Communicating Zero-Sum Product Stochastic Games

Tristan Garrec*

August 7, 2018

Abstract

We study two classes of zero-sum stochastic games with compact action sets and a finite product state space. These two classes assume a communication property on the state spaces of the players. For strongly communicating on one side games, we prove the existence of the uniform value. For weakly communicating on both sides games, we prove that the asymptotic value, and therefore the uniform value, may fail to exist.

1 Introduction

1.1 Problem

In a zero-sum stochastic game, two players interact repeatedly at discrete times, with opposite interests. At each stage, players face a zero-sum game given by the state of nature which evolves according to the current state, and the pair of actions players choose given the history. Therefore, the actions played at each stage impact both the payoff today and the law of the state of nature tomorrow. Players intend to optimize their expected overall payoff. The $n$-stage repeated and the $\lambda$-discounted games are the games in which the overall payoffs are respectively the Cesàro and Abel means of the stage payoffs. Under mild assumptions both games have a value denoted respectively $v_n$ and $v_\lambda$.

A fundamental question arising in the theory of dynamic games is the asymptotic behavior of these values. We shall focus on the two following approaches of this issue. The asymptotic approach studies the convergence of the values of the $n$-stage repeated game and the $\lambda$-discounted game, as $n$ goes to infinity and $\lambda$ goes to 0, that is as players become more patient. If these quantities converge and are equal, the game is said to have an

---

*Toulouse School of Economics, Université Toulouse 1 Capitole. E-mail address: tristan.garrec@ut-capitole.fr
asymptotic value. The uniform approach is dedicated to the existence, for both players, of strategies that are $\varepsilon$-optimal in every $n$-stage repeated game, provided that $n$ is large enough. If such strategies exist, players are able to play optimally in every game long enough without knowing the length of the game. In that case, the game is said to have a uniform value. While the existence of the uniform value implies the existence of the asymptotic value, the converse is not true.

The aim of this paper is to study the asymptotic and uniform values in two classes of zero-sum product stochastic games. A product state space is of the form $X \times Y$. Moreover, players control the transitions on their own components of the state space, that is the next state in $X$ only depends on the current state in $X$ and the action of player 1, and similarly for $Y$. We consider the case where $X$ and $Y$ are finite, and action sets are compact. The two classes we are interested in assume a communication property on the state spaces of the players. These are called the strong and the weak communication properties — the strong communication property implying the weak one.

The first class is the class of strongly communicating on one side zero-sum product stochastic games. In such games, for one player, there exists a fixed time $T$ such that independently of his choice of strategy, there is a positive probability of passing from any state to any other state in his component of the state space in exactly $T$ stages. This assumption, of ergodic nature, implies that the current state of the player having the strong communication property has in the long run little importance.

The second class is the class of weakly communicating on both sides zero-sum product stochastic games. In such games, for both players there exists a time $T$ and a strategy such that, for any two states in their components of the state space, they can move from one to the other with positive probability in exactly $T$ stages. Thus players totally control the dynamics on their components of the state space.

1.2 Contribution

Strongly communicating on one side and weakly communicating on both sides zero-sum product stochastic games have, to our knowledge, never been studied before.

For strongly communicating on one side games with finite state space and compact action sets, we prove the existence of the uniform value, which does not depend on the initial state of the player having the strong communication property. Furthermore we prove that this player has $\varepsilon$-optimal strategies that have a simple structure. We call them Markov periodic strategies. Under these strategies, the action chosen at each stage does not depend on the whole history but only on the current state and stage modulo the period. Hence they are a particular case of Markov strategies, for which
actions chosen depend only on the current state and stage, and are more
general than stationary strategies, which only depend on the current state.
Our proof is based on a classification of the state space of the player who
does not have the strong communication property. This decomposition relies
on recurrent classes induced by stationary strategies. It has been introduced
for Markov Decision Processes (MDP) by Ross and Varadarajan [18], similar
classifications have been used in Bather [2], Solan [20] and Flesch et al. [9].
Building on that classification, we consider a family of auxiliary stochastic
games and prove that they have a uniform value independent of the initial
state. Finally we build an auxiliary MDP for the player who does not have
the strong communication property, whose payoffs are the uniform values
of the previous auxiliary games. We conclude by proving that the uniform
value of the MDP is also the uniform value of the initial game.

Regarding weakly communicating on both sides games with finite state
space and compact action sets, we provide an example of a game which does
not have an asymptotic value (and hence neither has a uniform one). Our
proof is based on a reduction of this example to a simpler game of perfect
information with two absorbing and two non absorbing states introduced by
Renault [16]. A key element in the non-existence of the asymptotic value is
the non semi-algebraic aspect of the action set of one player.

1.3 Related literature
Zero-sum stochastic games were introduced by Shapley [19] in the finite
setting (finite state and action sets), for which he proved the existence of
the value in the $\lambda$-discounted game. Mertens and Neyman [13] proved the
existence of the uniform value. Their proof is based on the fact that the
value of the $\lambda$-discounted game has bounded variations in $\lambda$, as shown by
Bewley and Kohlberg [3]. It is a key question whether the existence of the
uniform value extends to non finite zero-sum stochastic games.

This question has been answered positively for several classes of zero-
sum stochastic games with a finite state space and non finite action sets, as
we consider in this paper. For absorbing games (Mertens et al. [14]) and
recursive games (Li and Sorin [12]), the proofs use the operator approach of
Rosenberg and Sorin [17] that relies on the Shapley operator which entirely
contains the dynamics of the game. Still for a finite state space, Bolte et
al. [5] showed that games with semi-algebraic (or more generally definable)
transitions and actions set have a uniform value. Finally, Renault [15] proved
the existence of the uniform value in MDPs with a finite state space and
arbitrary action set.

However, in the last few years, several counterexamples to the existence
of the asymptotic value in zero-sum stochastic games with finite state space
and compact action sets have been proposed (Vigeral [23], Sorin and Vigeral
[22] and Ziliotto [26]), ending the long standing idea that such games had an
asymptotic value. Laraki and Renault [11] provided such a counterexample with a product state space. All these counterexamples have in common non semi-algebraic transition probability or non semi-algebraic action sets.

Zero-sum stochastic games on a product state space have been introduced by Altman et al. [1], who examined the case where each player only observes his component of the state and his actions, and showed that these games can be solved by linear programming. Flesch et al. [6, 7] studied equilibria in N-players finite product stochastic games. They however considered an overall payoff which is the limit inferior of the n-stage repeated game payoff, while we are interested in the existence of the uniform value. Finally, Laraki and Renault [11] showed the existence of the asymptotic value in zero-sum product stochastic games under a strong acyclicity condition. This strong acyclicity condition encompasses the irreversibility in the transitions of several classes of repeated games for which the asymptotic value is known to exist. The strong acyclicity condition is incompatible with our weak communication property. It is important to understand in which classes of stochastic games that do not assume an irreversibility condition the asymptotic value may exist.

Gillette [9] introduced games in which independently of the choice of strategies, there is a positive probability of passing from any state to any other state in exactly T stages. He called them cyclic stochastic games. These were also investigated by Hoffman and Karp [10], Bewley and Kohlberg [4] and Vrieze [24]. A similar assumption has also been examined by Fudenberg and Yamamoto [8] for games where players observe the state and a public signal related to the actions played. However in these articles the property is considered on the whole state space and not only on one component of a product state space.

1.4 Organization

The article is organized as follows. In section 2 we describe the model of zero-sum product stochastic games and recall some elementary facts. In section 3 we give formal definitions of strongly and weakly communicating, and state the two main theorems. Finally, sections 4 and 5 are dedicated to the proofs of the two main theorems.

2 Preliminaries on zero-sum product stochastic games

2.1 Model and course of the game

Let X and Y be two nonempty finite sets. Let A and B be two nonempty compact metric sets endowed with their Borel \( \sigma \)-algebras. Let \( p : X \times A \to \Delta(X) \) and \( q : Y \times B \to \Delta(Y) \), be such that for all \( x, x' \in X \) and all \( y, y' \in Y \), \( p(x'|x, \cdot) \) and \( q(y'|y, \cdot) \) are continuous. \( \Delta(X) \) denotes the set of probability
measure over $X$, and similarly for $\Delta(Y)$. Let $u : X \times Y \times A \times B \to [0,1]$, be such that for all $(x, y) \in X \times Y$ and all $a \in A$ and $b \in B$, $u(x, y, \cdot, b)$ and $u(x, y, a, \cdot)$ are continuous.

$X$ is the state space of player 1, $Y$ is the state space of player 2. $A$ is the action set of player 1, $B$ is the action set of player 2. It is without loss of generality that the action sets do not depend on the current state. $p$ is the transition probability of player 1, $q$ is the transition probability of player 2. $u$ is the payoff to player 1.

Let $\Gamma = (X, Y, A, B, p, q, u)$. The game $\Gamma$ is played in stages as follows: an initial state $(x_1, y_1) \in X \times Y$ is given and known by the players. Inductively at stage $n$, knowing the past history $h_n = (x_1, y_1, a_1, b_1, \ldots, x_{n-1}, y_{n-1}, a_{n-1}, b_{n-1}, x_n, y_n)$, player 1 and 2 simultaneously choose an action, respectively $a_n \in A$ and $b_n \in B$. A new state $x_{n+1} \in X$ is selected according to the distribution $p(\cdot|x_n, a_n)$ on $X$ and a new state $y_{n+1} \in Y$ is selected according to the distribution $q(\cdot|y_n, b_n)$ on $Y$. The payoff to player 1 at stage $n$ is $u_n = u(x_n, y_n, a_n, b_n)$.

2.2 Strategies

Since the state space of the game $\Gamma$ is a product of two sets $X$ and $Y$, we distinguish strategies of the players that only depend on their own history, i.e. on their actions and component of the product state space, from strategies depending on the joint history. We denote strategies and set of strategies depending on the joint history with bold letters.

We denote the set of positive integers by $\mathbb{N}^*$. For $n \in \mathbb{N}^*$, let $H^1_n = X \times (A \times X)^{n-1}$ be the set of histories of player 1 at stage $n$ and $H^1_\infty = (A \times X)^\infty$ be the set of infinite histories. $H^1_n$ is endowed with the product $\sigma$-algebra $\mathcal{H}^1_n$, and $H^1_\infty$ with the product $\sigma$-algebra $\mathcal{H}^1_\infty$ spanned by $\bigcup_{n \geq 1} \mathcal{H}^1_n$. Let $\mathcal{S}$ denote the set of behavior strategies of player 1 depending only on his own history. A strategy $\sigma \in \mathcal{S}$ is a sequence $(\sigma_n)_{n \in \mathbb{N}^*}$, where $\sigma_n$ is a measurable map from $(H^1_n, \mathcal{H}^1_n)$ to $\Delta(A)$. A strategy $\sigma$ together with an initial state $x \in X$ define a unique probability distribution over $H^1_\infty$ which we denote $\mathbb{P}^\sigma_x$. We define analogous objects for player 2 and denote $\mathcal{T}$ the set of behavior strategies of player 1 depending only on his own history.

