ORIGINAL PAPER

Submixing and shift-invariant stochastic games

Hugo Gimbert[1] · Edon Kelmendi[2]

Abstract

We study optimal strategies in two-player stochastic games that are played on a finite graph, equipped with a general payoff function. The existence of optimal strategies that do not make use of memory and randomisation is a desirable property that vastly simplifies the algorithmic analysis of such games. Our main theorem gives a sufficient condition for the maximizer to possess such a simple optimal strategy. The condition is imposed on the payoff function, saying the payoff does not depend on any finite prefix (shift-invariant) and combining two trajectories does not give higher payoff than the payoff of the parts (submixing). The core technical property that enables the proof of the main theorem is that of the existence of ϵ-subgame-perfect strategies when the payoff function is shift-invariant. Furthermore, the same techniques can be used to prove a finite-memory transfer-type theorem: namely that for shift-invariant and submixing payoff functions, the existence of optimal finite-memory strategies in one-player games for the minimizer implies the existence of the same in two-player games. We show that numerous classical payoff functions are submixing and shift-invariant.

Keywords Stochastic games · Half positional · Epsilon subgame perfect strategies · Finite memory optimal strategies · Shift-invariant · Submixing

1 Introduction

The games that we study are played between two players on a finite graph. Every vertex of the graph belongs to one of the players, the one that decides which edge should be taken next. The result of such a play is an infinite path in the graph. The objective of the game is given using a payoff function, which maps infinite paths to

✉ Edon Kelmendi
  e.kelmendi@qmul.ac.uk

[1] CNRS, LaBRI, Université de Bordeaux, Bordeaux, France

[2] Queen Mary University of London, London, UK

real numbers. The maximizer or Player 1, wants to maximize the payoff, while his adversary (the minimizer) wants the opposite.

The study of such games has been an active area of research for a few decades, in a variety communities; especially in that of theoretical computer science and economics. They are used to model simplified adversarial (zero-sum) situations. In computer science they are used in verifying properties of systems, but also as a very beneficial theoretical tool in logic and automata theory.

In this paper we consider *stochastic* games, a more general model where in every step, after an action is chosen, there is a probability distribution on the set of vertices according to which the next vertex is chosen. In this scenario, Player 1 wants to maximize the *expected payoff*, and his adversary to minimize it.

Well-known examples of games played on graphs are the discounted games, mean-payoff games, games equipped with the limsup payoff function and parity games. These four classes of games share a common property: both players have very simple optimal strategies, namely optimal strategies that are both deterministic and stationary. These are strategies that guarantee maximal expected payoff and choose actions deterministically (without randomisation) and this deterministic choice depends only on the current vertex (it does not use memory). When games admit such strategies for the maximizer they are called *half-positional*, when they admit such strategies for both players they are called *positional*. This property is highly desirable and it is often the starting point for further algorithmic analysis.

The broad purpose of the present paper is to study what is the common quality of games that makes it possible for them to admit deterministic and stationary optimal strategies.

1.1 Context

There have 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 (Shapley 1953). Derman showed the positionality of one-player games with expected mean-payoff reward, using an Abelian theorem and a reduction to discounted games (Derman 1962). Gilette extended Derman's result to two-player games (Gilette 1957) but his proof was found to be wrong and corrected by Liggett and Lippman (1969). The positionality of one-player parity games was addressed in Courcoubetis and Yannakakis (1990) and later on extended to two-player games in Chatterejee et al. (2003); Zielonka (2004). Counter games were extensively studied in Brázdil et al. (2010) and several examples of positional counter games are given. There are also several examples of one-player and two-player positional games in Gimbert (2007); Zielonka (2010). A whole zoology of half-positional games is presented in Kopczynski (2009) and another example is given by mean-payoff co-Büchi games (Chatterjee et al. 2005). The proofs of these various results are quite

heterogeneous, making it difficult to find a common property that explains why they are positional or half-positional.

Some effort has been made to better understand conditions that make games (half) positional, which has made apparent that payoff functions that are shift-invariant and submixing play a crucial role. Our contributions lie in this direction.

1.2 Contributions

The results of the present paper can be summarised as follows.

First, the main theorem says that a sufficient condition for the game to be half-positional is for the payoff function to be shift-invariant and submixing. We give an informal explanation of this condition. Payoff functions f map infinite paths of the graph

$$s_0 s_1 s_2 s_3 \cdots$$

to real numbers. A payoff function is *shift-invariant* if it does not depend on finite prefixes, in other words

$$f(p\, s_0 s_1 s_2 s_3 \cdots) = f(s_0 s_1 s_2 s_3 \cdots),$$

for any finite prefix p, *i.e.* we can shift the trajectory to the left without changing the payoff. A payoff function is *submixing* on the other hand, if for any two infinite paths

$$s_0 s_1 s_2 s_3 \cdots$$
$$t_0 t_1 t_2 t_3 \cdots$$

shuffling (or combining) them such as

$$s_0 s_1 s_2 s_3 s_4 \quad s_5 s_6 s_7 s_8 \cdots$$
$$t_0 t_1 \qquad t_2 t_3 t_4 t_5 t_6 t_7 t_8 \cdots$$

does not give better payoff, that is:

$$f(s_0 s_1 s_2 t_0 t_1 s_3 s_4 t_2 t_3 t_4 t_5 t_6 s_5 s_6 s_7 s_8 t_7 t_8 \cdots) \leq \max\{f(s_0 s_1 s_2 \cdots), f(t_0 t_1 t_2 \cdots)\}.$$

Theorem 1.1 *Games equipped with a payoff function that is shift-invariant and submixing are half-positional.*

As mentioned above, half-positional games are those where the *maximizer* has a simple kind of strategy that is optimal. There is nothing special about this player, if instead of the submixing condition, we define an "inverse" submixing condition, namely one that requires that the combined payoff is larger than the minimum of the parts, we would have an analogous theorem that proves the existence of simple optimal strategies for the *minimizer*. Furthermore there are payoff functions for which both versions of the submixing condition hold, and for these games the theorem

proves positionality. The conditions in the statement of the theorem are not necessary; we will provide examples and discuss this fact.

The conclusion of the theorem, however, cannot be made stronger in the following sense: There are examples of games equipped with shift-invariant and submixing payoff functions where the minimizer *does not* have simple optimal strategies. For instance, consider the game with only two states s, t both controlled by the minimizer, and actions such that the underlying graph is a complete directed graph. The minimizer wins the game if and only if both states are visited infinitely often, and furthermore the blocks of contiguous visits to s have unbounded length. In other words, if the outcome is:

$$\underbrace{s \cdots s}_{n_1 \text{ times}} \quad \underbrace{t \cdots t}_{m_1 \text{ times}} \quad \underbrace{s \cdots s}_{n_2 \text{ times}} \cdots ,$$

then the sequence $n_1, n_2, \ldots$ has to be unbounded for the minimizer to win. The payoff function described above is shift-invariant and submixing, yet the optimal strategy of the minimizer either has to use randomisation, or infinite memory to remember how long was the longest block of contiguous visits to s.

The proof of Theorem 1.1 is by induction on number of edges, it uses Lévy's 0-1 law, as well as the following crucial property of the games under consideration. Namely that games equipped with a payoff function that is both bounded and Borel-measurable admit ϵ-subgame-perfect strategies, for every $\epsilon > 0$. A proof of this fact can be found in Mashiah-Yaakovi (2015).

A second contribution comes as a corollary of the techniques developed for the main theorem. It is a transfer-type theorem that lifts the existence of optimal finite-memory strategies in one-player games (also known as Markov decision processes, or MDPs) to the same for two-player games.

Theorem 1.2 *Let f be a payoff function that is both shift-invariant and submixing. Assume that in all games equipped with f and fully controlled by the minimizer, the minimizer has optimal strategies with finite memory. Then the minimizer has the same in all games equipped with f.*

Furthermore this theorem is proved by effectively constructing the optimal strategy which calls the optimal strategies in one-player games, thereby showing how to make optimal strategies for the minimizer in case of submixing payoffs, and maximizer in case of inverse-submixing payoffs in two-player games, by reusing optimal strategies in Markov decision processes, or one-player games.

1.3 Related work

For one-player games it was proved by the first author that every one-player game equipped with a payoff function that is both shift-invariant and submixing is positional (Gimbert 2007). This result was successfully used in Brázdil et al. (2010) to prove positionality of counter games. A weaker form of this condition was presented

in Gimbert and Zielonka (2004) to prove positionality of deterministic games (*i.e.* games where transition probabilities are equal to 0 or 1, not stochastic). Kopczyn-ski proved that two-player deterministic games equipped with a shift-invariant and submixing payoff function that takes only two values is half-positional (Kopczynski 2006).

A result of Zielonka (2010) provides a necessary and sufficient condition for the positionality of one-player games. The condition is expressed in terms of the exist-ence of particular optimal strategies in multi-armed bandit games. When trying to prove the positionality for a particular payoff function, the condition in Zielonka (2010) is harder to check than the submixing property which is purely syntactic.

Some results on finite-memory determinacy have been obtained in Bouyer et al. (2020), with different requirements: the size of the memory should be independent from the arena, whereas in this paper we do not make such an assumption.

The pre-print version of this present paper (Gimbert and Kelmendi 2014) has already been used in a number of works, mostly pertaining to algorithmic game theory community. We mention the papers that we are aware of. In Chatterjee and Doyen (2016), Chatterjee and Doyen study payoff functions that are a conjunction of mean-payoff objectives, and prove that they are in co-NP for finite-memory strat-egies. They use Theorem 1.1; and for Theorem 4.1 they observe that in the spe-cial case of finite-memory strategies there is a simple combinatorial proof, which bypasses the use of martingale theory. In Basset et al. (2018) the authors consider arbitrary boolean combination of expected mean-payoff objectives and the main theorem of the present paper appears as Theorem 1, and is the starting point of their further algorithmic analysis. Games played on finite graphs where the infor-mation flow is perturbed by non-deterministic signalling delays are considered in Berwanger and van den Bogaard (2015), where submixing and shift-invariant payoff functions play a central rôle. Our result is also used by Mayr, Schewe, Totzke and Wojtczak on their proof of the fact that games with energy-parity objectives and almost-sure semantics lie in NP ∩ co-NP (Mayr et al. 2021).

1.4 Organisation of the paper

We fix the notation and give the relevant definitions in Sect. 2, where one can also find an overview of the proof. We give examples of shift-invariant and submixing payoff functions in Sect. 3, as well as show how the Theorem 1.1 can be used to recover numerous classical determinacy results. In Sect. 4, we define reset strate-gies as a method of obtaining ϵ-subgame-perfect strategies, which exist due to Theo-rem 4.1. The proof of the main theorem, Theorem 1.1, is given in Sect. 5, and that of the transfer theorem for finite-memory strategies, Theorem 1.2, in Sect. 6.

2 Preliminaries

The purpose of this section is to introduce the basic notions that we need about stochastic games with perfect information, that is the definitions of: games, payoff functions, strategies and values.

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 the *objectives* that the players want to reach.

We use the following notations throughout the paper. Let $\mathbf{S}$ be a finite set. The set of finite (respectively infinite) sequences on $\mathbf{S}$ is denoted $\mathbf{S}^*$ (respectively $\mathbf{S}^\omega$). A *probability distribution* on $\mathbf{S}$ is a function $\delta : \mathbf{S} \to [0, 1]$ such that $\sum_{s \in \mathbf{S}} \delta(s) = 1$. The set of probability distributions on $\mathbf{S}$, we denote by $\Delta(\mathbf{S})$.

Definition 2.1 (*Arena*) A stochastic arena with perfect information is a tuple:

$$\left(\mathbf{S}, \ \mathbf{S}_1, \ \mathbf{S}_2, \ \mathbf{A}, \ (\mathbf{A}(s))_{s \in \mathbf{S}}, \ p \right)$$

where

- $\mathbf{S}$ is a finite set of states (that is nodes of the graph) partitioned in two sets $(\mathbf{S}_1, \mathbf{S}_2)$,
- $\mathbf{A}$ is a finite set of actions,
- 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 $p : \mathbf{S} \times \mathbf{A} \to \Delta(\mathbf{S})$.

An arena is *fully controlled by the minimizer* if $\mathbf{A}(s)$ is a singleton for every $s \in \mathbf{S}_1$.

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 $\mathbf{S(AS)}^*$ which is the prefix of an infinite play.

With each infinite play is associated a payoff computed by a *payoff function*. Player 1 (the maximizer) wants to maximize the expected payoff while Player 2 (the minimizer) 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 2.2 (*Stochastic game with perfect information*) A stochastic game with perfect information is a pair

$$(\mathcal{A}, f)$$

where $\mathcal{A}$ is an arena and f a payoff function for the arena $\mathcal{A}$.

2.2 Strategies

A *strategy* in an arena $\mathcal{A}$ for Player 1 is a function

$$\sigma \; : \; (\mathbf{SA})^*\mathbf{S}_1 \to \Delta(\mathbf{A})$$

such that for any finite play $s_0 a_1 \cdots s_n$, and every action $a \in \mathbf{A}$, if $\sigma(s_0 a_1 \cdots s_n)(a) > 0$ then the action a belongs to $\mathbf{A}(s_n)$, *i.e.* the played action is available. Strategies for Player 2 are defined similarly and are typically denoted τ. General strategies can have infinite memory as well as randomise among the available actions at every step. We are interested in a very simple sub-class of strategies, namely those that do not use any memory, or randomisation.

