Submixing and Shift-Invariant Stochastic Games

Hugo Gimbert[1] and Edon Kelmendi[2]

[1]CNRS, LaBRI, Université de Bordeaux, France
[2]Oxford University, UK

April 30, 2021

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 neither memory nor randomisation is a desirable property that vastly simplifies the algorithmic analysis of such games. Our main theorem gives a sufficient condition for the maximizer to possess such a simple optimal strategy. The condition is imposed on the payoff function, saying the payoff does not depend on any finite prefix (shift-invariant) and combining two trajectories does not give higher payoff than the payoff of the parts (submixing). The core technical property that enables the proof of the main theorem is that of the existence of ϵ-subgame-perfect strategies when the payoff function is shift-invariant. Furthermore, the same techniques can be used to prove a finite-memory transfer-type theorem: namely that for shift-invariant and submixing payoff functions, the existence of optimal finite-memory strategies in one-player games for the minimizer implies the existence of the same in two-player games. We show that numerous classical payoff functions are submixing and shift-invariant.

Contents

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 wants the opposite.

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

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

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

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

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

$$s_0 s_1 s_2 s_3 \cdots$$

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

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

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

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

shuffling (or combining) them such as

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

does not give better payoff, that is:

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

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

As mentioned above, half-positional games are those where the *maximizer* has a simple kind of strategy that is optimal. There is nothing special about this player, if instead of the submixing condition, we define an "inverse" submixing condition, namely one that requires that the combined payoff is larger than the minimum of the parts, we would have an analogous theorem that proves the existence of simple optimal strategies for the *minimizer*. Furthermore there are payoff functions for which both versions of the submixing condition hold, and for these games the theorem proves positionality. The conditions in the statement of the theorem are not necessary; we will provide examples and discuss this fact. The proof of Theorem 1.1 is by induction on number of edges, it uses Lévy's 0-1 law, as well as the following second contribution.

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

Theorem 1.2. *Games equipped with a payoff function that is shift-invariant, for every $\epsilon > 0$, admit ϵ-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 is a corollary of the techniques developed for the previous two theorems and it is a transfer-type theorem that lifts the existence of optimal finite-memory strategies in one-player games to the same for two-player games.

Theorem 1.3. *For every payoff function f that is both shift-invariant and submixing, if the minimizer has optimal strategies with finite memory in one player games equipped with f, then the minimizer has the same for two-player games equipped with f.*

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

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.

Theorem 1.2 was proved independently and in parallel by Mashiah-Yaakovi in [MY15, Proposition 11] for concurrent games (which subsume the turn-based games that we are considering). However our proof can be seen as a sharpening of the result of Mashiah-Yaakovi for turn-based games, since our construction makes more transparent the fact that building ϵ-subgame-perfect strategies from ϵ-strategies conserves important properties such as having finite memory and not using randomisation.

Some results on finite-memory determinacy have been obtained in [BRO+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; however for Theorem 1.2 they observe that in the special case of finite-memory strategies there is a simpler

combinatorial proof, which bypasses our 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 result is also used by Mayr, Schewe, Totzke and Wojtczak on their proof of the fact that games with energy-parity objectives and almost-sure semantics lie in NP ∩ co-NP [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. The proof of Theorem 1.2, that there are ϵ-subgame-perfect strategies is given in Section 3. Besides the construction of the strategy, the rest of the proof is independent of the other results and can be read out of order. We give the proof of the main theorem, Theorem 1.1, in Section 4, and the transfer theorem for finite-memory strategies, Theorem 1.3, in Section 5. By instantiating the main theorem, we recover a few classical determinacy results and give a few other applications in Section 6.

2 Preliminaries and Proof Overview

In this section, we introduce the basic notions we need about stochastic games with perfect information, namely games, payoff functions, strategies and values. In the end of the section we give an overview of the proofs of the theorems that were introduced in the previous section.

Games A game is specified by the *arena* and the *payoff function*. While the arena determines *how* the game is played, the payoff function specifies the *objectives* that the players want to reach.

