x^i, x^{-i})$. By the definitions of $Z$ and $x^{-i}$, the set $Z$ is closed for $(x^i, x^{-i})$ (i.e., play does not leave $Z$). Hence, there must exist an ergodic set $F \subset Z$ for $(x^i, x^{-i})$. Due to the choice of $x^i$ and the aperiodicity of $\mathcal{E}_k$, if $u \in F$, then $(t^i, u^{-i}) \in F$ for all states $t^i \in E_k$. Thus, $F$ must be the form $F = \tilde{F} \times E_k'$ for some nonempty $\tilde{F} \subset E_k$. Define a joint stationary strategy $y^{-i}$ for players $-i$ in $\tilde{G}_k$ as follows: let $y^{-i}_t = x^{-i}_t$ for all $t \in F$ and let $y^{-i}_t$ be an arbitrary completely mixed action on $\tilde{A}_k^{-i}$ for all $t \in (E_k - F)$. Now, take an arbitrary strategy $\pi^i$ for player $i$ in $\tilde{G}_k$ and consider $(\pi^i, y^{-i})$ with any initial state in $E_k$. Based on the aperiodicity of $\mathcal{E}_k$, for all $j \neq i$ and the choice of $y^{-i}$ outside $F$, players $-i$ eventually visit $\tilde{F}$, and hence play eventually visits $F$. Then, as $F$ is closed with respect to $(x^i, x^{-i})$ and $y^{-i}$ equals $x^{-i}$ on $F$, once play reaches $F$, it will never leave it and, in view of (8), player $i$’s reward will be at most $\alpha^i$. Consequently, the min max level $\tilde{v}^i_k$ of player $i$ in $\tilde{G}_k$ equals the constant $\alpha^i$, and $y^{-i}$ satisfies the second part of the lemma. $\square$

As an illustration, we now revisit the example in Theorem 1. First, take the restricted game $\tilde{G}_{(1,1)}$ consisting of cells (1, 1) and (2, 1) in state (1, 1) and cell (1, 1) in state (2, 1). Let us examine player 1’s min max level $\tilde{v}^1_{(1,1)}$ in $\tilde{G}_{(1,1)}$. In $\tilde{G}_{(1,1)}$, it is only player 1 who has a choice and only in state (1, 1). By choosing the first action, he receives payoff 2, while by playing the second one he receives payoff 0 and subsequently payoff 3 in state (2, 1), giving 3/2 on average, before returning to state (1, 1). Hence, player 1’s min max level is $\tilde{v}^1_{(1,1)} = 2$, whereas, for similar reasons, player 2’s min max level is $\tilde{v}^2_{(1,1)} = -2$.

For the other three restricted games, we have $\tilde{v}^1_{(1,1)} = \tilde{v}^1_{(1,2)} = 0$, $\tilde{v}^1_{(2,1)} = \tilde{v}^1_{(2,2)} = 0$, and $\tilde{v}^2_{(1,1)} = 1$, $\tilde{v}^2_{(1,2)} = -1$.

3.2.2. The min max levels in simple product-games. We call a product-game $G$ simple if it holds within any restricted game $\tilde{G}_k$ for any player $i$ that all payoffs to player $i$ are equal; i.e., for any $k \in K$ and for any player $i$, we have $r^i(a_t) = r^i(b_t)$ for any states $s, t \in E_k$ and for any joint actions $a_t = A_t, b_t = A_t$.

By a solitary move of player $i$ in state $s \in S$ we mean a step where only player $i$ chooses an action, say $a^i_t \in A^i_t$, and where the process moves to a new state of the form $(t', s')$, $t' \in S'$, with probability $p^i_{a^i_t}(t')$. A solitary move for players $-i$ is defined similarly.

The following lemma claims that player $i$ is unable to improve on his min max level $\nu^i$ in expectation by a solitary move. Similarly, solitary moves by players $-i$ cannot decrease player $i$’s min max level $\nu^i$ in expectation.

Lemma 3.2. Let $G$ be a simple aperiodic product-game. Take an arbitrary player $i$ and a state $s = (s', \ldots, s^g) \in S$. Then, for any action $a^i_t \in A^i_t$ of player $i$, it holds that $\nu^i_t = \tilde{v}^i_k$.

1. For any action $a^i_t \in A^i_t$ of player $i$, it holds that

\[ \sum_{t' \in S'} p^i_{a^i_t}(t') \nu^i_{(t', s')} \leq \nu^i_t. \]

2. For any joint action $a^{-i}_s \in A^{-i}_s$ of players $-i$, it holds that

\[ \sum_{t' \in S'} p^{-i}_{a^{-i}_s}(t') \nu^i_{(t', s')} \geq \nu^i_t. \]

Proof. We only show part (1) of the lemma; part (2) can be proven similarly. Let $\epsilon > 0$. The idea of the proof is as follows. We compare two specific games $\Omega$ and $\hat{\Omega}$. The game $\Omega$ is simply the original game $G$ with initial state $s$, while $\hat{\Omega}$ is the game that results from $G$ after the execution of the solitary move $a^i_t$ by player $i$.
in state \( s \); i.e., \( \bar{\Omega} \) starts in initial state \( (t', s^{-i}) \), \( t' \in S_i \), with probability \( p'_{s',i}(t') \). We define two joint strategies \( \pi \) for \( \Omega \) and \( \bar{\pi} \) for \( \bar{\Omega} \) in such a way that the following properties hold:

Property (a) for \( \Omega \): \( \gamma'_i(\pi) \leq v'_i + \epsilon \).

Property (b) for \( \bar{\Omega} \): \( \gamma'_{(t', s^{-i})}(\bar{\pi}) \geq v'_{(t', s^{-i})} - \epsilon \), for any \( t' \in S_i \).

Property (c): \( \pi \) and \( \bar{\pi} \) yield the same expected rewards in \( \Omega \) and \( \bar{\Omega} \), respectively; i.e.,

\[
\gamma'_i(\pi) = \sum_{t' \in S_i} p'_{s',i}(t') \gamma'_{(t', s^{-i})}(\bar{\pi}).
\]

It will then follow from properties (a), (b), and (c) that

\[
v'_i + \epsilon \geq \gamma'_i(\pi) = \sum_{t' \in S_i} p'_{s',i}(t') \gamma'_{(t', s^{-i})}(\bar{\pi}) \geq \sum_{t' \in S_i} p'_{s',i}(t') \cdot (v'_{(t', s^{-i})} - \epsilon) = \sum_{t' \in S_i} p'_{s',i}(t') \cdot v'_{(t', s^{-i})} - \epsilon.
\]

As \( \epsilon > 0 \) was arbitrary, the proof will then be complete.

**Step 1. The construction of \( \pi \) for \( \Omega \) and \( \bar{\pi} \) for \( \bar{\Omega} \).** For notational purposes, in this proof we assume that the histories also contain the present state, so a history up to stage \( m \) is in the form \( h_m = (s_1, a_1; \ldots; s_{m-1}, a_{m-1}; s_m) \). When \( h_m \) is used in the sequel, \( h_{m-1} \) will denote the restriction of \( h_m \) up to stage \( m - 1 \). It is useful to separate each \( h_m \) into the coordinates belonging to player \( i \) and to players \( -i \), respectively; i.e., let

\[
H_i = \{s_1^i, a_1^i; \ldots; s_{m-1}^i, a_{m-1}^i; s_m^i\}
\]

and let

\[
H^{-i} = \{s_1^{-i}, a_1^{-i}; \ldots; s_{m-1}^{-i}, a_{m-1}^{-i}; s_m^{-i}\}.
\]

Let \( H_m \) denote the set of histories for \( \Omega \) up to stage \( m \). Define \( \bar{H}_m \) similarly for \( \bar{\Omega} \). By separating the coordinates, we obtain the sets \( H_i^m, H^{-i}_m, \bar{H}_i^m, \bar{H}^{-i}_m \).

As players \(-i\) begin in \( s^{-i} \) in both \( \Omega \) and \( \bar{\Omega} \), we have \( H_i^{-i} = \bar{H}_i^{-i} \) for all \( m \).