For $n \in \mathbb{N}^*$, let $H_n = X \times Y \times (A \times B \times X \times Y)^{n-1}$ be the set of joint histories at stage $n$ and $H_\infty = (A \times B \times X \times Y)^\infty$ be the set of infinite joint histories. $H_n$ is endowed with the product $\sigma$-algebra $\mathcal{H}_n$, and $H_\infty$ with the product $\sigma$-algebra $\mathcal{H}_\infty$ spanned by $\bigcup_{n \geq 1} \mathcal{H}_n$. Let $\mathcal{S}$ and $\mathcal{T}$ denote the sets of behavior strategies of player 1 and player 2 respectively. A strategy $\sigma \in \mathcal{S}$ is a sequence $(\sigma_n)_{n \in \mathbb{N}^*}$, where $\sigma_n$ is a measurable map from $(H_n, \mathcal{H}_n)$ to $\Delta(A)$, and likewise for $\mathcal{T}$. A pair of strategies $(\sigma, \tau)$ together with an initial state $(x, y)$ define a unique probability distribution over $H_\infty$ which we denote $\mathbb{P}^x_{\sigma, \tau}$.

A strategy is a Markov strategy if the mixed action played at every stage
depends only on the current stage and state. Markov periodic strategies are Markov strategies depending on the stage modulo the period and on the current state. Let us give a formal definition.

**Definition 2.1.** For \( N \in \mathbb{N}^* \), a strategy \( \sigma \) of player 1 is called an \( N \)-periodic Markov strategy if there exists \( (\mu_n)_{n \in \{1, \ldots, N\}} \in \Delta(A)^{X \times Y \times N} \) such that for all \( n \in \mathbb{N}^* \) and all \( h_n = (x_1, y_1, a_1, b_1, \ldots, x_{n-1}, y_{n-1}, a_{n-1}, b_{n-1}, x_n, y_n) \in H_n, \sigma_n(h_n) = \mu'_n(x_n, y_n) \), where \( n' \) is equal to \( n \) modulo \( N \). Markov \( N \)-periodic strategies of player 2 are defined likewise.

Stationary strategies are Markov strategies depending only on the current stage, hence they are Markov 1-periodic strategies. Again, stationary strategies on the product state space \( X \times Y \) (elements of \( \Delta(A)^{X \times Y} \) and \( \Delta(B)^{X \times Y} \) for player 1 and 2 respectively), are denoted with bold letters. Stationary strategies of the players on their components of the state space (elements of \( \Delta(A)^{X} \) and \( \Delta(B)^{Y} \) respectively) are denoted with letters that are not bold.

Finally, a pair of stationary strategies \( (\mu, \nu) \in \Delta(A)^{X \times Y} \times \Delta(B)^{X \times Y} \) induces a Markov chain \( (X_n, Y_n)_{n \geq 1} \) over \( X \times Y \). A state \( (x', y') \in X \times Y \) is said to be accessible from \( (x, y) \) (in \( t \) stages) under \( (\mu, \nu) \) if \( \mathbb{P}_{\mu, \nu}((X_1, Y_1) = (x', y')) > 0 \). More generally, a property is said to hold under \( (\mu, \nu) \) if it holds for the Markov chain induced on \( X \times Y \). We use similar vocabulary from stationary strategies on \( X \) and on \( Y \).

### 2.3 \( N \)-stage and \( \lambda \)-discounted games

For all \( N \in \mathbb{N}^* \), the \( N \)-stage game \( \Gamma_N \) starting in \( (x, y) \in X \times Y \), is the game in which the payoff is \( \gamma_N(\sigma, \tau)(x, y) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}_{\sigma, \tau}^{x,y}(u_n) \) for all \( (\sigma, \tau) \in \mathcal{S} \times \mathcal{T} \). The value of the \( N \)-stage game starting at \( (x, y) \) is denoted \( v_N(x, y) \).

It is characterized by the following recursive equation:

\[
v_{N+1}(x, y) = \max_{\mu \in \Delta(A), \nu \in \Delta(B)} \left[ \frac{1}{N+1} u(x, y, \mu, \nu) + \frac{N}{N+1} \mathbb{E}_{\mu, \nu}^{x,y}(v_N) \right].
\]

Where

\[
\begin{align*}
&u(x, y, \mu, \nu) = \int_{A \times B} u(x, y, a, b) d\mu(a) d\nu(b), \\
&\mathbb{E}_{\mu, \nu}^{x,y}(v_N) = \sum_{x', y' \in X \times Y} v_N(x', y') \int_{A \times B} \int_{A \times B} p(x'|x, a) q(y'|y, b) d\mu(a) d\nu(b).
\end{align*}
\]

Moreover, by proposition 5.3 in Sorin [21], both players have optimal Markov strategies.

For all \( \lambda \in (0, 1] \), the \( \lambda \)-discounted game \( \Gamma_\lambda \) starting in \( (x, y) \in X \times Y \), is the game in which the payoff is \( \gamma_\lambda(\sigma, \tau)(x, y) = \lambda \sum_{n=1}^{\infty} (1-\lambda)^{n-1} \mathbb{E}_{\sigma, \tau}^{x,y}(u_n) \)

for all \((σ, τ) \in S \times T\). The value of the \(λ\)-discounted game starting at \((x, y)\) is denoted \(v_λ(x, y)\). It is characterized by the following fixed point equation:

\[
v_λ(x, y) = \text{val}_{μ}^{\text{val}_{ν}^{\text{val}_{ν}^{\mu}}^{\lambda u(\text{x}, \text{y}, \mu, \nu)} + (1 - \lambda) E_{\mu, \nu}^{\lambda}(v_λ)}.
\]

If \(v_n\) and \(v_λ\) converge as \(n\) goes to infinity and \(λ\) goes to 0, and the limits are equal, then the game is said to have an asymptotic value. The Tauberian theorem of Ziliotto [25] applies in this setting and the asymptotic value exists if \((v_λ)_{λ \in (0,1]}\) converges as \(λ\) goes to 0.

### 2.4 Uniform value and optimal strategies

Fix an initial state \((x, y) \in X \times Y\). Player 1 is said to uniformly guarantee \(v_∞ \in [0,1]\) if

\[
∀ \varepsilon > 0 \exists M ∈ N^* \forall τ ∈ T \forall N ≥ M \gamma_N(σ, τ)(x, y) ≥ v_∞ − \varepsilon.
\]

And similarly for player 2. If both players guarantee \(v_∞\), then it is called the uniform value of the game \(Γ\) starting at \((x, y)\).

Let \(ε ≥ 0\). A strategy \(σ \in S\) is said to be (uniformly) \(ε\)-optimal for player 1 if \(∃ M ∈ N^* \forall τ ∈ T \forall N ≥ M \gamma_N(σ, τ)(x, y) ≥ v_∞ − \varepsilon\). And similarly for player 2.

The next proposition states that if one of the players plays a stationary strategy, then the other player has an \(ε\)-optimal best reply which is also stationary.

**Proposition 2.2.** Let \(μ ∈ Δ(A)^X \times Y\) be a stationary strategy of player 1 in the game \(Γ\). Then, for all \(ε > 0\) there exist \(ν ∈ Δ(B)^X \times Y\) and \(M ∈ N^*\) such that for all \(N ≥ M\) and all \(τ ∈ T\):

\[
γ_N(μ, \nu)(x, y) ≤ γ_N(μ, \tau)(x, y) + ε.
\]

**Proof.** Let \(ε > 0\), \(μ ∈ Δ(A)^X \times Y\) and \((x, y) ∈ X \times Y\). When player 1 plays the stationary strategy \(μ\), player 2 faces a Markov decision process having a uniform value \(w_∞\), and for which he has a uniformly \(ε\)-optimal stationary strategy \(ν\) (consult Sorin [21], corollary 5.26).

Thus, there exists \(M ∈ N^*\) such that for all \(N ≥ M\) one has \(γ_N(μ, \nu)(x, y) ≤ w_∞ + \varepsilon\) and for all \(τ ∈ T\) one has \(γ_N(μ, \tau)(x, y) ≥ w_∞ − \varepsilon\). \(\square\)

### 3 Main results

In the present article, we examine two communication properties on state spaces of the players. In words, a player has the strong communication property if there exists a time \(T\) such that independently of his choice of strategy (depending only on his own history), there is a positive probability of moving from any initial state of his component of the state space to any other state in exactly \(T\) stages.
**Definition 3.1.** Player 1 has the strong communication property if there exists \( T \in \mathbb{N}^* \) such that for all strategies \( \sigma \in \mathcal{S} \) and all states \( x, x' \in X \), one has \( P_{\sigma}^x(X_T = x') > 0 \).

A similar definition stands for player 2. A game is strongly communicating on one side if either player has the strong communication property.

In words, a player has the weak communication property if there exists a time \( T \) and a strategy (depending only on his own history) such that, for any initial state, he can reach any other state in his component of the state space in exactly \( T \) stages, with positive probability.

**Definition 3.2.** Player 1 has the weak communication property if there exists \( T \in \mathbb{N}^* \) and a strategy \( \sigma \in \mathcal{S} \) such that for all states \( x, x' \in X \), one has \( P_{\sigma}^x(X_T = x') > 0 \).

A similar definition stands for player 2. A game is weakly communicating on both sides if both players have the weak communication property.

The next simple example illustrates the notions of strong and weak communication.

**Example 3.1.** In figure below we represent the state space and transition probabilities of player 1. The state space is \( X = \{x, y, z\} \). Transition probabilities are represented by arrows between states. In state \( x \) player 1 chooses \( \alpha \in A \). With probability \( \alpha \) the next state is \( x \) and with probability \( 1 - \alpha \) it is \( y \). In state \( y \) (resp. \( z \)) the next state is \( y \) (resp. \( x \)) with probability 1.

Hence if the action set \( A \) equals \([0,1]\), player 1 has the weak communication property. Whereas if \( A \) equals \([\varepsilon,1]\) for some \( \varepsilon > 0 \), player 1 has the strong communication property.

**Theorem 3.3.** Any strongly communicating on one side zero-sum stochastic game has a uniform value.
Moreover, assuming player 1 has the strong communication property, the uniform value only depends on the initial state of player 2 and for all $\varepsilon > 0$ player 1 has an $\varepsilon$-optimal Markov periodic strategy.

**Theorem 3.4.** There exists a weakly communicating on both sides zero-sum product stochastic game which does not admit an asymptotic value.

The strong communication property implies the weak one. It is sufficient that one player has the strong communication property for the uniform value to exist, while both players having the weak communicating property does not ensure the existence of the asymptotic value.

The proof of 3.3 relies on a classification of the state space of the player who does not have the strong communication property. We then consider a family of auxiliary stochastic games and prove that they have a uniform value independent of the initial state. Finally we build an auxiliary MDP for the player who does not have the strong communication property, whose payoffs are the uniform values of the previous auxiliary games. We conclude by proving that the uniform value of the MDP is also the uniform value of the initial game.

The proof of 3.4 relies on a counter-example. We show that the Shapley equations of this game are also the Shapley equations of a simpler game with two absorbing and two non absorbing states, for which the asymptotic value does not exist.

### 4 Proof of theorem 3.3

We assume in the proof of theorem 3.3 that player 1 has the strong communication property.

In subsection 4.1 we decompose the state space of player 2 in $L$ maximal communicating sets $C_i$ (definition 4.1). These sets, forming a partition of $Y$ (proposition 4.2), allow us in subsection 4.2 to introduce $L$ auxiliary stochastic games $\Gamma_i$ having uniform values $v^\infty_i$ which do not depend on the initial state (proposition 4.6).

