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

Hugo Gimbert[1] and Edon Kelmendi[2]

1 **CNRS, LaBRI, Université de Bordeaux, France**
hugo.gimbert@labri.fr
2 **LaBRI, Université de Bordeaux, France**
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 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.

1998 ACM Subject Classification F.3.1 Software/Program Verification.

Keywords and phrases Games, Stochastic, Perfect Information, Positional Strategies

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

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

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. Half-positionality is a very desirable property in a game, if the objective is to compute the optimal strategy.

* This work was partially supported by the French ANR project "Stoch-MC" and "LaBEX CPU" of Université de Bordeaux.

42nd Conference on Very Important Topics (CVIT 2016).
Editors: John Q. Open and Joan R. Acces; Article No. 23; pp. 23:1–23:14
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

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 [24]. Derman showed the positionality of one-player games with expected mean-payoff reward, using an Abelian theorem and a reduction to discounted games [8]. Gilette extended Derman's result to two-player games [9] but his proof was found to be wrong and corrected by Ligget and Lippman [19]. The positionality of one-player parity games was addressed in [7] and later on extended to two-player games in [3, 27]. Counter games were extensively studied in [2] and several examples of positional counter games are given. There are also several examples of one-player and two-player positional games in [13, 28]. A whole zoology of half-positional games is presented in [18] and another example is given by mean-payoff co-Büchi games [4]. 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 9 and 10.

For one-player games it was proved by the first author that every one-player game equipped with such a payoff function is positional [13]. This result was successfully used in [2] to prove positionality of counter games. A weaker form of this condition was presented in [11] 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* payoff function which takes only two values is half-positional [17].

A recent result [28] 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 [28] is much harder to check than the submixing property which is purely syntactic. In [15] a transfer theorem is proved: if all one-player games are positional then the two-player games are positional as well, but if all one-player games are positional for Player 1, the two-player game is not necessarily half-positional [14].

The main contribution of the paper is the following result: 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 11).

This result relies on the existence of ϵ-subgame-perfect strategies for shift-invariant payoff functions, a result published recently in [22, Proposition 11] and obtained independently by the authors [14].

The paper starts with preliminaries, then in Section 3 we provide several examples of payoff functions. In Section 4 we state and prove that games with shift-invariant and submixing payoff functions are half-positional, in Section 5 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 $\forall s \in \mathbf{S},\ 0 \le \delta(s) \le 1$ and $\sum_{s\in\mathbf{S}} \delta(s) = 1$. The set of probability distributions on $\mathbf{S}$ is denoted $\Delta(\mathbf{S})$.

▶ **Definition 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 $p = s_0a_1s_1a_2 \cdots \in (\mathbf{SA})^\omega$ 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 $f(h)$.

▶ **Definition 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 $s_0a_1 \cdots s_n$, and every action $a \in \mathbf{A}$, $(\sigma(s_0a_1 \cdots s_n)(a) > 0) \implies (a \in \mathbf{A}(s_n))$. 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 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:

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

By definition of maxmin and minmax, for every state $s \in \mathbf{S}$, $\text{maxmin}(\mathbf{G})(s) \leq \text{minmax}(\mathbf{G})(s)$. As a corollary of the Martin's second determinacy theorem [21], the converse inequality holds as well:

▶ **Theorem 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 $s \in \mathbf{S}$:*

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

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

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

▶ **Definition 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 $s \in \mathbf{S}$,

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

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 1. 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 $s \in \mathbf{S}$:

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

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 6** (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 $h = s_0 \cdots s_n$

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

The following result is a cornerstone for our main theorem. This result was proved by the authors in [14], and was independently published in [22, Proposition 11].

▶ **Theorem 7.** *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 $\mathbf{G}$.*

A game that is equipped with a payoff function that is not shift-invariant (Definition 9) does not necessarily admit ϵ-subgame-perfect strategies.

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 [9]. 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:

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

The *discounted payoff* has been introduced by Shapley [24]. 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:

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

The *parity condition* is used in automata theory and logics [16]. 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 [20]. 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 [2], 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 [2]:

$$\limsup_n \sum_{0 \le i \le n} c_i = +\infty, \quad \limsup_n \sum_{0 \le i \le n} c_i = -\infty, \quad \text{and} \quad f_{\text{mean}}(c_0 c_1 \ldots) > 0 \ .$$

Generalized mean payoff games were introduced in [6]. 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:

$$\forall 1 \le i \le k, f^i_{\text{mean}}\left(r^{(i)}_0 r^{(i)}_1 \ldots\right) > 0 \ . \tag{7}$$

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 8** (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 9** (Shift-invariant). A payoff function $f : \mathbf{C}^\omega \to \mathbb{R}$ is said to be a *shift-invariant* payoff function if:

$$\forall c \in \mathbf{C}, \forall u \in \mathbf{C}^\omega, f(cu) = f(u) \ . \tag{8}$$

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

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

is tail-measurable but not shift-invariant [26], and this matters for examples the result of [5] hold for shift-invariant winning objectives but may not hold for tail-measurable objectives.