Now we define a transformation \( \phi: H_i^m \to H_i^m \). For \( h_i^m \in \bar{H}_m \), with \( m \geq 2 \), let

\[
\phi(h_i^m) = (s_1', a_1'; \ldots; s_{m-1}', a_{m-1}'; s_m').
\]

Naturally, we define \( \phi(h_i^1) = s_i' \).

We start with the strategies for players \(-i\). By the definition of the min max level \( v'_i \), there exists a joint strategy \( \pi^{-i} \) of players \(-i\) such that player \( i \)'s reward cannot be more in \( \Omega \) than \( v'_i + \epsilon \); i.e., \( \gamma'_i(\pi^{-i}, \pi^{-i}) \leq v'_i + \epsilon \) for all strategies \( \pi^{-i} \) of player \( i \). Given \( \pi^{-i} \), define \( \bar{\pi}^{-i} \) in \( \bar{\Omega} \) for players \(-i\) by \( \bar{\pi}^{-i}(h_i^m, h_i^{-i}) = \pi^{-i}(\phi(h_i^m), h_i^{-i}) \) for all \((h_i^m, h_i^{-i}) \in \bar{H}_m^{-i} \).

Now we define the strategy for player \( i \). By the definition of the min max level again, there exists a \( \bar{\pi}' \) for player \( i \) in \( \bar{\Omega} \) such that \( \gamma'_{(t', s^{-i})}(\bar{\pi}', \pi^{-i}) \geq v'_i - \epsilon \) for all initial states \( u = (t', s^{-i}) \), \( t' \in S_i \), in \( \bar{\Omega} \). Finally, let \( \pi' \) be defined for \( \Omega \) as \( \pi'(h_i^m, h_i^{-i}) = \bar{\pi}'(\phi(h_i^m), h_i^{-i}) \), for all \((h_i^m, h_i^{-i}) \in H_m \) with \( m \geq 2 \), where

\[
\phi(h_i^m) = (s_1', a_1'; \ldots; s_{m-1}', a_{m-1}'; s_m')\] Naturally, we define \( \pi'(h_i^1, h_i^{-1}) = a_i' \).

Note that

\[
\pi'_{(\phi(h_i^m), h_i^{-i})} = \bar{\pi}'(h_i^m, h_i^{-i})
\]

for all \((h_i^m, h_i^{-i}) \in \bar{H}_m \).

**Step 2. Proving Properties (a), (b), and (c) for \( \pi \) and \( \bar{\pi} \).** Note that Properties (a) and (b) are satisfied by the definitions of \( \pi^{-i} \) and \( \bar{\pi}^{-i} \). So, it remains to verify Property (c).

We will show that the stochastic processes induced by \( \pi \) in \( \Omega \) and by \( \bar{\pi} \) in \( \bar{\Omega} \) are strongly related, namely

\[
\mathbb{P}_\pi(h_i^m, h_i^{-i}) = \mathbb{P}_{\bar{\pi}}(\phi(h_i^m+1), h_i^{-i})
\]

for all \((h_i^m+1, h_i^{-i}) \in \bar{H}_m+1 \). Here, the left-hand side should be interpreted as the probability, with respect to \( \pi \) in \( \bar{\Omega} \), that the \( i \)-coordinate of the history up to stage \( m \) coincides with \( h_i^m \) and the \(-i\)-coordinate of the history up to stage \( m + 1 \) coincides with \( h_i^{-i} \). The interpretation of the right-hand side is analogous.
The proof goes by complete induction.

For \( m = 1 \), with \( h_i^1 = (t^i) \) and \( h_{i'j}^{-1} = (s^{-i}, a_i^{-i}, s_i^{-i}) \), we have

\[
\Pr_\pi(h_i^1, h_{i'j}^{-1}) = p_i(t^i) \cdot \pi_i(t^i, h_i^1)(a_i^{-i}) \cdot p_j^{-i}(s_i^{-i})
\]

and also with \( \phi(h_{i'}^{j'}) = (s', a_i', t') \),

\[
\Pr_\pi(\phi(h_{i'}^{j'}), h_{i'j}^{-1}) = p_i(t') \cdot \pi_i(t', s') \cdot (a_i') \cdot p_j^{-i}(s_i^{-i} - t^i).
\]

Since \( \pi_i(t^i, h_i^1) = \pi_i(s', s^{-i}) = \pi_i(s', s^{-i}) \), the equality (11) holds for \( m = 1 \).

Suppose that equality (11) is valid for a certain \( m \). For \( m + 1 \) we obtain

\[
\Pr_\pi(h_i^{m+1}, h_{i'j}^{-1}) = \Pr_\pi(h_i^m, h_{i'j}^{-1}) \cdot \pi_i(h_i^m, h_{i'j}^{-1})(a_i^{-i}) \cdot p_j^{-i}(s_i^{-i} - t^i)
\]

and

\[
\Pr_\pi(\phi(h_{i'}^{j'}), h_{i'j}^{-1}) = \Pr_\pi(\phi(h_i^{m+1}, h_{i'j}^{-1}) \cdot \pi_i(\phi(h_i^{m+1}, h_{i'j}^{-1})(a_i') \cdot p_j^{-i}(s_i^{-i} - t^i)).
\]

From our assumption that (11) holds for \( m \), from equality (10), and from the definition of \( \tilde{\pi}^{-i} \), it follows that (11) holds for \( m + 1 \). Consequently, (11) holds for all \( m \).

Recall that, with respect to any initial state and any joint strategy, play eventually settles, with probability 1, in a restricted game. By equality (11), the probability that this restricted game is some \( \tilde{G}_i \) with respect to \( \pi \) in \( \tilde{\pi} \) equals the probability that this is \( \tilde{G}_i \) with respect to \( \tilde{\pi} \). Since the game is simple, the average reward is determined by this restricted game, and hence the expected reward for \( \pi \) in \( \tilde{\pi} \) equals the expected reward for \( \pi \) in \( \tilde{\pi} \). This proves Property (c), and therefore the proof is complete. \( \square \)

**Lemma 3.3.** Take a simple aperiodic product-game \( G \). Let \( E_{i'}^i \) denote a maximal communicating set for player \( i \), for some \( k' \in K \), and \( E_{i'}^{-i} \), denote a joint maximal communicating set for players \(-i\), for some \( k^{-i} \in K^{-i} \).

1. For any two states \( s', t' \in E_{i'}^i \) of player \( i \) and any joint state \( s^{-i} \in S^{-i} \) of players \(-i\), the min max level of player \( i \) satisfies \( v_i(s', t') \leq v_i(s', t') \).
2. For any two joint states \( s^{-i}, t^{-i} \in E_{i'}^{-i} \) of players \(-i\) and any state \( s' \in S' \) of player \( i \), the min max level of player \( i \) satisfies \( v_i(s', t^{-i}) \leq v_i(s', t^{-i}) \).

**Proof.** We will show part (1); the proof of part (2) is similar. Take an arbitrary \( s^{-i} \in S^{-i} \). Let \( F^i \) denote those states \( s' \in E_{i'}^i \) for which \( v_i'(s', s^{-i}) \leq v_i'(s', s^{-i}) \) for all \( s' \in E_{i'}^i \). Suppose by way of contradiction that \( E_{i'}^i - F^i \) is not empty. Take a state \( s' \in F^i \) and any action \( a_i' \in A_i' \) which moves from state \( s' \) to a state in \( E_{i'}^i - F^i \) with a positive probability. Then, the solitary move \( a_i' \) in state \( (s', s^{-i}) \) for player \( i \) would improve player \( i \)'s min max level in expectation, which contradicts part (1) of Lemma 3.2. Hence, \( F^i = E_{i'}^i \), and part (1) of the lemma follows. \( \square \)

**Corollary 3.1.** Let \( G \) be a simple aperiodic product-game, and let \( E_k \) be a joint maximal communicating set for some \( k \in K \). Then, the min max level \( v_i \) of any player \( i \) is constant on \( E_k \); i.e., \( v_i = v_i = v_i = v_i \) for all \( s, t \in E_k \).

