Two-Player Perfect-Information Shift-Invariant Submixing Stochastic Games Are Half-Positional

Hugo Gimbert, Edon Kelmendi

Two-Player Perfect-Information Shift-Invariant Submixing Stochastic Games Are Half-Positional

Hugo Gimbert*
CNRS, LaBRI, Université de Bordeaux, France
hugo.gimbert@cnrs.fr

Edon Kelmendi
LaBRI, Université de Bordeaux, France
edon.kelmendi@labri.fr

October 7, 2015

Abstract

We consider zero-sum stochastic games with perfect information and finitely many states and actions. The payoff is computed by a function which associates to each infinite sequence of states and actions a real number. We prove that if the the payoff function is both shift-invariant and submixing, then the game is half-positional, i.e. the first player has an optimal strategy which is both deterministic and stationary. This result relies on the existence of ϵ-subgame-perfect strategies in shift-invariant games, a second contribution of the paper.

1 Introduction.

We consider zero-sum stochastic games with finitely many states $\mathbf{S}$ and actions $\mathbf{A}$, perfect information and infinite duration. Each state is controlled by either Player 1 or Player 2. A play of the game is an infinite sequence of steps: at each step the game is in some state $s \in \mathbf{S}$ and the player who controls this state chooses an action, which determines a lottery that is used to randomly choose the next state. Players have full knowledge about the rules of the game (the states and actions sets, who controls which state and the lotteries associated to pairs of states and actions) and when they choose an action they have full knowledge of the actions played and the states visited so far.

Each player wants to maximize his expected payoff, and the game is zero-sum. The payoff of Player 1 (which is exactly the loss of Player 2) associated with an infinite play $s_0 a_1 s_1 \cdots \in (\mathbf{SA})^\omega$ is computed by a measurable and bounded payoff function $f : (\mathbf{SA})^\omega \to \mathbb{R}$.

*This work was supported by the ANR projet "Stoch-MC" and the LaBEX "CPU".

Well-known examples of payoff functions are the discounted payoff function, the mean-payoff function, the limsup payoff function and the parity payoff function. These four classes of games share a common property: in these games both players have *optimal* strategies and moreover these strategies can be chosen to be both *deterministic* and *stationary*: such strategies guarantee a maximal expected payoff and choose actions deterministically and this deterministic choice only depends on the current state.

When deterministic and stationary optimal strategies exist for Player 1 in any game equipped with f, we say f is *half-positional*, and we say f is *positional* if such strategies exist for both players.

There has been numerous papers about the existence of deterministic and stationary optimal strategies in games with different payoff functions. Shapley proved that stochastic games with discounted payoff function are positional using an operator approach [Sha53]. Derman showed the positionality of one-player games with expected mean-payoff reward, using an Abelian theorem and a reduction to discounted games [Der62]. Gilette extended Derman's result to two-player games [Gil57] but his proof was found to be wrong and corrected by Ligget and Lippman [LL69]. The positionality of one-player parity games was addressed in [CY90] and later on extended to two-player games in [CJH03, Zie04]. Counter games were extensively studied in [BBE10] and several examples of positional counter games are given. There are also several examples of one-player and two-player positional games in [Gim07, Zie10]. A whole zoology of half-positional games is presented in [Kop09] and another example is given by mean-payoff co-Büchi games [CHJ05]. The proofs of these various results are mostly ad-hoc and very heterogeneous.

Some research has been made to find a common property of these games which explains why they are positional or half-positional. It appears that *shift-invariant* and *submixing* payoff functions play a central role. In a nutshell, a payoff is *shift-invariant* if changing a finite-prefix of a play does not change the payoff of the play and it is *submixing* if the payoff associated to the shuffle of two plays cannot be greater than both payoffs of both shuffled plays, see Definitions 3.2 and 5.1.

For one-player games it was proved by the first author that every one-player game equipped with such a payoff function is positional [Gim07]. This result was successfully used in [BBE10] to prove positionality of counter games. A weaker form of this condition was presented in [GZ04] to prove positionality of deterministic games (i.e. games where transition probabilities are equal to 0 or 1). Kopczynski proved that two-player deterministic games equipped with a *shift-invariant* and *submixing* which takes only two values is half-positional [Kop06].

A recent result [Zie10] provides a necessary and sufficient condition for the positionality of one-player games. The condition is expressed in terms of the existence of particular optimal strategies in multi-armed bandit games. When trying to prove the positionality of a particular payoff function, the condition in [Zie10] is much harder to check than the submixing property which is purely syntactic.

The present paper provides two contributions.

First, in games whose payoff function is shift-invariant, both players have ϵ-subgame-perfect strategies, i.e. strategies that are ϵ-optimal not only when the game starts but also whatever finite play has already been played (Theorem 4.1).

Second, every two-player game equipped with a shift-invariant and submixing payoff function is half-positional, i.e. Player 1 has an optimal strategy which is both deterministic and stationary (Theorem 5.2).

In parallel to our work, the first result was obtained very recently in [MY15, Proposition 11], for a larger class of games and equilibria.

The paper starts with preliminaries, then in Section 3 we provide several examples of payoff functions. In Section 4 we state and prove the existence of ϵ-subgame-perfect strategies in shift-invariant games, in Section 5 we state and prove that games with shift-invariant and submixing payoff functions are half-positional, in Section 6 we present some applications.

2 Stochastic games with perfect information.

In this section, we present the notion of stochastic games with perfect information, of the value and determinacy of such games, as well as several results about martingales that are used in the next section.

2.1 Games

A game is specified by the *arena* and the *payoff function*. While the arena determines *how* the game is played, the payoff function specifies *what for* the players play.

We use the following notations throughout the paper. Let $\mathbf{S}$ be a finite set. The set of finite (resp. infinite) sequences on $\mathbf{S}$ is denoted $\mathbf{S}^*$ (resp. $\mathbf{S}^\omega$) and $\mathbf{S}^\infty = \mathbf{S}^* \cup \mathbf{S}^\omega$.

A *probability distribution* on $\mathbf{S}$ is a function $\delta : \mathbf{S} \to \mathbb{R}$ such that $\forall s \in \mathbf{S},\ 0 \leq \delta(s) \leq 1$ and $\sum_{s\in\mathbf{S}} \delta(s) = 1$. The set of probability distributions on $\mathbf{S}$ is denoted $\Delta(\mathbf{S})$.

Definition 2.1 (Arenas)**.** *A stochastic arena with perfect information* $\mathcal{A} = (\mathbf{S}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{A}, (\mathbf{A}(s))_{s\in\mathbf{S}}, p)$ *is made of:*

- *a set of states* $\mathbf{S}$ *partitioned in two sets* $(\mathbf{S}_1, \mathbf{S}_2)$*,*
- *a set of actions* $\mathbf{A}$*,*
- *for each state* $s \in \mathbf{S}$*, a non-empty set* $\mathbf{A}(s) \subseteq \mathbf{A}$ *of actions available in* s*,*
- *and transition probabilities* $p : \mathbf{S} \times \mathbf{A} \to \Delta(\mathbf{S})$ *.*

In the sequel, stochastic arenas with perfect information are simply called *arenas* and we only consider arenas with finitely many states and actions.

An *infinite play* in an arena $\mathcal{A}$ is an infinite sequence $p = s_0 a_1 s_1 a_2 \cdots \in (\mathbf{SA})^\omega$ such that for every $n \in \mathbb{N}$, $a_{n+1} \in \mathbf{A}(s_n)$. A *finite play* in $\mathcal{A}$ is a finite sequence in $\mathbf{S}(\mathbf{AS})^*$ which is the prefix of an infinite play. The first state of a play is called its *source*, the last state of a finite play is called its *target*.

With each infinite play is associated a payoff computed by a *payoff function*. Player 1 prefers strategies that maximize the expected payoff while Player 2 has the exact opposite preference.

Formally, a payoff function for the arena $\mathcal{A}$ is a bounded and Borel-measurable function $f : (\mathbf{SA})^\omega \to \mathbb{R}$ which associates with each infinite play h a payoff $f(h)$. In the next section, we will present several examples of such functions.

Definition 2.2 (Stochastic game with perfect information)**.** *A stochastic game with perfect information is a pair* $\mathbf{G} = (\mathcal{A}, f)$ *where* $\mathcal{A}$ *is an arena and* f *a payoff function for the arena* $\mathcal{A}$*.*

2.2 Strategies

A *strategy* in an arena $\mathcal{A}$ for Player 1 is a function $\sigma : (\mathbf{SA})^*\mathbf{S}_1 \to \Delta(\mathbf{A})$ such that for any finite play $s_0 a_1 \cdots s_n$, and every action $a \in \mathbf{A}$, $(\sigma(s_0 a_1 \cdots s_n)(a) > 0) \implies (a \in \mathbf{A}(s_n))$. Strategies for Player 2 are defined similarly and are typically denoted τ.

We are especially interested in a very simple class of strategies: deterministic and stationary strategies.

Definition 2.3 (Deterministic and stationary strategies)**.** *A strategy* σ *for Player* 1 *is* deterministic *if for every finite play* $h \in (\mathbf{SA})^*\mathbf{S}_1$ *and action* $a \in \mathbf{A}$*,*

$$(\sigma(h)(a) > 0) \iff (\sigma(h)(a) = 1) \ .$$

A strategy σ *is* stationary *if* $\sigma(h)$ *only depends on the target of* h*. In other words* σ *is stationary if for every state* $t \in \mathbf{S}_1$ *and for every finite play* h *with target* t*,*

$$\sigma(h) = \sigma(t) \ .$$

In the definition of a stationary strategy, remark that $t \in \mathbf{S}$ denotes at the same time the target of the finite play h as well as the finite play t of length 1.

Given an initial state $s \in \mathbf{S}$ and strategies σ and τ for players 1 and 2 respectively, the set of infinite plays with source s is naturally equipped with a sigma-field and a probability measure denoted $\mathbb{P}_s^{\sigma,\tau}$ that are defined as follows. Given a finite play h and an action a, the set of infinite plays $h(\mathbf{AS})^\omega$ and $ha(\mathbf{SA})^\omega$ are *cylinders* that we abusively denote h and ha. The sigma-field is the one generated by cylinders and $\mathbb{P}_s^{\sigma,\tau}$ is the unique probability measure on the set of infinite plays with source s such that for every finite play h with target t, for every action $a \in \mathbf{A}$ and state $r \in \mathbf{S}$,

$$\mathbb{P}_s^{\sigma,\tau}(ha \mid h) = \begin{cases} \sigma(h)(a) & \text{if } t \in S_1, \\ \tau(h)(a) & \text{if } t \in S_2, \end{cases} \tag{1}$$

$$\mathbb{P}_s^{\sigma,\tau}(har \mid ha) = p(r|t,a) \ . \tag{2}$$

For $n \in \mathbb{N}$, we denote S_n and A_n the random variables defined by $S_n(s_0a_1s_1\cdots) = s_n$ and $A_n(s_0a_1s_1\cdots) = a_n$.

2.3 Values and optimal strategies

Let $\mathbf{G}$ be a game with a bounded measurable payoff function $f : (\mathbf{SA})^\omega \to \mathbb{R}$. The expected payoff associated with an initial state s and two strategies σ and τ is the expected value of f under $\mathbb{P}_s^{\sigma,\tau}$, denoted $\mathbb{E}_s^{\sigma,\tau}[f]$.

The *maxmin* and *minmax* values of a state $s \in \mathbf{S}$ in the game $\mathbf{G}$ are:

$$\begin{aligned}\mathrm{maxmin}(\mathbf{G})(s) &= \sup_\sigma \inf_\tau \mathbb{E}_s^{\sigma,\tau}[f] \ , \\ \mathrm{minmax}(\mathbf{G})(s) &= \inf_\tau \sup_\sigma \mathbb{E}_s^{\sigma,\tau}[f] \ .\end{aligned}$$

By definition of maxmin and minmax, for every state $s \in \mathbf{S}$, $\mathrm{maxmin}(\mathbf{G})(s) \leq \mathrm{minmax}(\mathbf{G})(s)$. As a corollary of the Martin's second determinacy theorem [Mar98], the converse inequality holds as well:

Theorem 2.4 (Martin's second determinacy theorem)**.** *Let $\mathbf{G}$ be a game with a Borel-measurable and bounded payoff function f. Then for every state $s \in \mathbf{S}$:*

$$\mathrm{maxmin}(\mathbf{G})(s) = \mathrm{minmax}(\mathbf{G})(s) \ .$$

This common value is called the value of state s in the game $\mathbf{G}$ and denoted $\mathrm{val}(\mathbf{G})(s)$.

The existence of a value guarantees the existence of ϵ-optimal strategies for both players and every $\epsilon > 0$.

Definition 2.5 (optimal and ϵ-optimal strategies)**.** *Let $\mathbf{G}$ be a game, $\epsilon > 0$ and σ a strategy for Player* 1*. Then σ is ϵ-optimal if for every strategy τ and every state $s \in \mathbf{S}$,*

$$\mathbb{E}_s^{\sigma,\tau}[f] \geq \mathrm{minmax}(\mathbf{G})(s) - \epsilon \ .$$

The definition for Player 2 *is symmetric. A* 0*-optimal strategy is simply called* optimal.