In subsection 4.3 we introduce a Markov Decision Process $G$ for player 2 with the same transition probability $q$ as in the initial stochastic game, but with payoff $v^\infty_i$ when the state is in $C_i$. This MDP has a uniform value $w^\infty$, which is also the uniform value of the initial stochastic game $\Gamma$.

In subsection 4.4 we prove that player 2 uniformly guarantees $w^\infty$ in $\Gamma$ (proposition 4.7). The idea is for player 2 to first player optimally in the MDP $G$ (disregarding player 1) and then to switch to an optimal strategy in one of the games $\Gamma_i$.

Finally in subsection 4.5 we prove that player 1 uniformly guarantees $w^\infty$ (proposition 4.10). This is one of the main difficulties of the proof. The natural idea is for player 1 to play optimally in each game $\Gamma_i$. However player
1 does not control the transitions from one $C_i$ to another (which may happen infinitely often). We deal with this issue by letting player 2 play a Markov periodic best reply which prevents the state from jumping infinitely often between maximal communicating sets (lemma 4.9). We conclude showing that the payoff this yields can also be obtained by player 2 as a limit payoff in $G$ (proposition 4.10).

4.1 Classification of states

The state space $Y$ is classified in a similar way to Ross and Varadarajan [18], using recurrent classes induced by stationary strategies on $Y$.

**Definition 4.1.** A subset $C$ of $Y$ is said to be a maximal communicating set if

i) There exists a stationary strategy on $Y$ such that $C$ is a recurrent class of the induced Markov chain on $Y$. Such a strategy is said to be a stationary strategy associated to $C$.

ii) $C$ is maximal, i.e. if there exists $C'$ a subset of $Y$ such that i) holds for $C'$ and $C \subseteq C'$, then $C' = C$.

Let $C_1, \ldots, C_L$ denote the maximal communicating sets. $D$ denotes the set of transient states under every stationary strategy (sometimes just called transient states for short).

**Remark 4.1.** Player 1 having the strong communication property, there is only one maximal communicating set in $X$, which is the whole state space $X$ itself.

Proposition 4.2 below can be found in Ross and Varadarajan [18] when the action space $B$ is finite, since the proof is similar with a compact action space we omit it. This result, showing that the maximal communicating sets and the set of states that are transient under any stationary strategy form partition of $Y$, is fundamental in our proof of theorem 3.3 since it allows us (in the next section) to consider independent auxiliary games over each maximal communicating set.

**Proposition 4.2.** $\{C_1, \ldots, C_L, D\}$ is a partition of $Y$.

The two examples given below allow to show how the state space of player 2 decomposes in maximal communicating sets, and what behavior the state process may have with regard to this decomposition. In particular, they show behaviors of the state process that may happen in that compact action setting and cannot happen in the finite one.

**Example 4.1.** In figure 2 the state space of player 2 is $Y = \{x, y, z\}$ and the action space is $B = [0, 1/2]$. Again, the transitions probabilities are
represented on the arrows between states, e.g. playing $\beta \in B$ in state $y$, the next state is $x$ with probability $\beta$, $y$ with probability $1 - \beta - \beta^2$ and $z$ with probability $\beta^2$. The maximal communicating sets are \{y\} and \{z\} and the set of transient states under every stationary strategy is \{x\}.

![Figure 2: States, actions and transitions of player 2](image)

Remark that, if the initial state is $x$, playing $\beta = \frac{1}{2n}$ in state $y$ at stage $n \geq 1$, player 2 has a positive probability of switching infinitely often between the transient state $x$ and the maximal communicating set \{y\}. This cannot happen when the action space is finite. In that case, for any strategy of player 2, after finitely many stages the process $(Y_n)_{n \geq 1}$ remains forever in one of the maximal communicating sets with probability 1 (see Ross and Varadarajan [18], lemma 2 and proposition 2).

The next example is derived from the previous one and shows another possible behavior of the state process in the compact action setting.

**Example 4.2.** Here the state space of player 2 is $Y = \{x, y, z\}$ and the action space is $B = [0, 1/2]$. The maximal communicating sets are \{x\}, \{y\} and \{z\}.

![Figure 3: States, actions and transitions of player 2](image)

Again, remark that if the initial state is $x$, playing $\alpha = 1/2$ in state $x$ and $\beta = \frac{1}{2n}$ in state $y$ at stage $n \geq 1$, player 2 has a positive probability of switching infinitely often between maximal communicating sets \{x\} and \{y\}.

As we mentioned earlier, this behavior the state of player 2 may have gives rise to a difficulty since player 1 cannot control the jumps of player’s
2 state between maximal communicating sets or transient states. This difficulty is defused by letting the players play Markov periodic strategies.

Another way of defining maximal communicating sets is via pairs of stationary strategies of player 1 and 2 on $X \times Y$, as in the following definition.

**Definition 4.3.** A subset $C$ of $X \times Y$ is said to be a joint maximal communicating set if

i) There exists a pair of stationary strategies in $\Delta(A)^{X \times Y} \times \Delta(B)^{X \times Y}$ such that $C$ is a recurrent class of the induced Markov chain on $X \times Y$.

ii) $C$ is maximal, i.e. if there exists $C'$ a subset of $X \times Y$ such that i) holds for $C'$ and $C \subseteq C'$, then $C' = C$.

Advantageously, maximal communicating sets and joint maximal communicating sets match in following sense.

**Proposition 4.4.** The joint maximal communicating sets are $X \times C_1, \ldots, X \times C_L$, and the set of states that are transient under any pair of stationary strategies is $X \times D$.

**Proof.** Let $i \in \{1, \ldots, L\}$. $X \times C_i$ is a recurrent class in $X \times Y$ for some pair of stationary strategies of player 1 on $X$ and player 2 on $Y$ (just take the stationary strategy of player 2 on $Y$ associated to $C_i$ and any stationary strategy of player 1 on $X$).

Let us show that $X \times C_i$ is maximal. Suppose that there exists a subset $C$ of $X \times Y$ such that $C$ is a recurrent class under a pair of stationary strategies $(\mu, \nu) \in \Delta(A)^{X \times Y} \times \Delta(B)^{X \times Y}$ and $X \times C_i \subseteq C$.

Let $C' = \{y \in Y \mid \exists x \in X \ (x, y) \in C\}$ be the projection of $C$ over $Y$. For all $y \in C'$, define $C_y = \{x \in X \mid (x, y) \in C\}$. Finally, define $\nu \in \Delta(B)^Y$ by, for all $y \in C'$

$$\nu(y) = \frac{1}{|C_y|} \sum_{x \in C_y} \nu(x, y),$$

and arbitrarily outside $C'$.

$C'$ is closed under $\nu$ and every state in $C'$ is accessible from any other state. Hence $C'$ is a recurrent class. Thus $C' \subseteq C_i$, and $C = X \times C_i$.

Clearly $X \times D = X \times Y \setminus \left( \bigcup_{i=1}^L X \times C_i \right)$ is the set of transient states. □

### 4.2 The auxiliary games over each maximal communicating set

Consider now a family of $L$ zero-sum product stochastic games $(\Gamma_i)_{i \in \{1, \ldots, L\}}$. For all $i \in \{1, \ldots, L\}$, if $y \in C_i$, define the set of actions of player 2 at state $y$ under which the state has probability 1 of staying in $C_i$,

$$B_y = \{b \in B \mid q(C_i|y, b) = 1\}.$$
Remark that the set $B_y$ is closed. Then the game $\Gamma_i$ is given by $\Gamma_i = (X, C, A, B_y)_{y \in C_i}, p, q, u)$. The sets of strategies for player 1 and player 2 in $\Gamma_i$ are respectively denoted $S_i$ and $T_i$. The value of the $N$-stage game starting at $(x, y) \in X \times C_i$, is denoted $v_N^i(x, y)$.  

In lemma 4.5 and proposition 4.6 below we prove that each of these games have a uniform value that does not depend on the initial state of the game in $X \times C_i$.

**Lemma 4.5.** Let $v^i : X \times C_i \rightarrow [0, 1]$ be any uniform limit point of the sequence $(v_N^i)_{N \geq 1}$. Then $v^i$ is constant over $X \times C_i$.

**Proof.** Let $i \in \{1, \ldots, L\}$ and $(x_i^*, y_i^*) = \arg\max_{X \times C_i} v^i(\cdot)$. Since player 1 has the strong communication property

$$\forall t \geq T \forall x' \in X \forall \sigma \in S \mathbb{P}^x_{\sigma} (X_T = x') > 0,$$

(2)

and since $C_i$ is a maximal communicating set

$$\exists \nu \in \Delta(B)^Y \forall y' \in C_i \exists t \geq T \mathbb{P}^{y_i^*}_{\nu}(Y_t = y') > 0.$$

(3)

We now fix $(x', y') \in X \times C_i$ and let $\nu \in \Delta(B)^Y$ and $t \geq T$ be as in statement (2). We denote $(x_i^*, y_i^*) = (x', y')$. Then there exist $y_2^*, \ldots, y_{t-1}^* \in C_i$ such that

$$q(y_i^*|y_{i-1}^*, \nu(y_{i-1}^*)) \ldots q(y_2^*|y_1^*, \nu(y_1^*)) > 0.$$

Passing to the limit (uniformly) in the Shapley equation (1) one has

$$v^i(x_i^*, y_i^*) = \max_{\mu_1 \in \Delta(A)} \min_{\nu_1 \in \Delta(B_{y_i^*})} \sum_{x_2 \in X} p(x_2|x_i^*, \mu_1)q(y_2|y_i^*, \nu_1)v^i(x_2, y_2).$$

Hence

$$v^i(x_i^*, y_i^*) \leq \max_{\mu_1 \in \Delta(A)} \sum_{x_2 \in X} p(x_2|x_i^*, \mu_1)q(y_2|y_i^*, \nu(y_i^*))v^i(x_2, y_2) \leq v^i(x_i^*, y_i^*).$$

Thus there exists $\mu_i^* \in \Delta(A)$ such that

$$v^i(x_i^*, y_i^*) = \sum_{x_2 \in X} p(x_2|x_i^*, \mu_i^*)q(y_2|y_i^*, \nu(y_i^*))v^i(x_2, y_2).$$

