# 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
hugo.gimbert@labri.fr

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

***Abstract*—We consider zero-sum stochastic games with perfect information and finitely many states and actions. The payoff is computed by a payoff 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 $\epsilon$-subgame-perfect equilibria in shift-invariant games, a second contribution of the paper.**

***Index Terms*—Stochastic Games**

## I. 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 $f : (\mathbf{SA})^\omega \to \mathbb{R}$.

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 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 [1]. Derman showed the positionality of one-player games with expected mean-payoff reward, using an Abelian theorem and a reduction to discounted games [2]. His results was later on extended to two-player games by Bierth [3], see also [4] for a modern proof. Gilette extended Derman result to two-players games [5] but his proof was found wrong and corrected by Ligget and Lippman [6]. Note that the authors of the present paper are not convinced by the proof of Liggett and Lippman, we discuss this in Section VI-A. The positionality of one-player parity games was addressed in [7] and later on extended to two-player games in [8], [9]. Counter games were extensively studied in [10] and several examples of positional counter games are given. There are also several examples of one-player and two-player positional games in [11], [12]. A whole zoology of half-positional games is presented in [13] and another example is given by mean-payoff co-Büchi games [14]. 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. For one-player games it was proved by the first author that every one-player game equipped with such a payoff function is positional [11]. This result was successfully used in [10] to prove positionality of counter games. A weaker form of this condition was presented in [15] 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 [16].

The present paper provides two contributions.

First, in games whose payoff function is shift-invariant, both players have $\epsilon$-subgame-perfect strategies, i.e. strategies that are $\epsilon$-optimal not only when the game starts but also whatever finite play has already been played (Theorem IV.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 IV.1).

The paper starts with preliminaries then in Section III we provide several examples of payoff functions, in Section IV we state and prove the existence of $\epsilon$-subgame-perfect strategies in shift-invariant games, in Section V we state and prove that games with shift-invariant and submixing payoff functions are half-positional and finally in Section VI we present some applications. Proofs that are not provided in the paper can be found in the appendix.

## II. Stochastic games with perfect information.

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

### A. 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 $\mathbf{S}^\infty = \mathbf{S}^* \cup \mathbf{S}^\omega$. If $\mathbf{S}$ is an infinite set then $\mathbf{S}^\omega$ denotes the set of infinite sequences $u \in \mathbf{S}^\mathbb{N}$ such that $u \in \mathbf{S}_0^\mathbb{N}$ for some finite subset $\mathbf{S}_0 \subseteq \mathbf{S}$.

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 II.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 II.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 $\mathcal{A}$.*

### B. 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 often denoted $\tau$.

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

**Definition II.3** (Deterministic and stationary strategies)**.** *A strategy $\sigma$ 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 $\sigma$ is* stationary *if $\sigma(h)$ only depends on the target of $h$. In other words $\sigma$ 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 $\sigma$ and $\tau$ 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 for every 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 $S_n(s_0 a_1 s_1 \cdots) = s_n$ and $A_n(s_0 a_1 s_1 \cdots) = a_n$.

### C. 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 $\sigma$ and $\tau$ 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:

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

By definition of maxmin and minmax, for every state $s \in \mathbf{S}$, $\operatorname{maxmin}(\mathbf{G})(s) \leq \operatorname{minmax}(\mathbf{G})(s)$. As a corollary of the second Martins's determinacy theorem [17], the converse inequality holds as well:

**Theorem II.4** (Martin second determinacy theorem)**.** *Let* $\mathbf{G}$ *be a game with a Borel-measurable payoff function* $f$*. Then for every state* $s \in \mathbf{S}$*:*

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

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

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

**Definition II.5** (optimal and $\epsilon$-optimal strategies)**.** *Let* $\mathbf{G}$ *be a game,* $\epsilon > 0$ *and* $\sigma$ *a strategy for player* 1*. Then* $\sigma$ *is* $\epsilon$*-optimal if for every strategy* $\tau$ *and every state* $s \in \mathbf{S}$*,*

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

*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 II.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* $s \in \mathbf{S}$*:*

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

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

**Definition II.7** ($\epsilon$-subgame-perfect strategies)**.** *Let* $\mathbf{G}$ *be a game equipped with a payoff function* $f$ *and* $\sigma$ *a strategy for player* 1*. For every finite play* $p = s_0 \cdots s_n$ *we denote* $\sigma[p]$ *the shift of strategy* $\sigma$ *by* $p$ *as the strategy defined by*

$$\sigma[p](t_0 a_1 t_1 \cdots a_m t_m) = \begin{cases} \sigma(p a_1 t_1 \cdots a_m t_m) & \text{if } s_n = t_0, \\ \sigma(t_0 a_1 t_1 \cdots a_m t_m) & \text{otherwise.} \end{cases}$$

*Then* $\sigma$ *is said to be* $\epsilon$*-subgame-perfect if for every finite play* $p$ *the strategy* $\sigma[p]$ *is* $\epsilon$*-optimal.*

### *D. Martingales*

In the proofs of Section IV we use the following classical results about martingales.

**Definition II.8** (Martingale)**.** *A sequence of real-valued random variables* $X_0, X_1, \ldots$ *is called a martingale if for every time* $n$ *the following are satisfied:*

$$\mathbb{E}[|X_n|] < \infty$$
$$\mathbb{E}[X_{n+1} \mid X_1, \ldots, X_n] = X_n.$$

*In case of having* $\leq$*, or* $\geq$ *instead of the equality, then the process is called a supermartingale or submartingale respectively.*

Martingales do converge almost-surely.

**Lemma II.9** (Doob's convergence theorem for martingales)**.** *Let* $X_0, X_1, \ldots$ *be a (sub)(super)martingale such that* $(\mathbb{E}[|X_n|])_{n\in\mathbb{N}}$ *is bounded. Then almost surely the limit* $X = \lim_{n\to\infty} X_n$ *exists and is finite.*

We use stopping times as well.

**Definition II.10.** *A stopping time* $V$ *with respect to a sequence of random variables* $S_0, S_1, \ldots$ *is a random variable that takes values from* $\mathbb{N} \cup \{\infty\}$ *such that the event* $V = n$ *is* $(S_0, \ldots, S_n)$*-measurable. A stopping time* $V$ *is called almost surely finite, if* $\mathbb{P}(V < \infty) = 1$*.*

The following theorem is a variant of Doob's optional stopping theorem.

**Theorem II.11.** *Let* $T$ *be a stopping time with respect to a sequence of random variables* $S_0, S_1, \ldots$*. Let* $(X_n)_{n\in\mathbb{N}}$ *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 submartingale or a supermartingale, 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.*

Note that the definition of $X_T$ makes sense thanks to Lemma II.9.