**Lemma 3.4.** Let \( G \) be a simple aperiodic product-game. Then, for any player \( i \) the following properties hold.

1. Let \( s = (s^1, \ldots, s^n) \) be a state such that \( s^i \) belongs to a maximal communicating set \( E_{i'}^i \). Then, regardless of the joint mixed action \( x^{-i} \) played by players \(-i\) in state \( s \), all the actions in \( A_i' \) guarantee in expectation the best possible min max level for player \( i \) after transition; i.e., for any actions \( a_i' \in A_i' \) and \( b_i' \in A_i' \) it holds that

\[
\sum_{t \in S} p_{x_i'(a_i', b_i')} (t) v_i'(t) \geq \sum_{t \in S} p_{x_i'(a_i', b_i')} (t) v_i'.
\]

2. Let \( s = (s^1, \ldots, s^n) \) be a state such that \( s^i \) belongs to a maximal communicating set \( E_{i'}^i \) for all players \( j \neq i \). Then, all joint actions in \( A_i' \) for players \(-i\) in state \( s \) guarantee in expectation that player \( i \)'s min max level cannot increase after transition; i.e., for any joint action \( a_i^{-i} \in A_i^{-i} \) and for any action \( a_i' \in A_i' \) it holds that

\[
\sum_{t \in S} p_{x_i'(a_i', a_i')} (t) v_i'(t) \leq v_i'.
\]
The next lemma proposes a way for player \(i\) cannot decrease in expectation from state \(s\) (i.e., at least one player \(i\) regardless of the action played by player \(t\)). For some \(s\), therefore, in state \(s\), player \(i\) that if player \(s_i = \frac{1}{m^2} s_i + \frac{1}{m^2} s_i \) for some \(s\).

First we prove part (1). Take an arbitrary mixed action \(x_{s_i}^{-1}\) for players \(-i\), and actions \(a_{s_i}^j \in A_i^{-1}\) and \(b_{s_i}^j \in A_i^{-1}\) for player \(i\) in state \(s\). Then, by part (1) of Lemma 3.3 and part (1) of Lemma 3.2, we have for any \(t' \in S^{-1}\) that

\[
\sum_{r \in S} p_{s_i}^{-1} u_r(t') v_{s_i}(t', t') = v_{s_i}(t', t') \geq \sum_{r \in S} p_{s_i}^{-1} b_r(t') v_{s_i}(t', t')
\]

Hence

\[
\sum_{r \in S} p_{s_i}(-s_i, x_{s_i}^{-1})(t) v_{s_i}^i = \sum_{r \in S} p_{s_i}^{-1}(t') \left[ \sum_{r \in S} p_{s_i}^{-1} u_r(t') v_{s_i}(t', t') - \sum_{r \in S} p_{s_i}^{-1} b_r(t') v_{s_i}(t', t') \right] \\
\geq \sum_{r \in S} p_{s_i}^{-1}(t') \left[ \sum_{r \in S} p_{s_i}^{-1} u_r(t') v_{s_i}(t', t') - \sum_{r \in S} p_{s_i}^{-1} b_r(t') v_{s_i}(t', t') \right] \\
= \sum_{r \in S} p_{s_i}(-s_i, x_{s_i}^{-1})(t) v_{s_i}^i.
\]

As for part (2) of Lemma 3.4, take arbitrary joint actions \(a_{s_i}^{-1} \in \tilde{A}_i^{-1}\) and \(b_{s_i}^{-1} \in A_i^{-1}\) in state \(s\). It follows similarly to part (1) that for all mixed actions \(x_i^s\) of player \(i\),

\[
\sum_{r \in S} p_{s_i}(-s_i, x_{s_i}^{-1})(t) v_{s_i}^i \leq \sum_{r \in S} p_{s_i}(-s_i, b_{s_i}^{-1})(t) v_{s_i}^i.
\]

Therefore, in state \(s\), the infimum in equality (7) is attained at all \(a_{s_i}^{-1} \in \tilde{A}_i^{-1}\); hence

\[
\sum_{r \in S} p_{s_i}(-s_i, a_{s_i}^{-1})(t) v_{s_i}^i \leq v_{s_i}^i,
\]

which proves part (2). \(\square\)

Consider a restricted game \(\tilde{G}_s\) within a simple product-game, which is unsatisfactory to player \(i\) (i.e., \(i \in K[i]\)). The next lemma proposes a way for player \(i\) to exit \(\tilde{G}_s\). A similar result holds for players \(-i\).

**Lemma 3.5.** Let \(G\) be a simple aperiodic product-game, and let \(E_k\) be a joint maximal communicating set for some \(k = (k^1, \ldots, k^n) \in K\). Let \(z_{s_i}^k\) denote player \(i\)'s unique reward in the restricted game \(\tilde{G}_s\), and let \(v_{s_i}^k\) be player \(i\)’s min max level on \(E_k\) in the game \(G\) (a constant, cf., Corollary 3.1).

(1) Suppose \(z_{s_i}^k < v_{s_i}^k\). Then, there is a state \(s' \in E_{s_i}^k\) and an action \(a_{s_i}^j \in A_{s_i}^{-1} \rightarrow \tilde{A}_{s_i}^{-1}\) for player \(i\) in state \(s'\) such that if player \(i\) plays action \(a_{s_i}^j\) in state \(s'\), with \(s' \in E_{s_i}^{k-1}\), then player \(i\)’s min max level cannot decrease in expectation from state \(s\), regardless of the actions played by players \(-i\). More precisely, for any \(a_{s_i}^{-1} \in A_{s_i}^{-1}\) we have

\[
\sum_{r \in S} p_{s_i}(-s_i, a_{s_i}^{-1})(t) v_{s_i}^i \geq v_{s_i}^i.
\]

(2) Suppose \(z_{s_i}^k > v_{s_i}^k\). Then, there is a joint state \(s^{-1} \in E_{s_i}^{-1}\) of players \(-i\) and a joint action \(a_{s_i}^{-1} \in A_{s_i}^{-1} \rightarrow \tilde{A}_{s_i}^{-1}\) (i.e., at least one player \(j \neq i\) plays outside \(A_{s_i}^{-1}\)) such that if players \(-i\) play joint action \(a_{s_i}^{-1}\) in any state \(s = (s', s^{-1}) \in E_s\), with \(s' \in E_{s_i}^k\), then player \(i\)’s min max level cannot increase in expectation from state \(s\), regardless of the action played by player \(i\). More precisely, for any \(a_{s_i}^j \in A_{s_i}^j\) we have

\[
\sum_{r \in S} p_{s_i}(-s_i, a_{s_i}^{-1})(t) v_{s_i}^i \leq v_{s_i}^i.
\]

**Proof.** We will prove part (1); the proof of part (2) is similar.

Due to \(z_{s_i}^k < v_{s_i}^k\), when starting in \(E_k\), player \(i\) can only defend his min max level \(v_{s_i}^k\) in \(G\) if he leaves \(E_k\). Therefore, there must be at least one state \(s' \in E_{s_i}^k\) joint action \(b_{s_i}^j \in \tilde{A}_{s_i}^{-1}\), and action \(a_{s_i}^j \in A_{s_i}^{-1} \rightarrow \tilde{A}_{s_i}^{-1}\) such that

\[
\sum_{r \in S} p_{s_i}(-s_i, b_{s_i}^j)(t) v_{s_i}^i \geq v_{s_i}^i.
\]

Now we will show that state \(s'\) and action \(a_{s_i}^j\) satisfy part (1) of the Lemma 3.5.