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

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

A strategy σ is *stationary* if $\sigma(h)$ only depends on the last state of h. In other words σ is stationary if for every state $t \in \mathbf{S}_1$ and for every finite play $h = s_0 a_1 \cdots a_k t$,

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

Given an initial state $s \in \mathbf{S}$ and strategies σ and τ for players 1 and 2 respectively, the set of infinite plays that start at state 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 that start at s such that for every finite play h that ends in state t, for every action $a \in \mathbf{A}$ and state $r \in \mathbf{S}$,

$$\mathbb{P}_s^{\sigma,\tau}(ha \mid h) = \begin{cases} \sigma(h)(a) & \text{if } t \in \mathbf{S}_1, \\ \tau(h)(a) & \text{if } t \in \mathbf{S}_2, \end{cases} \qquad (1)$$

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

For $n \in \mathbb{N}$, we denote S_n and A_n the random variables defined by

$$S_n(s_0 a_1 s_1 \cdots) \stackrel{\text{def}}{=} s_n,$$
$$A_n(s_0 a_1 s_1 \cdots) \stackrel{\text{def}}{=} a_n.$$

2.3 Values and optimal strategies

Let $\mathbf{G}$ be a game with a bounded measurable payoff function f. 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:

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

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

By definition of maxmin and minmax, for every state $s \in \mathbf{S}$, $\mathrm{maxmin}\,(\mathbf{G})(s) \leq \mathrm{minmax}\,(\mathbf{G})(s)$. As a corollary of the Martin's determinacy theorem for Blackwell games (Martin 1998, Section 1), the converse inequality holds as well:

Theorem 2.4 (Martin's second determinacy theorem, (Martin 1998, Section 1)) *Let $\mathbf{G}$ be a game with a Borel-measurable and bounded payoff function f. Then for every state $s \in \mathbf{S}$*:

$$\mathrm{val}\,(\mathbf{G})(s) \overset{\text{def}}{=} \mathrm{maxmin}\,(\mathbf{G})(s) = \mathrm{minmax}\,(\mathbf{G})(s).$$

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

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

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

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

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

A stronger class of ϵ-optimal strategies are ϵ-subgame-perfect strategies, which are strategies that are not only ϵ-optimal from the initial state s but stay ϵ-optimal throughout the game. More precisely, given a finite play $h = s_0 \cdots s_n$ and a function g whose domain is the set of (in)finite plays, by $g[h]$ we denote the function g *shifted* by h:

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

Definition 2.6 (*ε-Subgame-perfect strategy*) Let **G** be a game equipped with a pay-off function f. A strategy $\hat{\sigma}$ for Player 1 is said to be ϵ-subgame-perfect if for every finite play $h := s_0 \cdots s_n$,

$$\inf_\tau \mathbb{E}_{s_n}^{\sigma[h],\tau} \big[f[h] \big] \geq \sup_\sigma \inf_\tau \mathbb{E}_{s_n}^{\sigma,\tau} \big[f[h] \big] - \epsilon.$$

2.4 Shift-invariant and submixing

Without loss of generality we can assume that there is a finite set **C** (colours assigned to the states of the game) such that the payoff function f is a function

$$f \; : \; \mathbf{C}^\omega \to \mathbb{R},$$

that is Borel-measurable and bounded. We define the two conditions with respect to such payoff functions.

Definition 2.7 (*Shift-invariant*) The payoff function f is shift-invariant if and only if for all finite prefixes $p \in \mathbf{C}^*$ and trajectories $u \in \mathbf{C}^\omega$,

$$f(p\,u) = f(u).$$

Note that shift-invariance is a stronger condition than saying: if one can get $u' \in \mathbf{C}^\omega$ from $u \in \mathbf{C}^\omega$ by replacing finitely many letters then $f(u) = f(u')$. Sometimes in the literature this stronger condition is called "prefix-independent" or "tail-measurable". Intuitively shift-invariant payoff functions are such that they only measure asymptotic properties, and do not talk about indices.

A factorisation of $u \in \mathbf{C}^\omega$ is a sequence $u_1, u_2, \ldots$ of non-empty finite words (*i.e.* elements of $\mathbf{C}^+$) such that

$$u = u_1 u_2 u_3 \cdots .$$

For $u, v, w \in \mathbf{C}^\omega$, we say that w is a shuffle of u and v if there are respective factorisations $u_1, u_2, \ldots$, and $v_1, v_2, \ldots$ such that

$$w = u_1 \, v_1 \, u_2 \, v_2 \cdots .$$

Definition 2.8 (*Submixing*) The payoff function f is submixing if and only if for all $u, v, w \in \mathbf{C}^\omega$ such that w is a shuffle of u and v we have

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

The submixing condition says that one cannot shuffle two losing trajectories to make a winning one. This requirement simplifies the kind of strategies that the players need.

The submixing condition is not symmetric over the players, and it implies different results for different players (notice the difference between Theorems 1.1 and 1.2). We define the inverse-submixing condition which is its reflection about the players:

Definition 2.9 (*Inverse-submixing*) The payoff function f is inverse-submixing if and only if for all $u, v, w \in \mathbf{C}^\omega$ such that w is a shuffle of u and v we have

$$f(w) \geq \min\{f(u), f(v)\}.$$

There are payoff functions that are both submixing and inverse-submixing (*e.g.* the parity function); for such payoffs Theorem 1.1 implies simple optimal strategies for both players, *i.e.* positionality.

3 Applications and examples

In this section we give a variety of examples of payoff functions that are shift-invariant and submixing, some of them very well-known, others less so. Thus we unify a number of classical positional determinacy results and also sketch how straightforward it is to apply Theorem 1.1 to novel payoff functions. Furthermore, we comment on the hypothesis of Theorem 1.1: Are the conditions necessary? What do they imply about the optimal strategies of the minimizer? Under what operations is this class of payoff functions closed? We start by listing a few well-known examples.

3.1 Unification of classical results

The *mean-payoff function* has been introduced by Gilette (1957). 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:

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

The *discounted payoff* has been introduced by Shapley (1953). 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:

$$f_{disc}\big((r_0, \lambda_0)(r_1, \lambda_1)\cdots\big) \stackrel{def}{=} r_0 + \lambda_0 r_1 + \lambda_0 \lambda_1 r_2 + \cdots.$$

The *parity condition* is used in automata theory and logics (Grädel et al. 2002). Each state s is labeled with some color $c(s) \in \{0, \ldots, d\}$. The payoff is 1 if the highest color seen infinitely often is even, and 0 otherwise. For $c_0 c_1 \cdots \in \{0, \ldots, d\}^\omega$,

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

The *limsup payoff function* has been used in the theory of gambling games (Maitra and Sudderth 1996). States are labeled with immediate rewards and the payoff is the limit supremum of the rewards:

$$f_{lsup}(r_0 r_1 \cdots) \stackrel{def}{=} \limsup_n r_n.$$

The *liminf payoff function* can be defined similarly.

The two following propositions follow easily from Theorem 1.1.

Proposition 3.1 *The payoff functions f_{lsup}, f_{linf}, f_{par} and f_{mean} are shift-invariant and submixing. Moreover f_{lsup}, f_{linf}, and f_{par} are inverse-submixing as well.*

Proposition 3.2 *In every two-player stochastic game equipped with the parity, limsup, liminf, mean or discounted payoff function, Player 1 has a deterministic and stationary strategy which is optimal. The same is true for Player 2 for the parity, limsup and liminf payoff.*

One comment should be made about the discounted payoff function: while it is not shift-invariant, it is possible to reduce games equipped with this function to games with the mean-payoff function, by interpreting discount factors as stopping probabilities as was done in the seminal paper of Shapley (1953).

Thus we have unified a number of classical results, thereby giving a common reason for the half-positionality of seemingly unrelated games. The approaches that can be found in the literature for proving that these games are (half-)positional are diverse, as one can see, for example, by consulting the papers (Courcoubetis and Yannakakis 1990) and Maitra and Sudderth (1996) that show positionality for parity games and limsup games, respectively. The existence of deterministic and stationary optimal strategies in mean-payoff games has a colourful history attached. The first proof was given by Gilette (1957) 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 (Liggett and Lippman 1969). For one-player games, Bierth (1987) gave a proof using martingales and elementary linear algebra while (Vrieze et al. 1983) provided a proof based on linear programming and a modern proof can be found in Neyman and Sorin (2003) based on a reduction to discounted games and the use of analytical tools.

For two-player games, a proof based on a transfer theorem from one-player to two-player games can be found in Gimbert (2006); Gimbert and Zielonka (2009, 2016).

3.2 Other examples

We mention a few more recent examples of games.

One-counter stochastic games have been introduced in Brázdil et al. (2010), in these games each state $s \in \mathbf{S}$ is labeled by a relative integer $c(s) \in \mathbb{Z}$. Three different winning conditions were defined and studied in Brázdil et al. (2010):

$$\limsup_n \sum_{0 \leq i \leq n} c_i = +\infty \tag{3}$$

$$\limsup_n \sum_{0 \leq i \leq n} c_i = -\infty \tag{4}$$

$$f_{\text{mean}}(c_0 c_1 \dots) > 0 \tag{5}$$

The winning conditions given by (3) and (4) are clearly shift-invariant, it is furthermore plain that they are also submixing. The *positive average condition* defined by (5) is a variant of mean-payoff payoff, which may be more suitable to model quality of service constraints or decision makers with a loss aversion. One can naturally define a payoff function f_{posavg}, that outputs 1 if the condition holds, and 0 otherwise.

Although f_{posavg} seems similar to the f_{mean} function, maximizing the expected value of f_{posavg} and doing the same for f_{mean}, are two different goals. For example, a positive average maximizer prefers seeing the sequence $1, 1, 1, \dots$ for sure rather than seeing with equal probability $\frac{1}{2}$ the sequences $0, 0, 0, \dots$ or $3, 3, 3, \dots$ while a mean-value maximizer prefers the second situation to the first one. To the best knowledge of the authors, the classical techniques developed in Bierth (1987); Neyman and Sorin (2003); Vrieze et al. (1983) cannot be used to prove positionality of games equipped with the positive average condition. However, since f_{posavg} can be defined as the composition of the submixing function f_{mean} with an increasing function, it is submixing itself. As a consequence of the main theorem of the present paper, it then follows that games that are equipped with f_{posavg} are half-positional.

Further relatively recent examples can be derived as variants of generalized mean payoff games, that were introduced in Chatterjee et al. (2010). Let us explain them in turn. In *mean-payoff co-Büchi games*, the states are labeled by immediate rewards, and a distinguished subset of the states are called the Büchi states. The payoff of Player 1 is $-\infty$ if Büchi states are visited infinitely often and the mean-payoff value of the rewards otherwise. One can easily check that such a payoff mapping is shift-invariant and submixing. Although we do not explicitly handle payoff mappings that take infinite values, 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.

Another variant is that of *optimistic generalized mean-payoff games*. In these games, each state is labeled by a fixed number of immediate rewards $(r^{(1)}, \dots, r^{(k)})$, which define as many mean payoff conditions $(f_{\text{mean}}^1, \dots, f_{\text{mean}}^k)$. The winning condition is:

$$\exists i \quad 1 \leq i \leq k \quad \text{and} \quad f_{\text{mean}}^i\left(r_0^{(i)} r_1^{(i)} \dots\right) > 0. \tag{6}$$

It is an exercise to show that this winning condition is submixing. More generally, if $f_1, \dots, f_n$ are submixing payoff mappings then $\max\{f_1, \dots, f_n\}$ is submixing as well. As a consequence of this observation and Theorem 1.1, games with the optimistic generalized mean-payoff condition are half-positional. Such games are not positional however. One can show that the minimizer requires (finite) memory. Intuitively, he needs to use the memory to remember which dimensions have to be decreased, in order to render the condition false.

The generalized mean-payoff games of Chatterjee et al. (2010), (where the winning condition is as in (6), but with a universal instead of a existential quantifier), however are not submixing.

One final but interesting example of a payoff function that is shift-invariant, submixing, and even inverse-submixing (hence positional for both players in two-players games) is the *positive frequency payoff*. Every state is labeled by a color from a finite set C, each of which has a payoff $u(c)$. An infinite play generates an infinite word of colors:

$$w \stackrel{\text{def}}{=} c_0 c_1 c_2 \cdots,$$

For a color c and $n \in \mathbb{N}$ define $\#(c, c_0 c_1 \cdots c_n)$ to be the number of occurrences of the color c in the prefix $c_0 c_1 \cdots c_n$. The frequency of the color c in w is defined as:

$$\text{freq}(c, w) \stackrel{\text{def}}{=} \limsup_{n \to \infty} \frac{\#(c, c_0 c_1 \cdots c_n)}{n},$$

and the payoff

$$f_{\text{freq}}(w) \stackrel{\text{def}}{=} \max\{u(c) \ : \ c \in C, \ \text{freq}(c, w) > 0\}.$$

Other examples can be found in Gimbert (2007); Kopczynski (2009); Gimbert (2006), and in the papers cited in the introduction.

3.3 The class of shift-invariant and submixing functions

In this section we have already used two operators under which the class of shift-invariant and submixing functions is closed:

- If $f_1, \dots, f_k$ are shift-invariant and submixing then so is

$$f(w) \stackrel{\text{def}}{=} \max\{f_1(w), \ldots, f_k(w)\}.$$

- If f is shift-invariant and submixing, and g is an increasing function then

$$g \circ f$$

  is shift-invariant and submixing.