The following proposition provides a link between the notion of optimal strategies and the notion of value.

Proposition 2.6. *Let $\mathbf{G}$ be a game and suppose that Player* 1 *has an optimal strategy $\sigma^\sharp$. Then $\mathbf{G}$ has a value and for every state $s \in \mathbf{S}$:*

$$\mathrm{val}(\mathbf{G})(s) = \inf_\tau \mathbb{E}_s^{\sigma^\sharp,\tau}[f] \ . \tag{3}$$

Proof. By definition of the maxmin, $\mathrm{maxmin}(\mathbf{G})(s) \geq \inf_\tau \mathbb{E}_s^{\sigma^\sharp,\tau}[f]$ and by definition of an optimal strategy, $\inf_\tau \mathbb{E}_s^{\sigma^\sharp,\tau}[f] \geq \mathrm{minmax}(\mathbf{G})(s)$. As a consequence, $\mathrm{maxmin}(s) \geq \mathrm{minmax}(s)$ thus s has a value and (3) holds. □

An even stronger class of ϵ-optimal strategies are ϵ-subgame-perfect strategies, i.e. strategies that are not only ϵ-optimal from the initial state s but stays also ϵ-optimal whatever the beginning of the play is.

Given a finite play $h = s_0 \cdots s_n$ and a function g whose domain is the set of infinite plays, or the set of finite plays, by $g[h]$ we denote the function g *shifted* by h:

$$g[h](t_0 a_1 t_1 \cdots) = \begin{cases} g(h a_1 t_1 \cdots) & \text{if } s_n = t_0, \\ g(t_0 a_1 t_1 \cdots) & \text{otherwise.} \end{cases}$$

Definition 2.7 (ϵ-subgame-perfect strategies)**.** *Let $\mathbf{G}$ be a game equipped with a payoff function f. A strategy for Player 1, σ is said to be ϵ-subgame-perfect if for every finite play $h = s_0 \cdots s_n$*

$$\inf_{\tau} \mathbb{E}_{s_n}^{\sigma[h],\tau}\left[f[h]\right] \geq \mathrm{val}(\mathbf{G})(s_n) - \epsilon$$

2.4 Martingales

In Section 4 we observe that the stochastic process, $\mathrm{val}(\mathbf{G})(S_0), \mathrm{val}(\mathbf{G})(S_1), \ldots$ is a martingale when both players choose only "value-preserving" actions, and make use of the following basic result about martingales.

Proposition 2.8. *Let T be a stopping time with respect to a sequence of random variables $S_0, S_1, \ldots$. Let $(X_n)_{n\in\mathbb{N}}$ be a martingale such that for every $n \in \mathbb{N}$, X_n is $(S_0, \ldots, S_n)$-measurable. Assume there exists $K > 0$ such that $\forall n \in \mathbb{N}, \mathbb{P}(|X_n| \leq K) = 1$. Let X_T be the random variable defined by:*

$$X_T = \begin{cases} X_n & \text{if } T \text{ is finite equal to } n, \\ \lim_{n\in\mathbb{N}} X_n & \text{if } T = \infty. \end{cases}$$

Then:

$$\mathbb{E}[X_T] = \mathbb{E}[X_0] \ .$$

Similarly if the process is a supermartingale or a submartingale, and the same conditions hold we have $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$ and $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$ respectively.

Proof. According to Doob's convergence theorem for martingales, X_T is well-defined. For every $k \in \mathbb{N}$ let $Y_k = X_{\min(T,k)}$. The stopped process $(Y_k)_{k\in\mathbb{N}}$ is also a martingale, a basic property of martingales. Since $(X_n)_{n\in\mathbb{N}}$ is bounded by K, $(Y_n)_{n\in\mathbb{N}}$ also is, and according to Doob's theorem it converges almost-surely. By definition of X_T the limit of $(Y_n)_{n\in\mathbb{N}}$ is X_T. By definition of martingales, for every $n \in \mathbb{N}, \mathbb{E}[Y_n] = \mathbb{E}[Y_0] = \mathbb{E}[X_0]$ thus according to Lebesgue dominated convergence theorem $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. A similar proof applies for the case of supermartingales or submartingales. □

3 Computing payoffs

In this section, we present several examples of payoff functions and generalize the definition of a payoff function to cover these examples.

3.1 Examples

Among the most well-known examples of payoff functions, are the mean-payoff and the discounted payoff functions, used in economics, as well as the parity condition, used in logics and computer science, and the limsup payoff function, used in game theory.

The *mean-payoff function* has been introduced by Gilette [Gil57]. Intuitively, it measures average performances. Each state $s \in \mathbf{S}$ is labeled with an immediate reward $r(s) \in \mathbb{R}$. With an infinite play $s_0 a_1 s_1 \cdots$ is associated an infinite sequence of rewards $r_0 = r(s_0), r_1 = r(s_1), \ldots$ and the payoff is:

$$f_{\text{mean}}(r_0 r_1 \cdots) = \limsup_n \frac{1}{n+1} \sum_{i=0}^{n} r_i \ . \tag{4}$$

The *discounted payoff* has been introduced by Shapley [Sha53]. Intuitively, it measures long-term performances with an inflation rate: immediate rewards are discounted. Each state s is labeled not only with an immediate reward $r(s) \in \mathbb{R}$ but also with a discount factor $0 \leq \lambda(s) < 1$. With an infinite play h labeled with the sequence $(r_0, \lambda_0)(r_1, \lambda_1) \cdots \in (\mathbb{R} \times [0, 1[)^\omega$ of daily payoffs and discount factors is associated the payoff:

$$f_{\text{disc}}((r_0, \lambda_0)(r_1, \lambda_1) \cdots) = r_0 + \lambda_0 r_1 + \lambda_0 \lambda_1 r_2 + \cdots \ . \tag{5}$$

The *parity condition* is used in automata theory and logics [GTW02]. Each state s is labeled with some priority $c(s) \in \{0, \ldots, d\}$. The payoff is 1 if the highest priority seen infinitely often is odd, and 0 otherwise. For $c_0 c_1 \cdots \in \{0, \ldots, d\}^\omega$,

$$f_{\text{par}}(c_0 c_1 \cdots) = \begin{cases} 0 \text{ if } \limsup_n c_n \text{ is even,} \\ 1 \text{ otherwise.} \end{cases} \tag{6}$$

The *limsup payoff function* has been used in the theory of gambling games [MS96]. States are labeled with immediate rewards and the payoff is the supremum limit of the rewards: $f_{\text{lsup}}(r_0 r_1 \cdots) = \limsup_n r_n$.

One-counter stochastic games have been introduced in [BBE10], in these games each state $s \in S$ is labeled by a relative integer from $c(s) \in \mathbb{Z}$. Three different winning conditions were defined and studied in [BBE10]:

$$\limsup_n \sum_{0 \leq i \leq n} c_i = +\infty \tag{7}$$

$$\limsup_n \sum_{0 \leq i \leq n} c_i = -\infty \tag{8}$$

$$f_{\text{mean}}(c_0 c_1 \ldots) > 0 \tag{9}$$

Generalized mean payoff games were introduced in [CDHR10]. Each state is labeled by a fixed number of immediate rewards $\left(r^{(1)}, \ldots, r^{(k)}\right)$, which define as many mean payoff conditions $\left(f^1_{\text{mean}}, \ldots, f^k_{\text{mean}}\right)$. The winning condition is:

$$\forall 1 \leq i \leq k, f^i_{\text{mean}}\left(r_0^{(i)} r_1^{(i)} \ldots\right) > 0 \ . \tag{10}$$

3.2 Payoff functions

In the four examples above, the way payoffs are computed is actually independent from the arena which is considered. To be able to consider a payoff function independently of the arenas equipped with this payoff function, we generalize the definition of payoff functions.

Definition 3.1 (Payoff functions)**.** *A* payoff function *is a bounded and measurable function $f : \mathbf{C}^\omega \to \mathbb{R}$ where $\mathbf{C}$ is a finite set called the set of colours. We say a game $\mathbf{G} = (\mathcal{A}, g)$ is* equipped *with f if there exists a mapping $r : \mathbf{S} \times \mathbf{A} \to \mathbf{C}$ such that for every infinite play $s_0 a_1 s_1 a_2 \cdots$ in the arena $\mathcal{A}$,*

$$g(s_0, a_1, s_1, a_2, \ldots) = f(r(s_0, a_1), r(s_1, a_2), \ldots) \ .$$

In the case of the mean payoff and the limsup payoff functions, colours are real numbers and $\mathbf{C} \subseteq \mathbb{R}$, whereas in the case of the discounted payoff colours are pairs $\mathbf{C} \subseteq \mathbb{R} \times [0, 1[$ and for the parity game colours are integers $\mathbf{C} = \{0, \ldots, d\}$.

Throughout this paper, we focus on shift-invariant payoff functions.

Definition 3.2 (Shift-invariant)**.** *A payoff function $f : \mathbf{C}^\omega \to \mathbb{R}$ is said to be a* shift-invariant *payoff function if:*

$$\forall c \in \mathbf{C}, \forall u \in \mathbf{C}^\omega, f(cu) = f(u) \ . \tag{11}$$

Note that shift-invariance is a strictly stronger property than tail-measurability. Tail-measurability means that for every $n \in \mathbb{N}$, the value of $f(c_0 c_1 \cdots)$ is independent of the coordinates $c_0, \ldots, c_n$. For example the function

$$f(c_0 c_1 \cdots) = \begin{cases} 1 & \text{if } \exists n \in \mathbb{N}, \forall k, k' \geq n, c_{2*k} = c_{2*k'} \\ 0 & \text{otherwise,} \end{cases}$$

is tail-measurable but not shift-invariant [Zie11]. It seems to the authors of the present paper that the results of [Cha07] hold for shift-invariant winning objectives but no proof is given under the weaker assumption of tail-measurability. Actually, tail-measurability and shift-invariance are presented as equivalent in the preliminaries ([Cha07, l. 28 p. 184]) and the shift-invariance hypothesis is used in the core of the proof ([Cha07, l. -1 p.190]).

4 On the existence of ϵ-subgame-perfect strategies

The following theorem is one of the two contributions of the paper, and is a cornerstone for the result of the next section.

In parallel to our work, this result was obtained very recently in [MY15, Proposition 11], for a larger class of games and equilibria.

Theorem 4.1. *Let $\mathbf{G}$ be a game equipped with a payoff function f and $\epsilon > 0$. Assume f is shift-invariant. Then both players have ϵ-subgame-perfect strategies in $\mathbf{G}$.*

The remainder of this section is devoted to proving this theorem. The proof is done from the point of view of Player 1, but it holds symmetrically for Player 2.

In order to avoid heavy notations in the proof of Theorem 4.1 we fix an arbitrary $\epsilon > 0$ for the entire Section 4 with the intention of proving the existence of 2ϵ-subgame perfect strategies.

We fix for the rest of the section a game $\mathbf{G}$ equipped with a shift-invariant payoff function f.

Section 4.1 introduces the central notion of weaknesses. In Section 4.2 we define a reset strategy which detects the weaknesses and responds to them by resetting the memory. Then the idea is to show that when the reset strategy is used there will be only finitely many weaknesses and hence only finitely many resets, almost surely. This is done in Section 4.4. But before this, in Section 4.3 we demonstrate that for all $\epsilon > 0$ there are ϵ-optimal strategies which play only *value-preserving* actions. And at last in Section 4.5 Theorem 4.1 is proved.

4.1 Weaknesses

Definition 4.2 (Weakness)**.** *Given a strategy σ for Player 1, a finite play $h \in \mathbf{S}(\mathbf{AS})^*$ is a σ-weakness if $\sigma[h]$ is not 2ϵ-optimal.*

Notice that a strategy is 2ϵ-subgame perfect if and only if there is no σ-weakness.

When playing with strategy σ we say that a weakness occurs in an infinite play if some finite prefix of it is a σ-weakness.

Let σ be a strategy for Player 1 and $h = s_0 a_1 s_1 \dots$ an infinite play of a game. Then h can be factorized in σ-weaknesses in the following way: a) let $h_0 = s_0 a_1 s_1 \cdots s_n$ be the shortest prefix of h such that h_0 is a σ-weakness, if no such prefix exists we are done, b) repeat step (a) for the infinite play $s_n a_{n+1} s_{n+1} \cdots$. In this way we produce the plays $h_0, h_1, \dots$ such that for all $i \in \mathbb{N}$:

1. if h_i is finite then h_i is a σ-weakness,
2. if h_i is finite then no strict prefix of h_i is a σ-weakness,
3. if h_i is infinite then $h = h_0 h_1 \cdots h_i$ and no prefix of h_i is a σ-weakness,

The reset strategy defined in the following section will reset the memory when a weakness occurs, hence it will namely reset it after seeing each factor $h_0, h_1, \dots$.

This factorization is made more formal with the following definition of the function δ, which gives the dates of the factorizations.