## III. Computing payoffs

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

### *A. 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 [5]. 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:

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

The *discounted payoff* has been introduced by Shapley [1]. 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:

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

The *parity condition* is used in automata theory and logics [18]. Each state $s$ is labeled with some priority $c(s) \in \{0, \ldots, d\}$. The payoff 1 if the highest priority seen infinitely often is odd, and 0 otherwise. For $c_0 c_1 \cdots \in \{0, \ldots, d\}^\omega$,

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

The *limsup payoff function* has been used in the theory of gambling games [19]. 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 [10], 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 [10]:

$$\limsup_n \sum_{0 \leq i \leq n} c_i = +\infty \quad (7)$$

$$\limsup_n \sum_{0 \leq i \leq n} c_i = -\infty \quad (8)$$

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

Generalized mean payoff games were introduced in [20]. 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:

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

### B. 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 III.1** (Payoff functions)**.** *A* payoff function *is any 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 focuse on shift-invariant payoff functions.

**Definition III.2** (Shift-invariant)**.** *A payoff function $f : \mathbf{C}^\omega \to \mathbb{R}$ is said to be a* shift-invariant *payoff function if:*

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

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 prefix-independent [21]. It seems to the authors of the present paper that the results of [22] 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 ( [22, l. 28 p. 184]) and the shift-invariance hypothesis is used in the core of the proof ( [22, l. -1 p.190]).

## IV. On the existence of $\epsilon$-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.

**Theorem IV.1.** *Let $\mathbf{G}$ be a game equipped with a payoff function $f$. Assume $f$ is shift-invariant. Then both players have $\epsilon$-subgame-perfect strategies in $\mathbf{G}$.*

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

### A. Weaknesses

For the remainder of Section IV fix $\epsilon > 0$.

In order to talk about unwanted behavior in the quest of proving that there exists a strategy that is $\epsilon$-subgame-perfect, we need the notion of a weakness.

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

When playing with strategy $\sigma$ we say that a weakness occurs in an infinite play if some finite prefix of it is a $\sigma$-weakness. Remark that a strategy is $2\epsilon$-subgame-perfect if and only if no $\sigma$-weakness can occur when playing with $\sigma$.

Every infinite play $q \in \mathbf{S}(\mathbf{AS})^\omega$ can be factorized uniquely as a finite or infinite sequence of plays $p_0, p_1, p_2, \ldots$ such that $q = p_0 p_1 p_2 \ldots$, and for every $p_n$,

1) if $p_n$ is finite then $p_n$ is a $\sigma$-weakness,
2) if $p_n$ is finite then no strict prefix of $p_n$ is a $\sigma$-weakness,
3) if $p_n$ is infinite then $q = p_0 p_1 \cdots p_n$ and no prefix of $p_n$ is a $\sigma$-weakness,

Having this factorization in mind helps understand better the results of this section. In the proofs we will rely upon the function $\delta$, defined as follows.

**Definition IV.3.** *Let $\sigma$ be a strategy for Player 1, and $s_0 \dots s_n \in \mathbf{S}(\mathbf{AS})^*$ define inductively on $n$*

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

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

The function $\delta$ give the cutting dates of the factorization described before the lemma.

### *B. The reset strategy*

Using the notion of a weakness above we define a strategy called the reset strategy, that resets the memory whenever a weakness occurs, and the sequel is devoted to proving that this strategy is subgame perfect.

**Definition IV.4** (The reset strategy)**.** *Let $\sigma$ be an $\epsilon$-optimal strategy. We define the reset strategy $\hat{\sigma}$ based on $\sigma$ 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 [22, Theorem 1]; while at most one memory reset is performed by the switching strategy, the reset strategy may reset its memory infinitely often.

The reset strategy has been used in [23] to prove the existence of optimal strategies in games with perfect information and two-valued payoff functions. Note that this property does not hold in the general case, for example there are in general no optimal strategies in games colored by $\{0, 1\}$ and equipped with the payoff function

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

### *C. Consistent $\epsilon$-optimal strategies*

We devote this subsection to proving that for every $\epsilon > 0$ there are $\epsilon$-optimal strategies that play consistently, that is they never play a suboptimal action.

*1) Consistent strategies:*

**Definition IV.5.** *The action $a \in A(s)$ is* optimal *in state $s \in \mathbf{S}$ if and only if $val(s) = \sum_{s' \in \mathbf{S}} p(s, a, s') val(s')$. If $a$ is not optimal it is said to be* suboptimal*.*

**Definition IV.6.** *A strategy is consistent if it plays only optimal actions.*

An important property of games when both players play only optimal actions is that the stochastic process $(val(S_n))_{n \in \mathbb{N}}$ is a martingale.

**Theorem IV.7.** *Let $\sigma$ and $\tau$ be strategies for player 1 and 2 and $s \in \mathbf{S}$. Assume $\sigma$ is consistent then $(val(S_n))_{n \in \mathbb{N}}$ is a sub-martingale for the probability measure $\mathbb{P}_s^{\sigma,\tau}$. If both $\sigma$ and $\tau$ are consistent then $(val(S_n))_{n \in \mathbb{N}}$ is a martingale.*

*2) Both players play ultimately consistently:* In this subsection we show that if Player 1 plays consistently then almost surely both players eventually will play only optimal actions.

We actually prove an even stronger fact: under the same hypothesis, both players eventually play only actions which do not change the value of the state. First we define such actions:

**Definition IV.8** (Value-neutral actions)**.** *We call an action $a \in A(s)$ value neutral if for all $s' \in \mathbf{S}$ we have:*

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

Note that a value-neutral action is optimal, and a suboptimal action is not value-neutral.

**Lemma IV.9.** *For every consistent strategy $\sigma$, strategy $\tau$ and $s \in \mathbf{S}$, $\mathbb{P}_s^{\sigma,\tau}(\exists n \in \mathbb{N}, \forall k \geq n, A_n$ is value-neutral $) = 1$.*

The proof is based on Doob's lemma (LemmaII.9), which implies that when Player 1 plays consistently, $(val(S_n))_{n \in \mathbb{N}}$ converges almost-surely, which is possible only when all actions played are ultimately value-neutral.

Since value-neutral actions are optimal, we get the it implies that when Player 1 plays consistently then suboptimal actions are played only finitely often.

**Corollary IV.10.** *For every consistent strategy $\sigma$, strategy $\tau$ and $s \in \mathbf{S}$, $\mathbb{P}_s^{\sigma,\tau}(\exists n \in \mathbb{N}, \forall k \geq n, A_n$ is optimal $) = 1$.*

*3) There are consistent $\epsilon$-optimal strategies:* It would simplify the proofs in the sequel if we knew that for every $\epsilon > 0$ we can choose an $\epsilon$-optimal strategy that plays consistently. In that way we know that Player 1 will never decrease his chances for a larger payoff on average. That is the purpose of the following lemma.

**Lemma IV.11.** *In game $\mathbf{G}$ equipped with the payoff function $f$, let $\epsilon' > 0$, and $a \in A(s)$ a suboptimal action in state $s \in \mathbf{S}_1$. Then there exists an $\epsilon'$-optimal strategy for Player 1 that never plays action $a$.*

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

**Corollary IV.12.** *In a game equipped with a shift-invariant function, for all $\epsilon' > 0$ both players have $\epsilon'$-optimal consistent strategies.*

The proof is by induction on the number of non-optimal actions, using Lemma IV.11.