The proofs are routine.

The class of shift-invariant and submixing functions does not seem to have any non-trivial closure property. For example, even though this class is closed under max above, it is not closed under addition. That is if f_1 and f_2 are submixing, then $f(w) := f_1(w) + f_2(w)$ need not be. To see this, consider the example with colors a and b, and f_1 such that it maps to 1 if a occurs infinitely often, and 0 otherwise, and f_2 defined symmetrically.

Furthermore, neither condition is necessary in Theorem 1.1: discounted games are positional but not shift-invariant, and f_{mean} with lim inf instead of lim sup is positional but not submixing. However, as we have seen, this class contains many interesting payoff functions, and it is the salient property that allows one to prove the existence of positional optimal strategies. Perhaps even more importantly, it is typically trivial to check whether a given payoff function is shift-invariant and submixing.

4 ϵ-Subgame-perfect strategies

The proof of Theorem 1.1 hinges on a crucial property of games with perfect information, namely the fact that they admit ϵ-subgame-perfect strategies, for all $\epsilon > 0$.

Theorem 4.1 *Games equipped with a payoff function that is shift-invariant, bounded and Borel-measurable, admit ϵ-subgame-perfect strategies, for every $\epsilon > 0$.*

In this section we sketch the proof of this theorem, by giving a construction that takes some ϵ-optimal strategy σ and turns it into a 2ϵ-subgame-perfect strategy $\hat{\sigma}$. A full proof can be found in Mashiah-Yaakovi's paper (Mashiah-Yaakovi 2015, Proposition 11) in a more general setting, in the preprint version of the current paper (Gimbert and Kelmendi 2014, Section 3), as well as in Flesch et al. (2021).

The theorem is completely symmetric about the players, so it is sufficient to show that, say, the maximizer has ϵ-subgame-perfect strategies.

An ϵ-optimal strategy σ is a strategy that ensures (within ϵ) the largest payoff that it can be ensured from the starting position of the game. As the game progresses, if the adversary makes a non-optimal move at some point, that allows

one to gain more, an ϵ-optimal strategy does not necessarily take advantage of this slip up. The ϵ-subgame-perfect strategies are that subset of strategies that do take advantage of any such non-optimal action of the adversary. How does one turn an ϵ-optimal strategy to an ϵ-subgame-perfect strategy? One simple observation suffices: if after the non-optimal action is played, σ simply forgets the past, *i.e.* resets its memory, then the payoff will be larger. We explain why the strategy that always resets the memory when the adversary has played a non-optimal action is 2ϵ-subgame-perfect. First, let us make this construction precise.

A strategy σ is not 2ϵ-subgame-perfect if and only if there exists some finite play $h := s_0 \cdots s_n$ such that
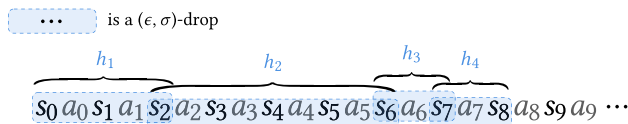
$$\inf_\tau \mathbb{E}_{s_n}^{\sigma[h],\tau}\big[f[h]\big] < \sup_{\sigma'} \inf_\tau \mathbb{E}_{s_n}^{\sigma',\tau}\big[f[h]\big] - 2\epsilon; \tag{7}$$

the reset strategy simply resets its memory when this happens. We give the formal definitions.

Definition 4.2 The finite play $h := s_0 \cdots s_n$ is called a (ϵ, σ)-*drop* if (7) holds. We write

$$\Delta(\epsilon, \sigma)(h) \qquad \Leftrightarrow \qquad h \text{ is a} (\epsilon, \sigma) - \text{drop}.$$

It is plain that one can factorise any infinite play that has infinitely many drops, into $h_1 h_2 \cdots$ where each h_i is a (ϵ, σ)-drop, but no strict prefix of h_i is (ϵ, σ)-drop. For example:

is a (ϵ, σ)-drop

$$\underbrace{s_0\, a_0\, s_1\, a_1\, s_2}_{h_1}\, a_2\, s_3\, a_3\, s_4\, a_4\, \underbrace{s_5\, a_5\, s_6}_{h_2}\, a_6\, \overbrace{s_7}^{h_3}\, a_7\, \overbrace{s_8}^{h_4}\, a_8\, s_9\, a_9 \cdots$$

Definition 4.3 We define the *date* of the most recent (or latest) drop for all $s_0 \cdots s_n$ inductively as:
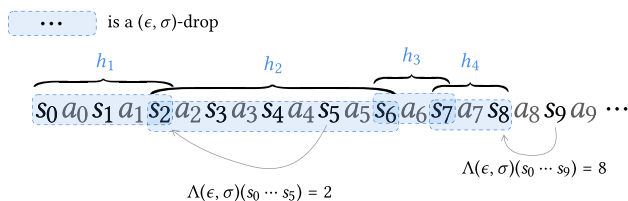
$$\Lambda(\epsilon, \sigma)(s_0) \stackrel{\text{def}}{=} 0$$

$$\Lambda(\epsilon, \sigma)(s_0 \cdots s_n) \stackrel{\text{def}}{=} \begin{cases} n & \text{if } h \text{ is a } (\epsilon, \sigma)\text{-drop} \\ \Lambda(\epsilon, \sigma)(s_0 \cdots s_{n-1}) & \text{otherwise}, \end{cases}$$

where

$$h \stackrel{\text{def}}{=} s_\ell \cdots s_n, \quad \text{and} \quad \ell \stackrel{\text{def}}{=} \Lambda(\epsilon, \sigma)(s_0 \cdots s_{n-1}).$$

The date of the most recent drop in the example above looks as follows:

The reset strategy resets its memory whenever a drop occurs, *i.e.* it keeps the memory since the most recent drop:

Definition 4.4 (*Reset Strategy*) For any strategy σ we define the reset strategy $\hat{\sigma}$ as:

$$\hat{\sigma}(s_0 \cdots s_n) = \sigma(s_\ell \cdots s_n),$$

where

$$\ell \overset{\text{def}}{=} \Lambda(\epsilon, \sigma)(s_0 \cdots s_n).$$

The crux of the proof of Theorem 4.1 is to show that under the strategy $\hat{\sigma}$ only finitely many resets occur, almost surely. Intuitively, this is because after every (ϵ, σ) -drop, by resetting the memory, the maximizer gains at least some amount δ that is bounded away from zero. But it is not possible to gain infinitely many times δ because the payoff function is assumed to be bounded.

One way of formally proving this observation is to use martingale theory. We start with a strategy σ that is ϵ-optimal. Then one shows that only finitely many drops occur almost surely with the strategy $\hat{\sigma}$, and that furthermore the strategy $\hat{\sigma}$ is itself ϵ-optimal. For the former, Doob's optional stopping theorem, and forward convergence theorem (Williams 1991, Theorem 11.5) are useful. Finally one proves that, by construction, if $\hat{\sigma}$ is ϵ-optimal, then it is also 2ϵ-subgame-perfect.

There is a technical detail which deserves some comments: to guarantee that $\hat{\sigma}$ is 2ϵ-subgame-perfect, we need the game to be value-preserving, in the following sense.

Definition 4.5 (*Value-preserving Game*) Let f be a shift-invariant payoff function. A game equipped with f is value-preserving for Player 1 if for every state s and every action a available in s,

$$\left(\sum_{t \in S} p(s, a, t) \operatorname{val}(\mathbf{G})(t) \right) \geq \operatorname{val}(\mathbf{G})(s). \tag{8}$$

The game is value-preserving for Player 2 if the converse inequality holds from every state of the game.

Since f is shift-invariant, inequality (8) holds when s is controlled by Player 2 (otherwise Player 2 could guarantee the expected payoff from s to be strictly smaller than the value of s). Any game equipped with f can be turned into a value-preserving game by deleting the actions available in Player 1 states in case they violate (8). This does not change the value of the state of the game, and moreover the ϵ-subgame perfect strategies of the new game are also ϵ-subgame perfect in the original game.

The ϵ-optimal strategies that we use, to turn into 2ϵ-subgame-perfect strategies $\hat{\sigma}$, are guaranteed to exist by Martin's theorem, Theorem 2.4. However, if in some game, one of the players, happens to have an optimal strategy (*i.e.* a 0-optimal strategy), then via the construction above one can prove the existence of a subgame-perfect strategy (*i.e.* a 0-subgame-perfect strategy).

Remark 4.6 In value-preserving games, if σ is ϵ-optimal, then the reset strategy $\hat{\sigma}$ is 2ϵ-subgame-perfect. Moreover, if σ is optimal, then $\hat{\sigma}$ is subgame-perfect.

Another property that is preserved by passing from σ to $\hat{\sigma}$ is that of finite memory, that is if the strategy σ has finite memory to begin with, so will the strategy $\hat{\sigma}$. First we define precisely what we mean by finite memory strategy.

A strategy σ is said to have *finite memory* if it is given using a transducer, namely it is a tuple:

$$\underbrace{\mathcal{M}}_{\text{a finite set}}, \quad \underbrace{\text{init} \;:\; \mathbf{S} \to \mathcal{M}}_{\text{memory initialiser}}, \quad \underbrace{\text{up} \;:\; \mathcal{M} \times \mathbf{A} \times \mathbf{S} \to \mathcal{M}}_{\text{update function}}, \quad \underbrace{\text{out} \;:\; \mathcal{M} \to \Delta(\mathbf{A})}_{\text{output function}}.$$

The map init and up are used to initialise the memory and update it, as the game unfolds: after the finite play $s_0 a_0 \cdots s_n$ has unfolded, the transducer reaches the memory state $m_n \in \mathcal{M}$ which is defined inductively as:

$$m_0 \overset{\text{def}}{=} \text{init}(s_0), \text{and}$$
$$m_{k+1} \overset{\text{def}}{=} \text{up}(m_k, a_{k+1}, s_{k+1}).$$

The output function is used to choose the action that the strategy plays, *i.e.*

$$\sigma(s_0 \cdots s_n) = \text{out}(m_n).$$

Proposition 4.7 *If σ is a finite memory strategy, so is $\hat{\sigma}$.*

Proof The reset strategy is constructed with respect to σ *and* some $\epsilon > 0$, since it depends on (ϵ, σ)-drops to reset the memory. We prove the proposition for any ϵ such that σ is ϵ-optimal.

Let σ be a finite memory strategy, that is given by the tuple

$$(\mathcal{M}, \text{init}, \text{up}, \text{out}),$$

and let ϵ be such that σ is ϵ-optimal, which fixes a reset strategy $\hat{\sigma}$.

Without loss of generality we can assume that the strategy is such that its memory state identifies the current state in the game, in other words assume that $\mathcal{M}$ can be partitioned into:

$$\mathcal{M} = \biguplus_{s \in \mathbf{S}} \mathcal{M}_s,$$

such that for any finite play $s_0 \cdots s_n$, if $m_1, \ldots, m_n$ is the sequence of memory states of the transducer of σ during this play, then

$$m_n \in \mathcal{M}_{s_n}.$$

We gather the subset of memory states where drops occur as follows. For $s \in \mathbf{S}$ and $m \in \mathcal{M}_s$, denote by σ_m the strategy that is the same as σ except that the initial memory state for s is m instead of $\text{init}(s)$. Define the subset of memory states where drops occur $\mathcal{D} \subset \mathcal{M}$ as

$$\mathcal{D} \stackrel{\text{def}}{=} \{m \in \mathcal{M}_s \; : \; s \in \mathbf{S} \text{ and } \sigma_m \text{is not } 2\epsilon - \text{optimal from state } s\}.$$

Construct the finite memory strategy σ' that avoids the memory states in $\mathcal{D}$ as follows. For any $s \in \mathbf{S}$ and $m \in \mathcal{M}_s \cap \mathcal{D}$, since σ is ϵ-optimal, $m \neq \text{init}(s)$. In the strategy σ' modify the function up in such a way that all the transitions that lead to m are redirected to the state $\text{init}(s)$ instead (the memory is reset). Do this simultaneously for any pair (s, m) as above. Comparing the definition of $\hat{\sigma}$ and σ' we conclude that they coincide.   □

This proposition, together with the salient property of the reset strategy, namely that it is ϵ-subgame-perfect, implies that if there are ϵ-optimal strategies with finite memory, then there are also *ϵ-subgame-perfect* strategies with finite memory. We make this statement more precise.