Definition 4.3. *Let σ be a strategy for Player 1, and $s_0 \dots s_n \in \mathbf{S}(\mathbf{AS})^*$ define inductively on n*

$$\delta_\sigma(s_0 \dots s_n) = \begin{cases} n & \text{if } h \text{ is a } \sigma\text{-weakness,} \\ \delta_\sigma(s_0 \dots s_{n-1}) & \text{otherwise} \end{cases}$$

where $h = s_{\delta_\sigma(s_0 \dots s_{n-1})} \dots s_n$*, and* $\delta_\sigma(s_0) = 0$*. For* $n > 0$*, we say that a weakness occurs at date* n *if* $\delta_\sigma(s_0 \dots s_n) = n$*.*

4.2 The reset strategy

Using the notion of a weakness, given some strategy σ, we define another strategy called the *reset* strategy, that resets the memory whenever a weakness occurs, and the rest of the Section 4 is devoted to proving that given a certain ϵ-optimal strategy σ then the reset strategy based on it is 2ϵ-subgame perfect.

Definition 4.4 (The reset strategy)**.** *Let* σ *be an* ϵ*-optimal strategy. We define the reset strategy* $\hat{\sigma}$ *based on* σ *as the strategy that resets the memory whenever a weakness occurs, that is*

$$\hat{\sigma}(s_0 \dots s_n) = \sigma(s_{\delta_\sigma(s_0 \dots s_n)} \dots s_n).$$

The construction of the reset strategy is an extension of the construction of the *switching strategy* in the proof of [Cha07, Theorem 1]; while at most one memory reset is performed by the switching strategy of [Cha07], the reset strategy of the present may reset its memory infinitely many times.

The reset strategy has been used in [HG08] to prove the existence of optimal strategies in games with perfect information and two-valued payoff functions. Note that in general the two-valued hypothesis is necessary: for example there is in general no optimal strategy in a game colored by $\{0, 1\}$ and equipped with the payoff function

$$f(c_0 c_1 \cdots) = \begin{cases} 0 & \text{if } \forall n \in \mathbb{N}, c_n = 0 \\ 1 - 2^{-\min\{n | c_n = 1\}} & \text{otherwise.} \end{cases}$$

4.3 Locally-optimal and ϵ-optimal strategies

An action is called *value-preserving* if it preserves the value of the state, namely the expected value of the future state will be equal to the value of the present state. Strategies that play only value-preserving actions are called locally-optimal strategies. In this section, we prove the somewhat intuitive fact that for every $\epsilon' > 0$ there are ϵ'-optimal strategies that are also locally-optimal, and more interestingly that if one player plays with a locally-optimal strategy, then the other one is also forced to do so eventually, in almost all trajectories.

4.3.1 Locally-optimal strategies

When we are obviously dealing with only one game we use the notation $val(s)$ instead of $val(\mathbf{G})(s)$.

Definition 4.5. *The action* $a \in A(s)$ *is* value-preserving *in state* $s \in \mathbf{S}$ *if and only if* $val(s) = \sum_{s' \in \mathbf{S}} p(s, a, s') val(s')$*. When an action is not value-preserving we call it* value-changing

Definition 4.6. *A strategy is* locally-optimal *if it plays only value-preserving actions.*

Note that a locally-optimal strategy is not necessarily optimal, but an optimal strategy is locally-optimal. When both players play only value-preserving actions, it is worth noting that the stochastic process $(val(S_n))_{n\in\mathbb{N}}$ is a martingale.

Theorem 4.7. *Let σ and τ be strategies for players 1 and 2 respectively, and $s \in \mathbf{S}$. Assume σ is locally-optimal then $(val(S_n))_{n\in\mathbb{N}}$ is a submartingale for the probability measure $\mathbb{P}_s^{\sigma,\tau}$. If both σ and τ are locally-optimal then $(val(S_n))_{n\in\mathbb{N}}$ is a martingale.*

Proof. By definition f is bounded hence for all $n \in \mathbb{N}$, all strategies σ, τ and states $s \in \mathbf{S}$, $\mathbb{E}_s^{\sigma,\tau}[|val(S_n)|] < \infty$. Then, for locally-optimal σ, strategies τ and states $s \in \mathbf{S}$, $\mathbb{E}_s^{\sigma,\tau}[val(S_{n+1}) \mid val(S_0), \ldots, \mathrm{val}(S_n)]$ is $\sum_{a\in\mathbf{A}(S_n)} \sigma(S_0 \cdots S_n)(a)\big(\sum_{s'\in\mathbf{S}} p(S_n, a, s')val(s')\big)$ if $S_n \in \mathbf{S}_1$ and $\sum_{a\in\mathbf{A}(S_n)} \tau(S_0 \cdots S_n)(a)\big(\sum_{s'\in\mathbf{S}} p(S_n, a, s')val(s')\big)$ otherwise. And we see that in both cases $\mathbb{E}_s^{\sigma,\tau}[val(S_{n+1}) \mid val(S_0), \ldots, \mathrm{val}(S_n)] \geq val(S_n)$ by definition of a locally-optimal strategy σ. Clearly the same proof holds when τ is also locally-optimal, making the inequality above an equality. □

4.3.2 Both players play ultimately only value-preserving actions

In this section we show that if Player 1 plays with a locally optimal strategy then almost surely both players eventually will play only value-preserving actions. We actually prove an even stronger fact: under the same hypothesis, both players eventually play only actions which cannot change the value of the state. First we define such actions:

Definition 4.8 (Stable actions)**.** *For some $s \in S$ we call an action $a \in A(s)$ stable if for all $s' \in \mathbf{S}$ we have:*

$$p(s, a, s') > 0 \implies val(s) = val(s').$$

Note that a stable action is necessarily value-preserving, and a value-changing action cannot be stable.

Lemma 4.9. *For every locally-optimal strategy σ for Player 1, strategy τ for Player 2 and $s \in \mathbf{S}$, $\mathbb{P}_s^{\sigma,\tau}(\exists n \in \mathbb{N}, \forall k \geq n, A_n$ is stable $) = 1$.*

Proof. In case every action is stable, we have nothing to prove. Assume that there exists an action $a \in A(s)$ which is not stable. Let $s' \in \mathbf{S}$ be such that $p(s, a, s') > 0$ and $val(s) \neq val(s')$. Denote the event "we see action a infinitely often" by E_{sa}, formally

$$E_{sa} := \{\forall m \in \mathbb{N},\ \exists n \geq m,\ S_n = s \wedge A_{n+1} = a\}.$$

The goal is to prove that for all locally-optimal strategies σ, $t \in \mathbf{S}$ and τ we have $\mathbb{P}_t^{\sigma,\tau}(E_{sa}) = 0$. Assume on the contrary that for some locally-optimal σ,

$t \in \mathbf{S}$ and τ we have $\mathbb{P}_t^{\sigma,\tau}(E_{sa}) > 0$. This implies that also $\mathbb{P}_t^{\sigma,\tau}(E_{sas'}) > 0$, where $E_{sas'} := \{\forall m \in \mathbb{N},\ \exists n \geq m,\ S_n = s \wedge A_{n+1} = a \wedge S_{n+1} = s'\}$. And this in turn implies that when playing with σ, τ and starting from state t, there is some non-zero probability that for infinitely many $n \in \mathbb{N}$,

$$| \ val(S_n) - val(S_{n+1}) \ | \geq val(s) - val(s') > 0. \tag{12}$$

Clearly a consequence of (12) is that there is some non-zero probability that the sequence $val(S_0), val(S_1), \ldots$ does not converge. But according to Theorem 4.7, the process $val(S_0), val(S_1), \ldots$ is a submartingale and it converges almost surely as a consequence of Doob's convergence theorem for martingales. Hence for all locally-optimal strategies σ all $t \in \mathbf{S}$ and τ we have $\mathbb{P}_t^{\sigma,\tau}(E_{sa}) = 0$. □

Since stable actions are value-preserving, it implies that when Player 1 plays with a locally-optimal strategy, value-changing actions are almost surely played only finitely often.

Corollary 4.10. *For every locally-optimal strategy σ, strategy τ and $s \in \mathbf{S}$,*

$$\mathbb{P}_s^{\sigma,\tau}(\exists n \in \mathbb{N}, \forall k \geq n, A_n \text{ is value-preserving in } S_n) = 1.$$

4.3.3 There exist an ϵ-optimal strategy which is locally-optimal

Playing value-changing actions strictly decreases (or increases, depending on which player plays it) the expected payoff hence it is intuitive that for every $\epsilon' > 0$ there are ϵ'-optimal strategies for both players, that play only value-preserving actions.

This is made rigorous in the following lemma.

Lemma 4.11. *Let $\epsilon' > 0$, and $a \in A(s)$ a value-changing action in state $s \in \mathbf{S}_1$. Then there exists an ϵ'-optimal strategy for Player 1 that never plays action a.*

The proof simply associates with every strategy using a, a better strategy that does not use a.

Proof. Let $\mathbf{G}'$ be the game identical to $\mathbf{G}$ except that $a \notin A(s)$ (action a is removed). For all $t \in S$ it is immediate that $val(\mathbf{G})(t) \geq val(\mathbf{G}')(t)$ since Player 1 has one more action to choose from in the game $\mathbf{G}$, while Player 2 has the same number of actions to choose from. Hence our goal is to prove the following:

$$\forall t \in S,\ val(\mathbf{G}')(t) \geq val(\mathbf{G})(t). \tag{13}$$

We split the proof of (13) in two cases, for $t = s$ and $t \neq s$.

- $val(\mathbf{G}')(s) \geq val(\mathbf{G})(s)$

  Let $d = val(\mathbf{G})(s) - \sum_{t \in S} p(s,a)(t) val(\mathbf{G})(t) > 0$, and τ the strategy for Player 2 in $\mathbf{G}$ that plays according to strategy τ' which is ϵ'-optimal in $\mathbf{G}'$,

as long as Player 1 does not choose the value-changing action a. In case she chooses it, τ switches definitely to the strategy τ'' that is $\frac{d}{2}$-optimal in $\mathbf{G}$. Let Opt be the event $(\forall n \in \mathbb{N},\ S_n = s \implies A_{n+1} \neq a)$, that is the event that Player 1 never chooses the value-changing action a.

When playing with the strategy τ we have the following properties, for all t and σ:

$$\mathbb{E}_t^{\sigma,\tau}[f \mid Opt] \leq val(\mathbf{G}')(t) + \epsilon' \tag{14}$$

$$\mathbb{E}_t^{\sigma,\tau}[f \mid \neg Opt] \leq val(\mathbf{G})(s) - d + \frac{d}{2} \tag{15}$$

To show (14), note because of the condition Opt the game is played only in $\mathbf{G}'$, hence the strategy σ even though it is a strategy in $\mathbf{G}$ it behaves like the strategy σ' in $\mathbf{G}'$ defined in the following way: for all $h = s_0 \dots s_n \in S(AS)^*$, and $b \in A(s_n)$, $\sigma'(h)(b) = \mathbb{P}_{s_0}^{\sigma,\tau}(hb \mid h \wedge Opt)$. That is $\mathbb{E}_t^{\sigma,\tau}[f \mid Opt] = \mathbb{E}_t^{\sigma',\tau'}[f]$, because τ never has to switch to τ''. Now (14) is a direct consequence of the ϵ'-optimality of τ'.

We get (15), because when Player 1 chooses the action a, τ switches to the τ'' strategy which is $\frac{d}{2}$-optimal in $\mathbf{G}$, and the decrease by d is a consequence of the choice of the action a, and the definition of d.

Since $\mathbb{E}_t^{\sigma,\tau}[f]$ is a convex combination of $\mathbb{E}_t^{\sigma,\tau}[f \mid Opt]$ and $\mathbb{E}_t^{\sigma,\tau}[f \mid \neg Opt]$, as a consequence of (14) and (15) we have that for all $t \in S$, $\epsilon' > 0$ and σ,

$$\mathbb{E}_t^{\sigma,\tau}[f] \leq \max\{val(\mathbf{G}')(t) + \epsilon', val(\mathbf{G})(s) - \frac{d}{2}\}.$$

Taking $t = s$ and the supremum over all σ, since the inequality above holds for any $\epsilon' > 0$, we get $val(\mathbf{G}')(s) \geq val(\mathbf{G})(s)$.

- $\forall t \in S,\ t \neq s,$ and $val(\mathbf{G}')(t) \geq val(\mathbf{G})(t)$

  Let Sw_σ be the event $(\exists n \in \mathbb{N}, S_n = s \wedge \sigma(S_0 \dots S_n)(a) > 0)$, that is the event that according to σ the value-changing action a is about to be played at some date n. For every strategy σ define σ_s as the strategy in $\mathbf{G}'$ that plays like σ as long as the latter does not choose the action a (with nonzero probability), and when it does, σ_s switches to the strategy σ' that is ϵ'-optimal in $\mathbf{G}'$. Let τ be the strategy for Player 2 which plays according to the strategy τ' which is ϵ'-optimal in $\mathbf{G}'$ as long as Player 1 does not choose the action a, otherwise it switches to strategy τ'' that is ϵ'-optimal in $\mathbf{G}$.