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

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

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

where

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

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

With each infinite play is associated a payoff computed by a *payoff function*. Player 1 (the maximizer) wants to maximize the expected payoff while Player 2 (the minimizer) has the exact opposite preference. Formally, a payoff function for the arena $\mathcal{A}$ is a bounded and Borel-measurable function

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

By definition of maxmin and minmax, for every state $s \in \mathbf{S}$, $\mathrm{maxmin}(\mathbf{G})(s) \leq \mathrm{minmax}(\mathbf{G})(s)$. 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* $s \in \mathbf{S}$*:*

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

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

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

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

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

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

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

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

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

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

This notion of ϵ-subgame-perfect strategy is not standard; usually one requires a much weaker condition, namely that for every finite play $h = s_0 \cdots s_n$,

$$\inf_{\tau} \mathbb{E}_{s_n}^{\sigma[h],\tau}[f[h]] \geq \sup_{\sigma'} \inf_{\tau} \mathbb{E}_{s_n}^{\sigma',\tau}[f[h]] - \epsilon. \tag{3}$$

Even though these two conditions coincide when the payoff function is shift-invariant, we prefer to use the stronger condition in Definition 2.6 for reasons that we will expound in Section 3.

Shift-invariant and submixing Without loss of generality we can assume that there is a finite set $\mathbf{C}$ (the colours) such that the payoff function f is a function

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

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

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

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

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

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

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

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

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

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

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

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

The submixing condition is not symmetric over the players, and it implies different results for different players (notice the difference between 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 $u, v, w \in \mathrm{C}^\omega$ 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.

2.1 Proof Overview

Subgame-Perfect Strategies In Section 3 we give the proof of Theorem 1.2. The proof of this theorem, while relatively technical, is conceptually simple. A strategy is ϵ-subgame-perfect if at whatever middle point we stop the game, the strategy is ϵ-optimal with respect to the state in the middle point as well. So the idea is to take some ϵ-optimal strategy σ, and if we happen to arrive at some juncture where by continuing to play with σ gives us smaller than ϵ-optimal payoff, with respect to the state in the juncture, we simply reset the memory. This causes no harm because the strategy σ (with its memory reset) is ϵ-optimal. The strategy based on σ, that behaves in such a manner is called the *reset* strategy. We prove that if we start with a strategy σ that is ϵ-optimal, and make a reset strategy based on it, then the latter will be 2ϵ-subgame-perfect. The main difficulty is showing that on every trajectory we only have to reset the memory finitely many times. For if we have this, intuitively for all junctures, after some point the strategy is behaving like a ϵ-subgame-perfect strategy.

Martingale theory proves to be useful for demonstrating that the memory needs to be reset only finitely many times. This is because under suitable strategies the stochastic process $\mathrm{val}(S_n)$, $n \in \mathbb{N}$ is a supermartingale, and martingale theory studies convergence and other properties of such processes.

The Main Theorem In Section 4 Theorem 1.1 is proved. The proof proceeds by induction on the number of actions of Player 1. In the base case he has no choice in any state, he has only one strategy that is optimal, so he certainly has a stationary and deterministic one. The induction step assumes the theorem for two smaller games $\mathbf{G}_1$ and $\mathbf{G}_2$, and we are to prove the theorem for the larger game $\mathbf{G}$. The smaller games are obtained by removing some actions of the maximizer in the larger game. We show that the value of the states in game $\mathbf{G}$ is no larger than that of the maximum of the values in the smaller games. For this the submixing property has to be used.

To show that Player 1 can gain no more in the large game than he would in the maximum of the smaller games, we employ a strategy for his opponent called the *merge strategy*. We start with ϵ-subgame-perfect strategies in each of the games $\mathbf{G}_1$, $\mathbf{G}_2$, and merge them together to form a strategy in the game $\mathbf{G}$, which makes sure that Player 1 will not receive more payoff in the larger game than he would in the smaller games.

Theorem 1.3 then comes almost as a corollary from the construction of the merge strategy that is used in the proof of the main theorem, together with the observations

that we make about the memory size of the reset strategy that is based on some strategy with finite memory.

3 ϵ-Subgame-Perfect Strategies

In this section we construct ϵ-subgame-perfect strategies in games with shift-invariant payoff functions. This property is crucial for the proof of the main theorem; furthermore it is interesting in itself.