Proposition 4.8 *Let $\mathbf{G}$ be a game and $\epsilon > 0$. Assume that the game is value-preserving for Player 1 and that Player 1 has an ϵ-optimal strategy σ with finite memory. Then Player 1 also has a 2ϵ-subgame-perfect strategy with finite memory, namely the reset strategy $\hat{\sigma}$. The same statement holds even if $\epsilon = 0$, i.e. for optimal strategies. The symmetric statements hold for Player 2.*

Proof From the discussion of this section the strategy $\hat{\sigma}$ is 2ϵ-subgame-perfect, and Proposition 4.7 implies that it has finite memory. The case $\epsilon = 0$ follows from Remark 4.6.   □

5 Half-positional games

We prove the main theorem:

Theorem 1.1 *Games equipped with a payoff function that is shift-invariant and submixing are half-positional.*

Neither of the conditions in the statement is necessary, as we saw from the examples given in Sect. 3. Necessary and sufficient conditions for positionality are known for deterministic games (Gimbert and Zielonka 2005). However the shift-invariant and submixing conditions are general enough to recover several known classical results, and to provide several new examples of games with deterministic stationary optimal strategies. Before we proceed with the proof we remark:

Corollary 5.1 *Games with payoff functions which are at the same time shift-invariant, submixing and inverse-submixing are positional.*

Proof A symmetric proof to that of Theorem 1.1, the subject of this section, can be used to prove a statement like that of Theorem 1.1, where Player 1 is replaced by Player 2 and submixing is replaced by inverse-submixing. □

Consider a game $\mathbf{G}$ fulfilling the conditions of the theorem. The proof proceeds by induction on the actions of the maximizer, that is on the quantity

$$N(\mathbf{G}) \stackrel{\text{def}}{=} \sum_{s \in \mathbf{S}_1} (|\mathbf{A}(s)| - 1).$$

It proceeds by removing more and more actions of the maximizer and showing that at every step the value has not decreased, until we are left with a single choice from every state that belongs to the maximizer. The unique choice will then be the positional optimal strategy.

If $N(\mathbf{G}) = 0$ there is no choice for the maximizer, hence he has a deterministic and stationary optimal strategy. If $N(\mathbf{G}) > 0$ there must be a state $\tilde{s} \in \mathbf{S}$ such that Player 1 has at least two actions in $\tilde{s}$, *i.e.* $\mathbf{A}(\tilde{s})$ has at least two elements. We split the game $\mathbf{G}$ in two strictly smaller subgames $\mathbf{G}_1$ and $\mathbf{G}_2$.

Definition 5.2 (*Split of a game*) Let $\mathbf{G}$ be a game with $N(\mathbf{G}) > 0$ and $\tilde{s} \in \mathbf{S}$ a state of $\mathbf{G}$ controlled by Player 1 in which there are at least two actions available, *i.e.* $\mathbf{A}(\tilde{s})$ has at least two elements. Partition $\mathbf{A}(\tilde{s})$ into two non-empty sets: $\mathbf{A}_1$ and $\mathbf{A}_2$. Let $\mathbf{G}_1$ and $\mathbf{G}_2$ be the games obtained from $\mathbf{G}$ by restricting the actions in the state $\tilde{s}$ to $\mathbf{A}_1$ and $\mathbf{A}_2$ respectively. Then $(\mathbf{G}_1, \mathbf{G}_2)$ is called a *split* of $\mathbf{G}$ on $\tilde{s}$.

The induction step relies on the two results stated in the next theorem. The first result says that the value of $\tilde{s}$ in the original game cannot be larger than that of the restricted games. The second result shows that Player 1 can play optimally in $\mathbf{G}$ by selecting one of the subgames and play optimally in it.

Theorem 5.3 *Let $\mathbf{G}$ be a game equipped with a payoff function that is shift-invariant and submixing. Let $(\mathbf{G}_1, \mathbf{G}_2)$ a split of $\mathbf{G}$ on $\tilde{s}$. Then*

$$\mathrm{val}\,(\mathbf{G})(\tilde{s}) = \max\{\ \mathrm{val}\,(\mathbf{G}_1)(\tilde{s}),\ \mathrm{val}\,(\mathbf{G}_2)(\tilde{s})\}. \tag{9}$$

Assume moreover that $\mathrm{val}\,(\mathbf{G}_1)(\tilde{s}) \geq\ \mathrm{val}\,(\mathbf{G}_2)(\tilde{s})$. *Then, for every* $s \in \mathbf{S}$,

$$\mathrm{val}\,(\mathbf{G})(s) =\ \mathrm{val}\,(\mathbf{G}_1)(s). \tag{10}$$

Theorem 1.1 is a simple corollary of Theorem 5.3.

Proof of Theorem 1.1 The proof is by induction on $N(G)$. If $N(\mathbf{G}) = 0$ there is no choice for the maximizer, hence he has a deterministic and stationary optimal strategy. If $N(\mathbf{G}) > 0$ then we choose a split $(\mathbf{G}_1, \mathbf{G}_2)$ of $\mathbf{G}$ on a pivot state $\tilde{s}$. By symmetry, we can choose a split such that $\mathrm{val}\,(\mathbf{G}_1)(\tilde{s}) \geq\ \mathrm{val}\,(\mathbf{G}_2)(\tilde{s})$. Then, according to (10) in Theorem 5.3, a strategy for Player 1 which is optimal in $\mathbf{G}_1$ is also optimal in $\mathbf{G}$. By induction hypothesis, there exists a positional optimal strategy in $\mathbf{G}_1$, thus $\mathbf{G}$ is half-positional. $\qquad\square$

The rest of the section is dedicated to the proof of Theorem 5.3. We fix a game $\mathbf{G}$ and a split $(\mathbf{G}_1, \mathbf{G}_2)$ of $\mathbf{G}$ on the state $\tilde{s}$. The inequality

$$\mathrm{val}\,(\mathbf{G})(\tilde{s}) \geq \max\{\ \mathrm{val}\,(\mathbf{G}_1)(\tilde{s}),\ \mathrm{val}\,(\mathbf{G}_2)(\tilde{s})\}$$

is clear, since Player 1 has more choice in $\mathbf{G}$ than he has in $\mathbf{G}_1$ and $\mathbf{G}_2$. We witness the converse inequality with a strategy for Player 2, called the *merge* strategy, which merges two ϵ-subgame-perfect strategies in the respective smaller games. This is done in Sect. 5.3. The definition of the merge strategy hinges on the projection of plays in the main game $\mathbf{G}$ to plays in the restricted games $\mathbf{G}_1$ and $\mathbf{G}_2$, which is done in Sect. 5.1. Then we analyse the two possible outcomes: (a) after some date the play remains only in game $\mathbf{G}_1$ (or only in game $\mathbf{G}_2$), (b) the play switches infinitely often between the two smaller games. This analysis is performed in Sects. 5.4 and 5.5. For the latter case (b) we use the submixing property to show that Player 1 cannot get a better payoff by switching between the two smaller games that he could get by staying in one of the subgames.

5.1 Projecting a play in G to a couple of plays in the subgames

There is a natural way to project a play h of the game $\mathbf{G}$ starting in $\tilde{s}$ to a couple of plays h_1 and h_2 in the restricted games $\mathbf{G}_1$ and $\mathbf{G}_2$ respectively, starting from $\tilde{s}$ as well. The two projections are computed simultaneously and inductively. Initially, $h = \tilde{s}$ and both projections h_1 and h_2 are also equal to $\tilde{s}$. Each step of the play in $\mathbf{G}$ is appended to either h_1 or h_2, depending on the action a played the last time the state $\tilde{s}$ was visited: if a belongs to $\mathbf{A}_1$ then the new step is appended to h_1, otherwise it is appended to h_2.

Before giving the formal definition of π_1 and π_2 below, we provide an example. Consider the game $\mathbf{G}$ with a single state s controlled by Player 1 (s is the pivot state $\tilde{s}$) and the set of actions $\mathbf{A} = \{a_1, a_2\}$ partitioned into $\mathbf{A}_1 = \{a_1\}$ and $\mathbf{A}_2 = \{a_2\}$. Then π_1 simply erases the loops on the action a_2: $\pi_1(s) = s$, $\pi_1(sa_2s) = s$, $\pi_1(sa_2sa_1s) = sa_1s$ and for every finite sequence of integers $k_0, \ldots, k_m$,

$\pi_1(s(a_2s)^{k_0}a_1s(a_2s)^{k_1} \dots a_1s(a_2s)^{k_m}) = s(a_1s)^m$. Symmetrically, π_2 erases the loops on the action a_2.

Formally, we define two maps π_1, π_2 from finite plays in $\mathbf{G}$ starting from $\tilde{s}$ to finite plays in $\mathbf{G}_1$ and $\mathbf{G}_2$ respectively, starting from $\tilde{s}$ as well. Let $h = s_0 a_0 s_1 \dots s_n$ be a finite play in $\mathbf{G}$ starting in $\tilde{s}$ and *has* a continuation of h in $\mathbf{G}$, with one more transition (s_n, a, s). Let last (has) be the action played in *has* after the last visit to $\tilde{s}$ i.e.

$$\text{last}\,(has) = a_{\max\{j \in 0 \dots n \mid s_j = \tilde{s}\}} = \begin{cases} a & \text{if } s_n = \tilde{s} \\ \text{last}\,(h) & \text{otherwise.} \end{cases}$$

Then

$$\pi_1(has) = \begin{cases} \pi_1(h)as & \text{if } \text{last}\,(has) \in \mathbf{A}_1 \\ \pi_1(h) & \text{if } \text{last}\,(has) \in \mathbf{A}_2. \end{cases}$$

And π_2 is defined symmetrically with respect to $\mathbf{A}_1$ and $\mathbf{A}_2$.

This definition can be extended to infinite plays in a natural way. Let $h = s_0 a_0 s_1 \dots$ be an infinite play in $\mathbf{G}$ starting in $\tilde{s}$. Then $\pi_1(h)$ is the limit of the sequence

$$\left(\pi_1(s_0 a_0 s_1 \dots s_n)\right)_{n \in \mathbb{N}}.$$

The projection $\pi_1(h)$ can be either finite or infinite, depending whether the play ultimately stays in $\mathbf{G}_2$ or not. If after some time the last action chosen in $\tilde{s}$ is always in $\mathbf{A}_2$, all subsequent moves in $\mathbf{G}$ are appended to the projection in $\mathbf{G}_2$, while the projection to $\mathbf{G}_1$ never gets updated and stays finite. In the single-state example provided above before the formal definition, $\pi_1(sa_1 sa_2 sa_1 sa_2 \dots)$ is the single possible infinite play $sa_1 sa_1 \dots$ in the subgame $\mathbf{G}_1$ while $\pi_1(sa_1 sa_2 sa_2 sa_2 \dots)$ is the finite play $sa_1 s$ in $\mathbf{G}_1$.

5.2 Linking the payoff in G to the payoffs in the subgames

The payoff in $\mathbf{G}$ can be related to the payoffs in the subgames $\mathbf{G}_1$ and $\mathbf{G}_2$. For that we introduce the following events in the game $\mathbf{G}$ (an event is a measurable subset of infinite plays). Recall that S_n and A_n are the random variables which output respectively the nth state s_n and action a_n when the play is $s_0 a_0 s_1 a_1 \cdots$.

$$\text{Stay}_{\geq n}(\mathbf{G}_2) \overset{\text{def}}{=} \{\forall m \geq n,\ \text{last}\,(S_0 A_0 \dots S_m A_m S_{m+1}) \in \mathbf{A}_2\}$$
$$\text{Stay}_{\omega}(\mathbf{G}_2) \overset{\text{def}}{=} \bigcup_{n \in \mathbb{N}} \text{Stay}_{\geq n}(\mathbf{G}_2).$$

If $\text{Stay}_{\geq n}(\mathbf{G}_2)$ holds, we say that the play *stays in* $\mathbf{G}_2$ *after step* n whereas if $\text{Stay}_{\omega}(\mathbf{G}_2)$ holds, we say that the play *ultimately stays in* $\mathbf{G}_2$.

Those two events can be described equivalently as a non-update of the projection to $\mathbf{G}_1$ after some point. For that, we make use of the random variables:

$$\Pi \stackrel{\text{def}}{=} S_0 A_0 S_1 \cdots \qquad \Pi_1 \stackrel{\text{def}}{=} \pi_1(S_0 A_0 S_1 \cdots), \qquad \Pi_2 \stackrel{\text{def}}{=} \pi_2(S_0 A_0 S_1 \cdots).$$

We see that Π is simply the identity map outputting the play in $\mathbf{G}$ while Π_i is essentially equivalent to π_i, it is a random variable that maps the infinite play in game $\mathbf{G}$ to its finite or infinite projection in game $\mathbf{G}_i$. Then

$$\text{Stay}_{\geq n}(\mathbf{G}_2) = \{\Pi_1 = \pi_1(S_0 A_1 \cdots S_n)\}$$
$$\text{Stay}_{\omega}(\mathbf{G}_2) = \{\Pi_1 \text{ is finite}\}.$$

The events $\text{Stay}_{\geq n}(\mathbf{G}_1)$ and $\text{Stay}_{\omega}(\mathbf{G}_1)$ are defined symmetrically. Define the event

$$\text{Switch} \stackrel{\text{def}}{=} \left(\neg \text{Stay}_{\omega}(\mathbf{G}_1) \wedge \neg \text{Stay}_{\omega}(\mathbf{G}_2)\right) = \{ \text{ both } \Pi_1 \text{ and } \Pi_2 \text{ are infinite } \}.$$

The following lemma shows that the payoff in $\mathbf{G}$ is tightly related to the payoffs in the subgames $\mathbf{G}_1$ and $\mathbf{G}_2$.