Finally there exists $\mu_i^* \in \Delta(A)$ such that

$$v^i(x_i^*, y_i^*) = \sum_{x_2 \in X} p(x_2|x_i^*, \mu_i^*)v^i(x_2, y_2).$$

Let $k \in \{2, \ldots, t-1\}$ and suppose there exist $\mu_i^* \in \Delta(A)$ and $\mu_2^*, \ldots, \mu_{k-1}^* \in \Delta(A)^X$ such that

$$v^i(x_i^*, y_i^*) = \sum_{x_2, \ldots, x_{k-1} \in X} p(x_{k-1}|x_{k-2}, \mu_{k-2}^*) \ldots p(x_2|x_1^*, \mu_1^*)v^i(x_{k-1}, y_{k-1}^*).$$
Then there exist $\mu^*_1 \in \Delta(A)$ and $\mu^*_2, \ldots, \mu^*_k \in \Delta(A)^X$ such that
\[
v^i(x^*_1, y^*_1) = \sum_{x_2, \ldots, x_{k-1} \in X} p(x_{k-1}|x_{k-2}, \mu^*_k) \ldots p(x_2|x^*_1, \mu^*_1) \\
\max_{\mu^*_{k-1} \in \Delta(A)} \min_{\nu_{k-1} \in \Delta(B_{y^*_k})} \sum_{y_k \in C_i} p(x_k|x_{k-1}, \mu^*_{k-1}) q(y_k|y^*_k, \nu_{k-1}) v^i(x_k, y_k)
\leq \sum_{x_2, \ldots, x_{k-1} \in X} p(x_{k-1}|x_{k-2}, \mu^*_k) \ldots p(x_2|x^*_1, \mu^*_1) \\
\max_{\mu^*_{k-1} \in \Delta(A)} \sum_{y_k \in C_i} p(x_k|x_{k-1}, \mu^*_{k-1}) q(y_k|y^*_k, \nu(y^*_k, \mu^*_{k-1})) v^i(x_k, y_k)
\leq v^i(x^*_1, y^*_1).
\]
Thus there exists $\mu^*_k \in \Delta(A)^X$ such that
\[
v^i(x^*_1, y^*_1) = \sum_{x_2, \ldots, x_{k-1} \in X} p(x_{k-1}|x_{k-2}, \mu^*_k) \ldots p(x_2|x^*_1, \mu^*_1) \\
\sum_{y_k \in C_i} p(x_k|x_{k-1}, \mu^*_{k-1}) q(y_k|y^*_k, \nu(y^*_k, \mu^*_{k-1})) v^i(x_k, y_k).
\]
Finally,
\[
v^i(x^*_1, y^*_1) = \sum_{x_2, \ldots, x_{k-1} \in X} p(x_{k-1}|x_{k-2}, \mu^*_k) p(x_{k-1}|x_{k-2}, \mu^*_k) \ldots p(x_2|x^*_1, \mu^*_1) v^i(x_k, y_k).
\]
Thus by induction, there exist $\mu^*_1 \in \Delta(A)$ and $\mu^*_2, \ldots, \mu^*_k \in \Delta(A)^X$ such that for all $x_2, \ldots, x_{k-1} \in X$
\[
0 = p(x^*_1|x_{l-1}, \mu^*_{k-1}(x_{l-1})) \ldots p(x_3|x_2, \mu^*_2(x_2)) p(x_2|x^*_1, \mu^*_1)(v^i(x^*_1, y^*_1) - v^i(x^*_1, y^*_1)).
\]
Let us define the strategy $\sigma \in S$ by $\sigma_1(x^*_1) = \mu^*_1$ and for all $k \in \{2, \ldots, t-1\}$ $\sigma_k(h^*_k) = \mu^*_k(x_k)$. Finally $\sigma$ is defined arbitrarily for all $k \geq t$. We have from statement (2) that there exists $x^*_2, \ldots, x^*_{t-1} \in X$ such that
\[
p(x^*_1|x^*_1, \mu^*_{t-1}) \ldots p(x^*_2|x^*_1, \mu^*_1) > 0.
\]
Finally, one has $v^i(x^*_1, y^*_1) = v^i(x^*_1, y^*_1)$. \qed

Since any limit point of $(v^*_N)_{N \geq 1}$ is constant over $X \times C_i$, we show that players guarantee $\limsup_{N \rightarrow +\infty} v^*_N$ by playing an $\varepsilon$-optimal Markov $N_0$-periodic strategy for a sufficiently large $N_0$.

**Proposition 4.6.** For all $i \in \{1, \ldots, L\}$ the game $\Gamma_i$ has a uniform value $v^*_\infty$, which is constant over $X \times C_i$.

Moreover, for all $\varepsilon > 0$ there exists $N_0 \in \mathbb{N}^*$ such that both players have an $\varepsilon$-optimal Markov $N_0$-periodic strategy in each $\Gamma_i$. 

14
Proof. Here we take $v^i = \limsup_{N \to +\infty} v^i_N$. Let $\varepsilon > 0$. Thanks to lemma \[\text{4.5}\] there exists $N_0^i \geq 1$ such that for all $(x, y) \in X \times C_i$ one has $|v^i_N(x, y) - v^i| \leq \varepsilon$. Hence

$$
\max_{\sigma \in \mathcal{S}_i} \min_{\tau \in \mathcal{T}_i} \mathbb{E}_{\sigma, \tau}^{x,y} \left[ \sum_{n=1}^{N_0^i} u_n \right] \geq N_0^i(v^i - \varepsilon).
$$

Recall that there exists a Markov strategy $\sigma^i = (\sigma^i_n)_{n \in [1, N_0^i]}$, independent of the initial state $(x, y)$, such that for all $\tau \in \mathcal{T}_i$

$$
\mathbb{E}_{\sigma^i, \tau}^{x,y} \left[ \sum_{n=1}^{N_0^i} u_n \right] \geq N_0^i(v^i - \varepsilon).
$$

Let $N = pN_0^i + r$, $p \geq 1$ and $r \leq N_0^i - 1$ be two integers. The following Markov $N_0^i$-periodic strategy is still denoted $\sigma^i$ : at stage $n \geq 1$, in state $(x, y) \in X \times C_i$, play $\sigma^i_n(x, y)$, where $n'$ equals $n$ modulo $N_0^i$. Then for all $\tau \in \mathcal{T}_i$

$$
\mathbb{E}_{\sigma^i, \tau}^{x,y} \left[ \sum_{n=1}^{N} u_n \right] = \sum_{k=0}^{p-1} \mathbb{E}_{\sigma^i, \tau}^{x,y} \left[ \sum_{n=kN_0^i+1}^{(k+1)N_0^i} u_n \right] + \mathbb{E}_{\sigma^i, \tau}^{x,y} \left[ \sum_{n=pN_0^i+1}^{N} u_n \right]
\geq pN_0^i(v^i - \varepsilon).
$$

Finally, for all $N \geq \frac{N_0^i}{\varepsilon}$,

$$
v^i_N(x, y) \geq \frac{pN_0^i(v^i - \varepsilon)}{N} \geq (1 - \varepsilon)(v^i - \varepsilon).
$$

That is, player 1 uniformly guarantees $v^i$ and $\sigma^i$ is an $\varepsilon$-optimal Markov $N_0^i$-periodic strategy. A similar proof shows that player 2 also uniformly guarantees $v^i$ and has an $\varepsilon$-optimal Markov $N_0^i$-periodic strategy $\tau^i$.

Note that $N_0^i$ can be taken uniformly over $\{1, \ldots, L\}$ by setting $N_0 = \prod_{i=1}^{L} N_0^i$. \[\square\]

4.3 The auxiliary Markov decision process $G$

Consider the Markov decision process $G = (Y, B, q, g)$, in which player 2 is the only decision maker and his aim is to minimize

$$
g : \ Y \rightarrow [0, 1]$$

$$y \mapsto \begin{cases} v^i_\infty & \text{if there exists } i \in \{1, \ldots, L\} \text{ such that } y \in C_i \\ 1/2 & \text{if } y \in D. \end{cases}
$$

Recall (see Sorin [21]) that, $G$ has a uniform value $w_\infty \in [0, 1]^Y$ and that for every $\varepsilon > 0$, player 2 has an $\varepsilon$-optimal stationary strategy.
The interpretation is the following. The objective of player 2 is to reach the maximal communicating set $C_i$ with corresponding auxiliary game $\Gamma_i$ having the lowest uniform value $v'_\infty$ possible, and stay in $C_i$. Note that the payoff of $1/2$ in $D$ is arbitrary and does not change the value of $w_\infty$.

We will prove that the uniform value $w_\infty$ of the MDP $\mathcal{G}$ is in fact also the uniform value of the initial game $\Gamma$.

### 4.4 Player 2 uniformly guarantees $w_\infty$ in $\Gamma$

**Proposition 4.7.** Player 2 uniformly guarantees $w_\infty$ in $\Gamma$.

To prove proposition 4.7 we show that player 2 has an $\varepsilon$-optimal strategy which has a rather simple structure: play according to some $\varepsilon$-optimal stationary strategy in $\mathcal{G}$ on $Y$, until reaching a recurrent class of the induced Markov chain on $Y$. Then switch to an $\varepsilon$-optimal Markov periodic strategy in some game $\Gamma_i$ having the property that the state remains in the corresponding maximal communicating set $C_i$ (indeed the $\varepsilon$-optimal strategies used in the proof of proposition 4.7 below can be taken Markov periodic thanks to proposition 4.6).

**Proof.** Let $\varepsilon > 0$, and let $\nu_\mathcal{G}$ be an $\varepsilon$-optimal stationary strategy of player 2 in $\mathcal{G}$. $\nu_\mathcal{G}$ induces $l$ recurrent classes $R_1, \ldots, R_l$ over $Y$. Moreover, by definition 4.1 there exists a mapping $\varphi : \{1, \ldots, l\} \to \{1, \ldots, l\}$ such that for all $i \in \{1, \ldots, l\}$, $R_i \subseteq C_{\varphi(i)}$.

For all $i \in \{1, \ldots, l\}$ define $T_i = \min\{n \in \mathbb{N}^* \mid Y_n \in R_i\}$ to be the hitting time of $R_i$ by $(Y_n)_{n \in \mathbb{N}^*}$. The minimum over an empty set is taken equal to $+\infty$.

We now consider the game $\Gamma$ with initial state $(x, y) \in X \times Y$. Let us define the strategy $\tau \in \mathcal{T}$ of player 2 as follows. Until there exists $i \in \{1, \ldots, l\}$ such that the state of player 2 is in $R_i$ play $\nu_\mathcal{G}$. Let $n \in \mathbb{N}^*$ be the first stage at which the state of player 2 reaches one of the recurrent classes $R_i$, $i \in \{1, \ldots, l\}$. From stage $n$ on, play $\tau_{\varphi(i)}$ which is an $\varepsilon$-optimal strategy in $\Gamma_{\varphi(i)}$.

Let $\sigma$ be any strategy of player 1 in $\Gamma$. Remark that under $(\sigma, \tau)$, the laws of the $T_i$’s are the same as under $\nu_\mathcal{G}$.
Let $N \geq 2$, 

$$
\frac{1}{N} \mathbb{E}^{x,y}_{\sigma,\tau} \left[ \sum_{n=1}^{N} u_n \right] = \frac{1}{N} \sum_{i=1}^{l} \mathbb{E}^{x,y}_{\sigma,\tau} \left[ \left( \sum_{n=1}^{T_i} u_n \right) \mathbb{1} \left( \sqrt{N} > T_i = \min_{k \in \{1,\ldots,l\}} T_k \right) \right] 
$$

(4)

$$
+ \frac{1}{N} \sum_{i=1}^{l} \mathbb{E}^{x,y}_{\sigma,\tau} \left[ \left( \sum_{n=T_i+1}^{N} u_n \right) \mathbb{1} \left( \sqrt{N} > T_i = \min_{k \in \{1,\ldots,l\}} T_k \right) \right] 
$$

(5)

$$
+ \frac{1}{N} \sum_{i=1}^{l} \mathbb{E}^{x,y}_{\sigma,\tau} \left[ \left( \sum_{n=1}^{N} u_n \right) \mathbb{1} \left( \sqrt{N} \leq T_i = \min_{k \in \{1,\ldots,l\}} T_k \right) \right] .
$$

(6)

Since the payoffs $u_n$ are at most 1 and $T_i$ is at most $\sqrt{N}$ in the indicator function, (4) is at most

$$
\frac{\sqrt{N}}{N} \sum_{i=1}^{l} \mathbb{E}^{x,y}_{\sigma,\tau} \left( \sqrt{N} > T_i = \min_{k \in \{1,\ldots,l\}} T_k \right),
$$

which itself is at most $\frac{l \sqrt{N}}{N}$ which is smaller than $\varepsilon$ for $N$ large enough.