It is useful to see the first weakness after some date $n$ as a stopping time and to see what we can say about the process $(val(S_n))_{n \in \mathbb{N}}$ stopped at such a time, which is a consequence of Doob's optional stopping theorem (Theorem II.11).

**Corollary IV.13.** *Let $T$ be a stopping time with respect to $(S_n)_{n \in \mathbb{N}}$. Then for all consistent strategy $\sigma$, and all strategies $\tau$ and state $s \in \mathbf{S}$, $(val(S_n))_{n \in \mathbb{N}}$ converges almost-surely and*

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

*In case we assume $\tau$ consistent instead of $\sigma$ the converse inequality holds.*

*Proof:* Immediate from Theorem II.11, since in a consistent game $(val(S_n))_{n\in\mathbb{N}}$ forms a martingale (Theorem IV.7), and from the hypothesis that the payoff function $f$ is bounded, we have that the values are also bounded. ∎

### D. Only finitely many weaknesses occur when playing $\hat{\sigma}$

The goal of this subsection is to show that when playing with the reset strategy $\hat{\sigma}$ (based on some $\epsilon$-optimal strategy $\sigma$) the set of paths with infinitely many weaknesses have measure 0, or equivalently almost surely there exists a last weakness. We start by proving that when Player 2 plays consistently, the probability that no weakness occur is strictly positive.

**Lemma IV.14.** *Let $\sigma$ be an $\epsilon$-optimal strategy, then there exists $\mu > 0$ such that for all strategy $\tau$ and $s \in S$, if $\tau$ is consistent,*

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

The proof of this lemma relies on the fact that each weakness can be exploited by player 2 and cause a loss of at least $\frac{3}{2}\epsilon$ because strategy $\tau$ is consistent. This can not happen almost-surely since $\sigma$ is both $\epsilon$-optimal.

The next step is to prove a statement similar to Lemma IV.14, which does assume consistency of $\tau$.

**Definition IV.15.** *We define the event*

$$\{\text{there is no } \sigma\text{-weakness after time } n\}$$

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

Notice that by definition of $\delta_\sigma$, if there is no deviation after date $n'$ then the sequence $(\delta_\sigma(S_0 \cdots S_m))_{m\in\mathbb{N}}$ is constant from date $n'$.

**Corollary IV.16.** *Let $\sigma$ be a consistent $\epsilon$-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it. For all strategies $\tau$ and states $s \in \mathbf{S}$ there exists $n \in \mathbb{N}$ such that*

$$\mathbb{P}_s^{\hat{\sigma},\tau}(\text{there is a } \sigma\text{-weakness after date } n) < 1.$$

*Proof (Sketch):* For every strategy $\tau$ and $n \in \mathbb{N}$ define a strategy $\tau_n$, which is sure to play consistently after date $n$. Thanks to Lemma IV.14, the corollary holds for this strategy. Use corollary IV.10 to prove the corollary for general $\tau$. ∎

Finally what is left to show is that when we are playing with the reset strategy, the number of weaknesses is a.s finite. Intuitively this is necessary because on the paths with infinitely many weaknesses the strategy might not be subgame-perfect, hence the need to show that the measure on these paths is 0.

**Lemma IV.17.** *Let $\sigma$ be an $\epsilon$-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it. Then for all strategies $\tau$ and states $s \in \mathbf{S}$,*

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

### E. The reset strategy is $2\epsilon$-subgame-perfect

The first goal is to prove that the reset strategy based on some $\epsilon$-optimal strategy is $\epsilon$-optimal itself. Let us first define the reset strategies that reset only up to some date and prove the $\epsilon$-optimality of them.

Let $\sigma$ be some $\epsilon$-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 IV.18.** *Let $\sigma$ be a consistent $\epsilon$-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it. Then for all strategies $\tau$, states $s \in \mathbf{S}$ and all $n \in \mathbb{N}$,*

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

*Proof (Sketch):* By induction on $n$, and decomposing the expected payoff on the event of a weakness at date $n+1$. ∎

Having shown that the strategies $\hat{\sigma}_n$ are $\epsilon$-optimal we can proceed into proving that the reset strategy $\hat{\sigma}$ itself is $\epsilon$-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 IV.19.** *Let $E$ be an event, and $\sigma_1$,$\sigma_2$ two strategies for Player 1 such that for all prefixes $p$ of an inifinite play in $E$, $\sigma_1(p) = \sigma_2(p)$. Then for all $\tau$ 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 IV.20.** *Let $\sigma$ be a consistent $\epsilon$-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it, then $\hat{\sigma}$ is $\epsilon$-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 IV.18 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 IV.19 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 IV.19 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 IV.17, $\lim_n (M - m)\mathbb{P}_s^{\hat{\sigma},\tau}(L > n) = 0$, hence

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

■

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

**Theorem IV.21.** *Let $\sigma$ be a consistent $\epsilon$-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it, then $\hat{\sigma}$ is $2\epsilon$-subgame-perfect.*

*Proof (Sketch):* Fix a prefix $p = s_0 \dots s_n \in \mathbf{S}(\mathbf{AS})^*$. If $p$ is a $\sigma$-weakness then $\hat{\sigma}[p] = \hat{\sigma}$ and since $\hat{\sigma}$ is $\epsilon$-optimal (Lemma IV.20), $\hat{\sigma}[p]$ is a*a fortiori* $\epsilon$-optimal. The case where $p$ is not a $\sigma$-weakness relies on on Lemma IV.20 and the consistency of $\hat{\sigma}$. ■

## V. Games with Shift-Invariant and Submixing Payoff Functions are Half-Positonal

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 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 V.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 $v \in \mathbf{C}^\omega$,*

$$f(w) \leq \max\{f(u), f(v)\} \ . \tag{12}$$

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

We can now state our main result.

**Theorem V.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 VI. Necessary and sufficient conditions for positionality are known for deterministic games [24].

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.

### A. Proof of Half-Positionality

We prove Theorem V.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 V.2 by induction on $N(\mathbf{G}) = \sum_s(|\mathbf{A}_1(s)| - 1)$. If $N(\mathbf{G}) = 0$ then in every state controlled by player 2 there is only one action available, thus player 2 has a unique strategy which is optimal, deterministic and stationary.

