



**HAL**  
open science

## Submixing and Shift-invariant Stochastic Games

Hugo Gimbert, Edon Kelmendi

► **To cite this version:**

Hugo Gimbert, Edon Kelmendi. Submixing and Shift-invariant Stochastic Games. 2014. hal-00936371v4

**HAL Id: hal-00936371**

**<https://hal.science/hal-00936371v4>**

Preprint submitted on 24 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Submixing and Shift-Invariant Stochastic Games

Hugo Gimbert<sup>1</sup> and Edon Kelmendi<sup>2</sup>

<sup>1</sup>CNRS, LaBRI, Université de Bordeaux, France

<sup>2</sup>Max Planck Institute for Software Systems, Germany

March 24, 2022

## 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  $\epsilon$ -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.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Preliminaries</b>	<b>6</b>
<b>3</b>	<b>Applications and Examples</b>	<b>10</b>
3.1	Unification of Classical Results . . . . .	10
3.2	Other Examples . . . . .	11
3.3	The Class of Shift-Invariant and Submixing Functions . . . . .	13
<b>4</b>	<b><math>\epsilon</math>-Subgame-Perfect Strategies</b>	<b>14</b>
4.1	Properties of the Reset Strategy . . . . .	17
4.1.1	Locally Optimal . . . . .	18
4.1.2	Finitely Many Drops . . . . .	21
4.1.3	$\epsilon$ -Optimal . . . . .	25
4.1.4	Remark on Optimal Strategies . . . . .	27
<b>5</b>	<b>Half-Positional Games</b>	<b>27</b>
5.1	Projecting a play in $G$ to a couple of plays in the subgames . . . . .	29
5.2	Linking the payoff in $G$ to the payoffs in the subgames . . . . .	30
5.3	The Merge Strategy . . . . .	31
5.4	On plays consistent with the merge strategy and ultimately staying in $G_2$	31
5.5	On plays consistent with the merge strategy and switching infinitely often between the two subgames . . . . .	34
5.6	Proof of Theorem 5.3 . . . . .	37
<b>6</b>	<b>From One-player Games to Two-player Games</b>	<b>40</b>
<b>7</b>	<b>The Finite Memory Transfer Theorem</b>	<b>41</b>

# 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 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 of 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.

**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 [Sha53]. Derman showed the positionality of one-player games with expected mean-payoff reward, using an Abelian theorem and a reduction to discounted games [Der62]. Gillette extended Derman's result to two-player games [Gil57] but his proof was found to be wrong and corrected by Liggett and Lippman [LL69]. The positionality of one-player parity games was addressed in [CY90] and later on extended to two-player games in [CJH03, Zie04]. Counter games were extensively studied in [BBE10] and several examples of positional counter games are given. There are also several examples of one-player and two-player positional games in [Gim07, Zie10]. A whole zoology of half-positional games is presented in [Kop09] and another example is given by mean-payoff co-Büchi games [CHJ05]. The proofs of these various results are 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.

**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

$$s_0 s_1 s_2 s_3 \dots$$

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

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

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

$$\begin{array}{l} s_0 s_1 s_2 s_3 \dots \\ t_0 t_1 t_2 t_3 \dots \end{array}$$

shuffling (or combining) them such as

$$\begin{array}{cccc} s_0 s_1 s_2 & s_3 s_4 & s_5 s_6 s_7 s_8 \dots & \\ t_0 t_1 & t_2 t_3 t_4 t_5 t_6 & t_7 t_8 \dots & \end{array}$$

does not give better payoff, that is:

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

**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 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  $\epsilon$ -subgame-perfect strategies, for every  $\epsilon > 0$ . A proof of this fact can be found in [MY15].

The second contribution says that having a shift-invariant payoff function is sufficient for the existence of  $\epsilon$ -subgame-perfect strategies.

**Theorem 1.2.** *Games equipped with a payoff function that is shift-invariant, for every  $\epsilon > 0$ , admit  $\epsilon$ -subgame-perfect strategies.*

The proof of this theorem uses martingale theory, and takes a large part of the paper, however it is independent of the rest.

A third 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) to the same for two-player games.

**Theorem 1.3.** *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, for every  $\epsilon > 0$ , the minimizer has an  $\epsilon$ -optimal strategy with finite memory. Then in every (two-player) game, for every  $\epsilon > 0$ , the minimizer has an  $\epsilon$ -subgame-perfect strategy that has finite memory.*

*The statement also holds for  $\epsilon = 0$ , that is: if the minimizer has an optimal strategy with finite memory in every game that he fully controls, then in every (two-player) game as well he has a subgame-perfect strategy with finite memory.*

Furthermore this theorem is proved by effectively constructing the  $\epsilon$ -subgame-perfect strategies in the two-player games. Those are obtained by combining and simplifying  $\epsilon$ -optimal strategies in one-player games.

A more general result about the transfer of simple class of strategies for the minimizer from one-player to two-player games is also formulated in Theorem 6.2.

**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 [Gim07]. This result was successfully used in [BBE10] to prove positionality of counter games. A weaker form of this condition was presented in [GZ04] to prove positionality of deterministic games (*i.e.* games where transition probabilities are equal to 0 or 1, not stochastic). Kopczynski proved that two-player deterministic games equipped with a shift-invariant and submixing payoff function that takes only two values is half-positional [Kop06].

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

Some results on finite-memory determinacy have been obtained in [BRO<sup>+</sup>20], 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 [GK14] has already been used in a number of works, mostly pertaining the algorithmic game theory community. We mention the papers that we are aware of. In [CD16], 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 strategies. They use Theorem 1.1; and for Theorem 1.2 they observe that in the special case of finite-memory strategies there is a simple

combinatorial proof, which bypasses the use of martingale theory. In [BKW18] 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 information flow is perturbed by non-deterministic signalling delays are considered in [BvdB15], where submixing and shift-invariant payoff functions play a central rôle. Our results and proof techniques were also used by Mayr, Schewe, Totzke and Wojtczak to establish a finite-memory transfer theorem analogous to the second part of Theorem 1.3 and to prove that games with energy-parity objectives and almost-sure semantics lie in  $\text{NP} \cap \text{co-NP}$  [MSTW21].

**Organisation of the paper.** We fix the notation and give the relevant definitions in Section 2, where one can also find an overview of the proof. We give examples of shift-invariant and submixing payoff functions in Section 3, as well as show how the Theorem 1.1 can be used to recover numerous classical determinacy results. In Section 4, we define reset strategies as a method of obtaining  $\epsilon$ -subgame-perfect strategies, which exist due to Theorem 1.2. The proof of the main theorem, Theorem 1.1, is given in Section 5, and that of the transfer theorem for finite-memory strategies, Theorem 1.3, in Section 7.

## 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.

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

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

$$(S, S_1, S_2, A, (A(s))_{s \in S}, p)$$

where

- $S$  is a finite set of states (that is nodes of the graph) partitioned in two sets  $(S_1, S_2)$ ,
- $A$  is a finite set of actions,
- for each state  $s \in S$ , a non-empty set  $A(s) \subseteq A$  of actions available in  $s$ ,
- and transition probabilities  $p : S \times A \rightarrow \Delta(S)$ .

An arena is *fully controlled by the minimizer* if  $A(s)$  is a singleton for every  $s \in S_1$ .

An *infinite play* in an arena  $\mathcal{A}$  is an infinite sequence  $p = s_0 a_1 s_1 a_2 \dots \in (SA)^\omega$  such that for every  $n \in \mathbb{N}$ ,  $a_{n+1} \in A(s_n)$ . A *finite play* in  $\mathcal{A}$  is a finite sequence in  $S(AS)^*$  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

$$f : (SA)^\omega \rightarrow \mathbb{R}$$

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  $\mathcal{A}$ .

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

$$\sigma : (SA)^* S_1 \rightarrow \Delta(A)$$

such that for any finite play  $s_0 a_1 \dots s_n$ , and every action  $a \in A$ , if  $\sigma(s_0 a_1 \dots s_n)(a) > 0$  then the action  $a$  belongs to  $A(s_n)$ , *i.e.* the played action is available. Strategies for Player 2 are defined similarly and are typically denoted  $\tau$ . 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  $\sigma$  for Player 1 is deterministic if for every finite play  $h \in (SA)^* S_1$  and action  $a \in A$ ,*

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

*A strategy  $\sigma$  is stationary if  $\sigma(h)$  only depends on the last state of  $h$ . In other words  $\sigma$  is stationary if for every state  $t \in S_1$  and for every finite play  $h = s_0 a_1 \dots a_k t$ ,*

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

Given an initial state  $s \in S$  and strategies  $\sigma$  and  $\tau$  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  $P_s^{\sigma, \tau}$  that are defined as follows. Given a finite play  $h$  and an action  $a$ , the set of infinite plays  $h(AS)^\omega$  and  $ha(SA)^\omega$  are *cylinders* that we abusively denote  $h$  and  $ha$ . The sigma-field is the one generated by cylinders and  $P_s^{\sigma, \tau}$



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  $r \in \mathbf{S}$ ,

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

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

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

$$S_n(s_0 a_1 s_1 \dots) \stackrel{\text{def}}{=} s_n,$$

$$A_n(s_0 a_1 s_1 \dots) \stackrel{\text{def}}{=} a_n.$$

**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  $\sigma$  and  $\tau$  is the expected value of  $f$  under  $\mathbb{P}_s^{\sigma, \tau}$ , denoted  $\mathbb{E}_s^{\sigma, \tau}[f]$ . The *maxmin* and *minmax* values of a state  $s \in \mathbf{S}$  in the game  $\mathbf{G}$  are:

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

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

By definition of maxmin and minmax, for every state  $s \in \mathbf{S}$ ,  $\text{maxmin}(\mathbf{G})(s) \leq \text{minmax}(\mathbf{G})(s)$ . As a corollary of the Martin's determinacy theorem for Blackwell games [Mar98, Section 1], the converse inequality holds as well:

**Theorem 2.4** (Martin's second determinacy theorem, [Mar98, Section 1]). *Let  $\mathbf{G}$  be a game with a Borel-measurable and bounded payoff function  $f$ . Then for every state  $s \in \mathbf{S}$ :*

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

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

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

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

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

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

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

$$g[h](t_0 a_1 t_1 \dots) \stackrel{\text{def}}{=} \begin{cases} g(ha_1 t_1 \dots) & \text{if } s_n = t_0, \\ g(t_0 a_1 t_1 \dots) & \text{otherwise.} \end{cases}$$

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

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

**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

$$f : C^\omega \rightarrow \mathbb{R},$$

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  $p \in C^*$  and trajectories  $u \in C^\omega$ ,*

$$f(pu) = f(u).$$

Note that shift-invariance is a stronger condition than saying: if one can get  $u' \in C^\omega$  from  $u \in C^\omega$  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  $u \in C^\omega$  is a sequence  $u_1, u_2, \dots$  of non-empty finite words (i.e. elements of  $C^+$ ) such that

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

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

$$w = u_1 v_1 u_2 v_2 \dots.$$

**Definition 2.8** (Submixing). *The payoff function  $f$  is submixing if and only if for all  $u, v, w \in C^\omega$  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 Theorem 1.1 and Theorem 1.3). 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  $u, v, w \in C^\omega$  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 Gillette [Gil57]. It measures average performances. Each state  $s \in S$  is labeled with an immediate reward  $r(s) \in \mathbb{R}$ . With an infinite play  $s_0 a_1 s_1 \dots$  is associated an infinite sequence of rewards  $r_0 = r(s_0), r_1 = r(s_1), \dots$  and the payoff is:

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

The *discounted payoff* has been introduced by Shapley [Sha53]. 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  $(r_0, \lambda_0)(r_1, \lambda_1) \dots \in (\mathbb{R} \times [0, 1])^\omega$  of daily payoffs and discount factors is associated the payoff:

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

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

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

The *limsup payoff function* has been used in the theory of gambling games [MS96]. States are labeled with immediate rewards and the payoff is the limit supremum of the

rewards:

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

The *liminf payoff function* can be defined similarly.

The two following propositions follow easily from Theorem 1.1.

**Proposition 3.1.** *The payoff functions  $f_{\text{lsup}}$ ,  $f_{\text{linf}}$ ,  $f_{\text{par}}$  and  $f_{\text{mean}}$  are shift-invariant and submixing. Moreover  $f_{\text{lsup}}$ ,  $f_{\text{linf}}$  and  $f_{\text{par}}$  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 [Sha53]. One can find details of this reduction in [Gim07, Gim06].

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 [CY90] and [MS96] 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 [Gil57] based on a variant of Hardy and Littlewood theorem. Later on, Ligget and Lippman found the variant to be wrong and proposed an alternative proof based on the existence of Blackwell optimal strategies plus a uniform boundedness result of Brown [LL69]. For one-player games, Bierth [Bie87] gave a proof using martingales and elementary linear algebra while [VTRF83] provided a proof based on linear programming and a modern proof can be found in [NS03] based on a reduction to discounted games and the use 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 [Gim06, GZ09, GZ16].

## 3.2 Other Examples

We mention a few more recent examples of games.

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

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

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

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

The *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  $f_{\text{posavg}}$ , that outputs 1 if the condition holds, and 0 otherwise.