4 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 $u = u_0 u_1 u_2 \cdots$. 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 $u = u_0 u_1 u_2 \cdots$ and $v = v_0 v_1 v_2 \cdots$ of u and v such that $w = u_0 v_0 u_1 v_1 u_2 v_2 \cdots$.

▶ **Definition 10** (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 $v \in \mathbf{C}^\omega$,

$$f(w) \leq \max\{f(u), f(v)\} \ . \tag{9}$$

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 11.** *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 5. Necessary and sufficient conditions for positionality are known for deterministic games [12].

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.

4.1 Proof of Half-Positionality

We prove Theorem 11. 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 11 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 11 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:

$$\mathrm{minmax}(\mathbf{G})(s) \leq \max\{\mathrm{val}_0(s), \mathrm{val}_1(s)\} \ , \tag{10}$$

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 (10) 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 (10) is enough to prove the inductive step.

4.2 The projection mapping

To prove (10), we make use of two mappings

$$\pi_0 : s(\mathbf{AS})^\infty \to s(\mathbf{AS})^\infty \tag{11}$$

$$\pi_1 : s(\mathbf{AS})^\infty \to s(\mathbf{AS})^\infty. \tag{12}$$

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 $h = s_0a_0s_1a_1\cdots s_n \in s(\mathbf{AS})^*$ 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 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:

$$\Pi = S_0A_1S_1\cdots \ , \tag{13}$$

$$\Pi_0 = \pi_0(S_0A_1S_1\cdots) \ , \tag{14}$$

$$\Pi_1 = \pi_1(S_0A_1S_1\cdots) \ . \tag{15}$$

4.3 The trigger strategy

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

According to Theorem 7, 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 ,

$$\mathbb{E}_s^{\sigma,\tau^\sharp}\left[\ f \mid \Pi_0 \text{ is finite}\right] \leq \mathrm{val}_1(s) + \epsilon\ \ , \tag{16}$$

$$\mathbb{E}_s^{\sigma,\tau^\sharp}\left[\ f \mid \Pi_1 \text{ is finite}\right] \leq \mathrm{val}_0(s) + \epsilon\ \ , \tag{17}$$

$$\mathbb{E}_s^{\sigma,\tau^\sharp}\left[\ f \mid \Pi_0 \text{ and } \Pi_1 \text{ are both infinite}\right] \leq \max\{\mathrm{val}_0(s), \mathrm{val}_1(s)\} + \epsilon\ \ . \tag{18}$$

4.4 Proof of inequalities (16) and (17)

To prove inequality (16), we introduce the probability measure μ_1 on plays in $\mathbf{G}_1$ defined as:

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

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

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

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,

$$\mu_1(ha \mid h) = \begin{cases} \sigma_1'(h)(a) & \text{if } t \in S_1, \\ \tau_1^\sharp(h)(a) & \text{if } t \in S_2\ , \end{cases} \tag{19}$$

$$\mu_1(has \mid ha) = p(s|t,a). \tag{20}$$

As a consequence of the equalities (19) and (20), 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 (16). The proof of (17) is symmetrical.

4.5 Proof of inequality (18)

The proof of (18) 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 $\mathbf{G}_0$,

$$\mathbb{P}_s^{\sigma_0,\tau_0^\sharp}(E) \geq \mathbb{P}_s^{\sigma,\tau^\sharp}(\Pi_0 \text{ is infinite and } \Pi_0 \in E) \quad . \tag{21}$$

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

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

if $\mathbb{P}_s^{\sigma,\tau^\sharp}(h \prec \Pi_0) > 0$ and otherwise $\sigma_0(h)$ is chosen arbitrarily. We prove that (21) holds. Let $\mathcal{E}$ be the set of measurable events $E \subseteq (\mathbf{SA})^\omega$ in $\mathbf{G}_0$ such that (21) is satisfied. First, $\mathcal{E}$ contains all cylinders $h_0(\mathbf{SA})^\omega$ of $\mathbf{G}_0$ with $h_0 \in (\mathbf{SA})^*$ because:

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

where the first inequality can be proved by induction on the size of h_0, using the definition of 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 (21).

Second step to obtain (18) is to prove that for every strategy σ_0 in $\mathbf{G}_0$:

$$\mathbb{P}_s^{\sigma_0,\tau_0^\sharp}(f \leq \mathrm{val}_0(s) + \epsilon \mid s \text{ is reached infinitely often}) = 1 \quad . \tag{22}$$

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 $n \in \mathbb{N}$,

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

because $\tau_0^\sharp$ is ϵ-subgame-perfect. As a consequence $\mathbb{P}_s^{\sigma,\tau^\sharp}(f \leq \liminf_n \mathrm{val}_0(S_n) + \epsilon) = 1$ hence (22).