Let $\mathbf{G}$ be a game $N(\mathbf{G}) > 0$ and suppose Theorem V.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 $s \in \mathbf{A}_1$ 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:

$$\mathrm{minmax}(\mathbf{G})(s) \leq \max\{\mathrm{val}_0(s), \mathrm{val}_1(s)\} \ , \tag{13}$$

Since every strategy of player 2 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 $\sigma_0$ and $\sigma_1$ in $\mathbf{G}_0$ and $\mathbf{G}_1$. Suppose that (13) holds, and without loss of generality suppose $\mathrm{minmax}(\mathbf{G})(s) \leq \mathrm{val}_0(s)$. Since the deterministic stationary $\sigma_0$ is optimal in $\mathbf{G}_0$, it guarantees for every $\tau$ and $s \in \mathbf{S}$, $\mathbb{E}_s^{\sigma_0,\tau}[f] \geq \mathrm{val}_0(s) \geq \mathrm{minmax}(\mathbf{G})(s)$, thus $\sigma_0$ is optimal not only in the game $\mathbf{G}_0$ but in the game $\mathbf{G}$ as well. Thus (13) is enough to prove the inductive step.

### B. The projection mapping

To prove (13), we make use of two mappings

$$\pi_0 : s(\mathbf{AS})^\infty \to s(\mathbf{AS})^\infty \tag{14}$$

$$\pi_1 : s(\mathbf{AS})^\infty \to s(\mathbf{AS})^\infty. \tag{15}$$

First $\pi_0$ and $\pi_1$ are defined on finite words. The mapping $\pi_0$ associates with each finite play $p \in (\mathbf{SA})^*$ in $\mathbf{G}$ with source $s$ a finite play $\pi_0(p)$ in $\mathbf{G}_0$.

Intuitively, play $\pi_0(p)$ is obtained by erasing from $p$ 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 $p$ is followed by an action $a$ which is either in $\mathbf{A}_0(s)$ or in $\mathbf{A}_1(s)$. To obtain $\pi_0(p)$ one erases from $p$ two types of subwords:

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

Formally, $\pi_0$ and $\pi_1$ are defined as follows. Let $p = s_0a_0s_1a_1\cdots s_n \in s(\mathbf{AS})^*$ 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 $p_l$ the $l$-th factor of $p$ defined

by $p_l = s_{i_l} a_{i_l} \cdots a_{i_{l+1}-1}$ and $p_k = s_{i_k} a_{i_k} \cdots a_{n-1} s_n$. Then for $j \in \{0,1\}$,

$$\pi_j(p) = \prod_{0 \le l \le k | a_{i_l} \in A_j} p_l$$

We extend $\pi_0$ and $\pi_1$ to infinite words in a natural way: for an infinite play $p \in s(\mathbf{AS})^\omega$ then $\pi_0(p)$ is the limit of the sequence $(\pi_0(p_n))_{n\in\mathbb{N}}$, where $p_n$ is the prefix of $p$ of length $2n+1$. Remark that $\pi_0(p)$ may be a finite play, in case play $p$ 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 $\pi_0$ and $\pi_1$. For every infinite play $p \in (\mathbf{SA})^\omega$,

(A) if $\pi_0(p)$ is finite then $p$ has a suffix which is an infinite play in $\mathbf{G}_1$ starting in $s$,
(B) if $\pi_1(p)$ is finite then $p$ has a suffix which is an infinite play in $\mathbf{G}_0$ starting in $s$,
(C) if both $\pi_0(p)$ and $\pi_1(p)$ are infinite then both $\pi_0(p)$ and $\pi_1(p)$ reach state $s$ infinitely often,
(D) if both $\pi_0(p)$ and $\pi_1(p)$ are infinite, then $p$ is a shuffle of $\pi_0(p)$ and $\pi_1(p)$.

We use the three following random variables:

$$\Pi = S_0 A_1 S_1 \cdots \ , \tag{16}$$
$$\Pi_0 = \pi_0(S_0 A_1 S_1 \cdots) \ , \tag{17}$$
$$\Pi_1 = \pi_1(S_0 A_1 S_1 \cdots) \ . \tag{18}$$

### C. The trigger strategy

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

According to Theorem IV.1, there exists $\epsilon$-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 $p$ be a finite play in $s(AS)^*$ and $last(p) \in A$ the action played after the last visit of $p$ to $s$ and $t$ the last state of $p$, then:

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

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

$$\mathbb{E}_s^{\sigma,\tau^\sharp}[\ f \mid \Pi_0 \text{ is finite}] \le \mathrm{val}_1(s) + \epsilon \ , \tag{19}$$
$$\mathbb{E}_s^{\sigma,\tau^\sharp}[\ f \mid \Pi_1 \text{ is finite}] \le \mathrm{val}_0(s) + \epsilon \ , \tag{20}$$
$$\mathbb{E}_s^{\sigma,\tau^\sharp}[\ f \mid \Pi_0 \text{ and } \Pi_1 \text{ are both infinite}] \le \max\{\mathrm{val}_0(s), \mathrm{val}_1(s)\} + \epsilon \ . \tag{21}$$

### D. Proof of inequalities (19) and (20)

To prove inequality (19), we introduce the probability measure $\mu_1$ 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 $\sigma_1'$ 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 $\mu_1$ has the following key properties. For every finite play $h$ in the game $\mathbf{G}_1$ whose finite state is $t$,

$$\mu_1(ha \mid h) = \begin{cases} \sigma_1'(h)(a) & \text{if } t \in S_1, \\ \tau_1^\sharp(h)(a) & \text{if } t \in S_2 \ , \end{cases} \tag{22}$$
$$\mu_1(has \mid ha) = p(s|t,a). \tag{23}$$

As a consequence of the equalities (22) and (23), and according to the characterisation given by (1) and (2) the probability measure $\mu_1$ coincides with the probability measure $\mathbb{P}_s^{\sigma_1',\tau_1^\sharp}$. Since $\tau_1^\sharp$ is $\epsilon$-optimal in the game $\mathbf{G}_1$, it implies (19). The proof of (20) is symmetrical.

### E. Proof of inequality (21)

The proof of (21) requires several steps.

First, we prove that for every strategy $\sigma$ in $\mathbf{G}$, there exists a strategy $\sigma_0$ in $\mathbf{G}_0$ such that for every measurable event $E \subseteq (\mathbf{SA})^\omega$ in $\mathbf{G}_0$,

$$\mathbb{P}_s^{\sigma_0,\tau_0^\sharp}(E) \ge \mathbb{P}_s^{\sigma,\tau^\sharp}(\Pi_0 \text{ is infinite and } \Pi_0 \in E) \ . \tag{24}$$

The strategy $\sigma_0$ 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 (24) holds. Let $\mathcal{E}$ be the set of measurable events $E \subseteq (\mathbf{SA})^\omega$ in $\mathbf{G}_0$ such that (24) is satisfied. First, $\mathcal{E}$ contains all cylinders $h_0(\mathbf{SA})^\omega$ of $\mathbf{G}_0$ with $h_0 \in (\mathbf{SA})^*$ because:

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

where the first inequality can be proved by induction on the size of $h_0$, using the definition of $\sigma_0$ 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 $\sigma$-field generated by cylinders, that is all measurable events $E$ in $\mathbf{G}_0$. This completes the proof of (24).