Take an arbitrary state \(s \in E_k\) with \(s' = s^{-1}\) and a joint action \(a_{s_i}^{-1}\) of players \(-i\) in state \(s\). Note that for any \(t' \in S'\),

\[
\sum_{r' \in S'} p_{s_i}^{-1} u_r(t') v_{s_i}(t', t') \geq v_{s_i}(t', t') = \sum_{r' \in S'} p_{s_i}^{-1} b_r(t') v_{s_i}(t', t')\]


where the inequality follows from part (2) of Lemma 3.2 and the equalities from part (2) of Lemma 3.3. Therefore,
\[
\sum_{r \in S} p_{s, (a^i_r, a^j_r)}(t) v^i_t = \sum_{r' \in S} p'_{s, a^i_r}(t') \left[ \sum_{r'' \in S} p^j_{s', a^j_r}(r'') v^j_{r'(t', r'')} \right] \\
\geq \sum_{r' \in S} p'_{s, a^i_r}(t') \left[ \sum_{r'' \in S} p^j_{s', a^j_r}(r'') v^j_{r'(t', r'')} \right] \\
= \sum_{r \in S} p_{s, (a^i_r, a^j_r)}(t) v^i_t \\
\geq v^i_t,
\]
which completes the proof of part (1).

**Example 2.** Consider the simple product-game \( G \) with two players given in Figure 2. This game is obtained from the game in Example 1 by replacing all payoffs for player 1 by 2 and for player 2 by \(-2\) in the restricted game \( \bar{G}_{t, t'} \), and all payoffs for either player by 0 in the restricted game \( \bar{G}_{t, t'} \). Hence, the only possible pair of payoffs is \((2, -2)\), in \( G_{t, t'} \) and \((0, 0)\) in \( \bar{G}_{t, t'} \). Finally, for restricted game \( \bar{G}_{t, t'} \), the only possible pair of payoffs remains \((0, 0)\), while in \( \bar{G}_{t, t'} \), it remains \((1, -1)\). So, the game is simple, indeed.

Let us examine the players’ min max levels in \( G \). For player 1, we will argue that
\[
v^1_{(1, 1)} = v^1_{(2, 1)} = v^1_{(2, 2)} = v^1_{(3, 2)} = 1, \quad v^1_{(3, 1)} = 0.
\]
Player 1’s min max level is clearly 0 for initial state \((3, 1)\), in view of player 2’s first action. Now consider another arbitrary initial state \( s \in S - \{(3, 1)\} \). By moving to his second state, player 2 can always make sure that player 1’s reward is at most 1. On the other hand, player 1 can guarantee reward 1 for state \( s \) by the pure stationary strategy \( x^1 \) defined as
\[
x^1_{(1, 1)} = (1, 0), \quad x^1_{(1, 2)} = (0, 1), \quad x^1_{(2, 1)} = (1, 0), \quad x^1_{(2, 2)} = (0, 1), \quad x^1_{(3, 1)} = x^1_{(3, 2)} = (1).
\]
Hence, player 1’s min max level equals 1 for all \( s \in S - \{(3, 1)\} \), indeed. We similarly find that \( v^1_2 = -v^1_1 \) for all \( s \in S \).

With the help of this example, we will now illustrate some lemmas in this section.

**Regarding Lemma 3.2:** Consider, for instance, a solitary move for player 1 in state \((2, 1)\). Then, player 1’s first action, as a solitary move, yields state \((1, 1)\), while the second action yields state \((3, 1)\). As \( v^1_{(2, 2)} = v^1_{(1, 1)} = 1 \) and \( v^1_{(3, 1)} = 0 \), player 1 is indeed unable to improve on his min max level by such a solitary move.

**Regarding Lemma 3.4:** Consider player 1 in state \((2, 1)\). If player 1 plays his first action (the action of the restricted game \( \bar{G}_{t, t'} \)), then his min max level will remain 1, regardless of the action chosen by player 2.

- **State (1, 1)**
  - 2, -2 → (1, 1)
  - 2, -2 → (2, 1)

- **State (1, 2)**
  - 0, 0 → (1, 2)
  - 0, 0 → (2, 2)

- **State (2, 1)**
  - 2, -2 → (1, 1)
  - 0, 0 → (3, 1)

- **State (2, 2)**
  - 0, 0 → (3, 2)

- **State (3, 1)**
  - 0, 0 → (3, 1)

- **State (3, 2)**
  - 1, -1 → (3, 2)

**Figure 2.** Game of Example 2.
Therefore, action 2 can never be better than action 1 for player 1, with respect to player 1’s expected min max level after transition.

Regarding Lemma 3.5: Consider the restricted game \( \tilde{G}_{(i,II)} \), in which player 1’s unique reward is 0 while his min max level \( v^1 \) is 1. Note that player 1 can leave \( \tilde{G}_{(i,II)} \) by moving to state (2, 2) and playing his second action there. By doing so, his min max level \( v^1 \) remains unchanged.

### 3.2.3. The min max levels of general product-games.

Take an arbitrary product-game \( G \). The next lemma presents a natural way of transforming \( G \) into a simple product-game \( \tilde{G} \), by replacing payoffs by min max levels, and claims that the min max levels of the players remain unchanged under this transformation. The idea to replace payoffs by min max levels also appeared in Solan [17] and in a more sophisticated way in Solan and Vohra [20].

**Lemma 3.6.** Take an arbitrary aperiodic product-game \( G \), with \( v^i \) denoting the min max level for every player \( i \) and for every state \( s \in S \). Let \( \bar{v}^i_s \) denote player \( i \)'s min max level in any restricted game \( \tilde{G}_k \) (which is constant; cf., Lemma 3.1). Let \( \tilde{G} \) denote the simple aperiodic product-game which is derived from \( G \) by replacing each player \( i \)'s payoffs in any restricted game \( \tilde{G}_k \) by his min max level \( \bar{v}^i_\cdot \). Further, let \( w^i_\cdot \) denote every player \( i \)'s min max level in \( \tilde{G} \) in state \( s \).

Then, the min max levels of the product-games \( G \) and \( \tilde{G} \) are equal; i.e., \( v^i = w^i_\cdot \) for all players \( i \) and for all states \( s \in S \).

**Proof.** Consider the original product-game \( G \) and take an arbitrary player \( i \). For this game \( G \), we will show below that players \( -i \) have a joint stationary strategy \( x^{i-} \) which guarantees that player \( i \)'s reward from any initial state \( s \in S \) is at most \( w^i_\cdot \); i.e., for all strategies \( \pi^i \) for player \( i \) we have

\[
\gamma^i(\pi^i, x^{i-}) \leq w^i_\cdot. \tag{12}
\]

This yields \( v^i_s \leq w^i_\cdot \) for all states \( s \in S \). Since \( v^i_s \geq w^i_\cdot \) for all \( s \in S \) follows in a similar fashion, we will have \( v^i_s = w^i_\cdot \) for all states \( s \) and the proof will then be complete.

Now we prove (12). Note first that \( w^i_\cdot \) is also a constant \( w^i_\cdot \) on any joint maximal communicating set \( E_k \), by Corollary 3.1 for the game \( \tilde{G} \). We construct the joint stationary strategy \( x^{i-} \) by distinguishing the following three mutually exclusive cases.

**Case 1:** States \( s = (s^1, \ldots, s^n) \in S \) such that \( s^i \) is of type 1 for at least one player \( j \) (possibly \( j = i \)). In any such a state \( s \), let \( x^{i-}_s \in X^{i-}_s \) be a joint mixed action for players \( -i \) such that for any mixed action \( x^i_s \in X^i_s \) of player \( i \) we have

\[
\sum_{s \in S} p_{s, (x^i_s, x^{i-}_s)}(t) u^i_s \leq w^i_\cdot. \]

Obviously, by expression (7) for player \( i \)'s min max level \( w^i_\cdot \) in \( \tilde{G} \), such a joint mixed action exists.

**Case 2:** States in a joint maximal communicating set \( E_k \) for which \( \bar{v}^i_s \leq w^i_\cdot \). Take a joint stationary strategy \( y^{i-} \) for player \( i \) in the corresponding restricted game \( \tilde{G}_k \) (which is a part of the original game \( G \)) in Lemma 3.1. Then, let \( x^{i-}_s = y^{i-}_s \) for all \( s \in E_k \).