  Since pairs of strategies σ, τ coincide to σ_s, τ' up to the date n in the event Sw_σ, we write $c = \mathbb{P}_t^{\sigma,\tau}(Sw_\sigma) = \mathbb{P}_t^{\sigma_s,\tau'}(SW_\sigma)$. From the definition of τ, σ_s and the fact that $val(\mathbf{G})(s) = val(\mathbf{G}')(s)$, shown above we have:

  $$\mathbb{E}_t^{\sigma,\tau}[f \mid Sw_\sigma] \leq val(\mathbf{G})(s) + \epsilon' = val(\mathbf{G}')(s) + \epsilon', \text{ and}$$
  $$val(\mathbf{G}')(s) - \epsilon' \leq \mathbb{E}_t^{\sigma_s,\tau}[f \mid Sw_\sigma].$$

Combining the two inequalities above we get

$$\mathbb{E}_t^{\sigma,\tau}[f \mid Sw_\sigma] \leq \mathbb{E}_t^{\sigma_s,\tau}[f \mid Sw_\sigma] + 2\epsilon'. \tag{16}$$

Keeping this in mind we proceed:

$$\begin{aligned}
\mathbb{E}_t^{\sigma,\tau}[f] &= c\mathbb{E}_t^{\sigma,\tau}[f \mid Sw_\sigma] + (1-c)\mathbb{E}_t^{\sigma,\tau}[f \mid \neg Sw_\sigma] \\
&\leq c(\mathbb{E}_t^{\sigma_s,\tau}[f \mid Sw_\sigma] + 2\epsilon') + (1-c)\mathbb{E}_t^{\sigma_s,\tau}[f \mid \neg Sw_\sigma] \\
&= \mathbb{E}_t^{\sigma_s,\tau}[f] + 2c\epsilon' \\
&= \mathbb{E}_t^{\sigma_s,\tau'}[f] + 2c\epsilon' \leq val(\mathbf{G}')(t) + \epsilon'(2c+1)
\end{aligned}$$

where the first equality is a basic property of expectations, the first inequality is from (16) and because on the paths of the event $\neg Sw_\sigma$ the strategies σ and σ_s coincide, the following equality is a basic property of expectations, while the second one and the last inequality are by definition of the strategy τ. We have $\mathbb{E}_t^{\sigma_s,\tau}[f] = \mathbb{E}_t^{\sigma_s,\tau'}[f]$ because by definition of the switch strategy the action a is never played hence τ never switches to the strategy τ''.

Since this holds for any $\epsilon' > 0$, taking the supremum over strategies σ we get $val(\mathbf{G}')(t) \geq val(\mathbf{G})(t)$ as desired.

Having proved that for all states, the values in both $\mathbf{G}$ and $\mathbf{G}'$ coincide, we have shown that there are ϵ'-optimal strategies for Player 1 that never play the value-changing action a. □

Corollary 4.12. *For all $\epsilon' > 0$ both players have ϵ'-optimal strategies that are locally-optimal.*

The proof is by induction on the number of actions which are not value-preserving, using Lemma 4.11. The argument for Player 2 is completely symmetric.

We finish this section with a corollary to Proposition 2.8. When Player 1 plays with a locally-optimal strategy, we are interested in what happens when the process $val(S_0), val(S_1), \ldots$ is stopped at random, in particular stopped at the *first σ-weakness after some date n*. The answer is provided by Doob's optional stopping theorem:

Corollary 4.13. *Let T be a stopping time with respect to $(S_n)_{n\in\mathbb{N}}$. Then for all locally-optimal strategies σ, all strategies τ and states $s \in \mathbf{S}$, $(val(S_n))_{n\in\mathbb{N}}$ converges almost-surely and*

$$\mathbb{E}_s^{\sigma,\tau}[\lim_n val(S_{\min(n,T)})] \geq val(s).$$

In case we assume τ to be locally-optimal instead of σ the converse inequality holds.

Proof. Immediate from Proposition 2.8, since in a game where both players play with a locally-optimal strategy $(val(S_n))_{n\in\mathbb{N}}$ forms a martingale (Theorem 4.7), and if only Player 1 plays with a locally-optimal strategy then $(val(S_n))_{n\in\mathbb{N}}$ is a submartingale. From the hypothesis that the payoff function f is bounded, we have that the values are also bounded.

□

4.4 Only finitely many weaknesses occur when playing $\hat{\sigma}$

The goal of this section is to demonstrate that when playing with the reset strategy based on some ϵ-optimal and locally-optimal strategy σ for Player 1 there are almost-surely only finitely many σ-weaknesses.

Definition 4.14. *We define the event*

$$\{\textit{there is no } \sigma\textit{-weakness after date } n\}$$

as the event $\{\forall m > n, \delta_\sigma(S_0 \cdots S_m) \leq n\}$. *And the event*

$$\{\textit{there are two } \sigma\textit{-weaknesses after date } n\}$$

as the event $\{\exists m, m', m > m' > n, \delta_\sigma(S_0 \cdots S_m) = m \wedge \delta_\sigma(S_0 \cdots S_{m'}) = m'\}$.

Lemma 4.15. *Let σ be a locally-optimal and ϵ-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it. For all strategies τ and states $s \in \mathbf{S}$ there exists $n \in \mathbb{N}$ such that*

$$\mathbb{P}_s^{\hat{\sigma},\tau}(\textit{there are two } \sigma\textit{-weaknesses after date } n) < 1.$$

For Lemma 4.15, we proceed in a couple of steps. First, we assume that Player 2 chooses a locally-optimal strategy and prove that the chance of having a weakness is bounded away from 1. This is shown by proving the existence of another strategy for Player 2 that takes advantage of the weakness and provides a contradiction of the hypothesis that σ is ϵ-optimal. This is Lemma 4.16.

In the next step, we fix a locally-optimal and ϵ-optimal strategy for Player 1 and build the reset strategy based on it. Then we prove the same statement as in Lemma 4.15 but we restrict Player 2 to strategies which are sure to be locally-optimal after some date n. This is Lemma 4.17. After these two facts we can proceed with the proof of Lemma 4.15.

Lemma 4.16. *Let σ be an ϵ-optimal strategy, then there exists $\mu > 0$ such that for all strategy τ and $s \in S$, if τ is locally-optimal,*

$$\mathbb{P}_s^{\sigma,\tau}(\exists n, S_0 \cdots S_n \textit{ is a } \sigma\textit{-weakness}) \leq 1 - \mu.$$

Proof. We define $F = \min\{n \in \mathbb{N} \mid S_0 \cdots S_n \text{ is a } \sigma\text{-weakness}\}$ with the convention $\min \emptyset = \infty$, and let σ be an ϵ-optimal strategy for Player 1, then for a given $n \in \mathbb{N}$, $m > n$ and prefix $s_0 \dots s_m \in \mathbf{S}(\mathbf{AS})^*$, define $\text{weak}(n, s_0 \dots s_m) := (\delta_\sigma(s_0 \dots s_m) = m) \wedge (\delta_\sigma(s_0 \dots s_{m-1}) \leq n)$, the boolean function characterizing the prefixes up to the first weakness after date n.

The event $F < \infty$ means that a "weakness occurs", and is equivalent to the event given in the statement of the lemma. Let τ be a strategy for Player 2 and $s \in \mathbf{S}$. Let M and m be upper and lower bound respectively of the payoff function f, and let τ' be the strategy that plays identically to τ as long as weakness does not occur, and when a weakness occurs it switches to a $\frac{\epsilon}{2}$-optimal response τ''. Since strategies τ and τ' coincide up to the first weakness let $c = \mathbb{P}_s^{\sigma,\tau'}(F = \infty) = \mathbb{P}_s^{\sigma,\tau}(F = \infty)$. From the ϵ-optimality of σ, and a basic property of conditional expectations:

$$\begin{aligned} val(s) - \epsilon &\leq \mathbb{E}_s^{\sigma,\tau'}[f] \\ &= (1-c) \cdot \mathbb{E}_s^{\sigma,\tau'}[f \mid F < \infty] + c \cdot \mathbb{E}_s^{\sigma,\tau'}[f \mid F = \infty] \qquad (17) \\ &\leq (1-c) \cdot \mathbb{E}_s^{\sigma,\tau'}[f \mid F < \infty] + cM. \end{aligned}$$

As a subsequence of the strategy τ' namely that it resets to the strategy τ'' if a weakness occurs, by shifting up to the first weakness we get:

$$\begin{aligned} \mathbb{E}_s^{\sigma,\tau'}[f \mid F < \infty] &= \sum_{\substack{s_0 \dots s_n \in \mathbf{S}(\mathbf{AS})^* \\ \text{weak}(0, s_0 \dots s_n)}} \mathbb{P}_s^{\sigma,\tau'}(s_0 \dots s_n | F < \infty) \mathbb{E}_{s_n}^{\sigma[s_0 \dots s_n],\tau''}[f] \\ &\leq \sum_{\substack{s_0 \dots s_n \in \mathbf{S}(\mathbf{AS})^* \\ \text{weak}(0, s_0 \dots s_n)}} \mathbb{P}_s^{\sigma,\tau'}(s_0 \dots s_n | F < \infty)(val(s_n) - 2\epsilon + \frac{\epsilon}{2}) \\ &= \mathbb{E}_s^{\sigma,\tau'}[val(S_F) \mid F < \infty] - \frac{3}{2}\epsilon, \end{aligned}$$

the inequality is a subsequence of the strategy τ' that takes advantage of the weakness, and by the definition of a weakness. Replacing this inequality in (17):

$$\begin{aligned} val(s) - \epsilon &\leq (1-c) \cdot \mathbb{E}_s^{\sigma,\tau'}[val(S_F) \mid F < \infty] - \frac{3}{2}\epsilon(1-c) + cM \\ &= \mathbb{E}_s^{\sigma,\tau'}[val(S_F)] + c(M - \mathbb{E}_s^{\sigma,\tau'}[val(S_F) \mid F = \infty]) - \frac{3}{2}\epsilon(1-c) \\ &\leq val(s) + c(M - \mathbb{E}_s^{\sigma,\tau'}[val(S_F) \mid F = \infty]) - \frac{3}{2}\epsilon(1-c) \\ &\leq val(s) + c(M - m) - \frac{3}{2}\epsilon(1-c) \end{aligned}$$

where in the equality we have decomposed $(1-c) \cdot \mathbb{E}_s^{\sigma,\tau'}[val(S_F) \mid F < \infty]$ to $\mathbb{E}_s^{\sigma,\tau'}[val(S_F)] - c \cdot \mathbb{E}_s^{\sigma,\tau'}[val(S_F) \mid F = \infty]$, and the second inequality is a consequence of the following: F is a stopping time, and $(val(S_n))_{n \in \mathbb{N}}$ is a martingale because Player 2 is playing only value-preserving actions, thus applying Corollary 4.13 we get $\mathbb{E}_s^{\sigma,\tau'}[val(S_F)] \leq \text{val}(s)$. Finally from above: $\mu = \frac{\epsilon}{2(M-m+3/2\epsilon)} \leq c$, a uniform bound, that does not depend on the choice of τ.

□

In the next step we approximate the strategy τ of player 2 by a sequence of strategies $(\tau_n)_n$ and give an upper bound on the probability that two σ-weaknesses occur in the play provided when player 1 plays the reset strategy

and player 2 plays strategy τ_n. The definition of τ_n relies on the notion of *value-preserving actions* (Definition, 4.5). For a state $s \in \mathbf{S}$ denote by $Vp(s)$ the set of value-preserving actions in s and for a finite set R denote $\mathcal{U}(R)$ the uniform distribution on R.

Lemma 4.17. *Let σ be a locally-optimal and ϵ-optimal strategy for player 1, and $\hat{\sigma}$ the reset strategy based on it. Let τ be any strategy for player 2. For every integer n, define τ_n to be the following strategy:*

$$\tau_n(s_0 \dots s_m) = \begin{cases} \tau(s_0 \dots s_m) & \text{if } m < n \\ & \text{or } \forall a \in A, (\tau(s_0 \dots s_m)a) > 0) \implies (a \in Vp(s_m)) \\ \mathcal{U}(Vp(s)) & \text{otherwise} \end{cases}$$

Then there exists $\mu > 0$ such that for all $s \in \mathbf{S}$ and $n \in \mathbb{N}$,

$$\mathbb{P}_s^{\hat{\sigma},\tau_n}(\text{ there are two } \sigma\text{-weakness after date } n\ ) \leq 1 - \mu.$$

Proof. For every $n \in \mathbb{N}$, define $F_n = \min\{m > n \mid \delta_\sigma(S_0 \dots S_m) = m\}$ and $\min \emptyset = \infty$, the date of the first weakness strictly after n, and $F_n^2 = F_{F_n}$ the date of the second weakness strictly after n. With the convention $F_\infty = \infty$.