Remark that in (5),

$$
\frac{1}{N} \leq \frac{1}{N - T_i},
$$

hence this quantity is less than or equal to

$$
\sum_{i=1}^{l} \mathbb{E}^{x,y}_{\sigma,\tau} \left[ \frac{1}{N - T_i} \left( \sum_{n=T_i+1}^{N} u_n \right) \mathbb{1} \left( \sqrt{N} > T_i = \min_{k \in \{1,\ldots,l\}} T_k \right) \right],
$$

the latter equals

$$
\sum_{i=1}^{l} \mathbb{E}^{x,y}_{\sigma,\tau} \left[ \frac{1}{N - T_i} \sum_{n=T_i+1}^{N} u_n \mathbb{1} \left( \sqrt{N} > T_i = \min_{k \in \{1,\ldots,l\}} T_k \right) \right] \mathbb{E}^{x,y}_{\sigma,\tau} \left( \sqrt{N} > T_i = \min_{k \in \{1,\ldots,l\}} T_k \right).
$$

Recall that for all $i \in \{1,\ldots,L\}$, $\tau^i$ is $\varepsilon$-optimal in $\Gamma_i$ which has uniform value $v^i_{\infty}$. For $N$ large enough the quantity above is less than or equal to

$$
\sum_{i=1}^{l} (v^i_{\infty} + \varepsilon) \mathbb{E}^{y}_{\nu_G} \left( T_i = \min_{k \in \{1,\ldots,l\}} T_k \right),
$$

because under $(\sigma,\tau)$, the laws of the $T_i$’s are the same as under $\nu_G$. And since $\nu_G$ is $\varepsilon$-optimal in $G$, $\sum_{i=1}^{l} v^i_{\infty} \mathbb{E}^{y}_{\nu_G} \left( T_i = \min_{k \in \{1,\ldots,l\}} T_k \right)$ is less than or equal to $w_{\infty}(y) + \varepsilon$.

Finally, since the payoffs $u_n$ are at most 1, (6) is less than or equal to

$$
\sum_{i=1}^{l} \mathbb{E}^{x,y}_{\sigma,\tau} \left( \sqrt{N} \leq T_i = \min_{k \in \{1,\ldots,l\}} T_k \right),
$$

17
which is at most
\[
\|P^y\|_v \left( \sqrt{N} \leq \min_{k \in \{1, \ldots, l\}} T_k \right),
\]
which itself is less than \(\varepsilon\) for \(N\) large enough. \(\square\)

4.5 Player 1 uniformly guarantees \(w_\infty\) in \(\Gamma\)

4.5.1 The auxiliary games \((\tilde{\Gamma}_i)_{i \in \{1, \ldots, L\}}\)

Let us fix \(\varepsilon > 0\) and \(N_0 \in \mathbb{N}^*\) accordingly as in proposition 4.6. Beware that the objects that we now introduce also depend on \(\varepsilon\).

We construct \(L\) auxiliary games \((\tilde{\Gamma}_i)_{i \in \{1, \ldots, L\}}\) which are copies of the games \((\Gamma_i)_{i \in \{1, \ldots, L\}}\) with an additional clock keeping track of the time modulo \(N_0\). The state of player 2 on \(C_i \times [1, N_0]\), where \([1, N_0] = \{1, \ldots, N_0\}\), moves on \(C_i\) as in \(\Gamma_i\) and on \([1, N_0]\) by adding 1 at each stage (and starting back to 1 when \(N_0 + 1\) is reached). For all \(i \in \{1, \ldots, L\}\),
\[
\tilde{\Gamma}_i = (X, C_i \times [1, N_0], A, (B_y)_{y \in C_i}, p, \tilde{q}, u),
\]
and
\[
\tilde{q} : C_i \times [1, N_0] \times (B_y)_{y \in C_i} \rightarrow \Delta(C_i \times [1, N_0])
\]
\[
(y, t, b) \mapsto q(y, b) \otimes \delta_{t+1},
\]
where \(t\) is taken modulo \(N_0\) in \(\delta_{t+1}\).

The purpose of these games is that they have the same uniform value as the \(\Gamma_i\)’s, but \(\varepsilon\)-optimal stationary strategies instead of \(\varepsilon\)-optimal Markov \(N_0\)-periodic strategies. Indeed, by proposition 4.6, let \(\sigma^i\) be an \(\varepsilon\)-optimal Markov \(N_0\)-periodic strategy in \(\Gamma_i\). Let \(\tilde{\mu}^i\) be the following stationary strategy in \(\tilde{\Gamma}_i\): for all \((x, y, t) \in X \times C_i \times [1, N_0]\), \(\tilde{\mu}^i(x, y, t) = \sigma^i_l(x, y)\). This defines a stationary strategy of player 1 in \(\tilde{\Gamma}_i\), which is \(\varepsilon\)-optimal.

Remark 4.2. Since we are dealing with Markov periodic strategies, one could think of defining maximal communicating sets with regard to this class of strategies rather than stationary strategies as in definition 4.1. However the construction we propose here appears to provide a simpler demonstration of player 1 uniformly guaranteeing \(w_\infty\) in \(\Gamma\).

4.5.2 The auxiliary game \(\tilde{\Gamma}\)

We now gather together the games \(\tilde{\Gamma}_i\) into one game \(\tilde{\Gamma}\). Let \(\tilde{Y} = Y \times [1, N_0]\). The game \(\tilde{\Gamma}\) is defined by \(\tilde{\Gamma} = (X, \tilde{Y}, A, B, p, \tilde{q}, u)\). Where \(\tilde{q}\) is extended on \(\tilde{Y} \times B\) as follows.
\[
\tilde{q} : \tilde{Y} \times B \rightarrow \Delta(\tilde{Y})
\]
\[
(y, t, b) \mapsto \tilde{q}(y, t, b),
\]
where \( \hat{q}(y', t + 1|y, t, b) = q(y'|y, b) \) if there exists \( i \in \{1, \ldots, L\} \) such that \( y, y' \in C_i \), and \( t \) is taken modulo \( N_0 \). 
\( \hat{q}(y', 1|y, t, b) = q(y'|y, b) \) if there exists \( i \in \{1, \ldots, L\} \) such that \( y \in C_i \) and \( y' \notin C_i \), or \( y \in D \). Otherwise \( \hat{q}(y', t'|y, t, b) = 0 \). In words, the clock on \([1, N_0]\) is incremented as in the games \( \Gamma_i \) as the state stays in \( C_i \), and is reset to 1 when the state jumps from one \( C_i \) to another or to a transient state, or from a transient state to any other state.

Let \( \mathcal{S} \) and \( \mathcal{T} \) be the set of strategies of player 1 and 2 respectively in \( \bar{\Gamma} \). It is important to note that any quantity guaranteed by a player in \( \bar{\Gamma} \) is also guaranteed in \( \Gamma \).

The following lemma states that the \( Y \) component of a recurrent class in \( X \times \bar{Y} \) cannot have nonempty intersection with two different maximal communicating sets in \( Y \).

**Lemma 4.8.** Let \( (\hat{\mu}, \hat{\nu}) \) be a pair of stationary strategies on \( X \times \bar{Y} \). Let \( R \subseteq X \times \bar{Y} \) be a recurrent class under \( (\hat{\mu}, \hat{\nu}) \). Then there exists \( i \in \{1, \ldots, L\} \) such that \( R \subseteq X \times C_i \times [1, N_0] \).

**Proof.** Let \( R' = \{(x, y) \in X \times Y \mid \exists t \in [1, N_0] \ (x, y, t) \in R\} \) be the projection of \( R \) over \( X \times Y \). For all \( (x, y) \in R' \) let \( R_{x,y} = \{t \in [1, N_0] \mid (x, y, t) \in R\} \).

Define the stationary strategies \( \mu \) and \( \nu \) on \( X \times Y \) by, for all \( (x, y) \in R' \),

\[
\mu(x, y) = \frac{1}{|R_{x,y}|} \sum_{t \in R_{x,y}} \hat{\mu}(x, y, t) \quad \text{and} \quad \nu(x, y) = \frac{1}{|R_{x,y}|} \sum_{t \in R_{x,y}} \hat{\nu}(x, y, t),
\]

and arbitrarily outside \( R' \). Under \( (\mu, \nu) \), \( R' \) is a recurrent class. Hence by proposition 4.4 there exists \( i \in \{1, \ldots, L\} \) such that \( R' \subseteq X \times C_i \). \(\Box\)

For all \( i \in \{1, \ldots, L\} \), let \( \hat{\mu}^i \) be an \( \varepsilon \)-optimal stationary strategy of player 1 in \( \bar{\Gamma}_i \). Define the stationary strategy \( \tilde{\mu} \) of player 1 in \( \bar{\Gamma} \) as follows. For all \( (x, y, t) \in X \times \bar{Y} \)

\[
\tilde{\mu}(x, y, t) = \begin{cases} 
\hat{\mu}^i(x, y, t) & \text{if there exists } i \in \{1, \ldots, L\} \text{ such that } y \in C_i \\
\text{arbitrary fixed action if } y \in D.
\end{cases}
\]

In the game \( \bar{\Gamma} \), by proposition 2.2, player 2 has a stationary strategy \( \hat{\nu} \in \Delta(B)^{X \times Y} \) and there exists \( M \in \mathbb{N}^* \) such that for all \( N \geq M \) and all \( \tau \in \bar{T} \)

\[
\gamma_N(\hat{\mu}, \hat{\nu})(x, y, t) \leq \gamma_N(\tilde{\mu}, \tau)(x, y, t) + \varepsilon.
\]

The pair \( (\hat{\mu}, \hat{\nu}) \) induces a Markov chain on \( X \times \bar{Y} \) with recurrent classes \( R_1, \ldots, R_m \). By lemma 4.5.8 there exists a mapping \( \psi : \{1, \ldots, m\} \to \{1, \ldots, L\} \) such that for all \( i \in \{1, \ldots, m\} \)

\[
R_i \subseteq X \times C_{\psi(i)} \times [1, N_0].
\]

For all \( i \in \{1, \ldots, m\} \) we define \( \bar{T}_i = \min\{n \in \mathbb{N}^* \mid (X_n, Y_n, t_n) \in R_i\} \) to be the hitting time of \( R_i \) by \( (X_n, Y_n, t_n)_{n \geq 1} \).
Lemma 4.9. In the game $\tilde{\Gamma}$ starting at $(x, y, 1) \in X \times \tilde{Y}$, player 1 uniformly guarantees
\[
\sum_{i=1}^{m} v_{\infty}^{(i)} \mathbb{P}^{x, y, 1}_{\tilde{\mu}, \tilde{\nu}} \left( \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right) - 3\varepsilon.
\]