**Case 3:** States in a joint maximal communicating set \( E_k \) for which \( \bar{v}^i_s > w^i_\cdot \). Take a joint state \( t^{i-}_s \in E^{i-}_k \) and a joint "exit" action \( a^{i-}_s \in A^{i-}_k \), with respect to the game \( \tilde{G} \) and its min max level \( w^i_\cdot \) for player \( i \), as in part (2) of Lemma 3.5. Then, for any \( s = (s^1, \ldots, s^n) \in E_k \), let \( x^{i-}_s = a^{i-}_s \) whenever \( s^{i-} = t^{i-} \), and let \( x^{i-}_s \) be an arbitrary joint completely mixed action on \( \tilde{G} \) whenever \( s^{i-} \neq t^{i-} \).

Take a stationary best reply \( x^i \) of player \( i \) in \( G \) against \( x^{i-} \). We will show that \( \gamma^i(x^i, x^{i-}) \leq w^i_\cdot \) for any initial state \( s \in S \).

First, consider an arbitrary ergodic set \( F \) for \( (x^i, x^{i-}) \). As players \( -i \) will leave any set \( E_k \) considered in Case 3, we conclude that \( F \subset E_k \) for some \( E_k \) in Case 2. Since \( x^i \) does not leave \( F \), the mixed action \( x^i \) only uses actions from \( A^i_k \) for all \( s \in F \), meaning that \( x^i \) behaves on \( F \) as a stationary strategy in the restricted game \( \tilde{G}_k \). Hence, by the choice of \( x^{i-} \) in Case 2, we have

\[
\gamma^i(x^i, x^{i-}) \leq \bar{v}^i_s \leq w^i_\cdot = w^i_\cdot \tag{13}
\]

for all \( s \in F \). As \( F \) was an arbitrary ergodic set, we have \( \gamma^i(x^i, x^{i-}) \leq w^i_\cdot \) for all states \( s \) that are recurrent for \( (x^i, x^{i-}) \).

Next, note that \( w^i \) cannot increase in expectation after transition with respect to \( (x^i, x^{i-}) \); i.e., \( P(x^i, x^{i-})w^i \leq w^i \). Indeed, for Cases 2 and 3 it is guaranteed by part (2) of Lemma 3.4 and by part (2) of
Lemma 3.5 (both applied to $w_i$ as the min max level of player $i$ in $\tilde{G}$), while it holds by construction for Case 1. Consequently, we also have $P^m(x', x^{-i})w_i \leq w_i$ for all $m \in \mathbb{N}$, yielding $Q(x', x^{-i})w_i \leq w_i$.

By applying equality (5), we now obtain

$$
\gamma'(x', x^{-i}) = Q(x', x^{-i}) \cdot \gamma'(x', x^{-i}) \leq Q(x', x^{-i}) \cdot w_i \leq w_i,
$$

where the first inequality follows from inequality (13) and from the fact that entry $(t, s)$ of the stochastic matrix $Q(x', x^{-i})$ is only positive if state $s$ is recurrent for $(x', x^{-i})$. Since $x'$ is a best reply to $x^{-i}$ in $G$, the proof of (12) is complete.

For an illustration of the above lemma, we refer to the games in Examples 1 and 2. Indeed, the product-game in Example 1 (which is now game $G$ with min max levels $v$) leads to the simple product-game in Example 2 (which is now game $F$ with min max levels $w$). Just as in the proof of the above lemma, we can construct a stationary strategy $v_i$ for player 1 ($v_i$ being $x^{-1}$ for players $-i$ with $i = 2$) which guarantees in $G$ that player 2’s reward is not more than $w_i$ for all initial states $s \in S$. Recall that $w_i^2 = 0$ while $w_i^1 = -1$ for all $s \in S - \{(3, 1)\}$, and that player 2’s min max levels within the restricted games are $v_i^2 = -2$, $v_i^1 = v_i(1, 1) = 0$, $v_i(1, 2) = -1$. Following the proof, as the joint maximal communicating sets $E_{(1, 1)}$, $E_{(1, 2)}$, and $E_{(1, 1)}$ all belong to Case 2 (i.e., $v_i^2 \leq w_i^1$ for $k = (1, 1), (1, 1), (1, 1)$), the strategy $v_i$ has to guarantee in the corresponding restricted games that player 2’s reward is not more than $v_i^1$. Also, $v_i$ has to leave $E_{(1, 1)}$, belonging to Case 3 (i.e., $v_i(1, 1) > w_i^2(1, 1)$). It is easy to see that

$$
y_{(1, 1)} = (1, 0), \quad y_{(1, 2)} = (0, 1), \quad y_{(2, 1)} = (1, 0), \quad y_{(2, 2)} = (0, 1), \quad y_{(3, 1)} = y_{(3, 2)} = (1)
$$

satisfies all these requirements.

The previous lemma (and its proof) has useful consequences.

**Corollary 3.2.** The results of Lemmas 3.2 to 3.4 in §3.2.2 for simple aperiodic product-games are also valid for any general aperiodic product-game $G$. Lemma 3.5 extends as well if one interprets $z_i^1$ as the min max level $v_i$ of player $i$ in the restricted game $\tilde{G}$.

Also, the infimum in expression (6) of the min max levels is attained at stationary strategies for all product-games. This is stated next.

**Corollary 3.3 (of the proof of Lemma 3.6).** Take an aperiodic product-game $G$ and an arbitrary player $i$. Then, players $-i$ have a joint stationary strategy $x^{-i}$ which guarantees that player $i$’s reward from any initial state $s \in S$ is at most his min max level $v_i$; i.e., for all strategies $\pi^i$ for player $i$ we have

$$
\gamma_i^i(\pi^i, x^{-i}) \leq v_i.
$$

With the help of this corollary, we are now ready to prove Main Theorem 2, which claimed that, in every two-player aperiodic zero-sum product-game, both players have a stationary 0-optimal strategy.

**Proof of Main Theorem 2.** Take an arbitrary two-player aperiodic zero-sum product-game, and take player $i = 1$. By Corollary 3.3, there exists a stationary strategy $x^{-1}$ for player 2 (as $-1 = \{2\}$) which guarantees that player 1’s reward is not more than $v_i$ for any initial state $s \in S$. Hence, $x^{-1}$ is 0-optimal for player 2. One finds similarly a stationary 0-optimal strategy for player 1, which completes the proof.

**3.2.4. The construction of 0-equilibria in product-games.** In the previous sections we achieved several results for the min max levels of aperiodic product-games. We will use this knowledge now to construct 0-equilibria in aperiodic product-games.

The following lemma states that, in any restricted game, there exists a 0-equilibrium in which, if no player deviates, the players’ future expectations remain unchanged during the whole play. The reasoning remains valid for all irreducible stochastic games. Note, however, that Flesch et al. [7] (with 3 players) and Simon [16] (with only 2 players) constructed examples proving that such a result would not hold for all stochastic games in general.

**Lemma 3.7.** Let $G$ be an arbitrary aperiodic product-game and consider the restricted game $\tilde{G}_k$, for any $k = (k^1, \ldots, k^k) \in K$. Then, there exists a 0-equilibrium $\pi$ in $\tilde{G}_k$ such that the corresponding rewards are independent of the initial state and all the continuation rewards remain unchanged with probability 1 during the whole play. More precisely, the reward $\gamma_i(\pi(h))$ is independent of the initial state $s \in E_k$ and the history $h$, given $h$ occurs with a positive probability with respect to $\pi$. Here $\gamma_i$ denotes the average reward for the restricted game $\tilde{G}_k$. 
Proof. Observe the following for the game \( \widetilde{G}_k \).

(i) The set of feasible rewards (i.e., the rewards that can be obtained by some joint strategy) is the same from any initial state in \( E_k \). This is an immediate consequence of the fact that, as \( E_k \) is an aperiodic joint maximal communicating set, the players can move from any state in \( E_k \) to any other one in \( E_k \), possibly in a number of steps.