Second step to obtain (21) is to prove that for every strategy $\sigma_0$ in $\mathbf{G}_0$:

$$\mathbb{P}_s^{\sigma_0,\tau_0^\sharp}\left(f \leq \mathrm{val}_0(s) + \epsilon \mid s \text{ is reached infinitely often}\right) = 1 \ . \tag{25}$$

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 $\epsilon$-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 (25).

Now we come to the end of the proof of (21). Let $\sigma_0$ be a strategy in $\mathbf{G}_0$ such that (24) holds for every measurable event $E$ in $\mathbf{G}_0$. According to (25), $\mathbb{P}_s^{\sigma_0,\tau^\sharp}\left(f > \mathrm{val}_0(s) + \epsilon \text{ and } s \text{ is reached infinitely often}\right) = 0$. For $i \in \{0,1\}$ denote $E_i$ and $F_i$ the events:

$$E_i = \{\Pi_i \text{ is infinite and reaches } s \text{ infinitely often}\}, \tag{26}$$

$$F_i = E_i \wedge \{f(\Pi_i) \leq \mathrm{val}_i(s) + \epsilon\}. \tag{27}$$

Remark that $F_i$ is well-defined since condition $E_i$ implies that $\Pi_i$ is infinite thus $f(\Pi_i)$ is well-defined in (27).

According to (25) 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 (24),

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

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

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

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

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

and according to (C) this terminates the proof of (21).

Since equations (19), (20) and (21) hold for every strategy $\sigma$ and every $\epsilon$, $\mathrm{minmax}(s) \leq \max\{\mathrm{val}_0(s), \mathrm{val}_1(s)\}$. W.l.o.g. assume $\mathrm{minmax}(s) \leq \mathrm{val}_0(s)$. Then the stationary deterministic strategy $\sigma_0$ optimal in $\mathbf{G}_0$ is a strategy in $\mathbf{G}$ as well and $\sigma_0$ 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 $\sigma_0$ is optimal in $\mathbf{G}$. This completes the proof of Theorem V.2. □

Note that the proof of Theorem V.2 does not rely on Martin's theorem.

## VI. Applications

### A. Unification of classical results

The existence of deterministic stationary optimal strategies in Markov decision processes with parity [7], limsup, liminf [19], mean-payoff [6], [4], [3], [25] or discounted payoff functions [1] is well-known. Theorem V.2 provides a unified proof of these five results, as a corollary of the following proposition.

**Proposition VI.1.** *The payoff functions $f_{lsup}$, $f_{linf}$, $f_{par}$ and $f_{mean}$ are shift-invariant and submixing.*

This was proved in [11], [26].

**Corollary VI.2.** *In every two-player stochastic game equipped with the parity, limsup, liminf, mean or discounted payoff function, both players have a deterministic and stationary strategy which is optimal.*

*Proof:* Except for the discounted payoff function, this is a direct consequence of Proposition VI.1 and Theorem V.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 [1]. Deshift-invariants follow. Every Markov decision process $\mathbf{G}$ with states $\mathbf{S}$ equipped with a discounted payoff function can be turned into a Markov decision process $\mathbf{G}'$ with states $\mathbf{S} \cup \mathbf{S} \times \{0\}$ equipped with a mean-payoff function such that for every strategy $\sigma'$ optimal in $\mathbf{G}'$ its restriction $\sigma$ to $\mathbf{S}$ is optimal in $\mathbf{G}$. States of $\mathbf{G}'$ are obtained by adding an extra absorbing state $(s,0)$ for each state $s \in \mathbf{S}$, the reward in state $(s,0)$ is $\frac{r(s)}{1-\lambda(s)}$. Whatever action is chosen in state $s$ there is probability $1-\lambda(s)$ to go to the absorbing state $(s,0)$ and stay there forever, whereas the original transition probabilities in $\mathbf{G}$ to states $s \in \mathbf{S}$ are multiplied by $\lambda(s)$. ■

Corollary VI.2 unifies and simplifies existing proofs of [7] for the parity game and [19] for the limsup game.

The existence of determinstic and stationary optimal strategies in mean-payoff games has attracted many attention. The first proof was given by Gilette [5] 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 [6]. Actually a careful inspection of their proof shows that it contains a flaw as well: first, it is not clear at all how to deduce (7) from (6) and (5) on line 20 of page 606 of [6], second there is no obvious reason for the hypotheses of Brown's result [27, Theorem 4.2] to be satisfied in the context of [6]. For one-player games, Bierth [3] gave a proof using martingales and elementary linear algebra while [25] provided a proof based on linear programming and a modern proof can be found in [4] based on a reduction to discounted games and the use analytical tools. For tw-player games, a correct proof of positionality of two-player mean-payoff games with perfect information can be found in [26], [28].

### B. 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 $f_{\text{posavg}}$ is very similar to the $f_{\text{mean}}$ function, maximizing the expected value of $f_{\text{posavg}}$ and $f_{\text{mean}}$ are two distinct goals. For example, a positive average maximizer prefers seeing the sequence $1, 1, 1, \ldots$ for sure rather than seeing with equal probability $\frac{1}{2}$ the sequences $0, 0, 0, \ldots$ or $10, 10, 10, \ldots$ while a mean-value maximizer prefers the second situation to the first one.

To the best knowledge of the author, the techniques used in [3], [4], [25] cannot be used to prove positionality of these games.

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

In mean-payoff co-Bchi games, a subset of the states are called Bchi states, and the payoff of player 1 is $-\infty$ if Bchi 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 explicitely 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-Bchi games.

### C. 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 $f_1, \ldots, f_n$ are submixing payoff mappings then $\max\{f_1, \ldots, f_n\}$ is submixing as well. Remark that optimistic generalized mean-payoff games are half-positional but not positional, this is a simple exercise.

Another examples are provided in [11], [13], [26].

## Conclusion

We would like to be able to generalize the submising condition in order to cover all positional counter-games presented in [10]. Although [12] provides a necessary and sufficient condition for the positionality of one-player games, it seems not obvious how this characterization can be used to prove positionality of counter games.

## References


[1] L. S. Shapley., "Stochastic games," *Proceedings of the National Academy of Science USA*, vol. 39, pp. 1095–1100, 1953.

[2] C. Derman, "On sequential decisions and markov chains," *Management Science*, vol. 9, pp. 16–24, 1962.

[3] K.-J. Bierth, "An expected average reward criterion," *Stochastic Processes and Applications*, vol. 26, pp. 133–140, 1987.

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

[5] D. Gilette, "Stochastic games with zero stop probabilities," vol. 3, 1957.

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

[7] C. Courcoubetis and M. Yannakakis, "Markov decision processes and regular events," in *Proceedings of ICALP'90*, ser. Lecture Notes in Computer Science, vol. 443. Springer, 1990, pp. 336–349.

[8] K. Chatterejee, M. Jurdziński, and T. Henzinger, "Quantitative stochastic parity games," in *SODA*, 2003, to appear.

[9] W. Zielonka, "Perfect-information stochastic parity games," in *FOSSACS 2004*, ser. Lecture Notes in Computer Science, vol. 2987. Springer, 2004, pp. 499–513.