Proof. Let $N \geq 2$ and $\varepsilon' > 0$,
\[
\frac{1}{N} \mathbb{E}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left[ \sum_{n=1}^{N} u_n \right] \geq \sum_{i=1}^{m} \frac{1}{N} \mathbb{E}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left[ \left( \sum_{n=\tilde{T}_i+1}^{N} u_n \right) \mathbb{1} \left( \sqrt{N} > \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right) \right].
\]
This last quantity is equal to
\[
\sum_{i=1}^{m} \left( 1 - \frac{\sqrt{N}}{N} \right) \mathbb{P}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left[ \frac{1}{N - \tilde{T}_i} \sum_{n=\tilde{T}_i+1}^{N} u_n \sqrt{N} > \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right] \mathbb{P}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left( \sqrt{N} > \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right).
\]
Since $\tilde{T}_i$ is taken less than $\sqrt{N}$ in the indicator function, the latter is greater than
\[
\sum_{i=1}^{m} \left( 1 - \frac{\sqrt{N}}{N} \right) \mathbb{P}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left[ \sqrt{N} > \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right].
\]
Recall that for all $i \in \{1, \ldots, L\}$, $\tilde{\mu}^i$ is $\varepsilon$-optimal in $\tilde{\Gamma}_i$, which has value $v_{\infty}^{(i)}$. Moreover, conditionally on $\tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k$, from $\tilde{T}_i + 1$ on, player 2 only plays actions that are in $(B_y)_{y \in R_i}$, otherwise the state process on $Y$ would have a positive probability of leaving $C_{\psi(i)}$ and $R_i$ would not be a recurrent class. Hence for $N$ large enough the quantity above is greater than or equal to
\[
\sum_{i=1}^{m} \left( 1 - \frac{\sqrt{N}}{N} \right) \left( v_{\infty}^{(i)} - \varepsilon \right) \mathbb{P}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left( \sqrt{N} > \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right),
\]
which, for $N$ large enough is greater than
\[
(1 - \varepsilon') \sum_{i=1}^{m} \left( v_{\infty}^{(i)} - \varepsilon \right) \mathbb{P}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left( \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right) - \varepsilon'.
\]
Hence there exists $\tilde{N} \in \mathbb{N}^*$ such that for all $N \geq \tilde{N}$ and all $\tau \in \tilde{\tau}$,
\[
\gamma_N(\tilde{\mu}, \tau)(x, y, 1) + \varepsilon \geq (1 - \varepsilon') \sum_{i=1}^{m} v_{\infty}^{(i)} \left( \mathbb{P}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left( \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right) - \varepsilon' \right) - \varepsilon(1 - \varepsilon').
\]
Finally, for $\varepsilon'$ small enough, the right-hand side is greater than
\[
\sum_{i=1}^{m} v_{\infty}^{(i)} \mathbb{P}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left( \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right) - 2\varepsilon,
\]
which concludes the proof. \(\square\)
To conclude, we prove that the payoff uniformly guaranteed (up to $3\varepsilon$) by player 1 in $\tilde{\Gamma}$ in lemma 4.9 can be obtained by player 2 as a limit payoff in the MDP $G$, and hence is greater than $w_\infty$, the uniform value of $G$.

**Proposition 4.10.** Player 1 uniformly guarantees $w_\infty$ in $\Gamma$.

**Proof.** By lemma 4.9 in the game $\tilde{\Gamma}$ starting at $(x, y, 1) \in X \times \tilde{Y}$ player 1 uniformly guarantees

$$\sum_{i=1}^{m} v_{\infty}^{(i)} \mathbb{P}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left( \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right) - 3\varepsilon.$$ 

In the MDP $G$, let $\tau_G$ be the following strategy of player 2. At stage 1, play $\tilde{\nu}(x, y, 1)$, a new state $(y_2, t_2) \in \tilde{Y}$ is selected according to $\tilde{q}(\cdot|y, 1, \tilde{\nu}(x, y, 1))$. At stage 2, play $\tilde{\nu}(x_2, y_2, t_2)$ where $x_2 \in X$ is selected according to $p(\cdot|x, \tilde{\mu}(x, y, 1))$. Inductively at stage $n > 2$, play $\tilde{\nu}(x_n, y_n, t_n)$ where $x_n \in X$ is selected according to $p(\cdot|x_{n-1}, \tilde{\mu}(x_{n-1}, y_{n-1}, t_{n-1}))$.

$\tau_G$ has the following interpretation. Player 2 plays in $G$ according to $\tilde{\nu} \in \Delta(B)^{X \times \tilde{Y}}$ by simulating at each stage a fictitious state of player 1 on $X$ that follows $\tilde{\mu} \in \Delta(B)^{X \times \tilde{Y}}$. This induces a fictitious hitting times $\tilde{T}_i$'s of the recurrent classes $R_1, \ldots, R_m$.

Under $\tau_G$, the laws of the $\tilde{T}_i$'s are the same as under $(\tilde{\mu}, \tilde{\nu})$. After the process $(X_n, Y_n, t_n)$ has reached $R_i$, the payoff to player 2 in $G$ is $v_i^{\infty}$ at each stage.

Therefore, $\tau_G$ yields in $G$ a limiting payoff of

$$\sum_{i=1}^{m} v_{\infty}^{(i)} \mathbb{P}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left( \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right).$$

Hence this quantity is greater than $w_\infty(y)$.

Since in the initial game $\Gamma$ starting at $(x, y)$, player 1 also uniformly guarantees

$$\sum_{i=1}^{m} v_{\infty}^{(i)} \mathbb{P}_{\tilde{\mu}, \tilde{\nu}}^{x, y, 1} \left( \tilde{T}_i = \min_{k \in \{1, \ldots, m\}} \tilde{T}_k \right) - 3\varepsilon,$$

he uniformly guarantees $w_\infty(y) - 3\varepsilon$.

This is true for all $\varepsilon > 0$. Hence player 1 uniformly guarantees $w_\infty$ in $\Gamma$.

\[\square\]

## 5 Proof of theorem 3.4

In subsection 5.1 we present the weakly communicating on both sides zero-sum product stochastic game used to prove theorem 3.4. The example proposed shows that in weakly communicating on both sides games, even if players can go from any state to any other state of their component in finite time, they can make mistakes that are irreversible with regards to the joint state. Subsection 5.2 is dedicated to the simplification of the Shapley equations obtained for the weakly communicating on both sides game.
Finally in subsection 5.3 we show that the Shapley equations obtained after simplification are also those of a simple game of perfect information, i.e., in each state only one player control the transition, with two absorbing and two non-absorbing states. The latter game does not have an asymptotic value (lemmas 5.3, 5.4, 5.5 and 5.6).

5.1 A counter-example

5.1.1 State spaces and action sets

The state space of player 1 is \( X = \{x, y\} \times C_8 \) where \( C_8 = \mathbb{Z}/8\mathbb{Z} \). The state space of player 2 is \( Y = \{x', y'\} \times C'_8 \) where \( C'_8 = \mathbb{Z}/8\mathbb{Z} \).

Let \( I = \{0\} \cup \{1/2^n \mid n \geq 1\} \), and \( J = [0, 1/4] \). Let \( A = I \times \{-1, +1\} \cup \{0, 1\} \times \{0\} \) and \( B = J \times \{-1, +1\} \cup \{0, 1\} \times \{0\} \) be the action sets of player 1 and 2 respectively.

5.1.2 Transitions

For \( i \in \{x, y\} \) we denote by \(-i\) the element of \( \{x, y\}\backslash \{i\} \).

In state \((i, k) \in X\) if player 1 plays \((\alpha, p) \in I \times \{-1, +1\}\) then with probability \(1 - \alpha - \alpha^2\) the new state is \((i, k + p)\), with probability \(\alpha\) the new state is \((-i, k + p)\), and with probability \(\alpha^2\) the new state is \((i, k - p)\) (see figure 4).

Still in state \((i, k) \in X\), if player 1 plays \((0, 0) \in \{0, 1\} \times \{0\}\), then with probability \(1 - \alpha\) the state remains in \((i, k)\) and with probability \(\alpha\) the new state is \((-i, k)\) (see figure 5).

Transitions for player 2 are analogous on \(Y\).

1 - \(\alpha - \alpha^2\)

Figure 4: Transition of player 1 when playing \((\alpha, +1), \alpha \in I\) in state \((x, 2)\)

Hence by playing \((0, p)\), player 1 totally controls the dynamics on \(C_8\), and by playing \((\alpha, 0)\) with \(\alpha\) equal to 0 or 1 he totally controls the dynamics.
Figure 5: Transition of player 1 when playing \((\alpha, 0)\), \(\alpha \in \{0, 1\}\) in state \((x, 2)\) on \(\{x, y\}\). And likewise for player 2.

5.1.3 Payoffs

Let \((i, k) \in X\) and \((i', k') \in Y\). We denote by \(d_{C_8}(k, k')\) the distance between player 1 and 2 on the circle \(C_8\).

The payoff function \(u\) is defined as follows. If \(d_{C_8}(k, k') \geq 3\) then \(u((i, k), (i', k')) = 1\). If \(d_{C_8}(k, k') \leq 1\) then \(u((i, k), (i', k')) = 0\). Otherwise, if \(d_{C_8}(k, k') = 2\) then \(u\) is defined by the following table:

| \(u(\cdot, \cdot)\) | \(x'\) | \(y'\) |
|---------------------|--------|--------|
| \(x\)               | 0      | 1      |
| \(y\)               | 1      | 0      |

The interpretation of the game is the following. Player 1 wants to maximize his distance to player 2 who wants to minimize his distance to player 1. If the distance between them on the circle is at most 1 or at least 3, then their positions in \(\{x, y\}\) and \(\{x', y'\}\) do not matter. Whereas if the distance between them on the circle is equal to 2, then player 1 wants to be in \(x\) (resp. \(y\)) when player 2 is in \(y'\) (resp. \(x'\)).

Note also that if the distance between the players is at least 3 (resp. at most 1), then player 1 (resp. player 2) can play such that the distance is always at least 3 (resp. at most 1), and this is optimal for him. Hence those joint states on \(X \times Y\) act as absorbing states with payoff 1 and 0 respectively.

5.1.4 Shapley equations

For \(p \in \{-1, 0, +1\}\) we denote as well \(p\) by the triplet \((p_{-1}, p_0, p_{+1}) \in \{0, 1\}^3\) with \(p_{-1} = 1\) if \(p = -1\) and \(p_{-1} = 0\) otherwise, and likewise for \(p_0\) and \(p_{+1}\).

We define similarly \(q = (q_{-1}, q_0, q_{+1})\).
Lemma 5.1. For all \( \lambda \in (0, 1) \) the following equations hold.

\[
x_\lambda > y_\lambda
\]

\[
y_\lambda = (1 - \lambda) \max_{(\alpha, p) \in A} \left( \alpha^2(p_{-1} - p_{+1})y_\lambda + \alpha p_{+1}(x_\lambda - y_\lambda) + p_{+1}y_\lambda \right)
\]

\[
x_\lambda = \lambda + (1 - \lambda) \min_{(\beta, q) \in B} \left( \beta^2(q_{-1} - q_{+1})(x_\lambda - 1) + \beta q_{+1}(y_\lambda - x_\lambda) + q_{+1}(x_\lambda - 1) + 1 \right).
\]
The proof of lemma 5.1 is postponed to the annex. In the next proposition, we show that \( x_\lambda \) and \( y_\lambda \) indeed have a rather simple expression.