(ii) The extreme points of the set of feasible rewards are induced by pure stationary strategies. Indeed, for any extreme point \( z \), Dutta [5] (cf., the appendix) shows that there exists a joint pure stationary strategy \( x \) which induces reward \( z \) for some fixed initial state \( s \in E_k \). As \( \gamma(x) = Q(x)\gamma(x) \) (as mentioned in (5)), \( \gamma(x) \) is a convex combination of the rewards for \( x \) within the ergodic sets which can be reached from state \( s \) with a positive probability. Take such an ergodic set \( F \subset E_k \). Then, because of property (i) and the choice of \( z \), the joint strategy \( x \) must yield reward \( z \) within \( F \). Since \( E_k \) is a joint maximal communicating set, there exists a joint pure stationary strategy \( y \) such that \( y \) coincides with \( x \) on \( F \), and moreover, play reaches \( F \) from any initial state in \( E_k \). Clearly, \( y \) induces reward \( z \) from all initial states in \( E_k \).

(iii) Each min max level in \( \widetilde{G}_k \) is a constant \( \bar{v}_k \), by Lemma 3.1.

Given these three observations, this game situation is almost identical to a repeated game. The following ideas and arguments are standard in Folk theorems for repeated games. For the context of stochastic games, we refer to Dutta [5]. Take an arbitrary feasible reward \( z_k = (z^1_k, \ldots, z^L_k) \) such that \( \bar{v}^i_k \geq \bar{v}_k \) for all players \( i \).

Let \( \sigma \) be the pure joint strategy which prescribes to play as follows: play \( a_i \) for \( d^1_i \) stages, then \( a_2 \) for \( d^2_i \) stages, \ldots, then \( a_{L_i} \) for \( d^L_i \) stages, and repeat this with lengths \( d^1_i, \ldots, d^3_i \), then with lengths \( d^1_i, \ldots, d^2_i \), and so on. The lengths \( d^m_i \) have to be chosen in such a way that, when \( m \) tends to infinity, then we have for each \( l \in \{1, \ldots, L_i\} \) that \( d^m_i \) goes to infinity, so that the expected average payoff when strategy \( a_i \) is played for \( d^m_i \) stages will approach \( \bar{y}(a_i) \); (b) \( d^m_i / (d^m_i + \cdots + d^m_{L_i}) \) tends to \( \alpha_i \), so that \( a_i \) is played in the right proportion of time; (c)

\[
\frac{d^1_i + \cdots + d^m_i}{d^1_i + \cdots + d^m_i + \cdots + (d^m_L + \cdots + d^m_{L_i}) + (d^m_i + \cdots + d^m_{L_i})} \to \frac{1}{L_i}
\]
tends to 0, so that the average payoffs will fluctuate less and less. Because of these three properties, \( \sigma \) induces reward \( z_k \), and moreover, any continuation reward is also \( z_k \); i.e., \( \bar{y}(\sigma[h]) = z_k \) for all states \( s \in E_k \) and for all histories \( h \). Let \( \pi \) be the joint strategy which prescribes to play \( \sigma \), unless some player \( i \) deviates from the action prescribed by \( \sigma^i \). In that case, from the new state, players \( i \) should switch to a joint stationary strategy \( x^i \) as in Lemma 3.1. Since the players receive \( z_k \geq \bar{v}_k \) according to \( \sigma \), while if a player \( i \) deviates then his reward is not more than \( \bar{v}_k \), the joint strategy \( \pi \) is a 0-equilibrium and satisfies the requirements of the lemma. ☐

Now we are sufficiently prepared to prove Main Theorem 1, which claimed that, in any aperiodic product-game \( G \), there exists a 0-equilibrium.

Proof of Main Theorem 1. For the idea of the construction, we refer to §3.1.

Let \( S^0 \) denote the set of states \( s = (s^1, \ldots, s^m) \in S \) in which \( s^i \) is of type 1 for at least one player \( i \). Recall that, on each \( E_k, k \in K \), the min max values in \( G \) and in \( \tilde{G} \) are both a constant \( \bar{v}_k \) and \( \bar{v}_k \), respectively, due to Lemma 3.1 and Corollary 3.1.

In every restricted game \( \tilde{G}_k \) with \( k \in K^* \), take a 0-equilibrium \( \sigma_k \) as in Lemma 3.7. Let \( z^i_k \) denote the corresponding reward for any player \( i \), which is independent of the initial state on \( E_k \). Then, as \( k \in K^* \), we have \( z^i_k \geq \bar{v}_k \geq \bar{v}_k \) for all players \( i \).

In every restricted game \( \tilde{G}_k \) with \( k \in K^* \), take a state \( s^i_k \in E^i_k \), and an “exit” action \( a^i_k \) for player \( i \) in state \( s^i_k \) as in part (1) of Lemma 3.5 (with Corollary 3.2).

The proof of Main Theorem 1 consists of the following steps. In Step 1, we construct a joint stationary strategy \( x^* \), which is used to reach the joint maximal communicating sets \( E_k \) with \( k \in K^* \). Then, in Step 2 we “extend” \( x^* \) to the joint strategy \( \pi \) according to which the players also receive rewards \( z_k \) in the sets \( E_k, k \in K^* \). Finally, in Step 3, we will complete the proof by showing that \( \pi \) supplemented with some joint stationary strategies \( y^i \), for all \( i \), forms a 0-equilibrium.

Step 1: The construction of the joint stationary strategy \( x^* \) and a number of properties of \( x^* \). First, two supplementary games \( G \) and \( G^* \) have to be constructed. Then, given \( G^* \), the joint strategy \( x^* \) will be found as a stationary 0-equilibrium in \( G^* \).
Step 1.1: The simple aperiodic product-game $\tilde{G}$. For $k \in K - K^*$, let $z^*_k := \bar{v}_k$ for all players $i$. Then, let $\tilde{G}$ denote the simple aperiodic product-game which is derived from $G$ by replacing each player $i$’s payoffs by $z^*_k$ in any restricted game $\tilde{G}_k$, $k \in K$.

Step 1.2: The stochastic game $G^*$. In this step, we will derive a stochastic game $G^*$ (not necessarily a product-game) from $\tilde{G}$. First, for every state $s$ which belongs to some joint maximal communicating set $E_k$, $k \in K$, fix an arbitrary completely mixed action $y^*_i$ for every player $i$ on $A_i$. Now, we define a subset $X^{*i}_s \subset X^{i}_s$ of mixed actions for every player $i$ in every state $s = (s^1, \ldots, s^k) \in S$ as follows:

Case (1): $s \in E_k$ with $k \in K^*$. In this case, we let $X^{*i}_s := \{y^*_i\}$ for all players $i$.

Case (2): $s \in E_k$ with $k \in K[i]$. In this case, for players $j \neq i$, we let $X^{*j}_s := \{y^*_i\}$. As for player $i$, if $s' = s^i$, then we let $X^{*i}_s := \{y^*_i\}$, while if $s' \neq s^i$, then we let $X^{*i}_s := \{y^*_i\}$.

Case (3): $s \in S^o$. In this case, we let $X^{*i}_s := X^{i}_s$ for all players $i$.

Note that $X^{*i}_s = X^{i}_s$ or $X^{*i}_s$ is a singleton, for any player $i$ in any state $s \in S$, and also that joint strategies in $X^*$ can only differ in states $s \in S^o$. Moreover, the ergodic sets for all $x \in X^*$ are precisely $E_k$, $k \in K^*$, due to the use of the “exit” actions, which eventually make play leave each $E_k$ belonging to Case (2).

Let $G^*$ denote the stochastic game which is derived from $\tilde{G}$ by restricting each player $i$ in each state $s \in S$ to the space $X^{*i}_s$ of mixed actions.

Step 1.3: Defining $x^*$ as a stationary 0-equilibrium of $G^*$ and proving a number of properties of $x^*$. As the ergodic sets are the same for all $x \in X^*$, Lemma A.1 in the appendix yields a stationary 0-equilibrium $x^* \in X^*$ for the game $G^*$. Obviously, $x^*$ is also a joint stationary strategy in the game $\tilde{G}$ and in the original game $G$, but not necessarily a 0-equilibrium.

We wish to point out three properties of $x^*$, with a proof.