Although  $f_{\text{posavg}}$  seems similar to the  $f_{\text{mean}}$  function, maximizing the expected value of  $f_{\text{posavg}}$  and doing the same for  $f_{\text{mean}}$ , are two different goals. For example, a positive average maximizer prefers seeing the sequence 1, 1, 1, ... for sure rather than seeing with equal probability  $\frac{1}{2}$  the sequences 0, 0, 0, ... or 3, 3, 3, ... 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 [Bie87, NS03, VTRF83] cannot be used to prove positionality of games equipped with the positive average condition. However, since  $f_{\text{posavg}}$  can be defined as the composition of the submixing function  $f_{\text{mean}}$  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  $f_{\text{posavg}}$  are half-positional.

Another recent example are the *generalized mean payoff games*, that were introduced in [CDHR10]. 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:

$$\forall 1 \leq i \leq k, f_{\text{mean}}^i(r_0^{(i)} r_1^{(i)} \dots) > 0. \quad (6)$$

In the special case of mean-payoff co-Büchi games, a subset of the states are called Büchi states, and the payoff of Player 1 is  $-\infty$  if Büchi states are visited infinitely often and the mean-payoff value of the rewards otherwise. 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.

The general payoffs captured by the condition in (6) are not submixing, however, a natural variant is: *Optimistic generalized mean-payoff games* are defined similarly except the winning condition is

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

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. There are even examples of shift-invariant and submixing payoff functions where the minimizer requires infinite memory to play optimally. Here is one of them.

The set of colours is  $\{a, b\}$ . The payoff function is equal to  $-1$  if and only if the word  $w \in \{a, b\}^\omega$  that it inputs contains infinitely many  $as$ , infinitely many  $bs$ , and

moreover

$$w = a^{n_1} b a^{n_2} b a^{n_3} b \dots,$$

is such that  $\liminf_{\ell} n_{\ell} = \infty$ , otherwise it is equal to 0.

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 set  $C$ , each of which has a payoff  $u(c)$ . An infinite play generates an infinite word of colors:

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

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

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

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 [Gim07, Kop09, Gim06], 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

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

- 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  $f_{\text{mean}}$  with  $\liminf$  instead of  $\limsup$  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 $\epsilon$ -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  $\epsilon$ -subgame-perfect strategies, for all  $\epsilon > 0$ .

**Theorem 1.2.** *Games equipped with a payoff function that is shift-invariant, for every  $\epsilon > 0$ , admit  $\epsilon$ -subgame-perfect strategies.*

Note that we cannot lift the shift-invariant hypothesis from Theorem 1.2. That is, one can easily find an example of a game where there are no  $\epsilon$ -subgame-perfect strategies, even a game with only one player.

Note that Theorem 1.2 is true for arbitrary payoff functions and the weaker notion of  $\epsilon$ -subgame-perfect strategy, requiring that for every finite play  $h = s_0 \dots s_n$ ,

$$\inf_{\tau} \mathbb{E}_{s_n}^{\sigma[h], \tau} [f[h]] \geq \sup_{\sigma'} \inf_{\tau} \mathbb{E}_{s_n}^{\sigma', \tau} [f[h]] - \epsilon. \quad (7)$$

Indeed, this was proved independently by Mashiah-Yaakovi, [MY15, Proposition 11] for concurrent games. That result implies Theorem 1.2, since for shift-invariant games, the condition (7) coincides with that of Definition 2.6.

The weak and strong notions of conditions coincide when the payoff function is shift-invariant. On the one hand, our proof only works for the strong notion of subgame-perfectness in Definition 2.6. On the other hand, our proof makes transparent how to construct  $\epsilon$ -subgame-perfect strategies from  $\epsilon/2$ -optimal ones, in a way that preserves some important properties of the strategy, notably its use of finite memory.

The proof of the theorem will be symmetric with respect to the players, so we will only show that Player 1 has  $\epsilon$ -subgame-perfect strategies. We will do this by taking an  $\epsilon$ -optimal strategy  $\sigma$  with some more structure, and using it to construct a *reset* strategy  $\hat{\sigma}$ , which will be  $2\epsilon$ -subgame-perfect. The reset strategy is conceptually very simple: a strategy  $\sigma$  is not  $2\epsilon$ -subgame-perfect if and only if there exists some finite play  $h := s_0 \dots s_n$  such that

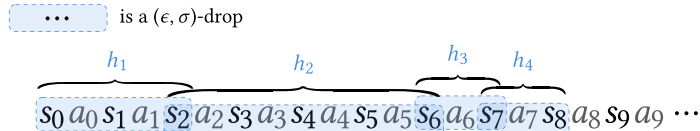
$$\inf_{\tau} \mathbb{E}_{s_n}^{\sigma[h], \tau} [f[h]] < \text{val}(s_n) - 2\epsilon; \quad (8)$$

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

**Definition 4.1.** *The finite play  $h := s_0 \dots s_n$  is called a  $(\epsilon, \sigma)$ -drop if (8) holds. We write*

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

It is plain that one can factorise any infinite play into  $h_1 h_2 \dots$  where each  $h_i$  is a  $(\epsilon, \sigma)$ -drop, but no strict prefix of  $h_i$  is  $(\epsilon, \sigma)$ -drop. For example:



**Definition 4.2.** We define the date of the most recent (or latest) drop for all  $s_0 \cdots s_n$  inductively as:

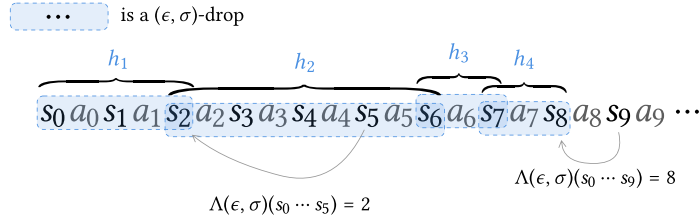
$$\Lambda(\epsilon, \sigma)(s_0) \stackrel{\text{def}}{=} 0$$

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

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

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

where

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

By construction, the reset strategy has the property that if it is  $\epsilon$ -optimal then it is also  $2\epsilon$ -subgame-perfect.

**Lemma 4.4.** Let  $\hat{\sigma}$  be a reset strategy that is  $\epsilon$ -optimal, then it is also  $2\epsilon$ -subgame-perfect.

*Proof.* Let  $s_0 \cdots s_n$  be a finite play, the goal is to show that:

$$\inf_{\tau} \mathbb{E}_{s_n}^{\hat{\sigma}[s_0 \cdots s_n], \tau} [f] \geq \text{val}(s_n) - 2\epsilon. \quad (9)$$

If there is a drop occurring in date  $n$ , that is  $\Lambda(\epsilon, \sigma)(s_0 \cdots s_n) = n$  then

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

by the definition of a reset strategy that is  $\epsilon$ -optimal. Assume then that the most recent drop is  $\ell < n$ , which means that:

$$\inf_{\tau} \mathbb{E}_{s_n}^{\sigma[s_\ell \cdots s_n], \tau} [f] \geq \text{val}(s_n) - 2\epsilon, \quad (10)$$



where  $\ell = \Lambda(\epsilon, \sigma)(s_0 \dots s_{n-1})$ . Towards a contradiction, assume that the goal (9) does not hold, *i.e.* there exists a strategy  $\tau$  that gives payoff strictly less than  $\text{val}(s_n) - 2\epsilon$ , then we will construct another strategy  $\tau'$  that will contradict (10).

Let  $\mathcal{D}$  be the set prefixes from  $s_\ell$  to the next  $(\epsilon, \sigma)$ -drop, that is

$$\mathcal{D} \stackrel{\text{def}}{=} \{s_\ell \dots s_{\ell'} : s_\ell \dots s_{\ell'} \text{ is a } (\epsilon, \sigma)\text{-drop but no strict prefix is}\},$$

and  $\overline{\mathcal{D}}$  the event that is generated by the cylinders in  $\mathcal{D}$  (note that the complement  $\overline{\mathcal{D}}$  is the event that no drop occurs). Define  $\tau'$  to be the strategy that plays like  $\tau$  except when a prefix in  $\mathcal{D}$  is met, in which case it switches to the  $\epsilon$ -response strategy  $\tau''$ . To simplify the notation let:

$$\sigma_1 \stackrel{\text{def}}{=} \hat{\sigma}[s_0 \dots s_n], \quad \sigma_2 \stackrel{\text{def}}{=} \sigma[s_\ell \dots s_n].$$

From the assumption that the goal does not hold we have the following inequality<sup>1</sup>

$$\begin{aligned} \text{val}(s_n) - 2\epsilon &> \mathbb{E}_{s_n}^{\sigma_1, \tau} [f \cdot \mathbb{1}_{\overline{\mathcal{D}}}] + \mathbb{E}_{s_n}^{\sigma_1, \tau} [f \cdot \mathbb{1}_{\mathcal{D}}] \\ &= \mathbb{E}_{s_n}^{\sigma_1, \tau} [f \cdot \mathbb{1}_{\overline{\mathcal{D}}}] + \mathbb{E}_{s_n}^{\sigma_2, \tau'} [f \cdot \mathbb{1}_{\overline{\mathcal{D}}}] \\ &= \mathbb{E}_{s_n}^{\sigma_1, \tau} [f \cdot \mathbb{1}_{\overline{\mathcal{D}}}] + \mathbb{E}_{s_n}^{\sigma_2, \tau'} [f] - \mathbb{E}_{s_n}^{\sigma_2, \tau'} [f \cdot \mathbb{1}_{\mathcal{D}}]. \end{aligned} \quad (11)$$

In the equality the strategy  $\sigma_1$ , respectively  $\tau$ , has been replaced by  $\sigma_2$ , respectively  $\tau'$  because on infinite plays without a drop they coincide.

For the first term above we have:

$$\begin{aligned} \mathbb{E}_{s_n}^{\sigma_1, \tau} [f \cdot \mathbb{1}_{\overline{\mathcal{D}}}] &= \sum_{t_0 \dots t_m \in \overline{\mathcal{D}}} \mathbb{P}_{s_n}^{\sigma_1, \tau}(t_0 \dots t_m) \mathbb{E}_{t_m}^{\hat{\sigma}, \tau[t_0 \dots t_m]} [f] \\ &\geq \sum_{t_0 \dots t_m \in \overline{\mathcal{D}}} \mathbb{P}_{s_n}^{\sigma_1, \tau}(t_0 \dots t_m) (\text{val}(t_m) - \epsilon), \end{aligned}$$

by definition of the  $\epsilon$ -optimal reset strategy and the fact that  $f$  is shift-invariant. For the last term on the other hand we have:

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

by construction of the  $\epsilon$ -response strategy  $\tau''$  and  $\tau'$ . The strategies  $\sigma_1$  and  $\sigma_2$  on one hand, and  $\tau$ ,  $\tau'$  on the other, coincide up to the first drop, consequently we can interchange them when measuring cylinders  $t_0 \dots t_m$ , which implies that the two inequalities above give:

$$\mathbb{E}_{s_n}^{\sigma_1, \tau} [f \cdot \mathbb{1}_{\overline{\mathcal{D}}}] \geq \mathbb{E}_{s_n}^{\sigma_2, \tau'} [f \cdot \mathbb{1}_{\overline{\mathcal{D}}}].$$

This contradicts (10) when plugged it in (11).  $\square$

As a consequence of this lemma, in order to prove Theorem 1.2, we only have to demonstrate that there exists a reset strategy that is  $\epsilon$ -optimal. In the rest of this section we will prove that there are strategies with more and more desirable properties, culminating in the proof that there is some  $\sigma$  whose reset strategy is  $\epsilon$ -optimal.

<sup>1</sup> $\mathbb{1}_{\mathcal{E}}$  is the indicator function of the event  $\mathcal{E}$ .

## 4.1 Properties of the Reset Strategy

We will show that there is a strategy  $\sigma$  with the following properties:

1.  $\sigma$  is  $\epsilon$ -optimal,
2.  $\sigma$  is locally optimal,<sup>2</sup>
3. for any  $\tau$  when playing with  $\hat{\sigma}$  and  $\tau$  almost surely there are only finitely many  $(\epsilon, \sigma)$ -drops, and
4.  $\hat{\sigma}$  is  $\epsilon$ -optimal.

We will do this in a manner that accumulates more structure, that is, for strategies with properties 1 and 2 we can prove the third property; and for strategies with all of the first three properties it is possible to prove that the reset strategy is  $\epsilon$ -optimal. Each subsection below corresponds to the proof of one of the last three properties (Property 1 is a consequence of Martin's theorem Theorem 2.4).

We are going to make use of some results from the theory of martingales<sup>3</sup>, which we introduce first.

**Definition 4.5** (Martingale). *A sequence of real-valued random variables  $X_0, X_1, \dots$  is called a martingale if for all  $n \in \mathbb{N}$*

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

*It is called a supermartingale, respectively submartingale, if instead of the equality we have  $\geq$ , respectively  $\leq$ .*

In our case the sequence  $\text{val}(S_0), \text{val}(S_1), \dots$  under suitable strategies will be a supermartingale, which will allow us to use in particular the following results.

**Theorem 4.6** (Doob's Forward Convergence Theorem, [Wil91, Theorem 11.5]). *Let  $X_0, X_1, \dots$  be a supermartingale such that the sequence  $(\mathbb{E} [|X_n|])_{n \in \mathbb{N}}$  is bounded. Then almost surely the limit*

$$\lim_{n \rightarrow \infty} X_n,$$

*exists and is finite.*

It follows from the definition of martingales that for all  $n \in \mathbb{N}$ , the expected value of  $X_n$  is equal to the expected value of  $X_0$ . In other words, the process that is stopped at time  $n$  is on average is equal to the process at time 0. The next theorem from martingale theory that we will make use of, has an analogous statement, namely that the process stopped at some random time  $T$  is on average equal to the process stopped at time zero. This theorem is known as Doob's optional stopping theorem. See for example Section 10.10 in [Wil91]. We give a variant of this theorem.