Lemma 5.4 *Let f be a shift-invariant and submixing payoff function. Every infinite play in $\mathbf{G}$ belongs to exactly one of the three events* $\{ \text{Stay}_{\omega}(\mathbf{G}_1), \text{Stay}_{\omega}(\mathbf{G}_2), \text{Switch} \}$. *Moreover,*

$$\text{if } \text{Stay}_{\omega}(\mathbf{G}_1) \text{ holds then } f(\Pi) = f(\Pi_1). \tag{11}$$

$$\text{If } \text{Stay}_{\omega}(\mathbf{G}_2) \text{ holds then } f(\Pi) = f(\Pi_2). \tag{12}$$

$$\text{If } \text{Switch} \text{ holds then } f(\Pi) \leq \max(f(\Pi_1), f(\Pi_2)). \tag{13}$$

Proof Since both projections in $\mathbf{G}_1$ and $\mathbf{G}_2$ cannot be finite at the same time then $(\text{Stay}_{\omega}(\mathbf{G}_1), \text{Stay}_{\omega}(\mathbf{G}_2), \text{Switch})$ is a partition of the infinite plays in $\mathbf{G}$. If Π_1 is finite then Π and Π_2 share an infinite suffix and the shift-invariance of f implies (11). The case where Π_2 is finite is symmetric, hence (12). If both Π_1 and Π_2 are infinite then the sequence of actions $(\text{last}(S_0 \ldots S_n A_n S_{n+1}))_{n \in \mathbb{N}}$ switches infinitely often between $\mathbf{A}_1$ and $\mathbf{A}_2$ thus $\tilde{s}$ is visited infinitely often. Moreover, in this case Π is a shuffle of Π_1 and Π_2 and since f is submixing, (13) follows. $\square$

5.3 The merge strategy

In light of Lemma 5.4, it is intuitively clear that to play well in $\mathbf{G}$, Player 2 has to play well in both subgames $\mathbf{G}_1$ and $\mathbf{G}_2$. Fix $\epsilon > 0$. The *merge strategy* for Player 2 is the composition of two strategies $\tau_1^{\sharp}$ and $\tau_2^{\sharp}$ for Player 2 in the subgames $\mathbf{G}_1$ and $\mathbf{G}_2$ respectively. We require $\tau_1^{\sharp}$ and $\tau_2^{\sharp}$ to be ϵ-subgame-perfect in the corresponding subgames; their existence is guaranteed by Theorem 4.1.

Definition 5.5 The merge strategy $\tau^{\sharp}$ is a strategy in $\mathbf{G}$ for Player 2 which ensures that Π_1 is consistent with $\tau_1^{\sharp}$ and Π_2 is consistent with $\tau_2^{\sharp}$ when the play starts from $\tilde{s}$. Let h be a finite play in $\mathbf{G}$ from $\tilde{s}$ and ending in a state controlled by Player 2, then

$$\tau^\sharp(h) = \begin{cases} \tau_1^\sharp(\pi_1(h)) & \text{if } \text{last}(h) \in \mathbf{A}_1, \\ \tau_2^\sharp(\pi_2(h)) & \text{if } \text{last}(h) \in \mathbf{A}_2. \end{cases}$$

The merge strategy is well-defined because in case $\text{last}(h) \in \mathbf{A}_i$, with $i \in \{1, 2\}$, then both h and $\pi_i(h)$ end with the same state, controlled by Player 2.

In the next two sections, we show that the merge strategy guarantees to Player 2 some upper-bounds on the expected payoffs, which reflect the bounds given in Lemma 5.4 for payoffs of individual plays.

5.4 On plays consistent with the merge strategy and ultimately staying in G_2

In this section, we show that in case the play ultimately stays in $\mathbf{G}_2$, then the expected payoff is upper-bounded by $\text{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon$.

For simplicity, we require ϵ to be small enough so that τ_2 does not select any *value-increasing action*, in the following sense.

Lemma 5.6 *In* $\mathbf{G}_2$, *fix a state s controlled by Player 2 and an action a available in that state. Denote*

$$\delta(s, a) = \left(\sum_{t \in S} p(s, a, t) \, \text{val}(\mathbf{G}_2)(t) \right) - \text{val}(\mathbf{G}_2)(s).$$

Then $\delta(s, a) \geq 0$.

In case $\delta(s, a) > 0$ *then a is said to be value-increasing in s. In that case, if more-over ϵ is strictly smaller that $\delta(s, a)$, then $\tau_2^\sharp$ never selects the action a in a play ending in state s.*

Proof Since the payoff function is shift-invariant, and s is controlled by Player 2, then $\delta(s, a) \geq 0$, because after Player 2 chooses a in s, he can proceed with an ϵ'-optimal strategy from the states t such that $p(s, a, t) > 0$, for an arbitrary $\epsilon' > 0$.

Assume ϵ strictly smaller that $\delta(s, a)$. Then $\tau_2^\sharp$ never selects a in state s, otherwise this would contradict the ϵ-subgame perfection of τ_2: Player 1 could proceed with some $(\delta(s, a) - \epsilon)/2$-optimal strategy in $\mathbf{G}_2$ and get an expected payoff strictly greater than $\text{val}(\mathbf{G}_2)(s) + \epsilon$. □

Lemma 5.7 *Assume that f is shift-invariant and ϵ is small enough to guarantee that $\tau_2^\sharp$ never selects any value-increasing action. Let σ be a strategy for Player 1 in* $\mathbf{G}$ *such that* $\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\left(\text{Stay}_\omega(\mathbf{G}_2) \right) > 0$. *Then*

$$\mathbb{E}_{\tilde{s}}^{\sigma, \tau^\sharp}\left[f \mid \text{Stay}_\omega(\mathbf{G}_2) \right] \leq \text{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon. \tag{14}$$

Proof The first ingredient of the proof is the sequence of random variables $(V_n)_{n \in \mathbb{N}}$, where V_n denotes the value in $\mathbf{G}_2$ of the last state of $\pi_2(S_0 A_1 \cdots S_n)$. Since the play starts in state $\tilde{s}$,

$$V_0 = \text{val}\,(\mathbf{G}_2)(\tilde{s}).$$

The value of V_n does not change unless the projection of the play to $\mathbf{G}_2$ via π_2 does. Since Π_2 is consistent with $\tau_2^\sharp$ and since $\tau_2^\sharp$ never selects any value-increasing action,

$$(V_n)_{n \in \mathbb{N}} \text{ is a super martingale.}$$

The second ingredient in the proof is a stopping time T, defined as follows. For every finite play $h = s_0 a_0 \ldots s_n$ in $\mathbf{G}$ starting in $\tilde{s}$ and consistent with σ and $\tau^\sharp$, denote

$$\phi(h) = \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\Big(\text{Stay}_{\geq n}(\mathbf{G}_2) \mid h \text{ is a prefix of the play}\Big),$$

in other words, $\phi(h)$ is the probability that, after the prefix h, the play stays forever in $\mathbf{G}_2$ i.e. π_1 is never updated anymore and stays equal to $\pi_1(h)$. Fix some $\epsilon' > 0$ and denote by T the stopping time

$$T = \min\left\{ n \in \mathbb{N} \mid \phi(S_0 A_0 \ldots S_n) \geq 1 - \epsilon' \right\},$$

with the usual convention $\min(\emptyset) = \infty$.

We use the event $\{T < \infty\}$ as an approximation of the event $\text{Stay}_\omega(\mathbf{G}_2)$ by proving

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\Big(\text{Stay}_\omega(\mathbf{G}_2) \mid T < \infty\Big) \geq 1 - \epsilon' \tag{15}$$

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\Big(T < \infty \mid \text{Stay}_\omega(\mathbf{G}_2)\Big) = 1. \tag{16}$$

The inequality (15) holds because by definition of ϕ, for every $n \in \mathbb{N}$,

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\Big(\text{Stay}_{\geq n}(\mathbf{G}_2) \mid T = n\Big) \geq 1 - \epsilon'.$$

We show (16). Fix $\epsilon'' > 0$. By definition of $\text{Stay}_\omega(\mathbf{G}_2)$, there exists $n_1 \in \mathbb{N}$ such that

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\Big(\text{Stay}_{\geq n_1}(\mathbf{G}_2) \mid \text{Stay}_\omega(\mathbf{G}_2)\Big) \geq 1 - \epsilon''. \tag{17}$$

According to Lévy's 0-1 law (see *e.g.* (Williams 1991, Theorem 14.4)), the sequence of random variables $\left(\mathbb{E}_{\tilde{s}}^{\sigma, \tau^\sharp}\big[\text{Stay}_{\geq n_1}(\mathbf{G}_2) \mid S_0, \ldots, S_n\big]\right)_{n \in \mathbb{N}}$ almost-surely converges to the indicator function $\mathbf{1}_{\text{Stay}_{\geq n_1}(\mathbf{G}_2)}$. Thus,

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\Big(\exists n_2 \geq n_1, \mathbb{E}_{\tilde{s}}^{\sigma, \tau^\sharp}\big[\text{Stay}_{\geq n_1}(\mathbf{G}_2) \mid S_0, \ldots, S_{n_2}\big] \geq 1 - \epsilon' \mid \text{Stay}_{\geq n_1}(\mathbf{G}_2)\Big) = 1.$$

Since $n_2 \geq n_1$ implies $\text{Stay}_{\geq n_2}(\mathbf{G}_2) \subseteq \text{Stay}_{\geq n_1}(\mathbf{G}_2)$,

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\Big(\exists n_2, \phi(S_0, \ldots, S_{n_2}) \geq 1 - \epsilon' \mid \text{Stay}_{\geq n_1}(\mathbf{G}_2)\Big) = 1.$$

Equivalently,

$$\mathbb{P}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big(T < \infty \mid \mathrm{Stay}_{\geq n_1}(\mathbf{G}_2)\big) = 1.$$

and with (17) we get

$$\mathbb{P}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big(T < \infty \mid \mathrm{Stay}_{\omega}(\mathbf{G}_2)\big) \geq 1 - \epsilon''.$$

This holds for every $\epsilon'' > 0$, hence (16).

Since $\epsilon' > 0$ can be chosen arbitrarily small, then according to (15) and (16), to show our goal (14), it is enough to establish:

$$\mathbb{E}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big[f \mid T < \infty\big] \leq \mathrm{val}\,(\mathbf{G}_2)(\tilde{s}) + \epsilon + 2\epsilon' \cdot ||f||_{\infty}. \tag{18}$$

This is well-defined, because (16) ensures $\mathbb{P}_{\tilde{s}}^{\sigma,\tau^{\sharp}}(T < \infty) \geq \mathbb{P}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big(\mathrm{Stay}_{\omega}(\mathbf{G}_2)\big) > 0$, and f is bounded.

Since $(V_n)_{n\in\mathbb{N}}$ is a bounded super martingale, we can deduce from Doob's Forward Convergence Theorem (Williams 1991, Theorem 11.5) that $(V_n)_{n\in\mathbb{N}}$ converges almost-surely. We denote V_T the random variable equal to $(\lim_n V_n)$ if $T = \infty$ and V_n if $T = n$.

We deduce (18) from the following three inequalities:

$$\mathbb{E}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big[V_T\big] \leq \mathrm{val}\,(\mathbf{G}_2)(\tilde{s}) \tag{19}$$

$$\mathbb{E}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big[V_T \mid T = \infty\big] = \mathrm{val}\,(\mathbf{G}_2)(\tilde{s}) \tag{20}$$

$$\mathbb{E}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big[f \mid T < \infty\big] \leq \mathbb{E}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big[V_T \mid T < \infty\big] + \epsilon + 2\epsilon' \cdot ||f||_{\infty}. \tag{21}$$

Assuming (19) and (20) do hold, then $\mathbb{E}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big[V_T \mid T < \infty\big] \leq \mathrm{val}\,(\mathbf{G}_2)(\tilde{s})$. Injecting this inequality in (21), we get (18), and the lemma is proved.

We prove the three inequalities (19)–(21). To obtain inequality (19) we apply Doob's Optional Stopping Theorem (Theorem 10.10 in Williams (1991)) to the bounded super-martingale $(V_n)_{n\in\mathbb{N}}$ and the stopping time $\min(T,k)$ for an arbitrary (large) $k \in \mathbb{N}$. This implies $\mathbb{E}_{\tilde{s}}^{\sigma,\tau^{\sharp}}[V_{\min(T,k)}] \leq V_0$. Then (19) follows by taking the limit of this inequality when $k \to \infty$ and using the equality $V_0 = \mathrm{val}\,(\mathbf{G}_2)(\tilde{s})$.

To prove (20), we prove an even stronger statement:

$$\mathbb{P}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big(V_T = \mathrm{val}\,(\mathbf{G}_2)(\tilde{s}) \mid T = \infty\big) = 1.$$

If $T = \infty$ then, according to (16), the event $\mathrm{Stay}_{\omega}(\mathbf{G}_2)$ does not hold. Thus, according to Lemma 5.4, either $\mathrm{Stay}_{\omega}(\mathbf{G}_1)$ or Switch holds. In the first case, Π_2 is ultimately constant equal to a play in $\mathbf{G}_2$ ending in $\tilde{s}$ and $(V_n)_{n\in\mathbb{N}}$ is ultimately constant equal to $\mathrm{val}\,(\mathbf{G}_2)(\tilde{s})$. In the second case, the play Π_2 visits $\tilde{s}$ infinitely often. Since $(V_n)_n$ converges almost-surely to V_T then $V_T = \mathrm{val}\,(\mathbf{G}_2)(\tilde{s})$.