Theorem 1.2. *Games equipped with a payoff function that is shift-invariant, for every $\epsilon > 0$, admit ϵ-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 ϵ-subgame-perfect strategies, even a game with only one player. However, for the weaker condition (3), the theorem is true for arbitrary payoff functions. 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 (3) coincides with that of Definition 2.6. Our proof on the other hand makes it transparent how to construct ϵ-subgame-perfect strategies from $\epsilon/2$-optimal ones, in a way that preserves some important properties of the strategy, namely 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 ϵ-subgame-perfect strategies. We will do this by taking an ϵ-optimal strategy σ with some more structure, and using it to construct a *reset* strategy $\hat{\sigma}$, which will be 2ϵ-subgame-perfect. The reset strategy is conceptually very simple: a strategy σ is not 2ϵ-subgame-perfect if and only if there exists some finite play $h := s_0 \cdots s_n$ such that

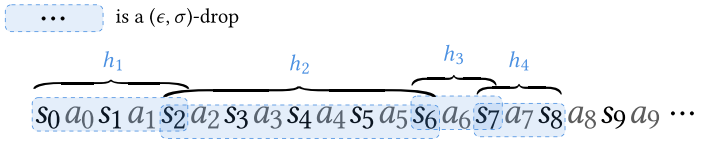
$$\inf_{\tau} \mathbb{E}_{s_n}^{\sigma[h],\tau}[f[h]] < \mathrm{val}(s_n) - 2\epsilon; \tag{4}$$

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

Definition 3.1. *The finite play $h := s_0 \cdots s_n$ is called a (ϵ, σ)-*drop *if* (4) *holds. We write*

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

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



Definition 3.2. *We define the* date *of the most recent (or latest) drop for all $s_0 \cdots s_n$ inductively as:*

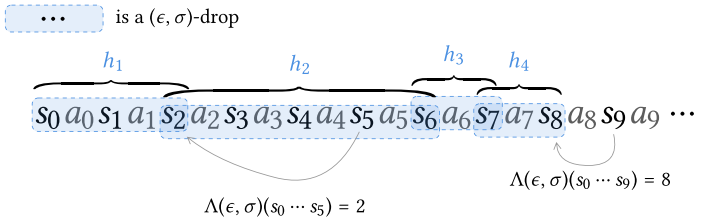
$$\Lambda(\epsilon,\sigma)(s_0) \stackrel{\text{def}}{=} 0$$

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

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

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

where

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

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

Lemma 3.4. *Let $\hat{\sigma}$ be a reset strategy that is ϵ-optimal, then it is also 2ϵ-subgame-perfect.*

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

$$\inf_\tau \mathbb{E}_{s_n}^{\hat{\sigma}[s_0 \cdots s_n],\tau}[f] \geq \text{val}(s_n) - 2\epsilon. \tag{5}$$

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 ϵ-optimal. Assume then that the most recent drop is $\ell < n$, which means that:

$$\inf_\tau \mathbb{E}_{s_n}^{\sigma[s_\ell \cdots s_n],\tau}[f] \geq \text{val}(s_n) - 2\epsilon, \tag{6}$$

where $\ell = \Lambda(\epsilon, \sigma)(s_0 \cdots s_{n-1})$. Towards a contradiction, assume that the goal (5) does not hold, *i.e.* there exists a strategy τ that gives payoff strictly less than $\mathrm{val}(s_n) - 2\epsilon$, then we will construct another strategy τ' that will contradict (6).

Let $\mathfrak{D}$ be the set prefixes from s_ℓ to the next (ϵ, σ)-drop, that is

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

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

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

From the assumption that the goal does not hold we have the following inequality[1]

$$\begin{aligned} \mathrm{val}(s_n) - 2\epsilon &> \mathbb{E}_{s_n}^{\sigma_1,\tau}\left[f \cdot \mathbb{1}_{\overline{\mathfrak{D}}}\right] + \mathbb{E}_{s_n}^{\sigma_1,\tau}\left[f \cdot \mathbb{1}_{\neg\overline{\mathfrak{D}}}\right] \\ &= \mathbb{E}_{s_n}^{\sigma_1,\tau}\left[f \cdot \mathbb{1}_{\overline{\mathfrak{D}}}\right] + \mathbb{E}_{s_n}^{\sigma_2,\tau'}\left[f \cdot \mathbb{1}_{\neg\overline{\mathfrak{D}}}\right] \\ &= \mathbb{E}_{s_n}^{\sigma_1,\tau}\left[f \cdot \mathbb{1}_{\overline{\mathfrak{D}}}\right] + \mathbb{E}_{s_n}^{\sigma_2,\tau'}[f] - \mathbb{E}_{s_n}^{\sigma_2,\tau'}\left[f \cdot \mathbb{1}_{\overline{\mathfrak{D}}}\right]. \end{aligned} \tag{7}$$

In the equality the strategy σ_1, respectively τ, has been replaced by σ_2, respectively τ' because on infinite plays without a drop they coincide.