<sup>2</sup>This means that it does not play an action that decreases the value on average, the precise definition will follow.

<sup>3</sup>As a general reference for this area one might use [Wil91].

**Definition 4.7** (Stopping Time). A random variable  $T$  taking values in  $\mathbb{N} \cup \{\infty\}$  is called a stopping time with respect to random variables  $S_0, S_1, \dots$  if the event  $\{T = n\}$  for  $n \in \mathbb{N}$  is  $(S_0, \dots, S_n)$ -measurable, meaning that it depends only on the random variables  $S_0, \dots, S_n$ .

**Theorem 4.8** (Doob's Optional Stopping Theorem). Let  $T$  be a stopping time with respect to the random variables  $S_0, S_1, \dots$  and  $(X_n)_{n \in \mathbb{N}}$  a uniformly bounded martingale such that for all  $n \in \mathbb{N}$ ,  $X_n$  is  $(S_0, \dots, S_n)$ -measurable. Define the random variable  $X_T$  which represents the process stopped at time  $T$  as:

$$X_T \stackrel{\text{def}}{=} \begin{cases} X_n & \text{if } T \text{ is finite and equal to } n, \\ \lim_{n \rightarrow \infty} X_n & \text{if } T = \infty. \end{cases}$$

Then the expectation of  $X_T$  is equal to that of  $X_0$ . Analogous statements hold for supermartingales and submartingales.

*Proof.* The random variable  $X_T$  is well-defined as a consequence of Theorem 4.6. For every  $k \in \mathbb{N}$  define:

$$Y_k \stackrel{\text{def}}{=} X_{\min(T, k)}.$$

The process  $(Y_k)_{k \in \mathbb{N}}$  is a uniformly bounded martingale that converges almost-surely, as well. By definition of martingales for all  $n \in \mathbb{N}$

$$\mathbb{E}[Y_n] = \mathbb{E}[Y_0] = \mathbb{E}[X_0].$$

Furthermore  $(Y_k)_{k \in \mathbb{N}}$  converges pointwise to  $X_T$ . One can now use Lebesgue's dominated convergence theorem (see for example [Wil91, Theorem 5.9]) to conclude that:

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

When the process is a supermartingale or a submartingale one can write an analogous proof.  $\square$

#### 4.1.1 Locally Optimal

An action is locally optimal if the average value of the successor states is equal to the value of the current state. Formally:

**Definition 4.9** (Locally Optimal Strategy). An action  $a \in \mathbf{A}(s)$  is called locally optimal if and only if

$$\text{val}(s) = \sum_{t \in \mathbf{S}} p(s, a, t) \text{val}(t).$$

A strategy that only plays locally optimal actions is called locally optimal.

The salient point is the following observation about the process  $\text{val}(S_0), \text{val}(S_1), \dots$  when players use locally optimal strategies.

**Observation 4.10.** *When Player 1 (respectively Player 2) uses a locally optimal strategy the process*

$$\text{val}(S_0), \text{val}(S_1), \dots$$

*is a supermartingale (respectively a submartingale).*

This observation readily follows from the definition above and the fact that the values are bounded.

One can get away with playing solely locally optimal actions in games with perfect information. In other words, suppose that the action  $a_0 \in A(s_0)$  (say belonging to Player 1) in game  $G$  is *not* locally optimal, and denote by  $G'$  the same game except that it does not have action  $a_0$  in state  $s_0$ . We will prove that the values of those two games coincide; this then clearly implies that Player 1 has  $\epsilon$ -optimal strategies that are locally optimal as well. The analogue fact for Player 2 can be proved symmetrically.

Player 1 has less choice in  $G'$ , so for every  $s \in S$

$$\text{val}(G')(s) \leq \text{val}(G)(s),$$

hence we only have to prove the inverse inequality. Towards this end, we first prove that:

$$\text{val}(G')(s_0) \geq \text{val}(G)(s_0). \quad (12)$$

Let

$$\delta \stackrel{\text{def}}{=} \text{val}(G)(s_0) - \sum_{t \in S} p(s_0, a_0, t) \text{val}(G)(t) > 0,$$

and  $\tau$  the strategy that plays according to the strategy  $\tau'$  that is  $\epsilon$ -optimal in  $G'$  – as long as the opponent does not choose the action  $a_0$ , in which case it switches definitely to the strategy  $\tau''$  which is  $\delta/2$ -optimal in  $G$ . Let  $\mathcal{Z}$  be the event that the action  $a_0$  is never chosen, *i.e.*

$$\mathcal{Z} \stackrel{\text{def}}{=} \{\forall n \ S_n = s_0 \Rightarrow A_{n+1} \neq a\}.$$

Then by construction of  $\tau$ , for all  $\sigma$  and  $s$ :

$$\begin{aligned} \mathbb{E}_s^{\sigma, \tau} [f \mid \mathcal{Z}] &\leq \text{val}(G')(s) + \epsilon, \text{ and} \\ \mathbb{E}_s^{\sigma, \tau} [f \mid \neg \mathcal{Z}] &\leq \text{val}(G)(s_0) - \delta + \delta/2, \end{aligned}$$

whence it follows that for all  $\sigma$ ,  $s$  and  $\epsilon > 0$

$$\mathbb{E}_s^{\sigma, \tau} [f] \leq \max\{\text{val}(G')(s) + \epsilon, \text{val}(G)(s_0) - \delta/2\}.$$

Taking  $s = s_0$  and the supremum over all  $\sigma$  gives (12).

Using (12), we prove now that for all  $s$

$$\text{val}(G')(s) \geq \text{val}(G)(s). \quad (13)$$

Define  $S(\sigma)$  to be the event that the action  $a_0$  is about to be played by strategy  $\sigma$ , that is

$$S(\sigma) \stackrel{\text{def}}{=} \{\exists n \ S_n = s_0 \text{ and } \sigma(S_0 \cdots S_n)(a_0) > 0\}.$$

Let  $\epsilon > 0$  and for any strategy  $\sigma$ , define  $\tilde{\sigma}$  to be the strategy that plays like  $\sigma$  unless the latter is about to play the action  $a_0$  in  $s_0$ , in which case it switches to the strategy  $\sigma'$  which is  $\epsilon$ -optimal in  $G'$ . Set  $\tau$  to be the strategy that plays according to some strategy  $\tau'$  which is  $\epsilon$ -optimal in  $G'$  as long as the opponent does not play the action  $a_0$ , otherwise it switches to some strategy that is  $\epsilon$ -optimal in  $G$ . By definitions of these strategies and (12) we have that for all  $\sigma$  and  $s$

$$\begin{aligned} \mathbb{E}_s^{\sigma, \tau} [f \mid S(\sigma)] &\leq \text{val}(G)(s_0) + \epsilon = \text{val}(G')(s_0) + \epsilon, \text{ and} \\ \mathbb{E}_s^{\tilde{\sigma}, \tau} [f \mid S(\sigma)] &\geq \text{val}(G')(s_0), \end{aligned}$$

a combination of which gives us

$$\mathbb{E}_s^{\sigma, \tau} [f \mid S(\sigma)] \leq \mathbb{E}_s^{\tilde{\sigma}, \tau} [f \mid S(\sigma)] + 2\epsilon. \quad (14)$$

The strategies  $\sigma$  and  $\tilde{\sigma}$  on one hand, and  $\tau$  and  $\tau'$  on the other, coincide up to the date when  $\sigma$  is about to play the action  $a_0$ , as a consequence:

$$P(\sigma, s) \stackrel{\text{def}}{=} \mathbb{P}_s^{\sigma, \tau} (S(\sigma)) = \mathbb{P}_s^{\tilde{\sigma}, \tau'} (S(\sigma)).$$

Now by construction of the strategies and (14), for all  $\sigma$  and  $s$  we have

$$\begin{aligned} \mathbb{E}_s^{\sigma, \tau} [f] &= P(\sigma, s) \mathbb{E}_s^{\sigma, \tau} [f \mid S(\sigma)] + (1 - P(\sigma, s)) \mathbb{E}_s^{\sigma, \tau} [f \mid \neg S(\sigma)] \\ &\leq P(\sigma, s) (\mathbb{E}_s^{\tilde{\sigma}, \tau} [f \mid S(\sigma)] + 2\epsilon) + (1 - P(\sigma, s)) \mathbb{E}_s^{\sigma, \tau} [f \mid \neg S(\sigma)] \\ &= \mathbb{E}_s^{\tilde{\sigma}, \tau} [f] + 2\epsilon P(\sigma, s) = \mathbb{E}_s^{\tilde{\sigma}, \tau'} [f] + 2\epsilon P(\sigma, s) \\ &\leq \text{val}(G')(s) + \epsilon(2P(\sigma, s) + 1). \end{aligned}$$

Since this holds for any  $\epsilon > 0$ , taking the supremum over all  $\sigma$  proves (13).

We have thus proved that for all  $\epsilon > 0$ , both players have strategies that are both

$$\text{locally optimal and } \epsilon\text{-optimal.} \quad (15)$$

We gather one more observation about games where at least one of the players utilises a locally optimal strategy. In this case, a stronger type of locally optimal action is the only one played infinitely many times.

**Definition 4.11** (Value-Conserving Action). *An action  $a \in \mathbf{A}(s)$  is called value-conserving in  $s$  if and only if for all  $t \in \mathbf{S}$ ,*

$$p(s, a, t) > 0 \quad \Rightarrow \quad \text{val}(s) = \text{val}(t).$$

**Proposition 4.12.** *For all strategies  $\sigma, \tau$  at least one of which is locally optimal and  $s \in \mathbf{S}$  we have*

$$\mathbb{P}_s^{\sigma, \tau} (\text{for all but finitely many } n, A_n \text{ is value-conserving in } S_n) = 1.$$

*Proof.* Fix  $\sigma$  and  $\tau$  and assume that  $\sigma$  is locally optimal, the other case is symmetrical. Suppose that  $a_0 \in A(s_0)$  is not value-conserving. It suffices to prove that the event

$$\{\text{for infinitely many } n, S_n = s_0 \text{ and } A_n = a_0\},$$

has measure zero. Assume towards a contradiction that the event above has non-zero probability, then the event which says that for infinitely many  $n$ , we have  $S_n = s_0$ ,  $A_n = a_0$  and  $S_{n+1} = t$  also has non-zero probability; where  $t \in S$  is a successor state of  $s_0$  under  $a_0$  that has value strictly smaller than that of  $s_0$  (its existence is guaranteed because  $a_0$  is not value-conserving). This means that there is non-zero probability that for infinitely many  $n$ ,

$$|\text{val}(S_n) - \text{val}(S_{n+1})| \geq \text{val}(s_0) - \text{val}(t) > 0,$$

which contradicts Theorem 4.6, since  $(\text{val}(S_n))$ ,  $n \in \mathbb{N}$  is a supermartingale as per Observation 4.10.  $\square$

#### 4.1.2 Finitely Many Drops

Recall that  $\Delta(\epsilon, \sigma)(\cdot)$  characterises finite plays that are  $(\epsilon, \sigma)$ -drops. We informally refer to the event

$$\text{for all } m > n, \neg \Delta(\epsilon, \sigma)(S_0 \cdots S_m),$$

as

$$\text{no } (\epsilon, \sigma)\text{-drops after date } n.$$

Similarly for events such as “there is a  $(\epsilon, \sigma)$ -drop” or “two  $(\epsilon, \sigma)$ -drops after date  $n$ ”. Our goal is to prove that for a reset strategy that is based on a  $\sigma$  that is both  $\epsilon$ -optimal and locally optimal (which exists because of (15)) almost surely there will only be finitely many  $(\epsilon, \sigma)$ -drops. To this end fix a  $\epsilon > 0$ , and  $\sigma$  a strategy that is both locally optimal and  $\epsilon$ -optimal, which allows us to simply say drop instead of  $(\epsilon, \sigma)$ -drop. The proof of the goal is relatively lengthy, however the idea and the plan is simple.

An intermediate fact that we have to prove is that when Player 1 plays with the reset strategy there is some  $n \in \mathbb{N}$  such that the probability that there is a drop after date  $n$  is bounded away from 1. This fact is easier to prove if we assume that the adversary is using a locally optimal strategy. Then Proposition 4.12 helps us lift this restriction on the strategies of Player 2. Therefore the plan is to prove this intermediate fact first (1) for locally optimal strategies, then (2) for strategies  $\tau_n$  that are locally optimal after date  $n$ , and finally (3) for general strategies. The intermediate fact then finalises the goal of the preset section, that is when Player 1 plays with the reset strategy  $\hat{\sigma}$  almost surely there will be only finitely many drops.