We prove that there exists $\mu > 0$ such that for all $n \in \mathbb{N}$, strategy τ, and state $s \in \mathbf{S}$,

$$\mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2 < \infty \mid F_n < \infty) \leq 1 - \mu. \tag{18}$$

Let μ be given by Lemma 4.16 and weak defined like in the proof of Lemma 4.16, that is $\text{weak}(n, s_0 \dots s_m) := (\delta_\sigma(s_0 \dots s_m) = m) \wedge (\delta_\sigma(s_0 \dots s_{m-1}) \leq n)$, the boolean function characterizing the prefixes up to the first weakness after date n. Then (18) is a consequence of:

$$\begin{aligned}
\mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2 < \infty \mid F_n < \infty) &= \sum_{\substack{h=s_0\dots s_m \in \mathbf{S}(\mathbf{AS})^* \\ \text{weak}(n,h)}} \mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2 < \infty \mid h \wedge F_n < \infty) \cdot \mathbb{P}_s^{\hat{\sigma},\tau_n}(h \mid F_n < \infty) \\
&= \sum_{\substack{h=s_0\dots s_m \in \mathbf{S}(\mathbf{AS})^* \\ \text{weak}(n,h)}} \mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2 < \infty \mid h) \cdot \mathbb{P}_s^{\hat{\sigma},\tau_n}(h \mid F_n < \infty) \\
&= \sum_{\substack{h=s_0\dots s_m \in \mathbf{S}(\mathbf{AS})^* \\ \text{weak}(n,h)}} \mathbb{P}_{s_m}^{\hat{\sigma},\tau_n[h]}(F_0 < \infty) \cdot \mathbb{P}_s^{\hat{\sigma},\tau_n}(h \mid F_n < \infty) \\
&= \sum_{\substack{h=s_0\dots s_m \in \mathbf{S}(\mathbf{AS})^* \\ \text{weak}(n,h)}} \mathbb{P}_{s_m}^{\sigma,\tau_n[h]}(F_0 < \infty) \cdot \mathbb{P}_s^{\hat{\sigma},\tau_n}(h \mid F_n < \infty) \\
&\leq 1 - \mu
\end{aligned}$$

where the first and second equalities hold because

$$\{F_n < \infty\} = \bigcup_{\substack{h=s_0\dots s_m \in \mathbf{S}(\mathbf{AS})^* \\ \text{weak}(n,h)}} h(AS)^\omega,$$

the third equality because if $\text{weak}(n, h)$ then $\hat{\sigma}[h] = \hat{\sigma}$, the fourth equality is a consequence of Lemma 4.20 since σ and $\hat{\sigma}$ coincide up to the first σ-weakness

and the last inequality holds by definition of τ_n and because $|h| \geq n$, as a consequence $\tau_n[h]$ is locally-optimal and we can apply Lemma 4.16.

Then we have $\mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2 < \infty) = \mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2 < \infty \mid F_n < \infty)\mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n < \infty) \leq 1 - \mu$. Because $\mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2 < \infty \mid F_n = \infty) = 0$, by definition of F_n. □

After establishing the previous two lemmas we are now ready to proceed with the proof of Lemma 4.15

Proof of Lemma 4.15. Let Ω be the random variable taking values in $\mathbb{N} \cup \{\infty\}$ that maps to the date of the last value-changing action played, if it exists, otherwise let it be ∞. Let τ_n be defined as in Lemma 4.17. Because τ and τ_n coincide on all paths where the last value-changing action is played before n, that is on the event $\{\Omega < n\}$, then for any $n \in \mathbb{N}$ and event E:

$$\mathbb{P}_s^{\hat{\sigma},\tau}(E) = \mathbb{P}_s^{\hat{\sigma},\tau}(\Omega < n)\mathbb{P}_s^{\hat{\sigma},\tau_n}(E \mid \Omega < n) + \mathbb{P}_s^{\hat{\sigma},\tau}(\Omega \geq n)\mathbb{P}_s^{\hat{\sigma},\tau}(E \mid \Omega \geq n).$$

Since $\hat{\sigma}$ is locally-optimal we can apply Corollary 4.10, we have $\lim_n \mathbb{P}_s^{\hat{\sigma},\tau}(\Omega < n) = 1$, therefore

$$\mathbb{P}_s^{\hat{\sigma},\tau_n}(E) \xrightarrow[n\to\infty]{} \mathbb{P}_s^{\hat{\sigma},\tau}(E).$$

Let $\mu > 0$ be the uniform bound in accord with Lemma 4.17, and fix $n \in \mathbb{N}$ such that for all events E we have $|\mathbb{P}_s^{\hat{\sigma},\tau}(E) - \mathbb{P}_s^{\hat{\sigma},\tau_n}(E)| < \mu$. For some $n' > n$ take E to be the event $\{$there is a weakness after date $n'\}$, and apply Lemma 4.17 to conclude this proof. □

To conclude this section what is left to show is that when we are playing with the reset strategy, the number of weaknesses is almost surely finite.

Lemma 4.18. *Let σ be a locally-optimal and ϵ-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it. Then for all strategies τ and states $s \in \mathbf{S}$,*

$$\mathbb{P}_s^{\hat{\sigma},\tau}(\exists n,\ \textit{there is no } \sigma\textit{-weakness after date } n) = 1\ .$$

Proof. Let $L = \lim_n \delta_\sigma(S_0 S_1 \dots S_n) \in \mathbb{N} \cup \{\infty\}$, which is well-defined since $(\delta_\sigma(S_0 S_1 \dots S_n))_{n\in\mathbb{N}}$ is pointwise increasing. Fix $\epsilon' > 0$, and choose τ and s such that:

$$\sup_{\tau',s'} \mathbb{P}_{s'}^{\hat{\sigma},\tau'}(L = \infty) \leq \mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty) + \epsilon'. \tag{19}$$

Let $n \in \mathbb{N}$ be such that according to Lemma 4.15 $\mu = \mathbb{P}_s^{\hat{\sigma},\tau}(F_n < \infty) < 1$. Then we have the following:

$$\begin{aligned}
\mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty) &= \mathbb{E}_s^{\hat{\sigma},\tau}[\mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty \mid F_n, S_0, \dots, S_{F_n})] \\
&= \mathbb{E}_s^{\hat{\sigma},\tau}[\mathbb{1}_{F_n<\infty} \cdot \mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty \mid F_n, S_0, \dots, S_{F_n})] \\
&= \mathbb{E}_s^{\hat{\sigma},\tau}[\mathbb{1}_{F_n<\infty} \cdot \mathbb{P}_{S_{F_n}}^{\hat{\sigma},\tau[S_0\dots S_{F_n}]}(L = \infty)] \\
&\leq \mathbb{E}_s^{\hat{\sigma},\tau}[\mathbb{1}_{F_n<\infty} \cdot (\mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty) + \epsilon')] \\
&= \mu(\mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty) + \epsilon')
\end{aligned}$$

First equality is a basic property of conditional expectations, the second one is a consequence of $\mathbb{P}_s^{\hat{\sigma},\tau}(F_n < \infty \mid L = \infty) = 1$, the third one is from the definition of the reset strategy $\hat{\sigma}$, the inequality is because of (19). Hence, because $\mu < 1$ we have $\mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty) \leq \frac{\mu}{1-\mu}\epsilon'$. And finally for all $s'' \in \mathbf{S}$ and τ'' we have

$$\begin{aligned}\mathbb{P}_{s''}^{\hat{\sigma},\tau''}(L = \infty) &\leq \sup_{\tau',s'} \mathbb{P}_{s'}^{\hat{\sigma},\tau'}(L = \infty) \\ &\leq \mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty) + \epsilon' \\ &\leq \frac{\epsilon'}{1-\mu}.\end{aligned}$$

Since this holds for any $\epsilon' > 0$, we get $\mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty) = 0$. □

4.5 The reset strategy is 2ϵ-subgame-perfect

The first goal is to prove the ϵ-optimality of the reset strategy, which is done first for a variant of the reset strategy that reset only up to some date.

Let σ be some ϵ-optimal strategy, we define $\hat{\sigma}_n$ as the strategy that resets only up to time $n \in \mathbb{N}$,

$$\hat{\sigma}_n(s_0 \dots s_m) = \sigma(s_{\delta_\sigma(s_0 \dots s_{m \wedge n})} \dots s_n)$$

where $m \wedge n = \min\{m, n\}$.

Lemma 4.19. *Let σ be a locally-optimal and ϵ-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it. Then for all strategies τ, states $s \in \mathbf{S}$ and all $n \in \mathbb{N}$,*

$$\mathbb{E}_s^{\hat{\sigma}_n,\tau}[f] \geq val(s) - \epsilon.$$

Proof. The proof is by induction on n, and decomposing the expected payoff on the event of a weakness at date $n+1$.

The base case is true by definition since $\hat{\sigma}_0 = \sigma$, and the induction hypothesis is that for all τ and $s \in \mathbf{S}$, $\mathbb{E}_s^{\hat{\sigma}_n,\tau}[f] \geq val(s) - \epsilon$. We have to show that the same holds when Player 1 plays with the strategy $\hat{\sigma}_{n+1}$. Fix a strategy τ and a state s. Let τ' be the strategy that plays like τ except if there is a weakness at date $n+1$ (in case of the event $\delta_\sigma(S_0 \dots S_{n+1}) = n+1$), then it resets to the $\frac{\epsilon}{2}$-optimal response τ''. Let $L_n = \delta_\sigma(S_0 \dots S_n)$. We decompose the expected values on the only event that matters, namely the event of a weakness at the date $n+1$. Let R be the set of all prefixes of length $n+1$ where a weakness occurs, that is $R = \{s_0 \dots s_{n+1} \in \mathbf{S}(\mathbf{AS})^* \mid \delta_{\sigma,\epsilon}(s_0 \dots s_{n+1}) = n+1\}$.

Then:

$$\{L_{n+1} = n+1\} = \bigcup_{h \in R} h(AS)^\omega$$

thus we have the two following inequalities. First,

$$\begin{aligned}\mathbb{E}_s^{\hat{\sigma}_{n+1},\tau}[f] &= \mathbb{E}_s^{\hat{\sigma}_{n+1},\tau}[\mathbb{1}_{L_{n+1}=n+1}\cdot f] + \mathbb{E}_s^{\hat{\sigma}_{n+1},\tau}[\mathbb{1}_{L_{n+1}\neq n+1}\cdot f] \\ &= \sum_{s_0\dots s_{n+1}\in R} \mathbb{P}_s^{\hat{\sigma}_{n+1},\tau}(s_0\dots s_{n+1})\mathbb{E}_{s_{n+1}}^{\sigma,\tau[s_0\dots s_{n+1}]}[f] + \mathbb{E}_s^{\hat{\sigma}_{n+1},\tau}[\mathbb{1}_{L_{n+1}\neq n+1}\cdot f] \\ &\geq \sum_{s_0\dots s_{n+1}\in R} \mathbb{P}_s^{\hat{\sigma}_{n+1},\tau}(s_0\dots s_{n+1})(val(s_{n+1})-\epsilon) + \mathbb{E}_s^{\hat{\sigma}_{n+1},\tau}[\mathbb{1}_{L_{n+1}\neq n+1}\cdot f],\end{aligned}$$

where the second equality holds because for every $s_0\cdots s_{n+1}\in R$ and $\hat{\sigma}_{n+1}[s_0\cdots s_{n+1}] = \sigma$, and the inequality by ϵ-optimality of σ. Second,

$$\begin{aligned}\mathbb{E}_s^{\hat{\sigma}_n,\tau'}[f] &= \mathbb{E}_s^{\hat{\sigma}_n,\tau'}[\mathbb{1}_{L_{n+1}=n+1}\cdot f] + \mathbb{E}_s^{\hat{\sigma}_n,\tau'}[\mathbb{1}_{L_{n+1}\neq n+1}\cdot f] \\ &= \sum_{s_0\dots s_{n+1}\in R} \mathbb{P}_s^{\hat{\sigma}_n,\tau'}(s_0\dots s_{n+1})\mathbb{E}_{s_{n+1}}^{\hat{\sigma}_n[s_0\dots s_{n+1}],\tau''}[f] + \mathbb{E}_s^{\hat{\sigma}_n,\tau'}[\mathbb{1}_{L_{n+1}\neq n+1}\cdot f] \\ &\leq \sum_{s_0\dots s_{n+1}\in R} \mathbb{P}_s^{\hat{\sigma}_n,\tau'}(s_0\dots s_{n+1})(val(s_{n+1})-2\epsilon+\frac{\epsilon}{2}) + \mathbb{E}_s^{\hat{\sigma}_n,\tau'}[\mathbb{1}_{L_{n+1}\neq n+1}\cdot f],\end{aligned}$$

where the second equality is by construction of τ' and the inequality because τ'' is chosen as the $\frac{\epsilon}{2}$-optimal response to $\sigma[s_0\dots s_{n+1}]$, which is not 2ϵ-optimal by definition of R.

We can combine the two inequalities above because both strategies $\hat{\sigma}_n$ and $\hat{\sigma}_{n+1}$ on one hand and both strategies τ and τ' on the other hand coincide upon all paths of length less than $n+1$ and also upon all paths where no weakness occurs at date $n+1$, therefore the second terms on the right hand side of the two inequalities above coincide and the same holds for the first terms (without the ϵ terms) hence

$$\mathbb{E}_s^{\hat{\sigma}_n,\tau'}[f] \leq \mathbb{E}_s^{\hat{\sigma}_{n+1},\tau}[f].$$

According to the induction hypothesis $\hat{\sigma}_n$ is ϵ-optimal thus

$$\mathrm{val}(s) - \epsilon \leq \mathbb{E}_s^{\hat{\sigma}_{n+1},\tau}[f].$$

Since this holds for every τ, $\hat{\sigma}_{n+1}$ is ϵ-optimal, which completes the proof of the inductive step. □

Having shown that the strategies $\hat{\sigma}_n$ are ϵ-optimal we can proceed into proving that the reset strategy $\hat{\sigma}$ itself is ϵ-optimal. First we need to give a link between the strategies $\hat{\sigma}$ and $\hat{\sigma}_n$, in the form of the following lemma.