Property (1): If $s \in S^o$, then no player $i$ can go to better states regarding his reward in $\tilde{G}$ by unilaterally deviating from $x^*_i$; i.e., for every action $b^r_i \in A^r_i$ we have

$$\sum_{t \in S} p_{s, (t', x^{r-1})}(t) \bar{v}_i(t) \leq \sum_{t \in S} p_{s, x^*_i}(t) \bar{v}_i(t),$$

where $\bar{v}_i$ denotes the average reward to player $i$ in the game $\tilde{G}$.

Property (2): If $s \in E_k$ for some $k \in K$, then no player $i$ can improve on his expected min max level in $G$ by unilaterally deviating from $x^*_i$; i.e., for every action $b^r_i \in A^r_i$ we have

$$\sum_{t \in S} p_{s, (t', x^{r-1})}(t) v^r_i \leq \sum_{t \in S} p_{s, x^*_i}(t) v^r_i.$$

Consequently, equality (7) also yields

$$v^r_i \leq \sum_{t \in S} p_{s, x^*_i}(t) v^r_i.$$  \hspace{1cm} (14)

Property (3): $x^*$ yields rewards in $\tilde{G}$ that are at least the min max levels in $G$; i.e., $\bar{v}(x^*) \geq v^r_i$ for all players $i$ and for all initial states $s \in S$.

Now, we prove the proofs for these properties.

Proof of Property (1): This property follows from the fact that $x^*$ is a 0-equilibrium in $G^*$, and no player is restricted in $G^*$ in state $s$.

Proof of Property (2): This property is a consequence of parts (1) of Lemmas 3.4 and 3.5 (with Corollary 3.2).

Proof of Property (3): This property requires a longer argument. By applying Equation (7) for the game $G$, in every state $s \in S^o$, there exists a mixed action $x^*_i \in X^*_i = X^i$ for player $i$ such that

$$\sum_{t \in S} p_{s, (t', x^{r-1})}(t) v^r_i \geq v^r_i.$$  \hspace{1cm} (15)

Given these mixed actions $x^*_i$ in all $s \in S^o$, there is a unique extension (with the mixed actions prescribed by $x^*$ in all $E_k$, $k \in K$) to a stationary strategy $x^*$ in $X^*$. Then, by inequality (14), we obtain $P(x^*, x^{r-1}) v^r_i \geq v^r_i$. Hence, $P^m(x^*, x^{r-1}) v^r_i \geq v^r_i$ for all $m \in \mathbb{N}$, and therefore $Q(x^*, x^{r-1}) v^r_i \geq v^r_i$.

Note that the ergodic sets are precisely $E_k$, $k \in K^*$, for both $(x^*, x^{r-1})$ and $x^*$, as both belong to $X^*$ (cf., Step 1.2). Hence, if $s$ is recurrent for $(x^*, x^{r-1})$, then (as $x^*$ equals $x^{r-1}$ on all recurrent states) we have

$$\bar{v}_i(x^*, x^{r-1}) = \bar{v}_i(x^*) = z^*_k \geq \bar{v}_k \geq v^r_i = v^r_i,$$

where $\bar{v}_i$ denotes the average reward to player $i$ in the game $\tilde{G}$. 


By applying equality (5), we now obtain
\[
\gamma_i^j(x^*, x^{*-}) = Q(x^*, x^{*-}) \cdot \gamma_i^j(x^*, x^{*-}) \geq Q(x^*, x^{*-}) \cdot v^j \geq v^j, 
\]
where the first inequality follows from (16) and from the fact that entry \((t, s)\) of the stochastic matrix \(Q(x^*, x^{*-})\) is only positive if state \(s\) is recurrent for \((x^*, x^{*-})\). Since \(x^{*-}\) is a best reply to \(x^{*-}\) in \(G^*\) and since \(x^* \in X^*\), we have
\[
\gamma_i^j(x^*) \geq \gamma_i^j(x^*, x^{*-}) \geq v^j
\]
for all initial states \(s \in S\), which proves Property (3).

**Step 2. The construction of the joint strategy \(\pi\) for the original game \(G\).** Given \(x^*\) from Step 1, the definition of \(\pi\) is easy. Let \(\pi\) be the joint strategy which prescribes to play as follows:

- **Case (1):** when play enters a set \(E_k\) with \(k \in K^*\). In this case, the players switch to the joint strategy \(\sigma_k\) (cf., begin of the proof).
- **Case (2):** when play enters a set \(E_k\) with \(k \in K[i]\). In this case, players \(-i\) switch to a joint stationary strategy as in Lemma 3.1, while player \(i\) follows \(x^*\); i.e., plays the mixed action \(x^*_i\) in state \(s \in E_k\).
- **Case (3):** in any state \(s \in S^*\). In this case, each player \(i\) follows \(x^*\); i.e., plays the mixed action \(x^*_i\).

Note that play according to \(\pi\) leaves all sets \(E_k\) belonging to Case (2), with the guidance of \(x^*\). On the other hand, in a set \(E_k\) in Case (1), by switching to \(\sigma_k\), each player \(i\) receives reward \(v^j\) in the game \(G\), which is exactly what the players would receive within \(E_k\) according to \(x^*\) in the game \(G\). So in some sense, \(x^*\) is used to reach the sets \(E_k\), \(k \in K^*\), and then \(\sigma_k\) is used to induce the “right” payoffs.

Thus, by Property (3) of Step 1.3
\[
\gamma_i^j(\pi) = \gamma_i^j(x^*) \geq v^j \tag{17}
\]
for all initial states \(s \in S\) and for all players \(i\), which means that \(\pi\) induces individually rational rewards in \(G\).

**Step 3. Proving that \(\pi\) supplemented with the joint stationary strategies \(y^{-i}\), for all players \(i\), is a 0-equilibrium.** For any player \(i\), in view of Corollary 3.3, we may take a joint stationary strategy \(y^{-i}\) for players \(-i\) such that for all initial states \(s \in S\) and for all strategies \(\tau^i\) for player \(i\) we have
\[
\gamma_i^j(\tau^i, y^{-i}) \leq v^j. 
\]

Let the joint strategy \(\eta\) be defined as in §3.1 Note that the expected rewards are equal with respect to \(\eta\) and with respect to \(\pi\), hence by (17),
\[
\gamma_i^j(\eta) = \gamma_i^j(\pi) = \gamma_i^j(x^*) \geq v^j 
\]
for all initial states \(s \in S\) and for all players \(i\). Notice also that if \(h\) denotes a history and \(s \in S\) a state such that, with a positive probability, \(h\) can occur and \(s\) can be the present state after \(h\) with respect to \(\eta\) (or equivalently with respect to \(\pi\)), then
\[
\gamma_i^j(\eta[h]) = \gamma_i^j(\pi[h]) = \gamma_i^j(\pi) = \gamma_i^j(x^*) \geq v^j, \tag{18}
\]
where for the second equality we used that for \(\sigma_k\), the “continuation rewards” remain the same due to Lemma 3.7. We may thus conclude that \(\eta\) yields individually rational rewards in \(G\), for all players \(i\), and for such histories \(h\) and states \(s \in S\).

It remains to show that \(\eta\) is a 0-equilibrium in \(G\).

**Deviations inside the support of \(\eta\):** Here, we only consider deviations by playing actions with a positive probability according to \(\eta\). We show that such deviations by a player cannot improve his expected reward. Indeed, (i) within a set \(E_k\), \(k \in K^*\) (cf., Case (1) in Step 2), the players play the 0-equilibrium \(\sigma_k\) in \(\bar{G}_k\), (ii) within a set \(E_k\), \(k \in K[i]\) (cf., Case (2) in Step 2), such a deviation by players \(-i\) (who do not make the “exit”) would not change the probability of eventually moving to another set \(E_{k'}\), (iii) within a set \(E_k\), \(k \in K[i]\) (cf., Case (2) in Step 2), player \(i\) has an incentive to “exit,” since within \(\bar{G}_k\), he can get at most \(v^j\), while \(v^j = v^j\) (iv) in states \(s \in S^*\) (cf., Case (3) in Step 2), no player \(i\) can go to better states regarding his reward according to the equalities in (18) and to Property (1) from Step 1.3.