**Lemma 4.13.** *There exists a  $c > 0$  such that for all  $s$  and locally optimal  $\tau$ ,*

$$\mathbb{P}_s^{\sigma, \tau}(\text{there is a drop}) \leq 1 - c.$$

*Proof.* Let  $T$  be the date of the first drop, that is

$$T \stackrel{\text{def}}{=} \min\{n : S_0 \cdots S_n \text{ is a drop}\},$$

with the convention that  $\min \emptyset = \infty$ . Notice that  $T$  is a stopping time with respect to the process  $(\text{val}(S_n))$ ,  $n \in \mathbb{N}$ . Let  $\tau'$  be a strategy that plays like  $\tau$  as long as no drop occurs, and once it does it switches to the strategy  $\tau''$  that is a  $\epsilon/2$ -optimal response. By construction,  $\tau$  and  $\tau'$  coincide on trajectories without drops so define:

$$P \stackrel{\text{def}}{=} \mathbb{P}_s^{\sigma, \tau} (\text{no drops}) = \mathbb{P}_s^{\sigma, \tau'} (\text{no drops}),$$

and let  $M$  respectively  $m$ , be an upper bound, respectively lower bound of the payoff function  $f$ . By  $\epsilon$ -optimality of  $\sigma$ , for all  $s$  we have:

$$\begin{aligned} \text{val}(s) - \epsilon &\leq (1 - P) \cdot \mathbb{E}_s^{\sigma, \tau'} [f \mid \text{there is a drop}] + P \cdot \mathbb{E}_s^{\sigma, \tau'} [f \mid \text{no drops}] \\ &\leq (1 - P) \cdot \mathbb{E}_s^{\sigma, \tau'} [f \mid \text{there is a drop}] + P \cdot M. \end{aligned}$$

Denote by  $\mathcal{D}$  the finite plays that are drops but that do not have a prefix that is a drop, *i.e.* it contains all the finite plays up to the first drop. Then by construction of  $\tau'$  we have for all  $s$ :

$$\begin{aligned} \mathbb{E}_s^{\sigma, \tau'} [f \mid \text{there is a drop}] &= \sum_{s_0 \cdots s_n \in \mathcal{D}} \mathbb{P}_s^{\sigma, \tau'} (s_0 \cdots s_n \mid \text{there is a drop}) \cdot \mathbb{E}_{s_n}^{\sigma[s_0 \cdots s_n], \tau''} [f] \\ &\leq \sum_{s_0 \cdots s_n \in \mathcal{D}} \mathbb{P}_s^{\sigma, \tau'} (s_0 \cdots s_n \mid \text{there is a drop}) \cdot (\text{val}(s_n) - 2\epsilon + \epsilon/2) \\ &= \mathbb{E}_s^{\sigma, \tau'} [\text{val}(S_T) \mid \text{there is a drop}] - \frac{3}{2}\epsilon. \end{aligned}$$

Replacing this inequality in the one above and decomposing the expectation of  $\text{val}(S_T)$ <sup>4</sup> we conclude that for all  $s$ :

$$\begin{aligned} \text{val}(s) - \epsilon &\leq \mathbb{E}_s^{\sigma, \tau'} [\text{val}(S_T)] + P \cdot \left( M - \mathbb{E}_s^{\sigma, \tau'} [\text{val}(S_T) \mid \text{no drops}] \right) - \frac{3}{2}\epsilon(1 - P) \\ &\leq \text{val}(s) + P \cdot (M - m) - \frac{3}{2}\epsilon(1 - P), \end{aligned}$$

where the expectation of  $\text{val}(S_T)$  is smaller than the  $\text{val}(s)$  for the following reason. Since  $T$  is a stopping time and  $\tau'$  plays like  $\tau$  before the first drop, hence it plays locally optimal actions, consequently the process  $\text{val}(S_n)$ ,  $n \in \mathbb{N}$  is a submartingale at least until the first drop<sup>5</sup>, so we can apply Theorem 4.8. Finally from the inequality above we have:

$$P \geq \frac{1}{2} \frac{\epsilon}{M - m + 3/2\epsilon} \stackrel{\text{def}}{=} c,$$

a uniform bound that does not depend on the choice of  $\tau$ . □

<sup>4</sup> Theorem 4.6 implies that this random variable is well-defined.

<sup>5</sup> formally one defines another process that *stops* after time  $T$ , that is a process  $\text{val}(S_{\min\{n, T\}})$ .

Next we approximate strategies  $\tau$  by a sequence  $\tau_n$  for every natural  $n$  as follows. The strategies  $\tau_n$  play like  $\tau$  only up to date  $n$ , otherwise they choose some locally optimal action, formally:

$$\tau_n(s_0 \cdots s_m) \stackrel{\text{def}}{=} \begin{cases} \tau(s_0 \cdots s_m) & \text{if } m < n \text{ or } \tau(s_0 \cdots s_m) \text{ chooses locally optimal actions,} \\ \text{some locally optimal action in } s_m & \text{otherwise.} \end{cases}$$

**Lemma 4.14.** *There is some  $c > 0$  such that for all strategies  $\tau$ ,  $s$  and  $n \in \mathbb{N}$ , we have*

$$\mathbb{P}_s^{\hat{\sigma}, \tau_n} (\text{there is a drop after date } n) \leq 1 - c.$$

*Proof.* For  $n \in \mathbb{N}$  define the stopping time  $T_n$  to be the date of the first drop after the date  $n$ , that is

$$T_n \stackrel{\text{def}}{=} \min\{m > n : S_0 \cdots S_m \text{ is a drop}\},$$

with the convention that  $\min \emptyset = \infty$ , and set  $T_n^2$  to be the date of the second drop after  $n$ , that is  $T_{T_n}$ . We prove that there is some  $c > 0$  such that for all  $n \in \mathbb{N}$ , strategy  $\tau$  and state  $s$  we have

$$\mathbb{P}_s^{\hat{\sigma}, \tau_n} (T_n^2 < \infty \mid T_n < \infty) \leq 1 - c. \quad (16)$$

The statement of the lemma then follows from (16) and sigma-additivity of measures. Define  $\mathcal{D}_n$  to be the set of finite plays, strictly longer than  $n$ , that are drops but such that they have no prefix longer than  $n$  that is a drop. In other words  $\mathcal{D}_n$  contains all the plays up to the first drop after the date  $n$ . Then by construction of the reset strategy:

$$\begin{aligned} \mathbb{P}_s^{\hat{\sigma}, \tau_n} (T_n^2 < \infty \mid T_n < \infty) &= \sum_{s_0 \cdots s_m \in \mathcal{D}_n} \mathbb{P}_s^{\hat{\sigma}, \tau_n} (T_n^2 < \infty \mid s_0 \cdots s_m) \mathbb{P}_s^{\hat{\sigma}, \tau_n} (s_0 \cdots s_m \mid T_n < \infty) \\ &= \sum_{s_0 \cdots s_m \in \mathcal{D}_n} \mathbb{P}_{s_m}^{\hat{\sigma}, \tau_n[s_0 \cdots s_m]} (T_0 < \infty) \mathbb{P}_s^{\hat{\sigma}, \tau_n} (s_0 \cdots s_m \mid T_n < \infty) \\ &= \sum_{s_0 \cdots s_m \in \mathcal{D}_n} \mathbb{P}_{s_m}^{\sigma, \tau_n[s_0 \cdots s_m]} (T_0 < \infty) \mathbb{P}_s^{\hat{\sigma}, \tau_n} (s_0 \cdots s_m \mid T_n < \infty), \end{aligned}$$

where in the last equality we have replaced the reset strategy by  $\sigma$ , because these two strategies are the same up to the first drop. Since  $m > n$ , by construction the strategy  $\tau_n[s_0 \cdots s_m]$  is locally optimal, consequently applying Lemma 4.13 gives

$$\mathbb{P}_{s_m}^{\sigma, \tau_n[s_0 \cdots s_m]} (T_0 < \infty) \leq 1 - c,$$

which when plugged into the equation above proves (16).  $\square$

In the third lemma there is no restriction upon the strategy  $\tau$ .

**Lemma 4.15.** *For all strategies  $\tau$  and  $s$  there is some  $n \in \mathbb{N}$  such that*

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



*Proof.* Fix a strategy  $\tau$  and a state  $s$ . Let  $T$  be the stopping time that gives the date of the last action that was played that is not value-conserving, if it exists, otherwise let it be  $\infty$ . Since the strategies  $\tau$  and  $\tau_n$  coincide on all paths where the last action that is not value-conserving is played before  $n$  (that is on the event  $T < n$ ), then for all  $n \in \mathbb{N}$  and events  $\mathcal{E}$  we have:

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

The strategy  $\sigma$  has been assumed to be locally optimal, and therefore the strategy  $\hat{\sigma}$  is locally optimal as well. As a consequence of Proposition 4.12 we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_s^{\hat{\sigma}, \tau}(T < n) = 1,$$

whence follows

$$\lim_{n \rightarrow \infty} \mathbb{P}_s^{\hat{\sigma}, \tau_n}(\mathcal{E}) = \mathbb{P}_s^{\hat{\sigma}, \tau}(\mathcal{E}),$$

for any event  $\mathcal{E}$ . The proof of the lemma now concludes by choosing the event “there is a drop after date  $n$ ” for  $\mathcal{E}$ , a suitable natural number  $n$  and applying Lemma 4.14.  $\square$

This lemma makes it possible now to prove the third property of the strategy  $\sigma$ , namely that for all strategies  $\tau$  and  $s$ ,

$$\mathbb{P}_s^{\hat{\sigma}, \tau}(\exists n \text{ no drops after date } n) = 1. \quad (17)$$

Let  $T$  be the stopping time that gives the date of the last drop, if it exists otherwise let it be equal to  $\infty$ . For a natural  $n$ , let  $F_n$  be the stopping time that gives the date of the first drop after  $n$  (same as  $T_n$  in the proof of Lemma 4.14) if it exists, otherwise say that it is equal to  $\infty$ .

Fix  $\delta > 0$  and choose the strategy  $\tilde{\tau}$  and state  $\tilde{s}$  such that

$$\sup_{\tau, s} \mathbb{P}_s^{\hat{\sigma}, \tau}(T = \infty) \leq \mathbb{P}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}}(T = \infty) + \delta. \quad (18)$$

Let  $\tilde{n} \in \mathbb{N}$  the number from the statement of Lemma 4.15, thus

$$d \stackrel{\text{def}}{=} \mathbb{P}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}}(F_{\tilde{n}} < \infty) < 1. \quad (19)$$

And from (18), some basic properties of expectations we deduce:

$$\begin{aligned} \mathbb{P}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}}(T = \infty) &= \mathbb{E}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}} \left[ \mathbb{P}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}}(T = \infty \mid F_{\tilde{n}}, S_0, \dots, S_{F_{\tilde{n}}}) \right] \\ &= \mathbb{E}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}} \left[ \mathbb{1}_{F_{\tilde{n}} < \infty} \cdot \mathbb{P}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}}(T = \infty \mid F_{\tilde{n}}, S_0, \dots, S_{F_{\tilde{n}}}) \right] \\ &= \mathbb{E}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}} \left[ \mathbb{1}_{F_{\tilde{n}} < \infty} \cdot \mathbb{P}_{S_{F_{\tilde{n}}}}^{\hat{\sigma}, \tilde{\tau}}[S_0 \dots S_{F_{\tilde{n}}}] (T = \infty) \right] \\ &\leq \mathbb{E}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}} \left[ \mathbb{1}_{F_{\tilde{n}} < \infty} \cdot \left( \mathbb{P}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}}(T = \infty) + \delta \right) \right] \\ &= d \cdot \left( \mathbb{P}_{\tilde{s}}^{\hat{\sigma}, \tilde{\tau}}(T = \infty) + \delta \right). \end{aligned}$$

The random variable  $S_{F_n}$  is well-defined because we are measuring the infinite plays where  $F_n$  is finite; on the third equality we have used the definition of the reset strategy and the last two (in)equalities we have used (18) and (19) respectively. Since  $d < 1$  then we have

$$\mathbb{P}_s^{\hat{\sigma}, \tilde{\tau}}(T = \infty) \leq \frac{d}{1-d} \delta,$$

so for all states  $s'$  and strategies  $\tau'$  it follows that

$$\mathbb{P}_{s'}^{\hat{\sigma}, \tau'}(T = \infty) \leq \sup_{\tau, s} \mathbb{P}_s^{\hat{\sigma}, \tau}(T = \infty) \leq \mathbb{P}_s^{\hat{\sigma}, \tilde{\tau}}(T = \infty) + \delta \leq \frac{\delta}{1-d}.$$

Since this holds for any  $\delta > 0$ , (17) follows.

### 4.1.3 $\epsilon$ -Optimal

The last property of  $\sigma$  that we have to prove is that if we assume that it has the previous properties, namely that it is  $\epsilon$ -optimal, locally optimal, and it has finitely many drops, then the reset strategy  $\hat{\sigma}$  is  $\epsilon$ -optimal as well. So fix an  $\epsilon > 0$  and a strategy  $\sigma$  that is both locally optimal and  $\epsilon$ -optimal, and for which (17) holds. We define for all naturals  $n$ , strategies  $\hat{\sigma}_n$  that reset only up to date  $n$ , and prove that they are  $\epsilon$ -optimal first.