Lemma 4.20. *Let E be an event, and σ_1,σ_2 two strategies for Player 1 such that for all prefixes p of an infinite play in E, $\sigma_1(p) = \sigma_2(p)$. Then for all payoff functions f, strategies for Player 2 τ, and $s\in\mathbf{S}$,*

$$\mathbb{E}_s^{\sigma_1,\tau}[f\cdot\mathbb{1}_E] = \mathbb{E}_s^{\sigma_2,\tau}[f\cdot\mathbb{1}_E].$$

Proof. This is obvious when f is the indicator of a cylinder, and the class of functions f with this property is closed by linear combinations and simple limits. □

Lemma 4.21. *Let σ be a locally-optimal and ϵ-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it, then $\hat{\sigma}$ is ϵ-optimal.*

Proof. Let m and M be the lower and upper bounds of the payoff function f respectively. Let $L = \lim_n \delta_\sigma(S_0 S_1 \dots S_n) \in \mathbb{N} \cup \{\infty\}$, which is well-defined since $(\delta_\sigma(S_0 S_1 \dots S_n))_{n\in\mathbb{N}}$ is pointwise increasing.

Applying Lemma 4.19 gives us

$$\begin{aligned} val(s) - \epsilon &\leq \mathbb{E}_s^{\hat{\sigma}_n,\tau}[f \cdot \mathbb{1}_{L\leq n}] + \mathbb{E}_s^{\hat{\sigma}_n,\tau}[f \cdot \mathbb{1}_{L>n}] \\ &\leq \mathbb{E}_s^{\hat{\sigma}_n,\tau}[f \cdot \mathbb{1}_{L\leq n}] + M\mathbb{P}_s^{\hat{\sigma}_n,\tau}(L > n). \end{aligned}$$

From here, since $\hat{\sigma}$ and $\hat{\sigma}_n$ coincide upon all plays in $\{L \leq n\}$, using Lemma 4.20 we get

$$\begin{aligned} \mathbb{E}_s^{\hat{\sigma},\tau}[f] - \mathbb{E}_s^{\hat{\sigma},\tau}[f \cdot \mathbb{1}_{L>n}] &= \mathbb{E}_s^{\hat{\sigma},\tau}[f \cdot \mathbb{1}_{L\leq n}] \\ &= \mathbb{E}_s^{\hat{\sigma}_n,\tau}[f \cdot \mathbb{1}_{L\leq n}] \\ &\geq val(s) - \epsilon - M\mathbb{P}_s^{\hat{\sigma}_n,\tau}(L > n). \end{aligned}$$

Since for any measurable function f, Lemma 4.20 holds, applying it to the constant function that maps to 1, gives us $\mathbb{P}_s^{\hat{\sigma},\tau}(L \leq n) = \mathbb{P}_s^{\hat{\sigma}_n,\tau}(L \leq n)$, hence

$$\mathbb{E}_s^{\hat{\sigma},\tau}[f] \geq val(s) - \epsilon - (M - m)\mathbb{P}_s^{\hat{\sigma},\tau}(L > n).$$

And the inequality above holds for any n, according to Lemma 4.18, $\lim_n (M - m)\mathbb{P}_s^{\hat{\sigma},\tau}(L > n) = 0$, hence

$$\mathbb{E}_s^{\hat{\sigma},\tau}[f] \geq val(s) - \epsilon.$$

□

Having shown that the reset strategy is ϵ-optimal, using similar ideas we prove that the reset strategy is also 2ϵ-subgame-perfect.

Theorem 4.22. *Let σ be a locally-optimal and ϵ-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it, then $\hat{\sigma}$ is 2ϵ-subgame-perfect.*

Proof. Let $s_0 \dots s_n \in \mathbf{S}(\mathbf{AS})^*$ be some finite prefix of an infinite play, then we have to show that

$$\inf_\tau \mathbb{E}_{s_n}^{\hat{\sigma}[s_0\dots s_n],\tau}[f] \geq val(s_n) - 2\epsilon.$$

In the case when $\delta_\sigma(s_0 \dots s_n) = n$ we have $\inf_\tau \mathbb{E}_{s_n}^{\hat{\sigma}[s_0\dots s_n],\tau}[f] = \inf_\tau \mathbb{E}_{s_n}^{\hat{\sigma},\tau}[f] \geq val(s_n) - \epsilon$ from the definition of $\hat{\sigma}$ and Lemma 4.21. Assume that $\delta_\sigma(s_0 \dots s_n) < n$, which gives us

$$\delta_\sigma(s_0 \dots s_n) = \delta_\sigma(s_0 \dots s_{n-1}),$$

and,

$$\inf_\tau \mathbb{E}_{s_n}^{\sigma[s_{\delta(s_0\dots s_{n-1})}\dots s_n],\tau}[f] \geq val(s_n) - 2\epsilon. \tag{20}$$

by definition of the function δ_σ, when $\delta_\sigma(s_0 \dots s_n) \neq n$. Assume that there exists a strategy τ such that $\mathbb{E}_{s_n}^{\hat{\sigma}[s_0 \dots s_n],\tau}[f] < val(s_n) - 2\epsilon$, then we will show that from τ we can build another strategy τ' such that

$$\mathbb{E}_{s_n}^{\sigma[s_{\delta(s_0 \dots s_{n-1}) \dots s_n}],\tau'}[f] < val(s_n) - 2\epsilon,$$

a contradiction of (20). Let τ' be the strategy that plays like τ as long as no weakness occurs, and in case it does, it switches to the ϵ-response strategy τ''. Let $L = \lim_n \delta_\sigma(S_0 S_1 \dots S_n) \in \mathbb{N} \cup \{\infty\}$, which is well-defined since $(\delta_\sigma(S_0 S_1 \dots S_n))_{n \in \mathbb{N}}$ is pointwise increasing. Define $F = \min\{n \in \mathbb{N} \mid S_0 \cdots S_n \text{ is a } \sigma\text{-weakness}\}$ with the convention $\min \emptyset = \infty$. Let $\hat{\sigma}_1 = \hat{\sigma}[s_0 \dots s_n]$ and $\sigma_2 = \sigma[s_{\delta(s_0 \dots s_{n-1}) \dots s_n}]$, then we have

$$\begin{aligned} val(s_n) - 2\epsilon &> \mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{L=0}] + \mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{F<\infty}] \\ &= \mathbb{E}_{s_n}^{\sigma_2,\tau'}[f \cdot \mathbb{1}_{L=0}] + \mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{F<\infty}] \\ &= \mathbb{E}_{s_n}^{\sigma_2,\tau'}[f] - \mathbb{E}_{s_n}^{\sigma_2,\tau'}[f \cdot \mathbb{1}_{F<\infty}] + \mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{F<\infty}], \end{aligned} \tag{21}$$

where the first inequality is by assumption on the strategy τ (and also because $\{L = 0\} = \{F = \infty\}$), the first equality is because both pairs of strategies $\hat{\sigma}_1, \sigma_2$ and τ, τ' coincide up to the first weakness. Let weak be the boolean function characterizing the prefixes up to first weakness that is $\text{weak}_\sigma(s_0 \dots s_n) := (\delta_\sigma(s_0 \dots s_n) = n) \wedge (\delta_\sigma(s_0 \dots s_{n-1}) \leq n)$. Then for the last two terms in (21) we have:

$$\begin{aligned} \mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{F<\infty}] &= \sum_{\substack{t_0 \dots t_m \in S(AS)^* \\ \text{weak}_\sigma(t_0 \dots t_m)}} \mathbb{P}_{s_n}^{\hat{\sigma}_1,\tau}(t_0 \dots t_m) \mathbb{E}_{t_m}^{\hat{\sigma},\tau[t_0 \dots t_m]}[f] \\ &\geq \sum_{\substack{t_0 \dots t_m \in S(AS)^* \\ \text{weak}_\sigma(t_0 \dots t_m)}} \mathbb{P}_{s_n}^{\hat{\sigma}_1,\tau}(t_0 \dots t_m)(val(t_m) - \epsilon), \end{aligned}$$

this is by the definition of $\hat{\sigma}$ and Lemma 4.21, while for the other term

$$\begin{aligned} \mathbb{E}_{s_n}^{\sigma_2,\tau'}[f \cdot \mathbb{1}_{F<\infty}] &= \sum_{\substack{t_0 \dots t_m \in S(AS)^* \\ \text{weak}_\sigma(t_0 \dots t_m)}} \mathbb{P}_{s_n}^{\sigma_2,\tau'}(t_0 \dots t_m) \mathbb{E}_{t_m}^{\sigma_2[t_0 \dots t_m],\tau'}[f] \\ &\leq \sum_{\substack{t_0 \dots t_m \in S(AS^*) \\ \text{weak}_\sigma(t_0 \dots t_m)}} \mathbb{P}_{s_n}^{\sigma_2,\tau'}(t_0 \dots t_m)(val(t_m) - 2\epsilon + \epsilon) \end{aligned}$$

in the probabilities of the cylinders $t_0 \dots t_m$ we can freely interchange the pairs of strategies $\hat{\sigma}_1, \sigma_2$ and τ, τ', since they coincide up to the first weakness. Therefore we get that

$$\mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{F<\infty}] \geq \mathbb{E}_{s_n}^{\sigma_2,\tau'}[f \cdot \mathbb{1}_{F<\infty}],$$

which is the promised contradiction of (20) when replacing it into (21) $\square$

5 Games with Shift-Invariant and Submixing Payoff Functions are Half-Positional

In this section, we introduce the class of *shift-invariant* and *submixing* payoff functions and we prove that in every game equipped with such a payoff function, Player 1 has a deterministic and stationary strategy which is optimal.

The definition of a submixing payoff function relies on the notion of the shuffle of two words. A *factorization* of a sequence $u \in \mathbf{C}^\omega$ is an infinite sequence $(u_n)_{n\in\mathbb{N}} \in (\mathbf{C}^*)^{\mathbb{N}}$ of finite sequences whose u is the concatenation i.e. such that $u = u_0u_1u_2\cdots$. A sequence $w \in \mathbf{C}^\omega$ is said to be a *shuffle* of $u \in \mathbf{C}^\omega$ and $v \in \mathbf{C}^\omega$ if there exists two factorizations $u = u_0u_1u_2\cdots$ and $v = v_0v_1v_2\cdots$ of u and v such that $w = u_0v_0u_1v_1u_2v_2\cdots$.

Definition 5.1 (Submixing payoff functions)**.** *A payoff function $f : \mathbf{C}^\omega \to \mathbb{R}$ is* submixing *if for every infinite words $u, v, w \in \mathbf{C}^\omega$ such that w is a shuffle of $u \in \mathbf{C}^\omega$ and $v \in \mathbf{C}^\omega$,*

$$f(w) \leq \max\{f(u), f(v)\} \ . \tag{22}$$

In other words, the submixing condition states that the payoff of the shuffle of two plays cannot be strictly greater than both payoffs of these plays.

We can now state our main result.

Theorem 5.2. *Let f be a payoff function and $\mathbf{G}$ a game equipped with f. Suppose that f is shift-invariant and submixing. Then the game $\mathbf{G}$ has a value and Player 1 has an optimal strategy which is both deterministic and stationary.*

The shift-invariant and submixing properties are sufficient but not necessary to ensure the existence of a pure and stationary optimal strategy for Player 1, there are counter-examples in Section 6. Necessary and sufficient conditions for positionality are known for deterministic games [GZ05].

However the shift-invariant and submixing conditions are general enough to recover several known results of existence of deterministic stationary optimal strategies, and to provide several new examples of games with deterministic stationary optimal strategies, as is shown in the two next sections.

5.1 Proof of Half-Positionality

We prove Theorem 5.2. Let $f : \mathbf{C}^\omega \to \mathbb{R}$ be a shift-invariant and submixing payoff function and $\mathbf{G}$ a game equipped with f. For the sake of simplicity we suppose without loss of generality that the alphabet of f is $\mathbf{C} = \mathbf{S} \times \mathbf{A}$.

We prove Theorem 5.2 by induction on $N(\mathbf{G}) = \sum_{s\in\mathbf{S}_1}(|\mathbf{A}(s)| - 1)$. If $N(\mathbf{G}) = 0$ then in every state controlled by Player 1 there is only one action available, thus Player 1 has a unique strategy which is optimal, deterministic and stationary.

Let $\mathbf{G}$ be a game $N(\mathbf{G}) > 0$ and suppose Theorem 5.2 has been proved for every game $\mathbf{G}'$ such that $N(\mathbf{G}') < N(\mathbf{G})$. Since $N(\mathbf{G}) > 0$ there exists a state $s \in \mathbf{S}_1$ such that $\mathbf{A}(s)$ has at least two elements. Let $(\mathbf{A}_0(s), \mathbf{A}_1(s))$

be a partition of $\mathbf{A}(s)$ in two non-empty sets. Let $\mathbf{G}_0$ and $\mathbf{G}_1$ be the two games obtained from $\mathbf{G}$ by restricting actions in state s to $\mathbf{A}_0(s)$ and $\mathbf{A}_1(s)$ respectively. According to the induction hypothesis, both $\mathbf{G}_0$ and $\mathbf{G}_1$ have values, let $\mathrm{val}_0(s)$ and $\mathrm{val}_1(s)$ denote the values of state s in $\mathbf{G}_0$ and $\mathbf{G}_1$.