Finally, we prove (21). Denote h_T the random variable defined when T is finite, which outputs the prefix of the play of length T, i.e.

$$h_T = S_0 A_0 \dots S_T,$$

and let h such that $\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}(h_T = h) > 0$. Denote t the last state of h. We modify the strategy $\sigma[h]$ in $\mathbf{G}$ to obtain a strategy σ_0 in $\mathbf{G}_2$, in the following way. Every finite play h_2 in $\mathbf{G}_2$ is also a finite play in $\mathbf{G}$. In case the last state of h_2 is the pivot state $\tilde{s}$, then the lottery $\sigma[h](h_2)$ selects an action which is either in $\mathbf{A}_2$ or $\mathbf{A}_1$. In the first case we say that $\sigma[h]$ keeps playing in $\mathbf{G}_2$ while in the second case we say that $\sigma[h]$ exits $\mathbf{G}_2$. The strategy σ_0 coincides with $\sigma[h]$ as long as it keeps playing in $\mathbf{G}_2$. Whenever $\sigma[h]$ exits $\mathbf{G}_2$ then σ_0 plays arbitrarily in $\mathbf{G}_2$, for the rest of the play. Then:

$$
\begin{aligned}
\mathbb{E}_{\tilde{s}}^{\sigma, \tau^\sharp}\left[f \mid h_T = h\right] &= \mathbb{E}_{\tilde{s}}^{\sigma, \tau^\sharp}\left[f \mid h \text{ is a prefix of the play}\right] \\
&= \mathbb{E}_t^{\sigma[h], \tau^\sharp[h]}\left[f\right] \\
&\leq \mathbb{E}_t^{\sigma_0, \tau^\sharp[h]}\left[f\right] + 2\epsilon' \cdot ||f||_\infty \\
&= \mathbb{E}_t^{\sigma_0, \tau_2^\sharp[\pi_2(h)]}\left[f\right] + 2\epsilon' \cdot ||f||_\infty \\
&\leq \mathrm{val}\,(\mathbf{G}_2)(t) + \epsilon + 2\epsilon' \cdot ||f||_\infty \\
&= \mathbb{E}_{\tilde{s}}^{\sigma, \tau^\sharp}\left[V_T \mid h_T = h\right] + \epsilon + 2\epsilon' \cdot ||f||_\infty.
\end{aligned}
$$

The first equality holds because $\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}(h_T = h) > 0$ thus, by definition of T, no strict prefix h' of h satisfies $\phi(h') \geq 1 - \epsilon$, hence h is a prefix of the play if and only if $h_T = h$. The second equality holds by shift-invariance of f. To prove the first inequality, we denote $X = \mathrm{Stay}_{\geq 0}(\mathbf{G}_2)$. Since $\sigma[h]$ and σ_0 coincide on X, and by definition of ϕ, $\mathbb{P}_t^{\sigma_0, \tau^\sharp[h]}(X) = \mathbb{P}_t^{\sigma[h], \tau^\sharp[h]}(X) = \phi(h) \geq 1 - \epsilon'$. As a consequence,

$$
\begin{aligned}
\mathbb{E}_t^{\sigma[h], \tau^\sharp[h]}\left[f\right] &\leq \mathbb{E}_t^{\sigma[h], \tau^\sharp[h]}\left[f 1_X\right] + \epsilon' ||f||_\infty \\
&= \mathbb{E}_t^{\sigma_0, \tau^\sharp[h]}\left[f 1_X\right] + \epsilon' ||f||_\infty \\
&\leq \mathbb{E}_t^{\sigma_0, \tau^\sharp[h]}\left[f\right] + 2\epsilon' ||f||_\infty.
\end{aligned}
$$

The third equality holds because σ_0 guarantees that the play stays in $\mathbf{G}_2$ (i.e. $\mathrm{Stay}_{\geq 0}(\mathbf{G}_2)$) and this implies that $\tau^\sharp[h]$ coincides with $\tau_2^\sharp[\pi_2(h)]$. The second inequality is by ϵ-subgame optimality of $\tau_2^\sharp$ in $\mathbf{G}_2$. The last equality holds by definition of V_T.

Since this holds for every possible value h of h_T when $T < \infty$, and there are at most countably many such values, the inequality (21) follows. $\qquad\square$

5.5 On plays consistent with the merge strategy and switching infinitely often between the two subgames

In this section, we provide an upper-bound on the payoff of plays which switch infinitely often between $\mathbf{G}_1$ and $\mathbf{G}_2$.

Lemma 5.8 *Assume that f is shift-invariant and submixing. For all strategies σ,*

$$\mathbb{P}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big(f \leq \max\{\,\mathrm{val}\,(\mathbf{G}_1)(\tilde{s}),\ \mathrm{val}\,(\mathbf{G}_2)(\tilde{s})\} + \epsilon \mid \mathrm{Switch}\,\big) = 1. \qquad (22)$$

Proof By definition of Switch, if Switch occurs then both projections Π_1 and Π_2 are infinite and visit $\tilde{s}$ infinitely often. According to the inequality (13) in Lemma 5.4, to prove (22) it is enough to show, for every $i \in \{1, 2\}$,

$$\mathbb{P}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big(f(\Pi_i) \leq \ \mathrm{val}\,(\mathbf{G}_i)(\tilde{s}) + \epsilon \mid \Pi_i \text{ is infinite and reaches } \tilde{s} \text{ infinitely often}\big) = 1. \qquad (23)$$

By symmetry, it is enough to show (23) when $i = 1$. For that, we define a strategy σ_1 in $\mathbf{G}_1$ such that for every measurable event $\mathcal{E}_1$ in the game $\mathbf{G}_1$,

$$\mathbb{P}_{\tilde{s}}^{\sigma_1,\tau_1^{\sharp}}\big(\mathcal{E}_1\big) \geq \mathbb{P}_{\tilde{s}}^{\sigma,\tau^{\sharp}}\big(\Pi_1 \text{ is infinite and } \Pi_1 \in \mathcal{E}_1\big). \qquad (24)$$

The definition of σ_1 is a cornerstone of the whole proof of Theorem 5.3 and to make this definition clear, we provide first before the formal definition of σ_1 an informal description of σ_1 and an example. Assume Player 1 has to choose the next action after a finite play h_1 in $\mathbf{G}_1$ between two actions a and b. Denote $p_a(h_1)$ (resp. $p_b(h_1)$) the probability that the projection Π_1 of the play from $\mathbf{G}$ on $\mathbf{G}_1$ admits $h_1 a$ (resp. $h_1 b$) as a prefix, when playing σ and $\tau^{\sharp}$. Then $\sigma_1(h_1)$ selects the action a with probability proportional to p_a i.e. equal to $p_a/(p_a + p_b)$, while b is selected with probability $p_b/(p_a + p_b)$.

As an example, consider a game with a single state s controlled by Player 1 (hence $\tau^{\sharp}$ is trivial) on which there are three loops on actions a, b, c which are partitioned into $\mathbf{A}_1 = \{a, b\}$ and $\mathbf{A}_2 = \{c\}$. Consider the strategy σ in $\mathbf{G}$ which plays the uniform lottery on $\{a, b, c\}$ for the first step, and then repeat the corresponding letter forever. The corresponding strategy σ_1 in $\mathbf{G}_1$ plays the uniform lottery on $\{a, b\}$ and then repeat the corresponding letter forever. Consider a more involved example in the same game but with a different strategy σ: this time σ initially plays the uniform lottery on $\{a, b, c\}$. If a (resp. b resp. c) is chosen, then letter b (resp. c resp. a) is repeated forever. What is the probability that σ_1 plays initially the action a? The letter a is a prefix of the projection Π_1 in $\mathbf{G}_1$ in exactly two cases: if in $\mathbf{G}$ either the letter a or the letter c is picked up first. Otherwise the projection to $\mathbf{G}_1$ starts with b. Hence $\sigma_1(s)(a) = 2/3$.

Formally, the definition of σ_1 relies on the prefix relation $\preceq$ and the *strict* prefix relation $\prec$ over finite or infinite plays. Given a finite play h_1 in $\mathbf{G}_1$, we consider the events $\{h_1 \preceq \Pi_1\}$ and $\{h_1 \prec \Pi_1\}$. The event $h_1 \prec \Pi_1$ means that not only h_1 appears as a prefix of the projection of the play on $\mathbf{G}_1$, but moreover at least one more action has been played in $\mathbf{G}_1$ after that, i.e. $\{h_1 \prec \Pi_1\} = \{\exists b \in \mathbf{A}, h_1 b \preceq \Pi_1\}$. The inclusion $\{h_1 \prec \Pi_1\} \subseteq \{h_1 \preceq \Pi_1\}$ implies

$$\mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 \prec \Pi_1\right) \leq \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 \preceq \Pi_1\right). \tag{25}$$

This inequality might be strict: for example if $\text{Stay}_{\geq 0}(\mathbf{G}_2)$ holds, i.e. if the play always stay in $\mathbf{G}_2$, then the event $\{\tilde{s} \prec \Pi_1\}$ has probability 0 while the event $\tilde{s} \preceq \Pi_1$ has probability 1.

The strategy σ_1 in $\mathbf{G}_1$ is defined as:

$$\sigma_1(h_1)(a) = \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 a \preceq \Pi_1 \mid h_1 \prec \Pi_1\right),$$

if $\mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 \prec \Pi_1\right) > 0$ and otherwise $\sigma_1(h_1)$ is chosen arbitrarily. Remark that in general, σ_1 is a mixed strategy.

We proceed with the proof of (24). Let $\mathfrak{E}$ be the set of measurable events $\mathcal{E}_1$ in $\mathbf{G}_1$ for which (24) holds. We prove first that $\mathfrak{E}$ contains all cylinders $h_1(\mathbf{SA})^\omega$ of $\mathbf{G}_1$, which relies on the following inequalities:

$$\begin{aligned}
\mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(\Pi_1 \text{ is infinite and } \Pi_1 \in h_1(\mathbf{SA})^\omega\right) &\leq \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 \preceq \Pi_1\right) \\
&\leq \mathbb{P}_{\tilde{s}}^{\sigma_1,\tau_1^\sharp}\left(h_1\right).
\end{aligned} \tag{26}$$

We abuse the notation and denote h_1 the event $\{h_1 \text{ is a prefix of the play}\}$. The first inequality is by definition of prefixes. Remark that this inequality might be strict, in case Π_1 is finite i.e. in case the play ultimately stays in $\mathbf{G}_2$. The second inequality (26) is proved by induction on the length of h_1. When h_1 is the single initial state $\tilde{s}$ then both terms in (26) are equal to 1, and the inequality is an equality. Let $h_1 ar$ be a finite play in $\mathbf{G}_1$ and assume that (26) holds for h_1. There are two cases, depending who controls the last state of h_1, denoted t. In case t is controlled by Player 1 then

$$\begin{aligned}
\mathbb{P}_{\tilde{s}}^{\sigma_1,\tau_1^\sharp}\left(h_1 ar\right) &= \mathbb{P}_{\tilde{s}}^{\sigma_1,\tau_1^\sharp}\left(h_1\right) \cdot \sigma_1(h_1)(a) \cdot p(t,a,r) \\
&\geq \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 \preceq \Pi_1\right) \cdot \sigma_1(h_1)(a) \cdot p(t,a,r) \\
&= \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 \preceq \Pi_1\right) \cdot \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 a \preceq \Pi_1 \mid h_1 \prec \Pi_1\right) \cdot p(t,a,r) \\
&\geq \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 \prec \Pi_1\right) \cdot \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 a \preceq \Pi_1 \mid h_1 \prec \Pi_1\right) \cdot p(t,a,r) \\
&= \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 a \preceq \Pi_1\right) \cdot p(t,a,r) \\
&= \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}\left(h_1 ar \preceq \Pi_1\right),
\end{aligned} \tag{27}$$

where the first and last equalities hold by definition of the probability measure, the first inequality by induction hypothesis and the second equality is by definition of σ_1.

The second inequality (27) holds because of the inclusion $\{h_1 \prec \Pi_1\} \subseteq \{h_1 \preceq \Pi_1\}$ already discussed previously [see (25)].