Define  $\mathfrak{T}_n$  to be the function that truncates finite plays to length  $n$ :

$$\mathfrak{T}_n(s_0 \cdots s_m) \stackrel{\text{def}}{=} \begin{cases} s_0 \cdots s_m & \text{if } m \leq n, \\ s_0 \cdots s_n & \text{otherwise.} \end{cases}$$

The reset strategy that resets only up to date  $n$  is then defined as:

$$\hat{\sigma}_n(s_0 \cdots s_m) \stackrel{\text{def}}{=} \sigma(s_\ell \cdots s_m),$$

where

$$\ell \stackrel{\text{def}}{=} \Lambda(\epsilon, \sigma)(\mathfrak{T}_n(s_0 \cdots s_m)).$$

**Lemma 4.16.** *For all  $n \in \mathbb{N}$ ,  $\hat{\sigma}_n$  is  $\epsilon$ -optimal.*

*Proof.* The proof is by induction on  $n$ . The base case is trivial since  $\hat{\sigma}_0 = \sigma$ , therefore assume that the lemma is true for  $n - 1$ , we prove that it is also true for  $n$ . Namely we fix a state  $s$  and a strategy  $\tau$  and prove that

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

Denote by  $\mathcal{E}$  the event that there is a drop at date  $n$ , and by  $\mathcal{D}$  the set of finite plays of length  $n$  that are drops. Then we have

$$\begin{aligned} \mathbb{E}_s^{\hat{\sigma}_n, \tau}[f] &= \mathbb{E}_s^{\hat{\sigma}_n, \tau}[\mathbb{1}_{\mathcal{E}} \cdot f] + \mathbb{E}_s^{\hat{\sigma}_n, \tau}[\mathbb{1}_{\neg \mathcal{E}} \cdot f] \\ &= \sum_{s_0 \cdots s_n \in \mathcal{D}} \mathbb{P}_s^{\hat{\sigma}_n, \tau}(s_0 \cdots s_n) \cdot \mathbb{E}_{s_n}^{\sigma, \tau[s_0 \cdots s_n]}[f] + \mathbb{E}_s^{\hat{\sigma}_n, \tau}[\mathbb{1}_{\neg \mathcal{E}} \cdot f] \\ &\geq \sum_{s_0 \cdots s_n \in \mathcal{D}} \mathbb{P}_s^{\hat{\sigma}_n, \tau}(s_0 \cdots s_n) \cdot (\text{val}(s_n) - \epsilon) + \mathbb{E}_s^{\hat{\sigma}_n, \tau}[\mathbb{1}_{\neg \mathcal{E}} \cdot f], \end{aligned}$$

where in the second equality we have used the definition of  $\hat{\sigma}_n$  and in the inequality the  $\epsilon$ -optimality of  $\sigma$ . Define the strategy  $\tau'$  to be the strategy that plays like  $\tau$  except if there is a drop at date  $n$ , in which case it resets to a  $\epsilon/2$ -response called  $\tau''$ . Then we have

$$\begin{aligned} \mathbb{E}_s^{\hat{\sigma}_{n-1}, \tau'} [f] &= \mathbb{E}_s^{\hat{\sigma}_{n-1}, \tau'} [\mathbb{1}_{\mathcal{E}} \cdot f] + \mathbb{E}_s^{\hat{\sigma}_{n-1}, \tau'} [\mathbb{1}_{\neg \mathcal{E}} \cdot f] \\ &= \sum_{s_0 \cdots s_n \in \mathfrak{D}} \mathbb{P}_s^{\hat{\sigma}_{n-1}, \tau'} (s_0 \cdots s_n) \cdot \mathbb{E}_{s_n}^{\hat{\sigma}_{n-1}[s_0 \cdots s_n], \tau''} [f] + \mathbb{E}_s^{\hat{\sigma}_{n-1}, \tau} [\mathbb{1}_{\neg \mathcal{E}} \cdot f] \\ &\leq \sum_{s_0 \cdots s_n \in \mathfrak{D}} \mathbb{P}_s^{\hat{\sigma}_{n-1}, \tau'} (s_0 \cdots s_n) \cdot (\text{val}(s_n) - 2\epsilon + \epsilon/2) + \mathbb{E}_s^{\hat{\sigma}_{n-1}, \tau} [\mathbb{1}_{\neg \mathcal{E}} \cdot f]. \end{aligned}$$

Now since the strategies  $\hat{\sigma}_{n-1}$  and  $\hat{\sigma}_n$  on one hand, and strategies  $\tau$  and  $\tau'$  on the other, behave the same for all plays of length smaller than  $n$  and on infinite plays where there is no drop at date  $n$ , it follows that the right-most terms in the two inequalities above, as well as the factors  $\mathbb{P}_s^{\sigma, \tau}$  on the left are equal. Consequently we can combine the two inequalities above to conclude that

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

This concludes the induction step and the proof of the lemma.  $\square$

We now prove that

$$\hat{\sigma} \text{ is } \epsilon\text{-optimal,} \tag{20}$$

the final property of  $\sigma$  given in the beginning of this section.

Let  $m$  respectively  $M$  be a lower bound, respectively upper bound of the payoff function. Define  $T$  to be the stopping time that is equal to the date of the last drop if it exists otherwise it is equal to  $\infty$ .

Applying Lemma 4.16 we have that for all  $n \in \mathbb{N}$ ,  $s$ , and  $\tau$

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

Since  $\hat{\sigma}$  and  $\hat{\sigma}_n$  behave the same on the plays in the event  $T \leq n$ , we have that for all  $n \in \mathbb{N}$ ,  $\tau$  and  $s$

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

The strategies  $\hat{\sigma}_n$  and  $\hat{\sigma}$  behave the same on plays in the event  $T \leq n$ , and therefore also on those in the event  $T > n$ ; consequently we can write

$$\mathbb{P}_s^{\hat{\sigma}_n, \tau} (T > n) = \mathbb{P}_s^{\hat{\sigma}, \tau} (T > n),$$

and from the inequality above we have:

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

From the sigma-additivity of measures and the property in (17) it follows that

$$\lim_{n \rightarrow \infty} (M - m) \cdot \mathbb{P}_s^{\hat{\sigma}, \tau} (T > n) = 0.$$

Since  $\tau$  and  $s$  are general, this proves  $\epsilon$ -optimality of  $\sigma$ , that is it proves the final property (20). Lemma 4.4 in conjunction with (20) implies Theorem 1.2.

#### 4.1.4 Remark on Optimal Strategies

Martin's theorem, Theorem 2.4 implies that the games that we are interested in have  $\epsilon$ -optimal strategies for every  $\epsilon > 0$ . We have then proved that there are locally optimal (2) strategies that are also  $\epsilon$ -optimal (1). We then showed that for strategies with properties (1) and (2), we can prove that they also possess the properties (3) and (4), which respectively stated that there are finitely many drops and that the reset strategy is also  $\epsilon$ -optimal. By inspection, in the proofs of

$$\begin{aligned} (1) \text{ and } (2) &\quad \Rightarrow \quad (3), \\ (1), (2) \text{ and } (3) &\quad \Rightarrow \quad (4), \end{aligned}$$

in Section 4.1.2 and Section 4.1.3 respectively, the variable  $\epsilon$  need not be strictly positive. Since optimal strategies are necessarily locally optimal, the following lemma follows from Lemma 4.4.

We can summarize our results for  $\epsilon = 0$  or  $\epsilon > 0$  as:

**Lemma 4.17.** *Let  $G$  be a game equipped with a shift-invariant payoff function. Let  $\epsilon \geq 0$  be a non-negative real number and  $\sigma$  be an  $\epsilon$ -optimal strategy in  $G$ . Assume that  $\sigma$  is locally optimal. Then the reset strategy  $\hat{\sigma}$  is  $2\epsilon$ -subgame-perfect in  $G$ .*

## 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 Section 3. Necessary and sufficient conditions for positionality are known for deterministic games [GZ05]. However the shift-invariant and submixing conditions are general enough to recover several known classical results, and to provide several new examples of games with deterministic stationary optimal strategies. Before we proceed with the proof we remark:

**Remark 5.1.** *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. A corollary of this is that games with shift-invariant, submixing and inverse-submixing payoff functions are positional.*

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

$$N(G) \stackrel{\text{def}}{=} \sum_{s \in S_1} (|A(s)| - 1).$$

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  $N(G) = 0$  there is no choice for maximizer, hence he has a deterministic and stationary optimal strategy. If  $N(G) > 0$  there must be a state  $\tilde{s} \in S$  such that Player 1 has at least two actions in  $\tilde{s}$ , i.e.  $A(\tilde{s})$  has at least two elements. We split the game  $G$  in two strictly smaller subgames  $G_1$  and  $G_2$ .

**Definition 5.2** (Split of a game). *Let  $G$  be a game with  $N(G) > 0$  and  $\tilde{s} \in S$  a state of  $G$  controlled by Player 1 in which there are at least two actions available, i.e.  $A(\tilde{s})$  has at least two elements. Partition  $A(\tilde{s})$  into two non-empty sets:  $A_1$  and  $A_2$ . Let  $G_1$  and  $G_2$  be the games obtained from  $G$  by restricting the actions in the state  $\tilde{s}$  to  $A_1$  and  $A_2$  respectively. Then  $(G_1, G_2)$  is called a split of  $G$  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  $G$  by selecting one of the subgames and play optimally in it.

**Theorem 5.3.** *Let  $G$  be a game equipped with a payoff function that is shift-invariant and submixing. Let  $(G_1, G_2)$  a split of  $G$  on  $\tilde{s}$ . Then*

$$\text{val}(G)(\tilde{s}) = \max\{\text{val}(G_1)(\tilde{s}), \text{val}(G_2)(\tilde{s})\}. \quad (21)$$

Assume moreover that  $\text{val}(G_1)(\tilde{s}) \geq \text{val}(G_2)(\tilde{s})$ . Then, for every  $s \in S$ ,

$$\text{val}(G)(s) = \text{val}(G_1)(s). \quad (22)$$

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  $N(G) = 0$  there is no choice for maximizer, hence he has a deterministic and stationary optimal strategy. If  $N(G) > 0$  then we choose a split  $(G_1, G_2)$  of  $G$  on a pivot state  $\tilde{s}$ . By symmetry, we can choose a split such that  $\text{val}(G_1)(\tilde{s}) \geq \text{val}(G_2)(\tilde{s})$ . Then, according to (22) in Theorem 5.3, a strategy for Player 1 which is optimal in  $G_1$  is also optimal in  $G$ . By induction hypothesis, there exists a positional optimal strategy in  $G_1$ , thus  $G$  is half-positional.  $\square$

The rest of the section is dedicated to the proof of Theorem 5.3. We fix a game  $G$  and a split  $(G_1, G_2)$  of  $G$  on the state  $\tilde{s}$ . The inequality

$$\text{val}(G)(\tilde{s}) \geq \max\{\text{val}(G_1)(\tilde{s}), \text{val}(G_2)(\tilde{s})\}$$

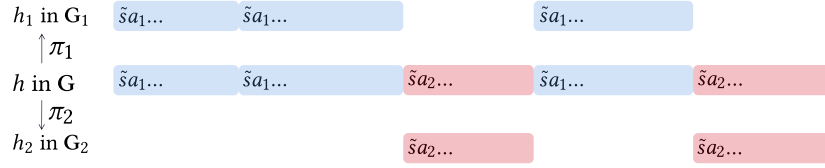


Figure 1: The play  $h$  is the concatenation of finite plays starting in  $\tilde{s}$ , represented by blocks whose colours depend on the first action played after  $\tilde{s}$ , blue if the action belongs to  $A_1$  and pink if it belongs to  $A_2$ . The projection  $h_1 = \pi_1(h)$  in  $G_1$  is the concatenation of the blue blocks while  $h_2 = \pi_2(h)$  is the concatenation of the pink blocks. The projections lose some information about the play in  $G$ : swapping two contiguous blocks of different colors in  $h$  does not modify the projections  $h_1$  and  $h_2$ .

is clear, since Player 1 has more choice in  $G$  than he has in  $G_1$  and  $G_2$ . We witness the converse inequality with a strategy for Player 2, called the *merge* strategy, which merges two  $\epsilon$ -subgame-perfect strategies in the respective smaller games. This is done in Section 5.3. The definition of the merge strategy hinges on the projection of plays in the main game  $G$  to plays in the restricted games  $G_1$  and  $G_2$ , which is done in section 5.1. Then we analyse the two possible outcomes: (a) after some date the play remains only in game  $G_1$  (or only in game  $G_2$ ), (b) the play switches infinitely often between the two smaller games. This analysis is performed in sections 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  $G$  starting in  $\tilde{s}$  to a couple of plays  $h_1$  and  $h_2$  in the restricted games  $G_1$  and  $G_2$  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  $G$  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  $A_1$  then the new step is appended to  $h_1$ , otherwise it is appended to  $h_2$ . The computation of  $h_1$  and  $h_2$  is illustrated on Figure 1.

Formally, we define two maps  $\pi_1, \pi_2$  from finite plays in  $G$  starting from  $\tilde{s}$  to finite plays in  $G_1$  and  $G_2$  respectively, starting from  $\tilde{s}$  as well. Let  $h = s_0 a_0 s_1 \dots s_n$  be a finite play in  $G$  starting in  $\tilde{s}$  and  $has$  a continuation of  $h$  in  $G$ , with one more transition  $(s_n, a, s)$ . Let  $\text{last}(has)$  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  $\pi_2$  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

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

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.