Now we come to the end of the proof of (18). Let σ_0 be a strategy in $\mathbf{G}_0$ such that (21) holds for every measurable event E in $\mathbf{G}_0$. According to (22),

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

For $i \in \{0, 1\}$ denote E_i and F_i the events:

$$E_i = \{\Pi_i \text{ is infinite and reaches } s \text{ infinitely often}\}, \tag{23}$$

$$F_i = E_i \wedge \{f(\Pi_i) \leq \mathrm{val}_i(s) + \epsilon\}. \tag{24}$$

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 (24).

According to (22) and the definition of $\tau^\sharp$,

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

and together with (21),

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

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 (18).

Since equations (16), (17) and (18) 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 11.

5 Applications

5.1 Unification of classical results

The existence of deterministic stationary optimal strategies in Markov decision processes with parity [7], limsup, liminf [20], mean-payoff [19, 23, 1, 25] or discounted payoff functions [24] is well-known. Theorem 11 provides a unified proof of these five results, as a corollary of the following proposition.

▶ Proposition 2. 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 [13, 10].

▶ **Corollary 12.** *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 2 and Theorem 11. 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 [24]. Details can be found in [13, 10]. ◀

Corollary 12 unifies and simplifies existing proofs of [7] for the parity game and [20] 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 [9] 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 [19]. For one-player games, Bierth [1] 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 [23] 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 [10, 15].

5.2 Variants of mean-payoff games

The positive average condition f_{posavg} is defined for a sequence of rewards $r_0, r_1, \ldots$, as the function that maps to 1 if $f_{\text{mean}}(r_0 r_1 \cdots) > 0$, and otherwise it maps to 0. It is a variant of the condition in 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 $1, 1, 1, \ldots$ for sure rather than seeing with equal probability $\frac{1}{2}$ the sequences $0, 0, 0, \ldots$ or $2, 2, 2, \ldots$ while a mean-value maximizer prefers the second situation to the first one.

To the best knowledge of the author, the techniques used in [1, 23, 25] 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.

5.3 New examples of positional payoff function

Although the generalized mean-payoff condition defined by (7) 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.

Other examples are provided in [13, 18, 10].

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

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

2 Tomás Brázdil, Václav Brozek, and Kousha Etessami. One-counter stochastic games. In *Proc. of FSTTCS'10*, pages 108–119, 2010.

3 K. Chatterejee, M. Jurdziński, and T.A. Henzinger. Quantitative stochastic parity games. In *proc. of SODA'03*, 2003.

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

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

6 Krishnendu Chatterjee, Laurent Doyen, Thomas A. Henzinger, and Jean-François Raskin. Generalized mean-payoff and energy games. In *Proc. of FSTTCS'10*, pages 505–516, 2010.

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

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

9 D. Gilette. Stochastic games with zero stop probabilities. *Contribution to the Theory of Games, Annals of Mathematics Studies*, 3, 1957.

10 H. Gimbert. *Jeux Positionnels*. PhD thesis, Université Denis Diderot, Paris, 2006.

11 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.

12 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.

13 Hugo Gimbert. Pure stationary optimal strategies in markov decision processes. In *Proceedings of STACS'07*, pages 200–211, 2007.

14 Hugo Gimbert and Edon Kelmendi. Two-player perfect-information shift-invariant submixing stochastic games are half-positional. *CoRR*, abs/1401.6575, 2014. URL: `http://arxiv.org/abs/1401.6575`.

15 Hugo Gimbert and Wieslaw Zielonka. Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games. In *HAL Archives ouvertes - http://hal.archives-ouvertes.fr/hal-00438359*, December 2009.

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

17 Eryk Kopczynski. Half-positional determinacy of infinite games. In *Proc. of ICALP'06*, pages 336–347, 2006.

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

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

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

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

22 Ayala Mashiah-Yaakovi. Correlated equilibria in stochastic games with borel measurable payoffs. *Dynamic Games and Applications*, 5(1):120–135, 2015. URL: `http://dx.doi.org/10.1007/s13235-014-0122-2`, `doi:10.1007/s13235-014-0122-2`.

23 A. Neyman and S. Sorin. *Stochastic games and applications*. Kluwer Academic Publishers, 2003.

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

25 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.

26 Wiesiek Zielonka. private communication, September 2011.

27 Wiesław Zielonka. Perfect-information stochastic parity games. In *Proc. of FOSSACS'04*, volume 2987 of *Lecture Notes in Computer Science*, pages 499–513. Springer, 2004.

28 Wieslaw Zielonka. Playing in stochastic environment: from multi-armed bandits to two-player games. In *Proc. of FSTTCS'10*, pages 65–72, 2010.