Now we prove the inequality (26), in case t is controlled by Player 2. For every finite play h'_1 in $\mathbf{G}_1$, denote $C(h'_1)$ the set of finite plays h' in $\mathbf{G}$ starting in $\tilde{s}$ and such that π_1 projects h' on h'_1, but no strict prefix of h' is projected on h'_1.

$$
\begin{aligned}
\mathbb{P}_{\tilde{s}}^{\sigma_1, \tau_1^\sharp}\big(h_1 ar\big) &= \mathbb{P}_{\tilde{s}}^{\sigma_1, \tau_1^\sharp}\big(h_1\big) \cdot \tau_1^\sharp(h_1)(a) \cdot p(t, a, r) \\
&\geq \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\big(h_1 \preceq \Pi_1\big) \cdot \tau_1^\sharp(h_1)(a) \cdot p(t, a, r) \\
&= \sum_{h' \in C(h_1)} \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\big(h'\big) \cdot \tau_1^\sharp(h_1)(a) \cdot p(t, a, r) \\
&= \sum_{h' \in C(h_1)} \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\big(h'\big) \cdot \tau^\sharp(h')(a) \cdot p(t, a, r) \\
&= \sum_{h' \in C(h_1)} \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\big(h' ar\big) \\
&\geq \sum_{h'' \in C(h_1 ar)} \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\big(h''\big) \\
&= \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\sharp}\big(h_1 ar \preceq \Pi_1\big).
\end{aligned}
\tag{28}
$$

The first equality is by definition of the probability measure. The first inequality is by induction hypothesis. The second equality holds because the event $h_1 \preceq \Pi_1$ is the disjoint union of the events $(h')_{h' \in C(h_1)}$: if the projection of an infinite play h to $\mathbf{G}_1$ starts with h_1, then there is a single prefix of this play in $C(h_1)$, this is the shortest (finite) prefix of h whose projection in $\mathbf{G}_1$ is h_1. The last equality holds by a similar argument. The third equality is by definition of $\tau^\sharp$, since $\forall h' \in C(h_1), \pi_1(h') = h_1$. The fourth equality is by definition of the probability measure. The second inequality (28) relies on the inclusion $C(h_1 ar) \subseteq \{h' ar \mid h' \in C(h_1)\}$ (the converse inclusion actually holds as well but there is no need to prove it). Take $h'' \in C(h_1 ar)$ and write $h'' = h' a' r'$, where a' and r' are the last action and state of h'' and h' is the remaining prefix. We prove first that $\pi_1(h') = h_1$. Denote b the action played in h'' after the last visit to $\tilde{s}$. Then b belongs to either $\mathbf{A}_1$ or $\mathbf{A}_2$ and by definition of π_1, in the first case $\pi_1(h'') = \pi_1(h') a' r'$ while in the second case $\pi_1(h'') = \pi_1(h')$. By minimality of h'' among plays projecting on $h_1 ar$, we can rule out the second case hence $\pi_1(h'') = \pi_1(h') a' r'$. Since $\pi_1(h'') = h_1 ar$ then $\pi_1(h') = h_1$ and $a = a'$ and $r = r'$. We prove by contradiction that $h' \in C(h_1)$. Assume otherwise and let $h'_\ell \in C(h_1)$ be the shortest prefix of h' which projects on h_1 as well and b, q an action and a state such that $h'_\ell bq \preceq h'$. Remark that the action after the last visit to $\tilde{s}$ in the play $h'_\ell bq$ is in $\mathbf{A}_2$ (because $\pi_1(h'_\ell bq) = \pi_1(h'_\ell)$) while in the play h'_ℓ it is in $\mathbf{A}_1$ (by minimality of h'_ℓ). As a consequence, this last visit of the play $h'_\ell bq$ to $\tilde{s}$ occurs at the end of h'_ℓ and $t = \tilde{s}$, a contradiction since t is controlled by Player 2. Thus $h' \in C(h_1)$. This holds for every $h'' \in C(h_1 ar)$, thus we have established the inclusion $C(h_1 ar) \subseteq \{h' ar \mid h' \in C(h_1)\}$ which completes the proof of the inequality (28).

This completes the proof of the inequality (26) in the second and last case, hence (24) holds when $\mathcal{E}_1$ is a cylinder.

Observe that $\mathfrak{E}$ is stable by finite disjoint unions, hence $\mathfrak{E}$ contains all finite disjoint unions of cylinders, which forms a boolean algebra. Moreover $\mathfrak{E}$ is a monotone class, so we can apply the monotone class theorem (see for example (Billingsley 2008, Theorem 3.4)). This implies that $\mathfrak{E}$ contains the sigma-field that is generated by cylinders, which by definition is the set of all measurable events in the game $\mathbf{G}_1$. This completes the proof of (24).

Next we prove that

$$\mathbb{P}_{\tilde{s}}^{\sigma_1, \tau_1^{\sharp}} \left(f \le \liminf_n \text{val} (\mathbf{G}_1)(S_n) + \epsilon \right) = 1. \tag{29}$$

Observe that due to the fact that $\tau_1^{\sharp}$ is ϵ-subgame-perfect and that f is shift-invariant, then for all $n \in \mathbb{N}$,

$$\mathbb{E}_{\tilde{s}}^{\sigma_1, \tau_1^{\sharp}} \left[f \mid S_0, A_0, \dots, S_n \right] = \mathbb{E}_{S_n}^{\sigma_1[S_0 \cdots S_n], \tau_1^{\sharp}[S_0 \cdots S_n]} \left[f \right] \le \text{val} (\mathbf{G}_1)(S_n) + \epsilon,$$

and as a consequence,

$$\liminf_n \mathbb{E}_{\tilde{s}}^{\sigma_1, \tau_1^{\sharp}} \left[f \mid S_0, A_0, \dots, S_n \right] \le \liminf_n \text{val} (\mathbf{G}_1)(S_n) + \epsilon. \tag{30}$$

According to Lévy's 0-1 law (see *e.g.* (Williams 1991, Theorem 14.4)), the sequence of random variables: $(\mathbb{E}_{\tilde{s}}^{\sigma_1, \tau_1^{\sharp}} \left[f \mid S_0, A_0, \dots, S_n \right])_{n \in \mathbb{N}}$ converges point-wise to the random variable $f(S_0 A_0 S_1 \cdots)$. As a consequence the left hand side of (30) is almost-surely equal to f and we get (29).

Denote $\mathcal{E}_1$ the event

$$\mathcal{E}_1 = \{ f > \text{val} (\mathbf{G}_1)(\tilde{s}) + \epsilon \text{ and } \tilde{s} \text{ is reached infinitely often} \}.$$

According to (29), $\mathbb{P}_{\tilde{s}}^{\sigma_1, \tau_1^{\sharp}} \left(\mathcal{E}_1 \right) = 0$. We apply (24) to $\mathcal{E}_1$ and get

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau^{\sharp}} \left(\Pi_1 \text{ is infinite and } \Pi_1 \in \mathcal{E}_1 \right) = 0.$$

By definition of $\mathcal{E}_1$, this last equality is equivalent to (23) with $i = 1$. The proof for the case $i = 2$ is symmetric. $\qquad\square$

5.6 Proof of Theorem 5.3

Proof of Theorem 5.3 To prove the first statement (9) in Theorem 5.3, we combine the two lemmas proved in the two previous sections in order to show:

$$\forall \sigma, \mathbb{E}_{\tilde{s}}^{\sigma, \tau^{\sharp}} \left[f \right] \le \max \{ \text{val} (\mathbf{G}_1)(\tilde{s}), \text{val} (\mathbf{G}_2)(\tilde{s}) \} + \epsilon. \tag{31}$$

The bound (31) can be obtained as follows. According to Lemma 5.4, the three events $\text{Stay}_\omega(\mathbf{G}_1)$, $\text{Stay}_\omega(\mathbf{G}_2)$ and Switch partition the set of infinite plays. In case $\text{Stay}_\omega(\mathbf{G}_1)$ occurs, Lemma 5.7 guarantees that the expected payoff is no more than $\text{val}(\mathbf{G}_1)(\tilde{s}) + \epsilon$. By symmetry, in case $\text{Stay}_\omega(\mathbf{G}_2)$ occurs, the expected payoff is no more than $\text{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon$. And in case Switch occurs, Lemma 5.8 guarantees that the payoff is almost-surely no more than $\max\{\text{val}(\mathbf{G}_1)(\tilde{s}),\ \text{val}(\mathbf{G}_2)(\tilde{s})\} + \epsilon$. Thus (31) holds. The inequality

$$\text{val}(\mathbf{G})(\tilde{s}) \geq \max\{\text{val}(\mathbf{G}_1)(\tilde{s}),\ \text{val}(\mathbf{G}_2)(\tilde{s})\}$$

is clear, since Player 1 has more choice in $\mathbf{G}$ than he has in $\mathbf{G}_1$ and $\mathbf{G}_2$. And ϵ can be chosen arbitrarily small in (31), hence the first statement (9) of Theorem 5.3.

We proceed with the second statement of Theorem 5.3. Assume that

$$\text{val}(\mathbf{G}_1)(\tilde{s}) \geq\ \text{val}(\mathbf{G}_2)(\tilde{s}). \tag{32}$$

We have to show (10), i.e.

$$\forall s \in \mathbf{S},\ \text{val}(\mathbf{G})(s) =\ \text{val}(\mathbf{G}_1)(s).$$

According to (9), we already know that this equality holds for $\tilde{s}$, and we shall extend it to all states $s \in \mathbf{S}$.

Recall that the merge strategy was defined only for plays that start in state $\tilde{s}$; we enlarge this definition, profiting from the assumption (32). First, extend the definition of $\text{last}(h)$ to any play h that has visited $\tilde{s}$ at least once, in which case $\text{last}(h)$ denotes the action that is played right after the last visit of h to $\tilde{s}$. Second, for all finite plays h that end in a state controlled by Player 2,

$$\tau^\sharp(h) \stackrel{\text{def}}{=} \begin{cases} \tau_1^\sharp(\pi_1(h)) & \text{if } h \text{ never visited } \tilde{s} \text{ or } \text{last}(h) \in \mathbf{A}_1 \\ \tau_2^\sharp(\pi_2(h)) & \text{if } h \text{ has visited } \tilde{s} \text{ at least once and } \text{last}(h) \in \mathbf{A}_2. \end{cases}$$

The merge strategy is well-defined because if h never visited $\tilde{s}$ or if $\text{last}(h) \in \mathbf{A}_1$ then both h and $\pi_1(h)$ end with the same state, controlled by Player 2. And if h has visited $\tilde{s}$ at least once and $\text{last}(h) \in \mathbf{A}_2$ then both h and $\pi_2(h)$ end with the same state, controlled by Player 2.

We prove that $\tau^\sharp$ guarantees a payoff smaller than $\text{val}(\mathbf{G}_1)(s) + \epsilon$ for every state s. Fix σ a strategy for Player 1 in $\mathbf{G}$, and define σ' to be the strategy that plays like σ as long as the play does not reach the pivot state $\tilde{s}$. As soon as the pivot state is reached, the strategy σ' switches definitively to a strategy $\sigma_1^\sharp$ that is optimal in the game $\mathbf{G}_1$, whose existence is guaranteed by the induction hypothesis. We shall prove that, for every $s \in S$,

$$\mathbb{E}_s^{\sigma,\tau^\sharp}[f] \leq \mathbb{E}_s^{\sigma',\tau^\sharp}[f] + \epsilon. \tag{33}$$

Since the strategies σ and σ' coincide on those plays that never reach $\tilde{s}$, to get (33) it is enough to prove that for every finite play h starting from s and reaching $\tilde{s}$ for the first time at the end of h,

$$\mathbb{E}_s^{\sigma,\tau^\sharp}\big[f \mid h\big] \leq \mathbb{E}_s^{\sigma',\tau^\sharp}\big[f \mid h\big] + \epsilon. \tag{34}$$

The inequality (34) holds because

$$\begin{aligned}
\mathbb{E}_s^{\sigma,\tau^\sharp}\big[f \mid h\big] &= \mathbb{E}_{\tilde{s}}^{\sigma[h],\tau^\sharp[h]}\big[f\big] \\
&\leq \text{val}\,(\mathbf{G}_1)(\tilde{s}) + \epsilon \\
&\leq \mathbb{E}_{\tilde{s}}^{\sigma_1^\sharp,\tau_1^\sharp[h]}\big[f\big] + \epsilon \\
&= \mathbb{E}_{\tilde{s}}^{\sigma'[h],\tau_1^\sharp[h]}\big[f\big] + \epsilon \\
&= \mathbb{E}_s^{\sigma',\tau^\sharp}\big[f \mid h\big] + \epsilon.
\end{aligned}$$

The first and third equalities hold because f is shift-invariant. The first inequality holds because the strategy $\tau^\sharp[h]$ is ϵ-optimal from state $\tilde{s}$, for the following reason. The strategy $\tau^\sharp[h]$ coincides with the strategy obtained by merging $\tau_1^\sharp[h]$ and $\tau_2^\sharp$ on the pivot state $\tilde{s}$, both of which are ϵ-subgame-perfect in the respective subgames. Since (31) was proved for any merge of two ϵ-subgame-perfect strategies, we can apply (31) to the strategy $\tau^\sharp[h]$, and conclude that the latter is ϵ-optimal from state $\tilde{s}$. The second inequality holds because $\sigma_1^\sharp$ is optimal in $\mathbf{G}_1$. The second equality holds because $\sigma'[h] = \sigma_1^\sharp$. Finally we have proved (33).

The plays consistent with σ' and $\tau^\sharp$ stay in the subgame $\mathbf{G}_1$. Since $\tau^\sharp$ coincides with $\tau_1^\sharp$ on plays staying in $\mathbf{G}_1$, and since $\tau_1^\sharp$ is ϵ-optimal in $\mathbf{G}_1$, we can write for all $s \in S$:

$$\mathbb{E}_s^{\sigma',\tau^\sharp}\big[f\big] = \mathbb{E}_s^{\sigma',\tau_1^\sharp}\big[f\big] \leq \text{val}\,(\mathbf{G}_1)(s) + \epsilon. \tag{35}$$

With (33) this shows that for all s,

$$\mathbb{E}_s^{\sigma,\tau^\sharp}\big[f\big] \leq \text{val}\,(\mathbf{G}_1)(s) + 2\epsilon. \tag{36}$$

This holds for every strategy σ and $\epsilon > 0$ arbitrarily small, thus $\text{val}\,(\mathbf{G})(s) \leq \text{val}\,(\mathbf{G}_1)(s)$, and moreover $\tau^\sharp$ is optimal in $\mathbf{G}$, from every state s. The converse inequality is obvious, because Player 1 has more freedom in $\mathbf{G}$ than in $\mathbf{G}_1$, hence the second statement (10) of Theorem 5.3. □

5.7 Remarks about the merge strategy

We observe a byproduct of inequality (36) in the proof of Theorem 5.3.