## 5.2 Linking the payoff in $\mathbf{G}$ to the payoffs in the subgames

The payoff in  $\mathbf{G}$  can be related to the payoff in the subgames. We introduce the events

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

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:

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

Recall that  $S_n$  and  $A_n$  are the random variables which output respectively the  $n$ -th state  $s_n$  and action  $a_n$  when the play is  $s_0 a_0 s_1 a_1 \dots$ . We see that  $\Pi$  is simply the identity map outputting the play in  $\mathbf{G}$  while  $\Pi_i$  is essentially equivalent to  $\pi_i$ , 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

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

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

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

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 prefix-independent 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,*

$$\text{if } \text{Stay}_\omega(\mathbf{G}_1) \text{ holds then } f(\Pi) = f(\Pi_1) . \quad (23)$$

$$\text{If } \text{Stay}_\omega(\mathbf{G}_2) \text{ holds then } f(\Pi) = f(\Pi_2) . \quad (24)$$

$$\text{If } \text{Switch} \text{ holds then } f(\Pi) \leq \max(f(\Pi_1), f(\Pi_2)) . \quad (25)$$

*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  $\Pi_1$  is finite then  $\Pi$  and  $\Pi_2$  share an infinite suffix and the prefix-independence of  $f$  implies (23). The case where  $\Pi_2$  is finite is symmetric, hence (24). If both  $\Pi_1$  and  $\Pi_2$  are infinite then the sequence of actions  $(\text{last}(S_0 \dots S_n A_n S_{n+1}))_{n \in \mathbb{N}}$  switches infinitely often between  $\mathbf{A}_1$  and  $\mathbf{A}_2$  thus  $\tilde{s}$  is visited infinitely often. Moreover, in this case  $\Pi$  is a shuffle of  $\Pi_1$  and  $\Pi_2$  and since  $f$  is submixing, (25) 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  $\tau_1^\#$  and  $\tau_2^\#$  for Player 2 in the subgames  $\mathbf{G}_1$  and  $\mathbf{G}_2$  respectively. We require  $\tau_1^\#$  and  $\tau_2^\#$  to be  $\epsilon$ -subgame-perfect in the corresponding subgames; their existence is guaranteed by Theorem 1.2.

**Definition 5.5.** *The merge strategy  $\tau^\#$  is a strategy in  $\mathbf{G}$  for Player 2 which ensures that  $\Pi_1$  is consistent with  $\tau_1^\#$  and  $\Pi_2$  is consistent with  $\tau_2^\#$  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*

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

The merge strategy is well-defined because if  $\text{last}(h) \in \mathbf{A}_1$  then both  $h$  and  $\pi_1(h)$  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 $\mathbf{G}_2$

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  $\epsilon$  to be small enough so that  $\tau_2$  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 moreover  $\epsilon$  is strictly smaller than  $\delta(s, a)$ , then  $\tau_2^\#$  never selects the action  $a$  in a play ending in state  $s$ .

*Proof.* Since the payoff function is prefix-independent, 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  $\epsilon'$ -optimal strategy from the states  $t$  such that  $p(s, a, t) > 0$ , for an arbitrary  $\epsilon' > 0$ . Assume  $\epsilon$  strictly smaller than  $\delta(s, a)$ . Then  $\tau_2^\#$  never selects  $a$  in state  $s$ , otherwise this would contradict the  $\epsilon$ -subgame perfection of  $\tau_2$ : 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  $\text{val}(\mathbf{G}_2)(s) + \epsilon$ .  $\square$

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

$$\mathbb{E}_{\tilde{s}}^{\sigma, \tau_2^\#} [f \mid \text{Stay}_\omega(\mathbf{G}_2)] \leq \text{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon . \quad (26)$$

*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 vertex of  $\pi_2(S_0 A_1 \dots S_n)$ . 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  $\pi_2$  does. Since  $\Pi_2$  is consistent with  $\tau_2^\#$  and since  $\tau_2^\#$  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  $h = s_0 a_0 \dots s_n$  in  $\mathbf{G}$  starting in  $\tilde{s}$  and consistent with  $\sigma$  and  $\tau^\#$ , denote

$$\phi(h) = \mathbb{P}_{\tilde{s}}^{\sigma, \tau_2^\#}(\text{Stay}_{\geq n}(\mathbf{G}_2) \mid h \text{ is a prefix of the play}) .$$

Fix some  $\epsilon' > 0$  and denote  $T$  the stopping time

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

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

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau_2^\#}(\text{Stay}_\omega(\mathbf{G}_2) \mid T < \infty) \geq 1 - \epsilon' \quad (27)$$

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau_2^\#}(T < \infty \mid \text{Stay}_\omega(\mathbf{G}_2)) = 1 . \quad (28)$$

The inequality (27) holds because by definition of  $\phi$ , for every  $n \in \mathbb{N}$ ,

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

We show (28). Fix  $\epsilon'' > 0$ . By definition of  $\text{Stay}_\omega(\mathbf{G}_2)$ , there exists  $n_1 \in \mathbb{N}$  such that

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (\text{Stay}_{\geq n_1}(\mathbf{G}_2) \mid \text{Stay}_\omega(\mathbf{G}_2)) \geq 1 - \epsilon'' . \quad (29)$$

According to Lévy's 0-1 law (see *e.g.* [Wil91, Theorem 14.4]), the sequence of random variables  $\left( \mathbb{E}_{\tilde{s}}^{\sigma, \tau^\#} [\text{Stay}_{\geq n_1}(\mathbf{G}_2) \mid S_0, \dots, S_n] \right)_{n \in \mathbb{N}}$  almost-surely converges to the indicator function  $\mathbf{1}_{\text{Stay}_{\geq n_1}(\mathbf{G}_2)}$ . Thus,

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

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

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

Equivalently,

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

and with (29) we get

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

This holds for every  $\epsilon'' > 0$ , hence (28).

Since  $\epsilon' > 0$  can be chosen arbitrarily small, then according to (27) and (28), to show our goal (26), it is enough to establish:

$$\mathbb{E}_{\tilde{s}}^{\sigma, \tau^\#} [f \mid T < \infty] \leq \text{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon + 2\epsilon' \cdot \|f\|_\infty . \quad (30)$$

This is well-defined, because (28) ensures  $\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (T < \infty) \geq \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (\text{Stay}_\omega(\mathbf{G}_2)) > 0$ , 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 [Wil91, 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 (30) from the following three inequalities:

$$\mathbb{E}_{\tilde{s}}^{\sigma, \tau^\#} [V_T] \leq \text{val}(\mathbf{G}_2)(\tilde{s}) \quad (31)$$

$$\mathbb{E}_{\tilde{s}}^{\sigma, \tau^\#} [V_T \mid T = \infty] = \text{val}(\mathbf{G}_2)(\tilde{s}) \quad (32)$$

$$\mathbb{E}_{\tilde{s}}^{\sigma, \tau^\#} [f \mid T < \infty] \leq \mathbb{E}_{\tilde{s}}^{\sigma, \tau^\#} [V_T \mid T < \infty] + \epsilon + 2\epsilon' \cdot \|f\|_\infty . \quad (33)$$

Assuming (31) and (32) do hold, then  $\mathbb{E}_{\tilde{s}}^{\sigma, \tau^\#} [V_T \mid T < \infty] \leq \text{val}(\mathbf{G}_2)(\tilde{s})$ . Injecting this inequality in (33), we get (30), and the lemma is proved.

We prove the three inequalities (31)- (33). The inequality (31) is obtained using the equality  $V_0 = \text{val}(\mathbf{G}_2)(\tilde{s})$  and applying Doob's Optional Stopping Theorem (Section 10.10 in [Wil91]) to the bounded super-martingale  $(V_n)_{n \in \mathbb{N}}$  and the stopping time  $T$ , which implies  $\mathbb{E}_{\tilde{s}}^{\sigma, \tau^\#} [V_T] \leq V_0$ .

To prove (32), we prove an even stronger statement:

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

If  $T = \infty$  then, according to (28), the event  $\text{Stay}_\omega(\mathbf{G}_2)$  does not hold. Thus, according to Lemma 5.4, either  $\text{Stay}_\omega(\mathbf{G}_1)$  or  $\text{Switch}$  holds. In the first case,  $(V_n)_{n \in \mathbb{N}}$  is ultimately constant equal to  $\text{val}(\mathbf{G}_2)(\tilde{s})$ . In the second case, the play  $\Pi_2$  visits  $\tilde{s}$  infinitely often. Since  $(V_n)_n$  converges almost-surely to  $V_T$  then  $V_T = \text{val}(\mathbf{G}_2)(\tilde{s})$ .

Finally, we prove (33). 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  $\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_T = h) > 0$ . Denote  $t$  the last state of  $h$ . Let  $\sigma_0$  be strategy in  $\mathbf{G}_2$  which coincides with  $\sigma[h]$  as long as the play stays in  $\mathbf{G}_2$ . Then:

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

The first equality holds because  $\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_T = h) > 0$  thus no strict prefix  $h'$  of  $h$  satisfies  $\phi(h') \geq 1 - \epsilon$ , and if  $h$  is a prefix of the play then  $h_T = h$ . The second equality holds by prefix-independence of  $f$ . The first inequality holds because  $\phi(h) \geq 1 - \epsilon'$  thus the strategies  $\sigma[h]$  and  $\sigma_0$  coincide with probability  $\geq 1 - \epsilon'$ , and when they do not the payoff difference is at most  $2\|f\|_\infty$ . The third equality holds because  $\tau[h]$  coincides with  $\tau_2^\#[\pi_2(h)]$  when the play stays in  $\mathbf{G}_2$ . The second inequality is by  $\epsilon$ -subgame optimality of  $\tau_2^\#$  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 (33) follows.  $\square$

## 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 prefix-independent and submixing. For all strategies  $\sigma$ ,*

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (f \leq \max\{\text{val}(\mathbf{G}_1)(\tilde{s}), \text{val}(\mathbf{G}_1)(\tilde{s})\} + \epsilon \mid \text{Switch}) = 1. \quad (34)$$

*Proof.* By definition of Switch, if Switch occurs then both projections  $\Pi_1$  and  $\Pi_2$  are infinite and visit  $\tilde{s}$  infinitely often. According to the inequality (25) in Lemma 5.4, to prove (34) it is enough to show, for every  $i \in \{1, 2\}$ ,

$$\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (f(\Pi_i) \leq \text{val}(\mathbf{G}_i)(s) + \epsilon \mid \Pi_i \text{ is infinite and reaches } \tilde{s} \text{ infinitely often}) = 1. \quad (35)$$

By symmetry, it is enough to show (35) when  $i = 1$ . For that, we define a strategy  $\sigma_1$  in  $\mathbf{G}_1$  such that for every measurable event  $\mathcal{E}_1$  in the game  $\mathbf{G}_1$ ,

$$\mathbb{P}_{\tilde{s}}^{\sigma_1, \tau_1^\#} (\mathcal{E}_1) \geq \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (\Pi_1 \text{ is infinite and } \Pi_1 \in \mathcal{E}_1). \quad (36)$$

Denote by  $\leq$  (respectively  $<$ ) the prefix relation (respectively strict prefix) over finite or infinite plays. The strategy  $\sigma_1$  in  $\mathbf{G}_1$  is defined as:

$$\sigma_1(h_1)(a) = \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 a \leq \Pi_1 \mid h_1 < \Pi_1),$$

if  $\mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 < \Pi_1) > 0$  and otherwise  $\sigma_1(h_1)$  is chosen arbitrarily. The event  $h_1 < \Pi_1$  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, so  $\sigma_1$  is equivalently defined as

$$\sigma_1(h_1)(a) = \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 a \leq \Pi_1 \mid \exists b \in \mathbf{A}, h_1 b \leq \Pi_1) .$$

Remark that in general,  $\sigma_1$  is a mixed strategy.

We proceed with the proof of (36). Let  $\mathfrak{E}$  be the set of measurable events  $\mathcal{E}_1$  in  $\mathbf{G}_1$  for which (36) 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:

$$\begin{aligned} \mathbb{P}_{\tilde{s}}^{\sigma_1, \tau_1^\#} (h_1) &\geq \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 \leq \Pi_1) \\ &\geq \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (\Pi_1 \text{ is infinite and } \Pi_1 \in h_1(\mathbf{SA})^\omega). \end{aligned} \quad (37)$$

We abuse the notation and denote  $h_1$  the event  $\{h_1 \text{ is a prefix of the play}\}$ . The second inequality is by definition of prefixes. The inequality (37) is proved by induction on the length of  $h_1$ . When  $h_1$  is the single initial state  $\tilde{s}$  then both terms in (37) 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 (37) 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

$$\begin{aligned} \mathbb{P}_{\tilde{s}}^{\sigma_1, \tau_1^\#} (h_1 ar) &= \mathbb{P}_{\tilde{s}}^{\sigma_1, \tau_1^\#} (h_1) \cdot \sigma_1(h_1)(a) \cdot p(t, a, r) \\ &\geq \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 \leq \Pi_1) \cdot \sigma_1(h_1)(a) \cdot p(t, a, r) \\ &= \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 \leq \Pi_1) \cdot \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 a \leq \Pi_1 \mid h_1 < \Pi_1) \cdot p(t, a, r) \\ &\geq \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 < \Pi_1) \cdot \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 a \leq \Pi_1 \mid h_1 < \Pi_1) \cdot p(t, a, r) \\ &= \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 a \leq \Pi_1) \cdot p(t, a, r) \\ &= \mathbb{P}_{\tilde{s}}^{\sigma, \tau^\#} (h_1 ar \leq \Pi_1) , \end{aligned} \quad (38)$$

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  $\sigma_1$ . The second inequality (38) holds because the event  $h_1 < \Pi_1$  is contained in the event  $h_1 \leq \Pi_1$ . This inclusion and the corresponding 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  $\tilde{s} < \Pi_1$  has probability 0 while the event  $\tilde{s} \leq \Pi_1$  has probability 1.