[10] T. Brázdil, V. Brozek, and K. Etessami, "One-counter stochastic games," in *FSTTCS*, 2010, pp. 108–119.

[11] H. Gimbert, "Pure stationary optimal strategies in markov decision processes," in *STACS*, 2007, pp. 200–211.

[12] W. Zielonka, "Playing in stochastic environment: from multi-armed bandits to two-player games," in *FSTTCS*, 2010, pp. 65–72.

[13] E. Kopczynski, "Half-positional determinacy of infinite games." Ph.D. dissertation, University of Warsaw, 2009.

[14] K. Chatterjee, T. Henzinger, and M. Jurdzinski, "Mean-payoff parity games," in *Proc. of LICS'05*. IEEE, 2005, pp. 178–187.

[15] H. Gimbert and W. Zielonka, "When can you play positionally?" in *Proc. of MFCS'04*, ser. Lecture Notes in Computer Science, vol. 3153. Springer, 2004, pp. 686–697.

[16] E. Kopczynski, "Half-positional determinacy of infinite games," in *ICALP (2)*, 2006, pp. 336–347.

[17] D. Martin, "The determinacy of Blackwell games," *Journal of Symbolic Logic*, vol. 63, no. 4, pp. 1565–1581, 1998.

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

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

[20] K. Chatterjee, L. Doyen, T. A. Henzinger, and J.-F. Raskin, "Generalized mean-payoff and energy games," in *FSTTCS*, 2010, pp. 505–516.

[21] W. Zielonka, septembre 2011.

[22] K. Chatterjee, "Concurrent games with tail objectives," *Theor. Comput. Sci.*, vol. 388, no. 1-3, pp. 181–198, 2007.

[23] F. Horn and H. Gimbert, "Optimal strategies in perfect-information stochastic games with tail winning conditions," *CoRR*, vol. abs/0811.3978, 2008.

[24] H. Gimbert and W. Zielonka, "Games where you can play optimally without any memory," in *Proceedings of CONCUR'05*, ser. Lecture Notes in Computer Science, vol. 3653. Springer, 2005, pp. 428–442.

[25] J. Vrieze, S. Tijs, T. Raghavan, and J. Filar, "A finite algorithm for switching control stochastic games," *O.R. Spektrum*, vol. 5, pp. 15–24, 1983.

[26] H. Gimbert, "Jeux positionnels," Ph.D. dissertation, Université Denis Diderot, Paris, 2006.

[27] B. Brown, "On the iterative method of dynamic programming on a finite space discrete time markov processs," *The annals of mathematical statistics*, vol. 36, pp. 1279–1285, 1965.

[28] H. Gimbert and W. Zielonka, "Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games," Dec. 2009. [Online]. Available: http://hal.archives-ouvertes.fr/hal-00438359

## APPENDIX

### PROOF OF PROPOSITION II.6

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.

### PROOF OF THEOREM II.11

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 LemmaII.9 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 $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. A similar proof applies for the case of supermartingales or submartingales.

### PROOF OF THEOREM IV.7

By definition $f$ is bounded hence for all $n \in \mathbb{N}$ and all strategies $\sigma$, $\tau$ and states $s \in \mathbf{S}$, $\mathbb{E}_s^{\sigma,\tau}[|val(S_n)|] < \infty$. Finally, for consistent $\sigma$ and, strategies $\tau$ 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 $\sum_{a\in\mathbf{A}(s)} \sigma(S_0 \ldots S_n)(a)\big(\sum_{s'\in\mathbf{S}} p(val(S_n), a, s')\big)$ if $S_n \in \mathbf{S}_1$ and $\sum_{a\in\mathbf{A}(s)} \tau(S_0 \ldots S_n)(a)\big(\sum_{s'\in\mathbf{S}} p(val(S_n), a, s')\big)$ 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 consistent strategy $\sigma$. Clearly the same proof holds when $\tau$ is also consistent, making the inequality above an equality.

### PROOF OF LEMMA IV.9

In case every action is value-neutral, we have nothing to prove. Assume that there exists an action $a \in A(s)$ which is not value neutral. 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 consistent strategies $\sigma$, $t \in \mathbf{S}$ and $\tau$ we have $\mathbb{P}_t^{\sigma,\tau}(E_{sa}) = 0$. Assume on the contrary that for some consistent $\sigma$, $t \in \mathbf{S}$ and $\tau$ 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 $\sigma$, $\tau$ and starting from state $t$, there is some non-zero probability that for infinitely many $n \in \mathbb{N}$,

$$\mid val(S_n) - val(S_{n+1}) \mid \geq val(s) - val(s') > 0. \tag{28}$$

Clearly a consequence of (28) is that there is some non-zero probability that the sequence $val(S_0), val(S_1), \ldots$ does not converge. But according to Lemma II.9, the submartingale $val(S_0), val(S_1), \ldots$ converges almost surely, hence for all consistent strategies $\sigma$ all $t \in \mathbf{S}$ and $\tau$ we have $\mathbb{P}_t^{\sigma,\tau}(E_{sa}) = 0$.

### PROOF OF LEMMA IV.11

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:

$$\forall t \in S,\ val(\mathbf{G}')(t) \geq val(\mathbf{G})(t). \tag{29}$$

We split the proof of (29) 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 $\tau$ the strategy for Player 2 in $\mathbf{G}$ that plays according to strategy $\tau'$ which is $\epsilon$-optimal in $\mathbf{G}'$, as long as Player 1 does not choose the suboptimal action $a$. In case he does choose it, $\tau$ switches definitely to the strategy $\tau''$ 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 suboptimal action $a$.
  When playing with the strategy $\tau$ we have the following properties, for all $t$ and $\sigma$:

$$\mathbb{E}_t^{\sigma,\tau}[f \mid Opt] \leq val(\mathbf{G}')(t) + \epsilon \tag{30}$$

$$\mathbb{E}_t^{\sigma,\tau}[f \mid \neg Opt] \leq val(\mathbf{G})(s) - d + \frac{d}{2} \tag{31}$$

  To show (30), note because of the condition $Opt$ the game is played only in $\mathbf{G}'$, hence the strategy $\sigma$ even though it is a strategy in $\mathbf{G}$ it behaves like the strategy $\sigma'$ in $\mathbf{G}'$ defined in the following way: for all $p = s_0 \ldots s_n \in S(AS)^*$, and $b \in A(s_n)$, $\sigma'(p)(b) = \mathbb{P}_{s_0}^{\sigma,\tau}(pb \mid p \wedge Opt)$. That is $\mathbb{E}_t^{\sigma,\tau}[f \mid Opt] = \mathbb{E}_t^{\sigma',\tau'}[f]$, because $\tau$ never has to switch to $\tau''$. Now (30) is a direct consequence of the $\epsilon$-optimality of $\tau'$.