To prove the existence of a deterministic stationary optimal strategy in $\mathbf{G}$ it is enough to prove:

$$\mathrm{minmax}(\mathbf{G})(s) \leq \max\{\mathrm{val}_0(s), \mathrm{val}_1(s)\} \ , \tag{23}$$

Since every strategy of Player 1 in $\mathbf{G}_0$ and $\mathbf{G}_1$ is a strategy in $\mathbf{G}$ as well, then $\mathrm{val}_0(s) \leq \mathrm{val}(s)$ and $\mathrm{val}_1(s) \leq \mathrm{val}(s)$. Moreover according to the induction hypothesis there exist deterministic stationary optimal strategies σ_0 and σ_1 in $\mathbf{G}_0$ and $\mathbf{G}_1$. Suppose that (23) holds, and without loss of generality suppose $\mathrm{minmax}(\mathbf{G})(s) \leq \mathrm{val}_0(s)$. Since the deterministic stationary σ_0 is optimal in $\mathbf{G}_0$, it guarantees for every τ and $s \in \mathbf{S}$, $\mathbb{E}_s^{\sigma_0,\tau}[f] \geq \mathrm{val}_0(s) \geq \mathrm{minmax}(\mathbf{G})(s)$, thus σ_0 is optimal not only in the game $\mathbf{G}_0$ but in the game $\mathbf{G}$ as well. Thus (23) is enough to prove the inductive step.

5.2 The projection mapping

To prove (23), we make use of two mappings

$$\pi_0 : s(\mathbf{AS})^\infty \to s(\mathbf{AS})^\infty \tag{24}$$

$$\pi_1 : s(\mathbf{AS})^\infty \to s(\mathbf{AS})^\infty. \tag{25}$$

First π_0 and π_1 are defined on finite words. The mapping π_0 associates with each finite play $h \in (\mathbf{SA})^*$ in $\mathbf{G}$ with source s a finite play $\pi_0(h)$ in $\mathbf{G}_0$.

Intuitively, play $\pi_0(h)$ is obtained by erasing from h some of its subwords. Remember that $(\mathbf{A}_0(s), \mathbf{A}_1(s))$ is a partition of $\mathbf{A}(s)$ hence every occurrence of state s in the play h is followed by an action a which is either in $\mathbf{A}_0(s)$ or in $\mathbf{A}_1(s)$. To obtain $\pi_0(h)$ one erases from h two types of subwords:

1. all simple cycles on s starting with an action in $\mathbf{A}_1(s)$ are deleted from h,
2. in case the last occurrence of s in h is followed by an action in $\mathbf{A}_1(s)$ then the corresponding suffix is deleted from h.

Formally, π_0 and π_1 are defined as follows. Let $h = s_0 a_0 s_1 a_1 \cdots s_n \in s(\mathbf{AS})^*$ and $i_0 < i_1 < \ldots < i_k = \{0 \leq i \leq n \mid s_i = s\}$ the increasing sequence of dates where the play reaches s. For $0 \leq l < k$ let h_l the l-th factor of h defined by $h_l = s_{i_l} a_{i_l} \cdots a_{i_{l+1}-1}$ and $h_k = s_{i_k} a_{i_k} \cdots a_{n-1} s_n$. Then for $j \in \{0, 1\}$,

$$\pi_j(h) = \prod_{\substack{0 \leq l \leq k \\ a_{i_l} \in A_j}} h_l \ ,$$

where $\prod$ denotes word concatenation.

We extend π_0 and π_1 to infinite words in a natural way: for an infinite play $h \in s(\mathbf{AS})^\omega$ then $\pi_0(h)$ is the limit of the sequence $(\pi_0(h_n))_{n\in\mathbb{N}}$, where h_n is the prefix of h of length $2n+1$. Remark that $\pi_0(h)$ may be a finite play, in case play h has an infinite suffix such that every occurrence of s is followed by an action of $\mathbf{A}_1(s)$.

We make use of the four following properties of π_0 and π_1. For every infinite play $h \in (\mathbf{SA})^\omega$,

(A) if $\pi_0(h)$ is finite then h has a suffix which is an infinite play in $\mathbf{G}_1$ starting in s,

(B) if $\pi_1(h)$ is finite then h has a suffix which is an infinite play in $\mathbf{G}_0$ starting in s,

(C) if both $\pi_0(h)$ and $\pi_1(h)$ are infinite then both $\pi_0(h)$ and $\pi_1(h)$ reach state s infinitely often,

(D) if both $\pi_0(h)$ and $\pi_1(h)$ are infinite, then h is a shuffle of $\pi_0(h)$ and $\pi_1(h)$.

We use the three following random variables:

$$\Pi = S_0A_1S_1\cdots \ , \tag{26}$$
$$\Pi_0 = \pi_0(S_0A_1S_1\cdots) \ , \tag{27}$$
$$\Pi_1 = \pi_1(S_0A_1S_1\cdots) \ . \tag{28}$$

5.3 The trigger strategy

We build a strategy $\tau^\sharp$ for Player 2 called the trigger strategy.

According to Theorem 4.1, there exists ϵ-subgame-perfect strategies $\tau_0^\sharp$ and $\tau_1^\sharp$ in the games $\mathbf{G}_0$ and $\mathbf{G}_1$ respectively. The strategy $\tau^\sharp$ is a combination of $\tau_0^\sharp$ and $\tau_1^\sharp$. Intuitively the strategy $\tau^\sharp$ switches between $\tau_0^\sharp$ and $\tau_1^\sharp$ depending on the action chosen at the last visit in s. Let h be a finite play in $s(AS)^*$ and $last(h) \in A$ the action played after the last visit of h to s and t the last state of h, then:

$$\tau^\sharp(h) = \begin{cases} \tau_0^\sharp(\pi_0(h)t) & \text{if } last(h) \in A_0 \\ \tau_1^\sharp(\pi_1(h)t) & \text{if } last(h) \in A_1. \end{cases}$$

We are going to prove that the trigger strategy $\tau^\sharp$ is ϵ-optimal for Player 2, thanks to three following key properties. For every strategy σ for Player 1 ,

$$\mathbb{E}_s^{\sigma,\tau^\sharp}[\ f \mid \Pi_0 \text{ is finite}] \leq \mathrm{val}_1(s) + \epsilon \ , \tag{29}$$
$$\mathbb{E}_s^{\sigma,\tau^\sharp}[\ f \mid \Pi_1 \text{ is finite}] \leq \mathrm{val}_0(s) + \epsilon \ , \tag{30}$$
$$\mathbb{E}_s^{\sigma,\tau^\sharp}[\ f \mid \Pi_0 \text{ and } \Pi_1 \text{ are both infinite}] \leq \max\{\mathrm{val}_0(s), \mathrm{val}_1(s)\} + \epsilon \ . \tag{31}$$

5.4 Proof of inequalities (29) and (30)

To prove inequality (29), we introduce the probability measure μ_1 on plays in $\mathbf{G}_1$ defined as:

$$\mu_1(E) = \mathbb{P}_s^{\sigma,\tau^\sharp}(\Pi_1 \in E \mid \Pi_0 \text{ is finite}) \ ,$$

and the strategy σ_1' for Player 1 in $\mathbf{G}_1$ defined for every play h controlled by Player 1 by:

$$\sigma_1'(h)(a) = \mathbb{P}_s^{\sigma,\tau^\sharp}(ha \preceq \Pi_1 \mid h \preceq \Pi_1 \text{ and } \Pi_0 \text{ is finite}) \ ,$$

where $\preceq$ denotes the prefix relation over words of $\mathbf{S}^\infty$:

$$\forall u \in \mathbf{C}^*, v \in \mathbf{C}^\infty, u \preceq v \iff \exists w \in \mathbf{C}^\infty, v = u \cdot w \ ,$$

and $\prec$ the strict prefix relation.

We abuse the notation and denote h and ha for the events $h(AS)^\omega$ and $ha(SA)^\omega$, so that

$$\sigma_1'(h)(a) = \mu_1(ha \mid h) \ .$$

The probability measure μ_1 has the following key properties. For every finite play h in the game $\mathbf{G}_1$ whose finite state is t,

$$\mu_1(ha \mid h) = \begin{cases} \sigma_1'(h)(a) & \text{if } t \in S_1, \\ \tau_1^\sharp(h)(a) & \text{if } t \in S_2 \ , \end{cases} \tag{32}$$

$$\mu_1(has \mid ha) = p(s|t,a). \tag{33}$$

As a consequence of the equalities (32) and (33), and according to the characterization given by (1) and (2) the probability measure μ_1 coincides with the probability measure $\mathbb{P}_s^{\sigma_1',\tau_1^\sharp}$. Since $\tau_1^\sharp$ is ϵ-optimal in the game $\mathbf{G}_1$, it implies (29). The proof of (30) is symmetrical.

5.5 Proof of inequality (31)

The proof of (31) requires several steps.

First, we prove that for every strategy σ in $\mathbf{G}$, there exists a strategy σ_0 in $\mathbf{G}_0$ such that for every measurable event $E \subseteq (\mathbf{SA})^\omega$ in $\mathbf{G}_0$,

$$\mathbb{P}_s^{\sigma_0,\tau_0^\sharp}(E) \geq \mathbb{P}_s^{\sigma,\tau^\sharp}(\Pi_0 \text{ is infinite and } \Pi_0 \in E) \ . \tag{34}$$

The strategy σ_0 in $\mathbf{G}_0$ is defined for every finite play controlled by Player 1 by:

$$\sigma_0(h)(a) = \mathbb{P}_s^{\sigma,\tau^\sharp}(ha \preceq \Pi_0 \mid h \prec \Pi_0) \ ,$$

if $\mathbb{P}_s^{\sigma,\tau^\sharp}(h \prec \Pi_0) > 0$ and otherwise $\sigma_0(h)$ is chosen arbitrarily. We prove that (34) holds. Let $\mathcal{E}$ be the set of measurable events $E \subseteq (\mathbf{SA})^\omega$ in $\mathbf{G}_0$

such that (34) is satisfied. First, $\mathcal{E}$ contains all cylinders $h_0(\mathbf{SA})^\omega$ of $\mathbf{G}_0$ with $h_0 \in (\mathbf{SA})^*$ because:

$$\begin{aligned}\mathbb{P}_s^{\sigma_0,\tau_0^\sharp}\left(h_0(\mathbf{SA})^\omega\right) &\geq \mathbb{P}_s^{\sigma,\tau^\sharp}\left(h_0 \preceq \Pi_0\right)\\ &\geq \mathbb{P}_s^{\sigma,\tau^\sharp}\left(\Pi_0 \text{ is infinite and } \Pi_0 \in h_0(\mathbf{SA})^\omega\right)\end{aligned}$$

where the first inequality can be proved by induction on the size of h_0, using the definition of σ_0 and where the second inequality is by definition of $\preceq$. Clearly $\mathcal{E}$ is stable by finite disjoint unions hence $\mathcal{E}$ contains all finite disjoint unions of cylinders, which form a boolean algebra. Moreover $\mathcal{E}$ is clearly a monotone class, hence according to the Monotone Class Theorem, $\mathcal{E}$ contains the σ-field generated by cylinders, that is all measurable events E in $\mathbf{G}_0$. This completes the proof of (34).

Second step to obtain (31) is to prove that for every strategy σ_0 in $\mathbf{G}_0$:

$$\mathbb{P}_s^{\sigma_0,\tau_0^\sharp}\left(f \leq \mathrm{val}_0(s) + \epsilon \mid s \text{ is reached infinitely often}\right) = 1 \ . \tag{35}$$

According to Levy's law, $(\mathbb{E}_s^{\sigma_0,\tau_0^\sharp}[f \mid S_0, A_1, \ldots, S_n])_{n\in\mathbb{N}}$ converges in probability to $f(S_0A_1S_1\cdots)$. Since f is a shift-invariant payoff function, for every $n \in \mathbb{N}$,

$$\begin{aligned}&\mathbb{E}_s^{\sigma_0,\tau_0^\sharp}[f \mid S_0, A_1, \ldots, S_n]\\ =&\mathbb{E}_s^{\sigma_0,\tau_0^\sharp}[f(S_nA_{n+1}S_{n+1}\cdots) \mid S_0, A_1, \ldots, S_n]\\ =&\mathbb{E}_{S_n}^{\sigma_0[S_0A_1\cdots S_n],\tau_0^\sharp[S_0A_1\cdots S_n]}[f]\\ \leq&\,\mathrm{val}_0(S_n) + \epsilon \ ,\end{aligned}$$

because $\tau_0^\sharp$ is ϵ-subgame-perfect. As a consequence $\mathbb{P}_s^{\sigma,\tau^\sharp}\left(f \leq \liminf_n \mathrm{val}_0(S_n) + \epsilon\right) = 1$ hence (35).