Now we prove inequality (37), 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  $\pi_1(h') = h'_1$  and  $\text{last}(h') \in \mathbf{A}_1$ . Equivalently,  $h'$  belongs  $C(h'_1)$  if and only  $\pi_1$  projects  $h'$  on  $h'_1$ , but no strict prefix of  $h'$  is projected on  $h'_1$ .

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

The first equality is by definition of the probability measure. The inequality is by induction hypothesis. The second equality holds because the event  $h_1 \leq \Pi_1$  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  $\tau^\#$ . The fourth equality is by definition of the probability measure. To show the fifth equality, we establish  $C(h_1 ar) = \{h' ar \mid h' \in C(h_1)\}$ . We start with the inclusion  $\{h' ar \mid h' \in C(h_1)\} \subseteq C(h_1 ar)$ . Let  $h' \in C(h_1)$ . Since  $\pi_1(h') = h_1$  and  $\text{last}(h') \in \mathbf{A}_1$  then  $h'$  and  $h_1$  have the same last state, i.e.  $t$ . And  $t$  is controlled by Player 2, hence  $t \neq \tilde{s}$ . Thus  $\text{last}(h' ar) = \text{last}(h') \in \mathbf{A}_1$  and  $\pi_1(h' ar) = \pi_1(h') ar$ . For the converse inclusion  $C(h_1 ar) \subseteq \{h' ar \mid h' \in C(h_1)\}$  take  $h'' \in C(h_1 ar)$  and write  $h'' = h' a' r'$ , where  $a'$  and  $r'$  are the last action and state of  $h''$ . Since  $\text{last}(h'') \in \mathbf{A}_1$  then  $\pi_1(h'') = \pi_1(h') a' r'$ . Since  $h'' \in C(h_1 ar)$  then  $\pi_1(h'') = h_1 ar$ , hence  $\pi_1(h') = h_1$  and  $a = a'$  and  $r = r'$  hence  $h'' \in C(h_1) ar$ . This completes the proof of the inequality (37).

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 [Bil08, 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  $G_1$ . This completes the proof of (36).

Next we prove that

$$\mathbb{P}_{\tilde{s}}^{\sigma_1, \tau_1^\#} \left( f \leq \liminf_n \text{val}(G_1)(S_n) + \epsilon \right) = 1. \quad (39)$$

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

$$\mathbb{E}_{\tilde{s}}^{\sigma_1, \tau_1^\#} [f \mid S_0, A_0, \dots, S_n] = \mathbb{E}_{S_n}^{\sigma_1[S_0 \cdots S_n], \tau_1^\#[S_0 \cdots S_n]} [f] \leq \text{val}(G_1)(S_n) + \epsilon,$$

and as a consequence,

$$\liminf_n \mathbb{E}_{\tilde{s}}^{\sigma_1, \tau_1^\#} [f \mid S_0, A_0, \dots, S_n] \leq \liminf_n \text{val}(G_1)(S_n) + \epsilon. \quad (40)$$

According to Lévy's 0-1 law (see e.g. [Wil91, Theorem 14.4]), the sequence of random variables:  $(\mathbb{E}_{\tilde{s}}^{\sigma_1, \tau_1^\#} [f \mid S_0, A_0, \dots, S_n])_{n \in \mathbb{N}}$  converges point-wise to the random variable  $f(S_0 A_0 S_1 \cdots)$ . As a consequence the left handside of (40) is almost-surely equal to  $f$  and we get (39).

Denote  $\mathcal{E}_1$  the event

$$\mathcal{E}_1 = \{f > \text{val}(G_1)(\tilde{s}) + \epsilon \text{ and } \tilde{s} \text{ is reached infinitely often}\}.$$

According to (39),  $\mathbb{P}_{\tilde{s}}^{\sigma_1, \tau_1^\#}(\mathcal{E}_1) = 0$ . We apply (36) to  $\mathcal{E}_1$  and get

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

By definition of  $\mathcal{E}_1$ , this last equality is equivalent to (35) with  $i = 1$ .  $\square$

## 5.6 Proof of Theorem 5.3

*Proof of Theorem 5.3.* To prove the first statement (21) in Theorem 5.3, we combine the two lemmas proved in the two previous sections in order to show:

$$\forall \sigma, \mathbb{E}_{\tilde{s}}^{\sigma, \tau^\#} [f] \leq \max\{\text{val}(G_1)(\tilde{s}), \text{val}(G_2)(\tilde{s})\} + \epsilon. \quad (41)$$

The bound (41) can be obtained as follows. According to Lemma 5.4, the three events  $\text{Stay}_\omega(G_1)$ ,  $\text{Stay}_\omega(G_2)$  and  $\text{Switch}$  partition the set of infinite plays. In case  $\text{Stay}_\omega(G_1)$  occurs, Lemma 5.7 guarantees that the expected payoff is no more than  $\text{val}(G_1)(\tilde{s}) + \epsilon$ . By symmetry, in case  $\text{Stay}_\omega(G_2)$  occurs, the expected payoff is no more than  $\text{val}(G_2)(\tilde{s}) + \epsilon$ . And in case  $\text{Switch}$  occurs, Lemma 5.8 guarantees that the payoff is almost-surely no more than  $\max\{\text{val}(G_1)(\tilde{s}), \text{val}(G_2)(\tilde{s})\} + \epsilon$ . Thus (41) holds. The inequality

$$\text{val}(G)(\tilde{s}) \geq \max\{\text{val}(G_1)(\tilde{s}), \text{val}(G_2)(\tilde{s})\}$$

is clear, since Player 1 has more choice in  $G$  than he has in  $G_1$  and  $G_2$ . And  $\epsilon$  can be chosen arbitrarily small in (41), hence the first statement (21) of Theorem 5.3.

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

$$\text{val}(\mathbf{G}_1)(\tilde{s}) \geq \text{val}(\mathbf{G}_2)(\tilde{s}) . \quad (42)$$

We have to show (22), i.e.

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

According to (21), we already now 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 (42). 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,

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

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  $\tau^\#$  guarantees a payoff smaller than  $\text{val}(\mathbf{G}_1)(s) + \epsilon$  for every state  $s$ . Fix  $\sigma$  a strategy for Player 1 in  $\mathbf{G}$ , and define  $\sigma'$  to be the strategy that plays like  $\sigma$  as long as the play does not reach the pivot state  $\tilde{s}$ . Whenever the pivot state is reached, the strategy  $\sigma'$  switches definitively to a strategy  $\sigma_1^\#$  that is optimal in the game  $\mathbf{G}_1$ , whose existence is guaranteed by the induction hypothesis. The plays consistent with  $\sigma'$  and  $\tau^\#$  stay in the subgame  $\mathbf{G}_1$ . Since  $\tau^\#$  coincides with  $\tau_1^\#$  on plays staying in  $\mathbf{G}_1$ , and since  $\tau_1^\#$  is  $\epsilon$ -optimal in  $\mathbf{G}_1$ , we can write for all  $s \in \mathbf{S}$ :

$$\mathbb{E}_s^{\sigma', \tau^\#} [f] = \mathbb{E}_s^{\sigma', \tau_1^\#} [f] \leq \text{val}(\mathbf{G}_1)(s) + \epsilon . \quad (43)$$

Let  $h$  be a finite play that is consistent with  $\sigma$  and  $\tau^\#$ , whose last state is  $\tilde{s}$  and which does not visit  $\tilde{s}$  before the last step. Then

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

The first and third equalities hold because  $f$  is prefix-independent. The first inequality holds because the strategy  $\tau^\#[h]$  is  $\epsilon$ -optimal from state  $\tilde{s}$ , for the following reason.

The strategy  $\tau^\# [h]$  coincides with the strategy obtained by merging  $\tau_1^\# [h]$  and  $\tau_2^\#$  on the pivot state  $\tilde{s}$ , both of which are  $\epsilon$ -subgame-perfect in the respective subgames. Since (41) was proved for any merge of two  $\epsilon$ -subgame-perfect strategies, we can apply (41) to the strategy  $\tau^\# [h]$ , and conclude that the latter is  $\epsilon$ -optimal from state  $\tilde{s}$ . The second inequality holds because  $\sigma_1^\#$  is optimal in  $G_1$ . The second equality holds because  $\sigma' [h] = \sigma_1^\#$ . Finally

$$\mathbb{E}_s^{\sigma, \tau^\#} [f | h] \leq \mathbb{E}_s^{\sigma', \tau^\#} [f | h] + \epsilon . \quad (44)$$

Since the strategies  $\sigma$  and  $\sigma'$  coincide on those plays that never reach  $\tilde{s}$ , and (44) holds for all finite plays reaching  $\tilde{s}$  for the first time, then for every  $s \in S$ ,

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

By using (43) now we have that for all  $s$ ,

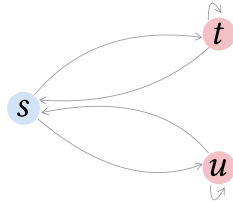
$$\mathbb{E}_s^{\sigma, \tau^\#} [f] \leq \text{val}(G_1)(s) + 2\epsilon . \quad (45)$$

This holds for every strategy  $\sigma$  and  $\epsilon > 0$  arbitrarily small, thus  $\text{val}(G)(s) \leq \text{val}(G_1)(s)$ . The converse inequality is obvious, because Player 1 has more freedom in  $G$  than in  $G_1$ , hence the second statement (22) of Theorem 5.3.  $\square$

**Remarks about the merge strategy.** We observe a byproduct of the proof of Theorem 5.3, namely that (45) has yielded  $2\epsilon$ -optimality of the merge strategy:

**Observation 5.9.** *The merge strategy  $\tau^\#$  constructed with  $\epsilon$ -subgame-perfect pieces is  $2\epsilon$ -optimal in the game  $G$ .*

After this observation, since the merge strategy is obtained by merging two  $\epsilon$ -subgame-perfect strategies, a natural question to ask is whether  $\tau^\#$  is  $2\epsilon$ -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 states are controlled by Player 1, and the red ones by his opponent. In the subgame  $G_1$  we remove the action  $s \rightarrow t$ . In particular in the game  $G_1$  the positional strategy  $\tau_1^\#$  which chooses  $u \rightarrow s$  and  $t \rightarrow s$  is subgame-perfect. We can therefore use it to construct a merge strategy  $\tau^\#$ . However this merge strategy is not  $2\epsilon$ -subgame-perfect, since in case Player 1 uses the suboptimal action  $s \rightarrow u$ , his opponent does not profit by taking the self-loop forever.



## 6 From One-player Games to Two-player Games

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.

A general result allows to lift the existence of  $\epsilon$ -optimal strategies from  $S$  in one-player games to two-player games.

An arena is said to be *fully controlled by the minimizer* if all states are controlled by Player 2. Fix a payoff function  $f$  that is both shift-invariant and submixing.

**Definition 6.1.** *Let  $S$  be a class of strategies for minimizer.*

*Say that the class  $S$  is stable by the reset operation if for every game  $G$  equipped with  $f$  and every strategy  $\tau$  of minimizer in  $G$ , if  $\tau$  belongs to  $S$  then the reset strategy  $\hat{\tau}$  belongs to  $S$  as well.*

*Say that the class  $S$  is stable by the merge operation if for every game  $G$  equipped with  $f$ , for every split  $(G_1, G_2)$  of  $G$  and for every strategies  $\tau_1^\#$  and  $\tau_2^\#$  in  $G_1$  and  $G_2$ , if both  $\tau_1^\#$  and  $\tau_2^\#$  belong to  $S$  then their merge  $\tau^\#$  belongs to  $S$  as well.*

Like in Proposition 7.1, say that the arena  $\mathcal{A}'$  is a *restriction* of the arena  $\mathcal{A}$  if one gets  $\mathcal{A}'$  from  $\mathcal{A}$  by erasing some actions from some states.

**Theorem 6.2.** *Let  $f$  a shift-invariant payoff function,  $\mathbb{A}$  a family of arenas that are closed under restrictions and  $\mathbb{S}$  a family of strategies for minimizer which are stable by both reset and merge operations.*

*Assume that in every game  $(A, f)$  with  $A \in \mathbb{A}$  that is fully controlled by the minimizer, for every  $\epsilon > 0$  the minimizer has an  $\epsilon$ -optimal strategy that belongs to  $\mathbb{S}$ . Then, in every two-player game  $(A, f)$  with  $A \in \mathbb{A}$ , the minimizer has an  $\epsilon$ -subgame perfect strategy that belongs to  $\mathbb{S}$ .*

*The statement holds for  $\epsilon = 0$  as well, that is: assume that in all games  $(A, f)$  with  $A \in \mathbb{A}$  that is fully controlled by the minimizer, the minimizer has an optimal strategy that belongs to  $\mathbb{S}$ . Then, in every two-player game  $(A, f)$  with  $A \in \mathbb{A}$ , the minimizer has a subgame perfect strategy that belongs to  $\mathbb{S}$ .*

*Proof.* Let  $G = (A, f)$  with  $A \in \mathbb{A}$ . The proof of both statements is by induction on  $N(G)$ , as in the proof of the main theorem in the previous section.