We get (31), because when Player 1 chooses the action $a$, $\tau$ switches to the $\tau''$ 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 (30) and (31) we have that for all $t \in S$, $\epsilon > 0$ and $\sigma$,

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

Taking $t = s$ and the supremum over all $\sigma$, 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 $Sw_\sigma$ 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 $\sigma$ the suboptimal action $a$ is about to be played at some date $n$. For every strategy $\sigma$ define $\sigma_s$ as the strategy in $\mathbf{G}'$ that plays like $\sigma$ as long as the latter does not choose the action $a$ (with nonzero probability), and when it does, $\sigma_s$ switches to the strategy $\sigma'$ that is $\epsilon$-optimal in $\mathbf{G}'$. Let $\tau$ be the strategy for Player 2 which plays according to the strategy $\tau'$ which is $\epsilon$-optimal in $\mathbf{G}'$ as long as Player 1 does not choose the action $a$, otherwise it switches to strategy $\tau''$ that is $\epsilon$-optimal in $\mathbf{G}$.
  Since pairs of strategies $\sigma, \tau$ coincide to $\sigma_s, \tau'$ up to the date $n$ in the event $Sw_\sigma$, we write $p = \mathbb{P}_t^{\sigma,\tau}(Sw_\sigma) = \mathbb{P}_t^{\sigma_s,\tau'}(SW_\sigma)$. From the definition of $\tau$, $\sigma_s$ and the fact that $val(\mathbf{G})(s) = val(\mathbf{G}')(s)$, shown above we have:

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

  Combining the two inequalities above we get

  $$\mathbb{E}_t^{\sigma,\tau}[f \mid Sw_\sigma] \leq \mathbb{E}_t^{\sigma_s,\tau}[f \mid Sw_\sigma] + 2\epsilon. \tag{32}$$

  Keeping this in mind we proceed:

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

  where the first equality is a basic property of expectations, the first inequality is from (32) and because on the paths of the event $\neg Sw_\sigma$ the strategies $\sigma$ and $\sigma_s$ coincide, the following equality is a basic property of expectations, while the second one and the last inequality are by definition of the strategy $\tau$. 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 $\tau$ never switches to the strategy $\tau''$.
  Since this holds for any $\epsilon > 0$, taking the supremum over strategies $\sigma$ we get $val(\mathbf{G}')(t) \geq val(\mathbf{G})(t)$ as desired.

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

## Proof of Lemma IV.14

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 $\sigma$ be a 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 $\tau$ 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 $\tau'$ be the strategy that plays identically to $\tau$ as long as weakness does not occur, and when a weakness occurs it switches to a $\frac{\epsilon}{2}$-optimal response $\tau''$. Since strategies $\tau$ and $\tau'$ coincide up to the first weakness let $p = \mathbb{P}_s^{\sigma,\tau'}(F = \infty) = \mathbb{P}_s^{\sigma,\tau}(F = \infty)$. From the $\epsilon$-optimality of $\sigma$, and a basic property of conditional expectations:

$$\begin{aligned}val(s) - \epsilon &\leq \mathbb{E}_s^{\sigma,\tau'}[f]\\ &= (1-p) \cdot \mathbb{E}_s^{\sigma,\tau'}[f \mid F < \infty]\\ &\quad + p \cdot \mathbb{E}_s^{\sigma,\tau'}[f \mid F = \infty]\\ &\leq (1-p) \cdot \mathbb{E}_s^{\sigma,\tau'}[f \mid F < \infty] + pM.\end{aligned} \tag{33}$$

As a subsequence of the strategy $\tau'$ namely that it resets to the strategy $\tau''$ 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 $\tau'$ that takes advantage of the weakness by definition. Plugging this inequality on (33):

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

where in the equality we have decomposed $(1-p)\cdot\mathbb{E}_s^{\sigma,\tau'}[val(S_F) \mid F<\infty]$ to $\mathbb{E}_s^{\sigma,\tau'}[val(S_F)] - p\cdot\mathbb{E}_s^{\sigma,\tau'}[val(S_F) \mid F=\infty]$, 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 consistently, thus applying Corollary IV.13 we get $\mathbb{E}_s^{\sigma,\tau'}[val(S_F)] \leq \text{val}(s)$. Finally from above: $\frac{\epsilon}{2(M-m+3/2)} \leq p < \mu$, a uniform lower bound for $\mu$, that does not depend on the choice of $\tau$.

**Definition A.1.** *For a state $s\in\mathbf{S}$ and a distribution $D\in\Delta(\mathbf{A}(s))$, let $opt(s,D) = (\forall a\in\mathbf{A}(s), D(a)>0 \implies a$ is optimal in $s)$. Denote by $\mathcal{U}$ the uniform distribution on a finite set, and for a state $s\in\mathbf{S}$ denote by $Op(s)$ the set of optimal actions in $s$. Given any strategy $\tau$ and natural number $n$, define $\tau_n$ to be the following strategy:*

$$\tau_n(s_0\dots s_m) = \begin{cases} \tau(s_0\dots s_m) & \text{if } m<n \\ & \text{or } opt(s_m,\tau(s_0\dots s_m)) \\ \mathcal{U}(Op(s_m)) & \text{otherwise} \end{cases}$$

**Lemma A.2.** *Let $\sigma$ be a consistent $\epsilon$-optimal strategy, and $\hat{\sigma}$ the reset strategy based on it. Then there exists $\mu>0$ such that for all $\tau$ and $s\in\mathbf{S}$ and $n\in\mathbb{N}$ there exists $n'>n$ such that,*

$$\mathbb{P}_s^{\hat{\sigma},\tau_n}(\text{ there is no } \sigma\text{-weakness after time } n'\ ) \geq \mu > 0,$$

*where $\tau_n$ is the strategy defined in Definition A.1.*

*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 $n$.

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

$$\mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2<\infty \mid F_n<\infty) \leq 1-\mu. \tag{34}$$

Let $\mu$ be given by Lemma IV.14 and weak defined like in the proof of Lemma IV.14. Then (34) is a consequence of:

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

where the first and second equalities hold because

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

the third equality because if $\mathrm{weak}(n,p)$ then $\hat{\sigma}[p] = \hat{\sigma}$, the fouth equality is a consequence of Lemma IV.19 since $\sigma$ and $\hat{\sigma}$ coincide up to the first $\sigma$-weakness and the last inequality holds because according to the definition of $\tau_n$, and since $|p| \geq n$ then $\tau_n[p]$ is consistent and we can apply Lemma IV.14.

Let $n' \geq n$ such that

$$\mathbb{P}_s^{\hat{\sigma},\tau_n}(F_{n'} < \infty) \leq \mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2 < \infty),$$

which exists because by definition of $F_n^2$,

$$\{F_n^2 < \infty\} \subseteq \bigcup_{n' \geq n} \{F_n^2 < \infty\}.$$

Then

$$\begin{aligned}
\mathbb{P}_s^{\hat{\sigma},\tau_n}(F_{n'} < \infty) &\leq \mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2 < \infty) \\
&= \mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n < \infty)\mathbb{P}_s^{\hat{\sigma},\tau_n}(F_n^2 < \infty \mid F_n < \infty) \qquad (35) \\
&\leq 1 - \mu,
\end{aligned}$$

where the first inequality is by choice of $n'$, the equality in (35) is because $\mathbb{P}(F_n^2 < \infty \mid F_n = \infty) = 0$, and the inequality is from (34). This completes the proof.