**Deviations outside the support of \(\eta\):** Consider now a deviation when, for the first time, say after history \(h\) in state \(s\), while the players should play a joint mixed action \(x^*_i\) according to \(\eta\), some player \(i\) deviates and plays an action \(b^j_i\) which has probability zero according to \(\eta^j\); i.e., \(x^j_i(b^j_i) = 0\). This deviation is immediately noticed by players \(-i\) and, according to \(\eta\), they switch to the joint stationary strategy \(y^{-i}\) from the next state, say state \(t\).
Consequently, player $i$’s reward will be at most $v'_i$ in expectation. Obviously, without deviation player $i$ would receive reward $γ'_i(η[h]) = \gamma'_i(x')$, in view of (18). Now, observe the following.

(A) Suppose $s ∈ S^0$. Then, $x'_i = x^*_i$, and player $i$’s expected reward after this deviation is at most

$$\sum_{t ∈ S} p_s(\theta_t, x'_i, x_i)(t)v'_i ≤ \sum_{t ∈ S} p_s(\theta_t, x'_i, x_i)(t)\gamma'_i(x'_i) ≤ \sum_{t ∈ S} p_s(\theta_t, x_i)(t)\gamma'_i(x_i) = γ'_i(η[h]),$$

where the inequalities follow from properties (3) and (1) in Step 1.3, respectively, while the equations follow from (4) and (18). Hence, the deviation is not profitable.

(B) Suppose $s ∈ E_k$ for some joint maximal communicating set $E_k$. Hence, player $i$’s expected reward after this deviation is at most

$$\sum_{t ∈ S} p_s(\theta_t, x'_i, x_i)(t)v'_i ≤ \sum_{t ∈ S} p_s(\theta_t, x_i)(t)v'_i ≤ γ'_i(η[h]).$$

Here, the first inequality follows from part (1) of Lemma 3.4 for $k ∈ K^*$, and from Property (2) in Step 1.3 if $k ∈ K − K^*$. The second inequality is a consequence of (18). Hence, the deviation is not profitable again.

In conclusion, no deviation is profitable, and $η$ is a 0-equilibrium in $G$. This completes the proof of Main Theorem 1. □

Remark 3.1. It remains unclear whether 0-equilibria always exist within the class of stationary strategies. This question is already challenging in the situation when each player $i$’s state space $S^i$ is just one maximal communicating set (precisely the situation we have in a restricted game). Even though, Corollary 3.2 would yield that all min max levels are constant on the whole state space $S$, it is still not evident how one should get a grip on the problem.

Finally, let us revisit Example 1. As we know, the min max levels of this game coincide with the min max levels of the game in Example 2, hence

$$v^1(1, 1) = v^1(1, 2) = v^1(2, 1) = v^1(2, 2) = v^1(3, 2) = 1, \quad v^1(3, 1) = 0,$$

while $v^2 = −v^1$. For the min max values of the restricted games, we refer to §3.2.1 Hence, as to which restricted games are satisfactory, we obtain $K^* = \{(II, I), (II, II)\}$, while $K[1] = \{(I, II)\}$ and $K[2] = \{(I, I)\}$.

Now consider the pure stationary strategy $x^1$ for player 1 defined as

$$x^1(1, 1) = (1, 0), \quad x^1(1, 2) = (0, 1), \quad x^1(2, 1) = (1, 0), \quad x^1(2, 2) = (0, 1), \quad x^1(3, 1) = x^1(3, 2) = (1),$$

and the pure stationary strategy $x^2$ for player 2 given as

$$x^2(1, 1) = x^2(2, 1) = (0, 1), \quad x^2(3, 1) = (1, 0), \quad x^2(1, 2) = x^2(2, 2) = x^2(3, 2) = (1).$$

This pair $(x^1, x^2)$ actually could play the role of $π$ in this example. Indeed, in each restricted game $G_k$, $k ∈ K^*$, the pair $(x^1, x^2)$ lets the players play a 0-equilibrium, while $x^1$ leaves $E_{(I, II)}$ and $x^2$ leaves $E_{(I, I)}$. Notice that no threat strategies are needed here, so $(x^1, x^2)$ is a 0-equilibrium.

### 4. Periodic product-games.

The previous sections dealt with aperiodic product-games. When we allow for periodic maximal communicating sets, the situation changes. Take for example a product-game with two players in which the Markov transition structure for either player is as follows: the state space is $\{1, 2\}$, there is only one action in either state, and this action leads to the other state with probability 1. So in the product-game, depending on the initial state, play moves back and forth either between states $(1, 1)$ and $(2, 2)$ or between states $(1, 2)$ and $(2, 1)$. This game is periodic, of course. Suppose the payoffs for either player are 1 in states $(1, 1)$ and $(2, 2)$, and 0 in states $(1, 2)$ and $(2, 1)$. Then, a solitary move for player 1 in state $(1, 2)$ would lead to state $(2, 2)$, improving player 1’s payoff. Hence, the important Lemma 3.2 is no longer valid for periodic product-games, and the proof in the previous sections are not directly applicable. Notice also that this game has two joint maximal communicating sets; i.e., $\{(1, 1), (2, 2)\}$ and $\{(1, 2), (2, 1)\}$, but neither of them can be written as a product of the form $E^1 × E^2$. This entails additional difficulties. Nevertheless, we conjecture that the main results of this paper extend to the periodic case as well.

### Appendix.

**Lemma A.1.** In a stochastic game, if the ergodic sets are the same for all joint stationary strategies, then there exists a stationary 0-equilibrium.
PROOF. For a joint stationary strategy \( x \in X \), consider the \( \beta \)-discounted reward, with \( \beta \in (0, 1) \), defined for player \( i \) and initial state \( s \in S \) as

\[
y^i_{\beta_m}(x) := (1 - \beta) \sum_{m=1}^{\infty} \beta^{m-1} \mathbb{E}_{x_s}(R^i_m),
\]

where \( R^i_m \) is the random variable for the payoff for player \( i \) at stage \( m \), and where \( \mathbb{E}_{x_s} \) stands for expectation with respect to initial state \( s \) and joint strategy \( x \). Fink [9] and Takahashi [22] showed that, for every \( \beta \in (0, 1) \), there exists a stationary 0-equilibrium with respect to the \( \beta \)-discounted rewards.

As the ergodic sets are the same for all joint stationary strategies, it is known (cf., Lemma 2.7.6 in Flesch [6]) that for any sequence of discount factors \( \beta_m \) converging to 1 and joint strategies \( x_m \) converging to \( x \) we have

\[
y^i_j(x) = \lim_{m \to \infty} y^i_{\beta_m}(x_m)
\]

for all states \( s \in S \) and players \( i \).

We will now work with a number of sequences in compact spaces. By taking subsequences, we may assume that all these sequences have limits. Let \( \beta_m \) be a sequence of discount factors converging to 1, and for any \( m \in \mathbb{N} \), let \( x_m \) be a stationary \( \beta_m \)-discounted 0-equilibrium. Let \( x = \lim_{m \to \infty} x_m \). We will show that \( x \) is a 0-equilibrium with respect to the average reward.

Take an arbitrary player \( i \) and a stationary best reply \( y^i \) to \( x^{-i} \). Then for any initial state \( s \in S \), from (19) and from the fact that \( x_m \) is a \( \beta_m \)-discounted 0-equilibrium, it follows that

\[
y^i_j(y^i, x^{-i}) = \lim_{m \to \infty} y^i_{\beta_m}(y^i, x_m^i) \leq \lim_{m \to \infty} y^i_{\beta_m}(x_m) = y^i(x).
\]

As \( y^i \) is a best reply to \( x^{-i} \), the joint strategy \( x \) is a stationary 0-equilibrium with respect to the average reward indeed. \( \square \)

Acknowledgments. The authors thank Eilon Solan and an anonymous referee for their constructive suggestions, which helped improve the presentation and relate the paper to existing literature.

References