The base of the induction follows from the assumption about games fully controlled by the minimizer, since we can give to the minimizer the control of states in which the maximizer has a single action, without changing the value of the game.

When  $N(G) > 0$ , the induction step is performed using a split  $(G_1, G_2)$  of  $G$  on a pivot state  $\tilde{s}$ . Remark that both arenas belong to  $\mathbb{A}$ , therefore the induction hypothesis for the first (resp. the second) statement says that for every  $\epsilon > 0$ , there are two strategies  $\tau_1^\#$  and  $\tau_2^\#$  in the games  $G_1$  and  $G_2$ , respectively, which belong to  $\mathbb{S}$  and are  $\epsilon$ -subgame perfect (resp. subgame perfect) in their respective subgames. Since  $\mathbb{S}$  is stable by the merge operation, then the strategy  $\tau^\#$  obtained by merging  $\tau_1^\#$  and  $\tau_2^\#$  also belongs to  $\mathbb{S}$ .

We carry over the induction step for the first (resp. the second) statement. According to Observation 5.9,  $\tau^\#$  is  $2\epsilon$ -optimal (resp. is optimal). We apply Lemma 4.17 to  $\tau^\#$  which guarantees that the reset strategy obtained from  $\tau^\#$  is  $4\epsilon$ -optimal (resp. is optimal). Moreover by hypothesis this strategy belongs to  $S$ .  $\square$

## 7 The Finite Memory Transfer Theorem

We give the proof of Theorem 1.3 that was announced in the introduction.

**Theorem 1.3.** *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, for every  $\epsilon > 0$ , the minimizer has an  $\epsilon$ -optimal strategy with finite memory. Then in every (two-player) game, for every  $\epsilon > 0$ , the minimizer has an  $\epsilon$ -subgame-perfect strategy that has finite memory.*

*The statement also holds for  $\epsilon = 0$ , that is: if the minimizer has an optimal strategy with finite memory in every game that he fully controls, then in every (two-player) game as well he has a subgame-perfect strategy with finite memory.*

Theorem 1.3 follows from Theorem 6.2 and the following results, which establish that the class of finite-memory strategies is stable by the reset (Proposition 7.1) and merge (Lemma 7.2) operations.

**Proposition 7.1.** *Let  $\mathcal{A}$  be a family of arenas that are closed under restrictions and  $f$  a shift-invariant payoff function. If for games whose arena is in  $\mathcal{A}$  and whose payoff function is  $f$ , and for every  $\epsilon > 0$ , Player 1 (respectively Player 2) has an  $\epsilon$ -optimal strategies  $\sigma$  with finite memory, then he also has an  $\epsilon$ -subgame-perfect strategies with finite memory, namely the reset strategies  $\hat{\sigma}$ . This holds as well for optimal strategies, i.e. if  $\epsilon = 0$ .*

*Proof.* Let  $\mathcal{A} \in \mathcal{A}$  be an arena. Remove the actions of Player 1 that are not locally optimal (with respect to the payoff function  $f$ ) to get a restriction  $\mathcal{A}'$ . From the hypothesis, it follows that there are  $\epsilon$ -optimal strategies in  $(\mathcal{A}', f)$  that have finite memory, and consequently there are  $\epsilon$ -optimal strategies in  $(\mathcal{A}, f)$  that have finite memory and are locally optimal. According to Lemma 4.17, the strategy  $\hat{\sigma}$  is  $2\epsilon$ -subgame-perfect, and Proposition 7.3 implies that it has finite memory.  $\square$

**Lemma 7.2.** *Let  $(G_1, G_2)$  be a split of a game  $G$  on a pivot state  $\tilde{s}$ . Let  $\tau_1^\#$  and  $\tau_2^\#$  two strategies for Player 2 in  $G_1$  and  $G_2$ , respectively. If both  $\tau_1^\#$  and  $\tau_2^\#$  have finite-memory then  $\tau^\#$  has finite memory as well.*

*Proof.* The strategies  $\tau_1^\#$  and  $\tau_2^\#$  with finite memory are given by the transducers:

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

for Player 2 in  $G_1$  and  $G_2$  respectively.

The strategy  $\tau^\#$  obtained by merging  $\tau_1^\#$  and  $\tau_2^\#$  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  $(1, \text{init}_1(s), \text{init}_2(s))$ . The updates on the components  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are performed with  $\text{up}_1$  and  $\text{up}_2$  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  $A_1$  or  $A_2$ . The choice of action, or the output, depends on the first component: in memory state  $(b, m_1, m_2)$  the action played by  $\tau^\#$  is  $\text{out}_b(m_b)$ .  $\square$

The finite-memory property is preserved when passing from  $\sigma$  to  $\hat{\sigma}$  that is if the strategy  $\sigma$  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  $\sigma$  is said to have *finite memory* if it is given using a transducer, namely it is a tuple:

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

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

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

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

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

**Proposition 7.3.** *Let  $\epsilon > 0$ . If  $\sigma$  is a finite memory strategy and is  $\epsilon$ -optimal then the  $\epsilon$ -reset of  $\sigma$  has finite-memory as well.*

*Proof.* Let  $\sigma$  be a finite memory strategy, that is given by the tuple

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

and let  $\epsilon$  be such that  $\sigma$  is  $\epsilon$ -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:

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

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

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

We gather the subset of memory states where drops occur as follows. For  $s \in S$  and  $m \in \mathcal{M}_s$ , denote by  $\sigma_m$  the strategy that is the same as  $\sigma$  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

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

Construct the finite memory strategy  $\sigma'$  that avoids the memory states in  $\mathcal{D}$  as follows. For any  $s \in \mathbb{S}$  and  $m \in \mathcal{M}_s \cap \mathcal{D}$ , since  $\sigma$  is  $\epsilon$ -optimal,  $m \neq \text{init}(s)$ . In the strategy  $\sigma'$  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  $\sigma'$  we conclude that they coincide.  $\square$

**On the size of the memory.** How large is the memory  $\mathcal{M}_G$  needed by Player 2 to play optimally in some  $G = (\mathcal{A}, f)$ ? Every deterministic and stationary strategy  $\sigma$  for Player 1 in  $G$  induces a game  $G_\sigma$  that is controlled by Player 2. Let  $\mathfrak{M}$  be the maximal memory size required by Player 2 to play optimally in the games  $G_\sigma$ . According to the proof of the theorem above, the memory  $\mathcal{M}_G$  needed by Player 2 to play optimally in  $G$  is of size  $2 \cdot |\mathcal{M}_{G_1}| \cdot |\mathcal{M}_{G_2}|$ . By induction we derive the following bound:

$$|\mathcal{M}_G| \leq (2\mathfrak{M})^{2^{\sum_s |\mathcal{A}(s)|}}.$$

When  $\mathfrak{M} = 1$ , *i.e.* when Player 2 has deterministic and stationary strategies in games he controls, then in [GZ05] it is shown that the same holds for two player games as well, hence the upper-bound can be downsized to 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.

## Acknowledgments

We are very grateful to Pierre Vandenholve for finding an error in the previous version of the proof of Lemma 5.7. This work was supported by the ANR projet "Stoch-MC" and the LaBEX "CPU".

## References

- [BBE10] Tomáš Brázdil, Václav Brozek, and Kousha Etessami. One-counter stochastic games. In *FSTTCS*, pages 108–119, 2010.
- [Bie87] K.-J. Bierth. An expected average reward criterion. *Stochastic Processes and Applications*, 26:133–140, 1987.
- [Bil08] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.
- [BKW18] N. Basset, M. Kwiatkowska, and C. Wiltsche. Compositional strategy synthesis for stochastic games with multiple objectives. *Information and Computation*, 261:536 – 587, 2018. Strategic Reasoning 2015.

- [BRO<sup>+</sup>20] Patricia Bouyer, Stéphane Le Roux, Youssef Oualhadj, Mickael Randour, and Pierre Vandenholte. Games where you can play optimally with arena-independent finite memory. In Igor Konnov and Laura Kovács, editors, *31st International Conference on Concurrency Theory, CONCUR 2020, September 1-4, 2020, Vienna, Austria (Virtual Conference)*, volume 171 of *LIPICs*, pages 24:1–24:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.
- [BvdB15] Dietmar Berwanger and Marie van den Bogaard. Games with delays - A frankenstein approach. In Prahladh Harsha and G. Ramalingam, editors, *35th IARCS Annual Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS 2015, December 16-18, 2015, Bangalore, India*, volume 45 of *LIPICs*, pages 307–319. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015.
- [CD16] Krishnendu Chatterjee and Laurent Doyen. 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*, page 247–256, New York, NY, USA, 2016. Association for Computing Machinery.
- [CDHR10] Krishnendu Chatterjee, Laurent Doyen, Thomas A. Henzinger, and Jean-François Raskin. Generalized mean-payoff and energy games. In *FSTTCS*, pages 505–516, 2010.
- [CHJ05] K. Chatterjee, T.A. Henzinger, and M. Jurdzinski. Mean-payoff parity games. In *Proc. of LICS'05*, pages 178–187. IEEE, 2005.
- [CJH03] K. Chatterjee, M. Jurdzinski, and T.A. Henzinger. Quantitative stochastic parity games. In *SODA*, 2003.
- [CY90] C. Courcoubetis and M. Yannakakis. Markov decision processes and regular events. In *Proceedings of ICALP'90*, volume 443 of *Lecture Notes in Computer Science*, pages 336–349. Springer, 1990.
- [Der62] Cyrus Derman. On sequential decisions and markov chains. *Management Science*, 9:16–24, 1962.
- [FHMP21] János Flesch, P Herings, Jasmine Maes, and Arkadi Predtetchinski. Subgame maxmin strategies in zero-sum stochastic games with tolerance levels. *Dynamic Games and Applications*, 11(4):704–737, 2021.
- [Gil57] D. Gillette. *Stochastic Games with Zero Stop Probabilities*, volume 3. Princeton University Press, 1957.
- [Gim06] H. Gimbert. *Jeux Positionnels*. PhD thesis, Université Denis Diderot, Paris, 2006.
- [Gim07] Hugo Gimbert. Pure stationary optimal strategies in markov decision processes. In *STACS*, pages 200–211, 2007.

- [GK14] Hugo Gimbert and Edon Kelmendi. Two-player perfect-information shift-invariant submixing stochastic games are half-positional. *CoRR*, abs/1401.6575, 2014.
- [GTW02] E. Grädel, W. Thomas, and T. Wilke. *Automata, Logics and Infinite Games*, volume 2500 of *Lecture Notes in Computer Science*. Springer, 2002.
- [GZ04] H. Gimbert and W. Zielonka. When can you play positionally? In *Proc. of MFCS'04*, volume 3153 of *Lecture Notes in Computer Science*, pages 686–697. Springer, 2004.
- [GZ05] H. Gimbert and W. Zielonka. Games where you can play optimally without any memory. In *Proceedings of CONCUR'05*, volume 3653 of *Lecture Notes in Computer Science*, pages 428–442. Springer, 2005.
- [GZ09] Hugo Gimbert and Wieslaw Zielonka. Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games. *HAL archives ouvertes*, hal-00438359, December 2009.
- [GZ16] Hugo Gimbert and Wieslaw Zielonka. Pure and stationary optimal strategies in perfect-information stochastic games with global preferences. *CoRR*, abs/1611.08487, 2016.
- [Kop06] Eryk Kopczynski. Half-positional determinacy of infinite games. In *ICALP (2)*, pages 336–347, 2006.
- [Kop09] Eryk Kopczynski. *Half-positional determinacy of infinite games*. PhD thesis, University of Warsaw, 2009.
- [LL69] T.S. Liggett and S.A. Lippman. Stochastic games with perfect information and time average payoff. *SIAM Review*, 11(4):604–607, 1969.
- [Mar98] D.A. Martin. The determinacy of Blackwell games. *Journal of Symbolic Logic*, 63(4):1565–1581, 1998.
- [MS96] A.P. Maitra and W.D. Sudderth. *Discrete gambling and stochastic games*. Springer-Verlag, 1996.
- [MSTW21] Richard Mayr, Sven Schewe, Patrick Totzke, and Dominik Wojtczak. Simple Stochastic Games with Almost-Sure Energy-Parity Objectives are in NP and coNP. *arXiv e-prints*, page arXiv:2101.06989, January 2021.
- [MY15] Ayala Mashiah-Yaakovi. Correlated equilibria in stochastic games with borel measurable payoffs. *Dynamic Games and Applications*, 5(1):120–135, 2015.
- [NS03] A. Neyman and S. Sorin. *Stochastic games and applications*. Kluwer Academic Publishers, 2003.
- [Sha53] L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Science USA*, 39:1095–1100, 1953.

- [VTRF83] O.J. Vrieze, S.H. Tijs, T.E.S. Raghavan, and J.A. Filar. A finite algorithm for switching control stochastic games. *O.R. Spektrum*, 5:15–24, 1983.
- [Wil91] David Williams. *Probability with martingales*. Cambridge university press, 1991.
- [Zie04] Wiesław Zielonka. Perfect-information stochastic parity games. In *FOS-SACS 2004*, volume 2987 of *Lecture Notes in Computer Science*, pages 499–513. Springer, 2004.
- [Zie10] Wieslaw Zielonka. Playing in stochastic environment: from multi-armed bandits to two-player games. In *FSTTCS*, pages 65–72, 2010.