For the first term above we have:

$$\begin{aligned} \mathbb{E}_{s_n}^{\sigma_1,\tau}\left[f \cdot \mathbb{1}_{\overline{\mathfrak{D}}}\right] &= \sum_{t_0\cdots t_m \in \mathfrak{D}} \mathbb{P}_{s_n}^{\sigma_1,\tau}(t_0 \cdots t_m)\, \mathbb{E}_{t_m}^{\hat{\sigma},\tau[t_0\cdots t_m]}[f] \\ &\geq \sum_{t_0\cdots t_m \in \mathfrak{D}} \mathbb{P}_{s_n}^{\sigma_1,\tau}(t_0 \cdots t_m)\,(\mathrm{val}(t_m) - e), \end{aligned}$$

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

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

by construction of the ϵ-response strategy τ'' and τ'. The strategies σ_1 and σ_2 on one hand, and τ, τ' on the other, coincide up to the first drop, consequently we can interchange them when measuring cylinders $t_0 \cdots t_m$, which implies that the two inequalities above give:

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

This contradicts (6) when plugged it in (7). □

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 ϵ-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 σ whose reset strategy is ϵ-optimal.

[1] $\mathbb{1}_{\mathcal{E}}$ is the indicator function of the event $\mathcal{E}$.

3.1 Properties of the Reset Strategy

We will show that there is a strategy σ with the following properties:

1. σ is ϵ-optimal,
2. σ is locally optimal,[2]
3. for any τ when playing with $\hat{\sigma}$ and τ almost surely there are only finitely many (ϵ, σ)-drops, and
4. $\hat{\sigma}$ is ϵ-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 ϵ-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[3], which we introduce first.

Definition 3.5 (Martingale)**.** *A sequence of real-valued random variables $X_0, X_1, \ldots$ is called a martingale if for all $n \in \mathbb{N}$*

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

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

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

Theorem 3.6 (Doob's Forward Convergence Theorem, [Wil91, Theorem 11.5])**.** *Let $X_0, X_1, \ldots$ be a supermartingale such that the sequence $(\mathbb{E}\left[|X_n|\right])_{n\in\mathbb{N}}$ is bounded. Then almost surely the limit*

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

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.

[2] This means that it does not play an action that decreases the value on average, the precise definition will follow.

[3] As a general reference for this area one might use [Wil91].

Definition 3.7 (Stopping Time)**.** *A random variable T taking values in $\mathbb{N} \cup \{\infty\}$ is called a* stopping time *with respect to random variables $S_0, S_1, \ldots$ if the event $\{T = n\}$ for $n \in \mathbb{N}$ is $(S_0, \ldots, S_n)$-measurable, meaning that it depends only on the random variables $S_0, \ldots, S_n$.*

Theorem 3.8 (Doob's Optional Stopping Theorem)**.** *Let T be a stopping time with respect to the random variables $S_0, S_1, \ldots$ and $(X_n)_{n\in\mathbb{N}}$ a uniformly bounded martingale such that for all $n \in \mathbb{N}$, X_n is $(S_0, \ldots, S_n)$-measurable. Define the random variable X_T which represents the process stopped at time T as:*

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

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 3.6. 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. □

3.1.1 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 3.9 (Locally Optimal Strategy)**.** *An action $a \in \mathbf{A}(s)$ is called locally optimal if and only if*

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

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

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

Observation 3.10. *When Player 1 (respectively Player 2) uses a locally optimal strategy the process*

$$\mathrm{val}(S_0), \mathrm{val}(S_1), \ldots$$

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 $a_0 \in \mathbf{A}(s_0)$ (say belonging to Player 1) in game $\mathbf{G}$ is *not* locally optimal, and denote by $\mathbf{G}'$ 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 ϵ-optimal strategies that are locally optimal as well. The analogue fact for Player 2 can be proved symmetrically.

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

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

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

$$\mathrm{val}(\mathbf{G}')(s_0) \geq \mathrm{val}(\mathbf{G})(s_0). \tag{8}$$

Let

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

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

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

Then by construction of τ, for all σ and s:

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

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

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

Taking $s = s_0$ and the supremum over all σ gives (8).