Observation 5.9 The merge strategy $\tau^\sharp$ constructed with ϵ-subgame-perfect pieces is 2ϵ-optimal in the game $\mathbf{G}$.

After this observation, since the merge strategy is obtained by merging two ϵ-subgame-perfect strategies, a natural question to ask is whether $\tau^\sharp$ is 2ϵ-subgame-perfect in the **G**? The answer is negative; consider the following simple example:
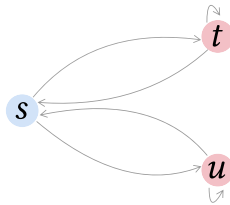
The goal of Player 1 is to visit the state t infinitely often (say that if he achieves this goal he receives a payoff 1, otherwise 0), and every action is deterministic. The blue state is controlled by Player 1, and the red ones by his opponent. In the subgame $\mathbf{G}_1$ we remove the action $s \to t$. In particular in the game $\mathbf{G}_1$ the positional strategy $\tau_1^\sharp$ which chooses $u \to s$ and $t \to s$ is subgame-perfect. We can therefore use it to construct a merge strategy $\tau^\sharp$. However this merge strategy is not 2ϵ-subgame-perfect, since in case Player 1 uses the sub-optimal action $s \to u$, his opponent does not profit by taking the self-loop forever.

6 Finite memory transfer theorem

The construction of the merge strategy in the previous section reveals that games that are equipped with shift-invariant and submixing payoffs have the following interesting property. While they yield very simple optimal strategies for Player 1, they allow his opponent to recombine strategies that work for one-player games (also known as Markov decision processes) and use them in a two-player game! We give the proof of this theorem that was announced in the "Introduction":

Theorem 1.2 *Let f be a payoff function that is both shift-invariant and submixing. Assume that in all games equipped with f and fully controlled by the minimizer, the minimizer has optimal strategies with finite memory. Then the minimizer has the same in all games equipped with f.*

We prove a slightly stronger theorem and derive Theorem 1.2 as a corollary. An arena is said to be *controlled by Player 2* if in every state that belongs to his opponent there is only one action available. (In other words these arenas are one-player games, or Markov decision processes).

Theorem 6.1 *Let f be a shift-invariant and submixing payoff function. If for all $\epsilon > 0$, Player 2 has an ϵ-optimal strategy with finite memory in every game controlled by*

himself, then in every (two-player) game he has an ϵ-subgame-perfect strategy that has finite memory.

The statement also holds for $\epsilon = 0$, that is: if Player 2 has a finite-memory optimal strategy in every game controlled by himself, then in every (two-player) game he has a subgame-perfect strategy with finite memory.

Proof The proof is by induction on actions of **G**, by defining the smaller games $\mathbf{G}_1$ and $\mathbf{G}_2$ as in the proof of the main theorem in the previous section.

We assume w.l.o.g. that **G** is value-preserving for Player 2, in the sense of Definition 4.5. If this is not the case initially then we simply remove from **G** the actions available in the states of Player 2 which are not value-preserving, which does not change the value of the states of the game. The ϵ-subgame-perfect strategies of Player 2 in the value-preserving game are ϵ-subgame-perfect in the original game.

The base of the induction follows from the assumption of the theorem, the induction hypothesis says that there are two ϵ-subgame perfect strategies $\tau_1^{\sharp}$ and $\tau_1^{\sharp}$ with finite memory, given by the transducers:

$$(\mathcal{M}_1, \text{init}_1, \text{up}_1, \text{out}_1) \qquad \text{and} \qquad (\mathcal{M}_2, \text{init}_2, \text{up}_2, \text{out}_2),$$

for Player 2 in $\mathbf{G}_1$ and $\mathbf{G}_2$ respectively. The strategy $\tau^{\sharp}$ obtained by merging $\tau_1^{\sharp}$ and $\tau_2^{\sharp}$ is also a finite-memory strategy, whose memory is

$$\mathcal{M} \stackrel{\text{def}}{=} \{1, 2\} \times \mathcal{M}_1 \times \mathcal{M}_2.$$

The initial memory state in state s is $(0, \text{init}_1(s), \text{init}_2(s))$. The updates on the components $\mathcal{M}_1$ and $\mathcal{M}_2$ are performed with up_1 and up_2 respectively. The first component is updated only when the play leaves the pivot state $\tilde{s}$; it is switched to 1 or 2 depending whether Player 1 chooses an action in $\mathbf{A}_1$ or $\mathbf{A}_2$. The choice of action, or the output, depends on the first component: in memory state (b, m_1, m_2) the action played by the strategy is $\text{out}_b(m_b)$.

According to Observation 5.9, $\tau^{\sharp}$ is 2ϵ-optimal. According to Proposition 4.8, there exists a finite-memory 4ϵ-subgame-perfect strategy in **G**. Since this holds for any $\epsilon > 0$, it concludes the induction step.

The second part of the theorem follows similarly thanks to Proposition 4.8. $\square$

6.1 On the size of the memory

How large is the memory $\mathcal{M}_{\mathbf{G}}$ needed by Player 2 to play optimally in some $\mathbf{G} = (\mathcal{A}, f)$? Every deterministic and stationary strategy $\sigma : S_1 \rightarrow A$ for Player 1 in **G** induces a game $\mathbf{G}_{\sigma}$ that is controlled by Player 2. Let $\mathfrak{M}$ be the maximal memory size required by Player 2 to play optimally in the games G_{σ}. According to the proof of the theorem above, the memory $\mathcal{M}_{\mathbf{G}}$ needed by Player 2 to play optimally in **G** is of size $2 \cdot |\mathcal{M}_{\mathbf{G}_1}| \cdot |\mathcal{M}_{\mathbf{G}_2}|$. We can iterate the construction on the same state $\tilde{s}$ and

partition $A(\tilde{s})$ until all actions sets are singletons. For every action $a \in A(\tilde{s})$ there is a subgame $\mathbf{G}_a$ where Player 1 always play action a when reaching $\tilde{s}$. Denote $\mathcal{M}_{\mathbf{G}_a}$ the memory needed Player 2 to play optimally in the subgame $\mathbf{G}_a$. To play optimally in $\mathbf{G}$, it is enough for Player 2 to remember which action in $A(\tilde{s})$ was chosen by Player 1 last time the play reached $\tilde{s}$, plus the memory states of its finite memory strategies in the subgames $\mathbf{G}_a, a \in A(\tilde{s})$, which can be implemented with the memory states

$$\mathcal{M}_{\mathbf{G}} = A(\tilde{s}) \times \prod_{a \in A(\tilde{s})} \mathcal{M}_{\mathbf{G}_a}.$$

Inductively, Player 2 can play optimally by remembering the last action choices in every state controlled by Player 1 as a mapping $\sigma : S_1 \to A$, such that $\forall s \in S_1, \sigma(s) \in A(s)$, and for every possible such mapping σ, the memory state in $1 \ldots \mathfrak{M}$ of an optimal strategy in the one-player game $\mathbf{G}_\sigma$. That leads to the following bound on the memory of Player 2. Enumerate S_1 as $s_1, s_2, \ldots, s_m$ then

$$|\mathcal{M}_{\mathbf{G}}| \leq |A(s_1)| \cdot |A(s_2)| \cdots |A(s_m)| \cdot |\mathfrak{M}^{A(s_1) \times A(s_2) \ldots \times A(s_m)}| \leq |A|^{|S_1|} \cdot \mathfrak{M}^{|A|^{|S_1|}}.$$

When $\mathfrak{M} = 1$, *i.e.* when Player 2 has deterministic and stationary strategies in games he controls, then in Gimbert and Zielonka (2005) it is shown that the same holds for two player games as well, hence the upper-bound can be downsized to 1 instead of $|A|^{|S_1|}$. In the general case where $\mathfrak{M} \geq 2$, we do not have examples where the memory size required by Player 2 to play optimally has the same order of magnitude as the upper bound above.

References

Basset N, Kwiatkowska M, Wiltsche C (2018) Compositional strategy synthesis for stochastic games with multiple objectives. Inf Comput 261:536 – 587 (2015) (strategic reasoning)

Berwanger D, van den Bogaard M (2015) Games with delays—a Frankenstein approach. In: Harsha P, Ramalingam G (eds) 35th IARCS annual conference on foundation of software technology and theoretical computer science, FSTTCS 2015, December 16–18, 2015, Bangalore. LIPIcs, vol 45, pp 307–319. Schloss Dagstuhl-Leibniz-Zentrum für Informatik

Bierth K-J (1987) An expected average reward criterion. Stoch Process Appl 26:133–140

Billingsley Patrick (2008) Probability and measure. Wiley, New York

Bouyer P, Le Roux S, Oualhadj Y, Randour M, Vandenhove P (2020) Games where you can play optimally with arena-independent finite memory. In: Konnov I, Kovács L (eds) 31st international

conference on concurrency theory, CONCUR 2020, September 1–4, 2020, Vienna, Austria (virtual conference). LIPIcs, vol 171, pp 24:1–24:22. Schloss Dagstuhl-Leibniz-Zentrum für Informatik

Brázdil Tomás, Brozek Václav, Etessami Kousha (2010) One-counter stochastic games. In FSTTCS, pp 108–119

Chatterjee K, Doyen L (2016) Perfect-information stochastic games with generalized mean-payoff objectives. In: Proceedings of the 31st annual ACM/IEEE symposium on logic in computer science, LICS '16, pp 247–256. Association for Computing Machinery, New York

Chatterjee K, Jurdziński M, Henzinger TA (2003) Quantitative stochastic parity games. In: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms, pp 121–130

Chatterjee K, Henzinger TA, Jurdzinski M (2005) Mean-payoff parity games. In: Proceedings of the LICS'05, pp 178–187. IEEE, New York

Chatterjee K, Doyen L, Henzinger TA, Raskin J-F (2010) Generalized mean-payoff and energy games. In: FSTTCS, pp 505–516

Courcoubetis C, Yannakakis M (1990) Markov decision processes and regular events. In Proceedings of ICALP'90. Lecture notes in computer science, vol 443, pp 336–349. Springer, New York

Derman Cyrus (1962) On sequential decisions and Markov chains. Manag Sci 9:16–24

Flesch János, Herings P, Maes Jasmine, Predtetchinski Arkadi (2021) Subgame maxmin strategies in zero-sum stochastic games with tolerance levels. Dyn Games Appl 11(4):704–737

Gilette D (1957) Stochastic games with zero stop probabilities, vol 3

Gimbert H (2006) Jeux positionnels. PhD thesis. Université Denis Diderot, Paris

Gimbert H (2007) Pure stationary optimal strategies in Markov decision processes. In STACS, pp 200–211

Gimbert H, Kelmendi E (2014) Two-player perfect-information shift-invariant submixing stochastic games are half-positional. CoRR: arXiv:1401.6575

Gimbert H, Zielonka W (2004) When can you play positionally? In: Proceedings of the MFCS'04. Lecture notes in computer science, vol 3153, pp 686–697. Springer, New York

Gimbert H, Zielonka W (2005) Games where you can play optimally without any memory. In: Proceedings of CONCUR'05. Lecture notes in computer science, vol 3653, pp 428–442. Springer, New York

Gimbert H, Zielonka W (2009) Pure and stationary optimal strategies in perfect-information stochastic games. In: HAL archives ouvertes, HAL-00438359, December

Gimbert H, Zielonka W (2016) Pure and stationary optimal strategies in perfect-information stochastic games with global preferences. CoRR: arXiv:1611.08487

Grädel E, Thomas W, Wilke T (2002) Automata, logics and infinite games. In: Lecture notes in computer science, vol 2500. Springer, New York

Kopczynski Eryk (2006) Half-positional determinacy of infinite games. ICALP 2:336–347

Kopczynski E (2009) Half-positional determinacy of infinite games. PhD thesis, University of Warsaw

Liggett TS, Lippman SA (1969) Stochastic games with perfect information and time average payoff. SIAM Rev 11(4):604–607

Martin DA (1998) The determinacy of Blackwell games. J Symbol Logic 63(4):1565–1581

Maitra AP, Sudderth WD (1996) Discrete gambling and stochastic games. Springer, New York

Mayr R, Schewe S, Totzke P, Wojtczak D (2021) Simple stochastic games with almost-sure energy-parity objectives are in NP and coNP. arXiv e-prints: arXiv:2101.06989

Mashiah-Yaakovi Ayala (2015) Correlated equilibria in stochastic games with Borel measurable payoffs. Dyn Games Appl 5(1):120–135

Neyman A, Sorin S (2003) Stochastic games and applications. Kluwer Academic Publishers, Dordrecht

Shapley LS (1953) Stochastic games. Proc Natl Acad Sci USA 39:1095–1100

Vrieze OJ, Tijs SH, Raghavan TES, Filar JA (1983) A finite algorithm for switching control stochastic games. O.R. Spektrum 5:15–24

Williams David (1991) Probability with martingales. Cambridge University Press, Cambridge

Zielonka W (2004) Perfect-information stochastic parity games. In: FOSSACS 2004. Lecture notes in computer science, vol 2987, pp 499–513. Springer, New York

Zielonka W (2010) Playing in stochastic environment: from multi-armed bandits to two-player games. In: FSTTCS, pp 65–72