Now we come to the end of the proof of (31). Let σ_0 be a strategy in $\mathbf{G}_0$ such that (34) holds for every measurable event E in $\mathbf{G}_0$. According to (35), $\mathbb{P}_s^{\sigma_0,\tau^\sharp}\left(f > \mathrm{val}_0(s) + \epsilon \text{ and } s \text{ is reached infinitely often}\right) = 0$. For $i \in \{0,1\}$ denote E_i and F_i the events:

$$E_i = \{\Pi_i \text{ is infinite and reaches } s \text{ infinitely often}\}, \tag{36}$$

$$F_i = E_i \wedge \{f(\Pi_i) \leq \mathrm{val}_i(s) + \epsilon\}. \tag{37}$$

Remark that F_i is well-defined since condition E_i implies that Π_i is infinite thus $f(\Pi_i)$ is well-defined in (37).

According to (35) and the definition of $\tau^\sharp$,

$$\mathbb{P}_s^{\sigma_0,\tau_0^\sharp}\left(f > \mathrm{val}_0(s) + \epsilon \wedge s \text{ is reached infinitely often}\right) = 0$$

and together with (34),

$$\mathbb{P}_s^{\sigma,\tau^\sharp}\left(f(\Pi_0) > \mathrm{val}_0(s) + \epsilon \text{ and } E_0\right) = 0.$$

Symmetrically, $\mathbb{P}_s^{\sigma,\tau^\sharp}(f(\Pi_1) > \mathrm{val}_1(s) + \epsilon \text{ and } E_1) = 0$, and this proves

$$\mathbb{P}_s^{\sigma,\tau^\sharp}(F_0 \text{ and } F_1 \mid E_0 \text{ and } E_1) = 1.$$

Together with (D) and because f is submixing this implies

$$\mathbb{P}_s^{\sigma,\tau^\sharp}(f \leq \max\{\mathrm{val}_0(s), \mathrm{val}_1(s)\} + \epsilon \mid E_0 \text{ and } E_1) = 1$$

and according to (C) this terminates the proof of (31).

Since equations (29), (30) and (31) hold for every strategy σ and every ϵ, $\mathrm{minmax}(s) \leq \max\{\mathrm{val}_0(s), \mathrm{val}_1(s)\}$. W.l.o.g. assume $\mathrm{minmax}(s) \leq \mathrm{val}_0(s)$. Then the stationary deterministic strategy σ_0 optimal in $\mathbf{G}_0$ is a strategy in $\mathbf{G}$ as well and σ_0 ensures an expected income of $\mathrm{val}_0(s)$ thus $\mathrm{minmax}(s) \leq \mathrm{val}_0(s) \leq \mathrm{maxmin}(s)$. As a consequence, the state s has value $\mathrm{val}_0(s)$ in the game $\mathbf{G}$ and σ_0 is optimal in $\mathbf{G}$. This completes the proof of Theorem 5.2.

6 Applications

6.1 Unification of classical results

The existence of deterministic stationary optimal strategies in Markov decision processes with parity [CY90], limsup, liminf [MS96], mean-payoff [LL69, NS03, Bie87, VTRF83] or discounted payoff functions [Sha53] is well-known. Theorem 5.2 provides a unified proof of these five results, as a corollary of the following proposition.

Proposition 6.1. *The payoff functions f_{lsup}, f_{linf}, f_{par} and f_{mean} are shift-invariant and submixing.*

The proof of this proposition is an elementary exercise, details are provided in [Gim07, Gim06].

Corollary 6.2. *In every two-player stochastic game equipped with the parity, limsup, liminf, mean or discounted payoff function, player 1 has a deterministic and stationary strategy which is optimal.*

Proof. Except for the discounted payoff function, this is a direct consequence of Proposition 6.1 and Theorem 5.2. The case of the discounted payoff function can be reduced to the case of the mean-payoff function, interpreting discount factors as stopping probabilities as was done in the seminal paper of Shapley [Sha53]. Details can be found in [Gim07, Gim06]. □

Corollary 6.2 unifies and simplifies existing proofs of [CY90] for the parity game and [MS96] for the limsup game.

The existence of deterministic and stationary optimal strategies in mean-payoff games has attracted much attention. The first proof was given by Gilette [Gil57] and based on a variant of Hardy and Littlewood theorem. Later on, Ligget and Lippman found the variant to be wrong and proposed an alternative proof based

on the existence of Blackwell optimal strategies plus a uniform boundedness result of Brown [LL69]. For one-player games, Bierth [Bie87] gave a proof using martingales and elementary linear algebra while [VTRF83] provided a proof based on linear programming and a modern proof can be found in [NS03] based on a reduction to discounted games and the use analytical tools. For two-player games, a proof based on a transfer theorem from one-player to two-player games can be found in [Gim06, GZ09].

6.2 Variants of mean-payoff games

The positive average condition defined by (9) is a variant of mean-payoff games which may be more suitable to model quality of service constraints or decision makers with a loss aversion.

Albeit function f_{posavg} is very similar to the f_{mean} function, maximizing the expected value of f_{posavg} and f_{mean} are two distinct goals. For example, a positive average maximizer prefers seeing the sequence $1, 1, 1, \dots$ for sure rather than seeing with equal probability $\frac{1}{2}$ the sequences $0, 0, 0, \dots$ or $2, 2, 2, \dots$ while a mean-value maximizer prefers the second situation to the first one.

To the best knowledge of the author, the techniques used in [Bie87, NS03, VTRF83] cannot be used to prove positionality of these games.

Since the positive average condition is the composition of the submixing function f_{mean} with an increasing function it is submixing as well, hence it is half-positional.

In mean-payoff co-Büchi games, a subset of the states are called Büchi states, and the payoff of player 1 is $-\infty$ if Büchi states are visited infinitely often and the mean-payoff value of the rewards otherwise. It is easy to check that such a payoff mapping is shift-invariant and submixing. Notice that in the present paper we do not explicitly handle payoff mappings that take infinite values, but it is possible to approximate the payoff function by replacing $-\infty$ by arbitrary small values to prove half-positionality of mean-payoff co-Büchi games.

6.3 New examples of positional payoff function

Although the generalized mean-payoff condition defined by (10) is not submixing a variant is. *Optimistic generalized mean-payoff games* are defined similarly except the winning condition is

$$\exists i, f^i_{\text{mean}} \geq 0.$$

It is a basic exercise to show that this winning condition is submixing. More generally, if $f_1, \dots, f_n$ are submixing payoff mappings then $\max\{f_1, \dots, f_n\}$ is submixing as well. Remark that optimistic generalized mean-payoff games are half-positional but not positional, this is a simple exercise.

Other examples are provided in [Gim07, Kop09, Gim06].

7 Comments

An anonymous reviewer proposed the following conjecture: given a payoff function f, if all one-player games equipped with f with the objective of maximizing the expected payoff are positional, then all two-player games equipped with f are half-positional. If it were true, then [Gim07] would imply the main theorem of the present paper. We provide a counter-example to this conjecture.

We give a payoff function such that all one-player games equipped with it are stationary with the objective of maximizing the expected payoff, and then we give a two-player game equipped with the same payoff function where it is necessary to use the memory, or to randomize in order to play optimally.

Consider the payoff function $f : \{a, b\}^* \to \{0, 1\}$, given as:

$$f(u) = \begin{cases} 0 & \text{if } u \text{ and } pab^2ab^4\cdots \text{ share a common suffix} \\ 1 & \text{otherwise} \end{cases},$$

where p is some finite word over the alphabet $\{a, b\}$. For every game, a colouring function $c : \mathbf{S} \to \{a, b, \epsilon\}$ is given.

It is fairly easy to see that on any one-player game equipped with f with the objective of maximizing the expected payoff any stationary strategy is optimal.

Now consider the two-player game given in Fig.1.

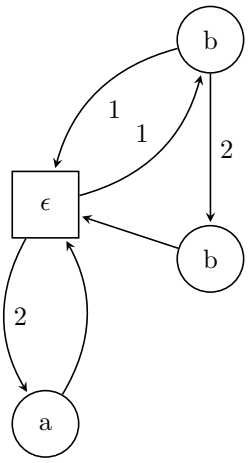

Figure 1: A two-player game counter-example

The states are labeled by the colouring function, and the circle states are controlled by Player 1 and the sole square state is controlled by Player 2. Both players have a single state with two actions, which are labeled by 1 and 2. One can see that when Player 1 chooses any stationary strategy, Player 2 can force the sequence $ab^2ab^4ab^6\cdots$. But Player 1 can win the game by either randomizing over the actions 1 and 2 or using the memory.

Unfortunately, the submixing property is not necessary for a shift-invariant payoff function to be half-positional. An example of a half-positional and shift-invariant though not sub-mixing game is provided in [BBE10]: states are labeled by integers and player 1 wins an infinite play labeled by $a_0a_1a_2 \in \mathbb{Z}^{\mathbb{N}}$ if $\liminf_n \sum_{i \leq n} a_i = -\infty$.

Conclusion

We have given a sufficient condition for a payoff function $f : C^{\mathbb{N}} \to \mathbb{R}$ to be half-positional, i.e. to guarantee the existence of a pure and stationary optimal strategy for the maximizer in any stochastic game with perfect information and finitely many states and actions. The existence of a sufficient and necessary condition for half-positionality expressed by equations on f remains an open problem.

References

[BBE10] Tomás Brázdil, Václav Brozek, and Kousha Etessami. One-counter stochastic games. In *FSTTCS*, pages 108–119, 2010.

[Bie87] K.-J. Bierth. An expected average reward criterion. *Stochastic Processes and Applications*, 26:133–140, 1987.

[CDHR10] Krishnendu Chatterjee, Laurent Doyen, Thomas A. Henzinger, and Jean-François Raskin. Generalized mean-payoff and energy games. In *FSTTCS*, pages 505–516, 2010.

[Cha07] Krishnendu Chatterjee. Concurrent games with tail objectives. *Theor. Comput. Sci.*, 388(1-3):181–198, 2007.

[CHJ05] K. Chatterjee, T.A. Henzinger, and M. Jurdzinski. Mean-payoff parity games. In *Proc. of LICS'05*, pages 178–187. IEEE, 2005.

[CJH03] K. Chatterejee, M. Jurdziński, and T.A. Henzinger. Quantitative stochastic parity games. In *SODA*, 2003.

[CY90] C. Courcoubetis and M. Yannakakis. Markov decision processes and regular events. In *Proceedings of ICALP'90*, volume 443 of *Lecture Notes in Computer Science*, pages 336–349. Springer, 1990.

[Der62] Cyrus Derman. On sequential decisions and markov chains. *Management Science*, 9:16–24, 1962.

[Gil57] D. Gilette. Stochastic games with zero stop probabilities. 3, 1957.

[Gim06] H. Gimbert. *Jeux Positionnels*. PhD thesis, Université Denis Diderot, Paris, 2006.

[Gim07] Hugo Gimbert. Pure stationary optimal strategies in markov decision processes. In *STACS*, pages 200–211, 2007.

[GTW02] E. Grädel, W. Thomas, and T. Wilke. *Automata, Logics and Infinite Games*, volume 2500 of *Lecture Notes in Computer Science*. Springer, 2002.

[GZ04] H. Gimbert and W. Zielonka. When can you play positionally? In *Proc. of MFCS'04*, volume 3153 of *Lecture Notes in Computer Science*, pages 686–697. Springer, 2004.

[GZ05] H. Gimbert and W. Zielonka. Games where you can play optimally without any memory. In *Proceedings of CONCUR'05*, volume 3653 of *Lecture Notes in Computer Science*, pages 428–442. Springer, 2005.

[GZ09] Hugo Gimbert and Wieslaw Zielonka. Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games. December 2009.

[HG08] Florian Horn and Hugo Gimbert. Optimal strategies in perfect-information stochastic games with tail winning conditions. *CoRR*, abs/0811.3978, 2008.

[Kop06] Eryk Kopczynski. Half-positional determinacy of infinite games. In *ICALP (2)*, pages 336–347, 2006.

[Kop09] Eryk Kopczynski. *Half-positional determinacy of infinite games.* PhD thesis, University of Warsaw, 2009.

[LL69] T.S. Liggett and S.A. Lippman. Stochastic games with perfect information and time average payoff. *SIAM Review*, 11(4):604–607, 1969.

[Mar98] D.A. Martin. The determinacy of Blackwell games. *Journal of Symbolic Logic*, 63(4):1565–1581, 1998.

[MS96] A.P. Maitra and W.D. Sudderth. *Discrete gambling and stochastic games*. Springer-Verlag, 1996.

[MY15] Ayala Mashiah-Yaakovi. Correlated equilibria in stochastic games with borel measurable payoffs. *Dynamic Games and Applications*, 5(1):120–135, 2015.

[NS03] A. Neyman and S. Sorin. *Stochastic games and applications*. Kluwer Academic Publishers, 2003.

[Sha53] L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Science USA*, 39:1095–1100, 1953.

[VTRF83] 0.J. Vrieze, S.H. Tijs, T.E.S. Raghavan, and J.A. Filar. A finite algorithm for switching control stochastic games. *O.R. Spektrum*, 5:15–24, 1983.

[Zie04] Wiesław Zielonka. Perfect-information stochastic parity games. In *FOSSACS 2004*, volume 2987 of *Lecture Notes in Computer Science*, pages 499–513. Springer, 2004.

[Zie10] Wieslaw Zielonka. Playing in stochastic environment: from multi-armed bandits to two-player games. In *FSTTCS*, pages 65–72, 2010.

[Zie11] Wiesiek Zielonka. private communication, September 2011.