■

## PROOF OF COROLLARY IV.16

We use Definition A.1.

Let $\Omega$ be the RV taking values in $\mathbb{N} \cup \{\infty\}$ that maps to the date of the last suboptimal action played, if it exists, otherwise let it be $\infty$. Because $\tau$ and $\tau_n$ coincide on all paths where the last suboptimal action is played before $n$, that is on the event $\{\Omega < n\}$, then for any $n \in \mathbb{N}$ and event $E$:

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

Since $\hat{\sigma}$ is consistent we can apply Corollary IV.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 A.2, 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 A.2 to conclude the proof.

## PROOF OF LEMMA IV.17

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 $\tau$ and $s$ such that:

$$\sup_{\tau',s'} \mathbb{P}_{s'}^{\hat{\sigma},\tau'}(L = \infty) \leq \mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty) + \epsilon'. \tag{36}$$

Let $n \in \mathbb{N}$ be such that according to Corollary IV.16 $\mu = \mathbb{P}_s^{\hat{\sigma},\tau}(F_n < \infty) < 1$. 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 (36). 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 $\tau''$ 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 $\mathbb{P}_s^{\hat{\sigma},\tau}(L = \infty) = 0$.

## PROOF OF LEMMA IV.18

We proceed by induction on $n$, the base case being true by definition since $\hat{\sigma}_0 = \sigma$, and the induction hypothesis being that for all $\tau$ 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 $\tau$ and a state $s$. Let $\tau'$ be the strategy that plays like $\tau$ except if there is a weakness at date $n+1$ (in case of the event $\delta_\sigma(S_0 \dots S_{n+1}) = n+1$) in which case it resets to the $\frac{\epsilon}{2}$-optimal response $\tau''$. Let $L_n = \delta_\sigma(S_0 \dots S_n)$ By decomposing 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:

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

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 $\epsilon$-optimality of $\sigma$. 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 secodn equality is by construction of $\tau'$ and the inequality because $\tau''$ is chosen as the $\frac{\epsilon}{2}$-optimal to $\sigma[s_0 \dots s_{n+1}]$, which is not $2\epsilon$-optimal by definitino 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 $\tau$ and $\tau'$ 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 $\epsilon$ terms) and

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

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

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

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

## PROOF OF THEOREM IV.21

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 IV.20. Assume that $\delta_\sigma(s_0 \dots s_n) < n$, from which we get that

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

and,

$$\inf_\tau \mathbb{E}_{s_n}^{\sigma[s_{\delta(s_0 \dots s_{n-1}) \dots s_n],\tau}}[f] \geq val(s_n) - 2\epsilon. \tag{37}$$

by definition of the function $\delta_\sigma$, when $\delta_\sigma(s_0 \dots s_n) \neq n$. Assume that there exists a strategy $\tau$ 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 $\tau$ we can build another strategy $\tau'$ 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 (37). Let $\tau'$ be the strategy that plays like $\tau$ as long as no weakness occurs, and in case it does, it switches to the $\epsilon$-response strategy $\tau''$. 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

$$\begin{aligned}
&val(s_n) - 2\epsilon \\
&> \mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{L=0}] + \mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{F<\infty}] \\
&= \mathbb{E}_{s_n}^{\sigma_2,\tau'}[f \cdot \mathbb{1}_{L=0}] + \mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{F<\infty}] \\
&= \mathbb{E}_{s_n}^{\sigma_2,\tau'}[f] - \mathbb{E}_{s_n}^{\sigma_2,\tau'}[f \cdot \mathbb{1}_{F<\infty}] \\
&\quad + \mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{F<\infty}],
\end{aligned} \tag{38}$$

where the first inequality is by assumption on the strategy $\tau$ (and also because $(L = 0) \iff (F = \infty)$), the first equality is because both pairs of strategies $\hat{\sigma}_1, \sigma_2$ and $\tau, \tau'$ coincide up to the first weakness. Let weak be the boolean function characterizing the prefixes up to first weakness after some specified date, that is $\mathrm{weak}_\sigma(n, s_0 \dots s_m) := (\delta_\sigma(s_0 \dots s_m) = m) \wedge (\delta_\sigma(s_0 \dots s_{m-1}) \leq n)$. Let

$$\begin{aligned}
R_1 = &\{t_0 \dots t_m \in \mathbf{S}(\mathbf{AS})^* \mid \\
&\mathrm{weak}_{\hat{\sigma}_1}(\delta_\sigma(s_0 \dots s_n), s_0 \dots s_{n-1} t_0 \dots t_m) \\
&\wedge s_n = t_0\}, \\
R_2 = &\{t_0 \dots t_m \in \mathbf{S}(\mathbf{AS})^* \mid \\
&\mathrm{weak}_{\sigma_2}(\delta_\sigma(s_0 \dots s_n), s_0 \dots s_{n-1} t_0 \dots t_m) \\
&\wedge s_n = t_0\},
\end{aligned}$$

sets of finite continuations of $s_0 \dots s_n$ with a weakness, for both strategies $\hat{\sigma}_1$ and $\sigma_2$ respectively. Then for the last two terms we have:

$$\begin{aligned}
&\mathbb{E}_{s_n}^{\hat{\sigma}_1,\tau}[f \cdot \mathbb{1}_{F<\infty}] \\
&= \sum_{t_0 \dots t_m \in R_1} \mathbb{P}_{s_n}^{\hat{\sigma}_1,\tau}(t_0 \dots t_m) \mathbb{E}_{t_m}^{\hat{\sigma},\tau[t_0 \dots t_m]}[f] \\
&\geq \sum_{t_0 \dots t_m \in R_1} \mathbb{P}_{s_n}^{\hat{\sigma}_1,\tau}(t_0 \dots t_m)(val(t_m) - \epsilon),
\end{aligned}$$

this is by the definition of $\hat{\sigma}$ and Lemma IV.20, while for the other term

$$\begin{aligned}
&\mathbb{E}_{s_n}^{\sigma_2,\tau'}[f \cdot \mathbb{1}_{F<\infty}] \\
&= \sum_{t_0 \dots t_m \in R_2} \mathbb{P}_{s_n}^{\sigma_2,\tau'}(t_0 \dots t_m) \mathbb{E}_{t_m}^{\sigma_2[t_0 \dots t_m],\tau'}[f] \\
&\leq \sum_{t_0 \dots t_m \in R_2} \mathbb{P}_{s_n}^{\sigma_2,\tau'}(t_0 \dots t_m)(val(t_m) - 2\epsilon + \epsilon)
\end{aligned}$$

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 $\tau, \tau'$, 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 (37) when plugging into (38)