Using (8), we prove now that for all s

$$\mathrm{val}(\mathbf{G}')(s) \geq \mathrm{val}(\mathbf{G})(s). \tag{9}$$

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

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

Let $\epsilon > 0$ and for any strategy σ, define $\tilde{\sigma}$ to be the strategy that plays like σ unless the latter is about to play the action a_0 in s_0, in which case it switches to the strategy σ' which is ϵ-optimal in $\mathbf{G}'$. Set τ to be the strategy that plays according to some strategy τ' which is ϵ-optimal in $\mathbf{G}'$ as long as the opponent does not play the action a_0, otherwise it switches to some strategy that is ϵ-optimal in $\mathbf{G}$. By definitions of these strategies and (8) we have that for all σ and s

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

a combination of which gives us

$$\mathbb{E}_s^{\sigma,\tau}[f \mid \mathcal{S}(\sigma)] \leq \mathbb{E}_s^{\tilde{\sigma},\tau}[f \mid \mathcal{S}(\sigma)] + 2\epsilon. \tag{10}$$

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

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

Now by construction of the strategies and (10), for all σ and s we have

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

Since this holds for any $\epsilon > 0$, taking the supremum over all σ proves (9).

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

$$\text{locally optimal and } \epsilon\text{-optimal.} \tag{11}$$

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

$$p(s,a,t) > 0 \qquad \Longrightarrow \qquad \mathrm{val}(s) = \mathrm{val}(t).$$

Proposition 3.12. *For all strategies σ, τ at least one of which is locally optimal and $s \in \mathbf{S}$ we have*

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

Proof. Fix σ and τ and assume that σ is locally optimal, the other case is symmetrical. Suppose that $a_0 \in \mathbf{A}(s_0)$ 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 $t \in \mathbf{S}$ 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,

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

which contradicts Theorem 3.6, since $(\operatorname{val}(S_n)), n \in \mathbb{N}$ is a supermartingale as per Observation 3.10. □

3.1.2 Finitely Many Drops

Recall that $\Delta(\epsilon, \sigma)(\cdot)$ characterises finite plays that are (ϵ, σ)-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 (ϵ, σ)-drop" or "two (ϵ, σ)-drops after date n". Our goal is to prove that for a reset strategy that is based on a σ that is both ϵ-optimal and locally optimal (which exists because of (11)) almost surely there will only be finitely many (ϵ, σ)-drops. To this end fix a $\epsilon > 0$, and σ a strategy that is both locally optimal and ϵ-optimal, which allows us to simply say drop instead of (ϵ, σ)-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 3.12 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 τ_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 3.13. *There exists a $c > 0$ such that for all s and locally optimal τ,*

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

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 τ' be a strategy that plays like τ as long as no drop occurs, and once it does it switches to the strategy τ'' that is a $\epsilon/2$-optimal response. By construction, τ and τ' 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 ϵ-optimality of σ, 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 $\boldsymbol{D}$ 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 τ' we have for all s:

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

Replacing this inequality in the one above and decomposing the expectation of $\text{val}(S_T)$[4] 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 τ' plays like τ 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[5], so we can apply Theorem 3.8. 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 τ. □

[4] Theorem 3.6 implies that this random variable is well-defined.

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

Next we approximate strategies τ by a sequence τ_n for every natural n as follows. The strategies τ_n play like τ 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 3.14. *There is some $c > 0$ such that for all strategies τ, s and $n \in \mathbb{N}$, we have*

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

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 τ and state s we have

$$\mathbb{P}_s^{\hat{\sigma},\tau_n}\left(T_n^2 < \infty \mid T_n < \infty\right) \leq 1 - c. \tag{12}$$

The statement of the lemma then follows from (12) 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:

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

where in the last equality we have replaced the reset strategy by σ, 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 3.13 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 (12). □

In the third lemma there is no restriction upon the strategy τ.

Lemma 3.15. *For all strategies τ and s there is some $n \in \mathbb{N}$ such that*

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

Proof. Fix a strategy τ 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 ∞. Since the strategies τ and τ_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 σ has been assumed to be locally optimal, and therefore the strategy $\hat{\sigma}$ is locally optimal as well. As a consequence of Proposition 3.12 we have

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

whence follows

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

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 3.14. □

This lemma makes it possible now to prove the third property of the strategy σ, namely that for all strategies τ and s,

$$\mathbb{P}_s^{\hat{\sigma},\tau}(\exists n \text{ no drops after date } n) = 1. \tag{13}$$

Let T be the stopping time that gives the date of the last drop, if it exists otherwise let it be equal to ∞. 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 3.14) if it exists, otherwise say that it is equal to ∞.