**Proposition 5.2.** For all \( \lambda \in (0, 1) \),

\[
\lambda y_\lambda = (1 - \lambda) \max_{\alpha \in I} \left( -\alpha^2 y_\lambda + \alpha (x_\lambda - y_\lambda) \right) \quad (10)
\]

\[
\lambda x_\lambda = \lambda + (1 - \lambda) \min_{\beta \in J} \left( \beta^2 (1 - x_\lambda) + \beta (y_\lambda - x_\lambda) \right). \quad (11)
\]

The proof of proposition 5.2 is postponed to the annex.

5.3 A simple counterexample to the convergence of \((x_\lambda)_{\lambda \in (0, 1)}\) and \((y_\lambda)_{\lambda \in (0, 1)}\)

Equations (10) and (11) are in fact the Shapley equations of the following game, non product, non weakly communicating on both sides, described in figure 6. This game does not have a product state space, neither has it a weakly communicating property. This example was introduced by Renault [16]. The state space is \( \Omega = \{0, 1, 0^*, 1^*\} \). The action space of player 1 (resp. 2) is \( I \) (resp. \( J \)). Player 1 (resp. 2) plays in state 0 (resp. 1). The states \( 0^* \) and \( 1^* \) are absorbing. The payoff in states 0 and \( 0^* \) (resp. 1 and \( 1^* \)) is 0 (resp. 1).

![Figure 6: A simpler game](image)

We conclude with the next lemmas showing that \((x_\lambda)_{\lambda \in (0, 1)}\) and \((y_\lambda)_{\lambda \in (0, 1)}\) do not converge. The proofs are provided for completeness.

**Lemma 5.3.** For all \( \lambda \leq 1/17 \), \( \beta_\lambda = \frac{x_\lambda - y_\lambda}{2(1 - x_\lambda)} \) is optimal for player 2. Moreover

\[
4\lambda(1 - x_\lambda)^2 = (1 - \lambda)(x_\lambda - y_\lambda)^2. \quad (12)
\]

Hence \( x_\lambda - y_\lambda \) goes to 0 as \( \lambda \) goes to 0.

**Proof.** Suppose that \( \beta_\lambda \leq 1/4 \), then it is easy to verify that the lemma holds. Suppose now that \( \beta_\lambda > 1/4 \). Then the minimum in equation (11) is
Lemma 5.4.

Proof. Assume that, up to some subsequence, \( v \) and \( \lambda_n \) converge to some \( v \in [0,1] \). Then \( v \leq 1/2 \). Moreover \( x_{\lambda_n} - y_{\lambda_n} \sim 2\sqrt{\lambda_n}(1 - v) \) and \( \beta_{\lambda_n} \sim \sqrt{\lambda_n} \) as \( n \) goes to \( +\infty \).

Proof. Let \( \alpha \) be an optimal strategy of player 1. By equation (12) one gets

\[
2\alpha \lambda y_{\alpha} \leq y_{\alpha}(1 + \alpha_1^2) = \lambda y_{\alpha} \alpha_1^2 + 2\alpha \sqrt{\lambda} \sqrt{1 - \lambda(1 - \alpha_1)}.
\]

Dividing by \( \alpha \sqrt{\lambda} \) and passing to the limit yields \( v \leq 1/2 \). \( \square \)

Lemma 5.5.

If for all \( n \in \mathbb{N} \), \( \sqrt{\lambda_n} \in I \) then

\[
\lim_{n \to +\infty} x_{\lambda_n} = 1/2.
\]

Proof. Assume that, up to some subsequence, \( x_{\lambda_n} \) and \( y_{\lambda_n} \) converge to some \( v \in [0,1] \). If player 1 plays \( \alpha = \sqrt{\lambda} \), equation (10) yields

\[
\lambda y_{\alpha} \geq (1 - \lambda) \sqrt{\lambda}(x_{\lambda} - y_{\lambda}) - (1 - \lambda) \lambda y_{\alpha}.
\]

Dividing by \( \lambda \) and passing to the limit, one gets \( v \geq 1/2 \). By lemma 5.4 \( v = 1/2 \). \( \square \)

Lemma 5.6.

If for all \( n \in \mathbb{N} \), \( (1/2\sqrt{\lambda_n}, 2\sqrt{\lambda_n}) \cap I = \emptyset \) then \( \limsup_{n \to +\infty} x_{\lambda_n} \leq 4/9 \).

Proof. Suppose that up to some subsequence, \( x_{\lambda_n} \) and \( y_{\lambda_n} \) converge to some \( v \geq 4/9 \). By lemma 5.4 \( v \leq 1/2 \). Let \( \alpha_1^{*} = \frac{\lambda y_{\alpha}^*}{2\sqrt{\lambda}} > 0 \) be the argument of the maximum of the unconstrained problem associated to equation (10). Then \( \alpha_1^{*} \sim \sqrt{\lambda} \frac{1 - v}{v} \). Hence for \( \lambda \) small enough in the sequence, \( 1/2 \sqrt{\lambda} \leq \alpha_1^{*} \leq 2\sqrt{\lambda} \).

The open interval \((1/2\sqrt{\lambda_n}, 2\sqrt{\lambda_n})\) does not contain any point in \( I \). Furthermore the objective function of player 1 is increasing between 0 and \( \alpha_1^{*} \), and decreasing after.

First case, \( \alpha_1 \leq 1/2 \sqrt{\lambda} \). Then

\[
\lambda y_{\alpha} \leq 1/2(1 - \lambda) \sqrt{\lambda}(x_{\lambda} - y_{\lambda}) - 1/4(1 - \lambda) \lambda y_{\alpha}.
\]

Dividing by \( \lambda \) and passing to the limit yields \( v \leq 4/9 \).

Second case, \( \alpha_1 \geq 2 \sqrt{\lambda} \). Then

\[
\lambda y_{\alpha} \leq 2(1 - \lambda) \sqrt{\lambda}(x_{\lambda} - y_{\lambda}) - 4(1 - \lambda) \lambda y_{\alpha},
\]

and again \( v \leq 4/9 \). \( \square \)

Thus taking the sequences \( \lambda_n = \frac{1}{2n} \) and \( \lambda'_n = \frac{1}{2n+2} \) one has from lemmas 5.5 and 5.6 that \( (x_{\lambda})_{\lambda \in [0,1]} \) and \( (y_{\lambda})_{\lambda \in [0,1]} \) do not converge as \( \lambda \) goes to 0.
Annex: omitted proofs of section 5

Proof of lemma 5.1. We consider the game starting in state \((x, x', 2)\). Player 2 can play \((\beta, q) = (0, +1)\), hence

\[
y_\lambda \leq (1 - \lambda) \max_{(\alpha, p) \in A} \left( \alpha^2 (p - p + 1)y_\lambda + \alpha p + 1(x_\lambda - y_\lambda) + p + 1y_\lambda \right).
\]

Suppose

\[
\{0\} \in \arg\max_p \left( \max_{\alpha \in \{0, 1\}} \left( \alpha^2 (p - p + 1)y_\lambda + \alpha p + 1(x_\lambda - y_\lambda) + p + 1y_\lambda \right) \right).
\]

Then \(y_\lambda \leq (1 - \lambda) \max_{\alpha \in \{0, 1\}} 0\), hence \(y_\lambda = 0\), contradiction.

Suppose

\[
\{-1\} \in \arg\max_p \left( \max_{\alpha \in I} \left( \alpha^2 (p - p + 1)y_\lambda + \alpha p + 1(x_\lambda - y_\lambda) + p + 1y_\lambda \right) \right).
\]

Then \(y_\lambda \leq (1 - \lambda) \max_{\alpha \in I} \alpha^2 y_\lambda = \frac{1}{16} y_\lambda\). Thus \(y_\lambda = 0\), contradiction. Thus

\[
\arg\max_p \left( \max_{\alpha} \left( \alpha^2 (p - p + 1)y_\lambda + \alpha p + 1(x_\lambda - y_\lambda) + p + 1y_\lambda \right) \right) = \{+1\}.
\]

Hence \(\lambda y_\lambda \leq (1 - \lambda) \max_{\alpha \in I} (-\alpha^2 + \alpha(x_\lambda - y_\lambda))\). Thus \(x_\lambda > y_\lambda\), and inequality (7) is proved.

To prove equation 8 we show that \((0, +1)\) is a dominant strategy of player 2, i.e.

\(\forall (\alpha, p) \in A \forall (\beta, q) \in B \ h(y_\lambda, x_\lambda, \alpha, p, \beta, q) - h(y_\lambda, x_\lambda, \alpha, p, 0, +1) \geq 0\).

There are 9 cases to test, corresponding to the different values of \((p, q) \in \{-1, 0, +1\}^2\).

1st case: \(p_{+1} = q_{+1} = 1\)

\[
\left((1 - \alpha - \alpha^2)(-\beta - \beta^2) + \alpha \beta + \alpha^2 \beta^2\right) y_\lambda +
\left((1 - \alpha - \alpha^2)\beta + (-\beta - \beta^2)\alpha\right) x_\lambda + \beta^2(1 - \alpha^2)
= \left((1 - 2\alpha - \alpha^2 + \alpha \beta)(x_\lambda - y_\lambda) + \beta(1 - \alpha^2)(1 - y_\lambda) + \beta \alpha^2 y_\lambda\right) \beta \geq 0.
\]

2nd case: \(p_{+1} = q_0 = 1 \quad 1 - \alpha^2 \geq 0\)
3rd case: $p_{+1} = q_{-1} = 1$

\[
\left((1 - \alpha - \alpha^2)\beta^2 + (1 - \beta - \beta^2)\alpha^2 - (1 - \alpha - \alpha^2)\right) y_\lambda + \\
(\alpha\beta^2 + \beta\alpha^2 - \alpha)x_\lambda + (1 - \alpha^2)(1 - \beta^2)
\]

\[
= \left((\beta^2 - 1)(1 - \alpha - 2\alpha^2) - \beta\alpha^2\right) y_\lambda + \\
(\beta\alpha^2 + \alpha(\beta^2 - 1))x_\lambda + (1 - \alpha^2)(1 - \beta^2)
\]

\[
= \beta\alpha^2 (x_\lambda - y_\lambda) + q(1 - \beta^2)(1 - x_\lambda) + \\
(1 - \beta^2)(1 - \alpha - \alpha^2)(1 - y_\lambda) + (1 - \beta^2)\alpha^2 y_\lambda \geq 0
\]

4th case: $p_{0} = q_{+1} = 1 \quad \beta^2 \geq 0$

5th case: $p_{0} = q_{0} = 1$

\[
((1 - \alpha)(1 - \beta) + \alpha\beta) y_\lambda + ((1 - \alpha)\beta + (1 - \beta)\alpha) x_\lambda \geq 0
\]

6th case: $p_{0} = q_{-1} = 1 \quad 1 - \beta^2 \geq 0$

7th case: $p_{-1} = q_{+1} = 1$

\[
\left((1 - \alpha - \alpha^2)\beta^2 + (1 - \beta - \beta^2)\alpha^2 - \alpha^2\right) y_\lambda + (\alpha^2 \beta + \alpha^2 \beta^2)x_\lambda + \alpha^2 \beta^2
\]

\[
= \left((1 - \alpha - \alpha^2)\beta^2 + (-\beta - \beta^2)\alpha^2\right) y_\lambda + (\alpha^2 \beta + \alpha^2 \beta^2)x_\lambda + \alpha^2 \beta^2
\]

\[
= (1 - 2\alpha^2)\beta^2 y_\lambda + (\alpha^2 \beta + \alpha^2 \beta^2)(x_\lambda - y_\lambda) + \alpha^2 \beta^2 \geq 0
\]

8th case: $p_{-1} = q_{0} = 1 \quad \alpha^2 \geq 0$

9th case: $p_{-1} = q_{-1} = 1$

\[
\left((1 - \alpha - \alpha^2)(1 - \beta - \beta^2) + \alpha\beta + \alpha^2 \beta^2 - \alpha^2\right) y_\lambda + \\
(1 - \alpha - \alpha^2)\beta + (1 - \beta - \beta^2)\alpha x_\lambda + \alpha^2 (1 - \beta^2)
\]

\[
= \left((1 - \alpha - \alpha^2)(1 - \beta - \beta^2) + \alpha\beta\right) y_\lambda + \\
(1 - \alpha - \alpha^2)\beta + (1 - \beta - \beta^2)\alpha x_\lambda + \alpha^2 (1 - \beta^2)(1 - y_\lambda) \geq 0
\]

Equality \[\text{[3]}\] is thus proved.

We now consider the game starting in state $(x, y', 2)$. To prove equation \[\text{[9]}\] we show that $(0, +1)$ is a dominant strategy of player 1, i.e.

\[
\forall(\alpha, p) \in A \quad \forall(\beta, q) \in B \quad h(x_\lambda, y_\lambda, \alpha, p, \beta, q) - h(x_\lambda, y_\lambda, 0, +1, \beta, q) \leq 0.
\]

There are again 9 cases to test, corresponding to the different values of $(p, q) \in \{-1, 0, +1\}^2$.
1st case: \( p_{+1} = q_{+1} = 1 \)

\[
\begin{align*}
&\left( (1 - \alpha - \alpha^2)(1 - \beta - \beta^2) + \alpha \beta + \alpha^2 \beta^2 - (1 - \beta - \beta^2) \right) x_\lambda + \\
&\left( (1 - \alpha - \alpha^2)\beta + (1 - \beta - \beta^2)\alpha - \beta \right) y_\lambda + \beta^2(1 - \alpha^2) - \beta^2 \\
= &\left( (1 - \alpha^2)(1 - \beta - \beta^2) + \alpha \beta + \alpha^2 \beta^2 \right) x_\lambda + \\
&\left( (1 - \alpha^2)\beta + (1 - \beta - \beta^2)\alpha \right) y_\lambda - \alpha^2 \beta^2 \\
= &\alpha(1 - 2\beta - \beta^2 - \alpha\beta)(y_\lambda - x_\lambda) - \alpha^2(1 - \beta^2)x_\lambda + \alpha^2 \beta^2(x_\lambda - 1) \leq 0.
\end{align*}
\]

2nd case: \( p_{+1} = q_{+1} = 1 \quad 1 \leq 0 \)

3rd case: \( p_{+1} = q_{-1} = 1 \)

\[
\begin{align*}
&\left( (1 - \alpha - \alpha^2)\beta^2 + (1 - \beta - \beta^2)\alpha^2 - \beta^2 \right) x_\lambda + \\
&(\alpha \beta^2 + \beta \alpha^2 y_\lambda + (1 - \alpha^2)(1 - \beta^2) - (1 - \beta^2) \\
= &\left( \beta^2(1 - \alpha - \alpha^2) + \alpha^2(1 - \beta - \beta^2) \right) x_\lambda + \left( \beta \alpha^2 + \alpha \beta^2 \right) y_\lambda - \alpha^2(1 - \beta^2) \\
= &\left( \beta \alpha^2 + \alpha \beta^2 \right)(y_\lambda - x_\lambda) + \alpha^2(1 - \beta^2)(x_\lambda - 1) - \alpha^2 \beta^2 x_\lambda \leq 0
\end{align*}
\]

4th case: \( p_0 = q_+ = 1 \quad -(1 - \beta - \beta^2)x_\lambda - \beta y_\lambda + \beta^2 - \beta^2 \leq 0 \)

5th case: \( p_0 = q_0 = 1 \)

\[
\begin{align*}
&((1 - \alpha)(1 - \beta) + \alpha \beta) x_\lambda + ((1 - \alpha)\beta + (1 - \beta)\alpha) y_\lambda - 1 \\
= &((1 - \alpha)\beta + \alpha(1 - \beta))(y_\lambda - x_\lambda) + x_\lambda - 1 \leq 0
\end{align*}
\]

6th case: \( p_0 = q_{-1} = 1 \quad 1 - \beta^2 - (1 - \beta^2) - \beta^2 x_\lambda \leq 0 \)

7th case: \( p_{-1} = q_{+1} = 1 \)

\[
\begin{align*}
&\left( (1 - \alpha - \alpha^2)\beta^2 + (1 - \beta - \beta^2)\alpha^2 - (1 - \beta - \beta^2) \right) x_\lambda + \\
&(\alpha^2 \beta + \alpha \beta^2 - \beta) y_\lambda + \alpha^2 \beta^2 - \beta^2 \\
= &\alpha^2 - 1)\beta y_\lambda + \alpha \beta^2(y_\lambda - x_\lambda) + \\
&\alpha^2 - 1)2\beta^2 - 1)x_\lambda + (\alpha^2 - 1)\beta(1 - x_\lambda) \leq 0
\end{align*}
\]

8th case: \( p_{-1} = q_0 = 1 \quad \alpha^2 - 1 \leq 0 \)
9th case: $p_{-1} = q_{-1} = 1$

\[
(1 - \alpha - \alpha^2)(1 - \beta - \beta^2) + \alpha \beta + \alpha^2 \beta^2 - \beta^2 \] \begin{aligned}
&= \left(1 - \alpha - \alpha^2\right)\beta + (1 - \beta - \beta^2)\alpha \right) y_\lambda + \alpha^2(1 - \beta^2) - (1 - \beta^2) \\
&= \left(1 - \alpha - \alpha^2\right)\beta + (1 - \beta - \beta^2)\alpha \right) (y_\lambda - \lambda) + \\
&= \beta^2(\alpha^2 - 1)\lambda + (1 - \alpha^2)(1 - \beta^2)(\lambda - 1) \leq 0
\end{aligned}

Thus equality (9) is proved.

\textit{Proof of proposition 5.2.} It has been seen in the proof of lemma 5.1 that

\[
\arg\max_p \left( \max_{\alpha} \left( \alpha^2(p_{-1} - p_{+1})y_\lambda + \alpha p_{+1}(x_\lambda - y_\lambda) + p_{+1}y_\lambda \right) \right) = \{+1\}.
\]

Thus one deduces equality (10). Likewise,

\[
\arg\min_q \left( \min_{\beta} \left( \beta^2(q_{-1} - q_{+1})(x_\lambda - 1) + \beta q_{+1}(y_\lambda - x_\lambda) + p_{+1}y_\lambda \right) \right) = \{+1\}.
\]

And equality (11) is proved.

\textbf{Acknowledgments}

The author wishes to express his gratitude to his Ph.D. advisor Jérôme Renault, as well as Fabien Gensbittel, Xavier Venel and Bruno Ziliotto for meaningful discussions.

\textbf{References}

[1] E. Altman, K. Avrachenkov, R. Marquez, and G. Miller, \textit{Zero-sum constrained stochastic games with independent state processes}, Mathematical Methods of Operations Research \textbf{6} (2005), 375–386.

[2] J. Bather, \textit{Optimal decision procedures for finite markov chains. Part III: General convex systems}, Advances in Applied Probability \textbf{5} (1973), 541–553.

[3] T. Bewley and E. Kohlberg, \textit{The asymptotic theory of stochastic games}, Mathematics of Operations Research \textbf{1} (1976), 197–208.

[4] T. Bewley and E. Kohleberg, \textit{On stochastic games with stationary optimal strategies}, Mathematics of Operations Research \textbf{3} (1978), no. 2, 104–125.

[5] J. Bolte, S. Gaubert, and G. Vigeral, \textit{Definable zero-sum stochastic games}, Mathematics of Operations Research (2014).
[6] J. Flesch, G. Schoenmakers, and K. Vrieze, *Stochastic games on a product state space*, Mathematics of Operations Research 33 (2008), 403–420.

[7] ______, *Stochastic games on a product state space: The periodic case*, International Journal of Game Theory 38 (2009), 263–289.

[8] D. Fudenberg and Y. Yamamoto, *The folk theorem for irreducible stochastic games with imperfect public monitoring*, Journal of Economic Theory 146 (2011), 1664–1683.

[9] D. Gillette, *Contributions to the theory of games*, vol. 3, ch. Stochastic Games with Zero Stop Probabilities, pp. 179–187, Princeton University Press, 1957.

[10] A. J. Hoffman and R. M. Karp, *On nonterminating stochastic games*, Management Science 12 (1966), no. 5, 359–370.

[11] R. Laraki and J. Renault, *Acyclic gambling games*, (2017), arXiv:1702.06866.

[12] X. Li and S. Sorin, *Uniform value for recursive games with compact action sets*, Operations Research Letters 44 (2016), 575–577.

[13] J.-F. Mertens and A. Neyman, *Stochastic games*, International Journal of Game Theory 10 (1981), 53–66.

[14] J.-F. Mertens, A. Neyman, and D. Rosenberg, *Absorbing games with compact action spaces*, Mathematics of Operations Research 34 (2009), no. 2, 257–262.

[15] J. Renault, *Uniform value in dynamic programming*, Journal of the European Mathematical Society 13 (2010), 309–330.

[16] ______, *Zero-sum stochastic games*, Lecture notes, available at https://sites.google.com/site/jrenaultsite/lecturenotes, 2015.

[17] D. Rosenberg and S. Sorin, *An operator approach to zero-sum repeated games*, Israel Journal of Mathematics 121 (2001), 221–246.

[18] K.W. Ross and R. Varadarajan, *Multichain markov decision processes with a sample path constraint: a decomposition approach*, Mathematics of Operations Research 16 (1991), no. 1, 195–207.

[19] L.S. Shapley, *Stochastic games*, Proceeding of the National Academy of Sciences of the United States of America 39 (1953), no. 10, 1095–1100.

[20] E. Solan, *Stochastic games and applications*, 570, ch. Perturbations of Markov chains with applications to stochastic games, pp. 265–280, Springer Netherlands, 2003.
[21] S. Sorin, *A first course on zero-sum repeated games*, Mathématiques et Applications, no. 37, Springer, 2002.

[22] S. Sorin and G. Vigeral, *Reversibility and oscillations in zero-sum discounted stochastic games*, Journal of Dynamics and Games 2 (2015), no. 1, 103–115.

[23] G. Vigeral, *A zero-sum stochastic game with compact action sets and no asymptotic value*, Dynamic Games and Applications 3 (2013), no. 2, 172–186.

[24] O.J. Vrieze, *Stochastic games and applications*, 570, ch. Stochastic Games and Stationary Strategies, pp. 37–50, Springer Netherlands, 2003.

[25] B. Ziliotto, *A Tauberian theorem for nonexpansive operators and applications to zero-sum stochastic games*, Mathematics of Operations Research 41 (2016), no. 4, 1522 – 1534.

[26] ______, *Zero-sum repeated games: counterexamples to the existence of the asymptotic value and the conjecture maxmin=lim v(n)*, Annals of Probability 44 (2016), no. 2, 1107–1133.