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

$$\sup_{\tau,s} \mathbb{P}_s^{\hat{\sigma},\tau}(T = \infty) \leq \mathbb{P}_{\tilde{s}}^{\hat{\sigma},\tilde{\tau}}(T = \infty) + \delta. \tag{14}$$

Let $\tilde{n} \in \mathbb{N}$ the number from the statement of Lemma 3.15, thus

$$d \overset{\text{def}}{=} \mathbb{P}_{\tilde{s}}^{\hat{\sigma},\tilde{\tau}}(F_{\tilde{n}} < \infty) < 1. \tag{15}$$

And from (14), some basic properties of expectations we deduce:

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

The random variable $S_{F_{\tilde{n}}}$ is well-defined because we are measuring the infinite plays where $F_{\tilde{n}}$ is finite; on the third equality we have used the definition of the reset strategy and the last two (in)equalities we have used (14) and (15) respectively. Since $d < 1$ then we have

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

so for all states s' and strategies τ' it follows that

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

Since this holds for any $\delta > 0$, (13) follows.

3.1.3 ϵ-Optimal

The last property of σ that we have to prove is that if we assume that it has the previous properties, namely that it is ϵ-optimal, locally optimal, and it has finitely many drops, then the reset strategy $\hat{\sigma}$ is ϵ-optimal as well. So fix an $\epsilon > 0$ and a strategy σ that is both locally optimal and ϵ-optimal, and for which (13) holds. We define for all naturals n, strategies $\hat{\sigma}_n$ that reset only up to date n, and prove that they are ϵ-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

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

Lemma 3.16. *For all $n \in \mathbb{N}$, $\hat{\sigma}_n$ is ϵ-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 τ and prove that

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

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

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

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

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

Now since the strategies $\hat{\sigma}_{n-1}$ and $\hat{\sigma}_n$ on one hand, and strategies τ and τ' 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. □

We now prove that

$$\hat{\sigma} \text{ is } \epsilon\text{-optimal}, \tag{16}$$

the final property of σ 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 ∞.

Applying Lemma 3.16 we have that for all $n \in \mathbb{N}$, s, and τ

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

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

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

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:

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

From the sigma-additivity of measures and the property in (13) it follows that

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

Since τ and s are general, this proves ϵ-optimality of σ, that is it proves the final property (16). Lemma 3.4 in conjunction with (16) implies Theorem 1.2.

3.1.4 Remark on Optimal Strategies

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

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

in Section 3.1.2 and Section 3.1.3 respectively, the variable ϵ need not be strictly positive. Since optimal strategies are necessarily localy optimal, the following remark follows from Lemma 3.4.

Remark 3.17. *If σ is an optimal strategy then $\hat{\sigma}$ is subgame-perfect.*

3.2 Preserving Finite Memory

We conclude this section by showing that the construction of the reset strategy preserves finite memory, that is if the strategy σ has finite memory to begin with, so will the strategy $\hat{\sigma}$. Let us first define precisely what we mean by finite memory strategy.

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

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

The map init and up are used to initialise the memory and update it, as the game unfolds: after the finite play $s_0a_0\cdots s_n$ has unfolded, the transducer reaches the memory state $m_n\in\mathcal{M}$ which is defined inductively as:

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

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

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

Proposition 3.18. *If σ is a finite memory strategy, so is $\hat{\sigma}$.*

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

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

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

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

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

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

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

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

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

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

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

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

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.

Proposition 3.19. *Let $\mathbb{A}$ be a family of arenas that are closed under restrictions and f a payoff function. If for games in $\mathbb{A}$ (with f), Player 1 (respectively Player 2) has ϵ-optimal strategies σ with finite memory. Then he also has an ϵ-subgame-perfect strategies with finite memory, namely the reset strategies $\hat{\sigma}$.*

Proof. Let $\mathcal{A} \in \mathbb{A}$ 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 ϵ-optimal strategies in $(\mathcal{A}', f)$ that have finite

memory, and consequently there are ϵ-optimal strategies in $(\mathcal{A}, f)$ that have finite memory and are locally optimal. From the discussion of this section the strategy $\hat{\sigma}$ is 2ϵ-subgame-perfect, and Proposition 3.18 implies that it has finite memory. □

Remark 3.17 and Proposition 3.18 imply the following observation.

Remark 3.20. *If a player has an optimal strategy with finite memory σ, then he also has a subgame-perfect strategy with finite memory, namely the reset strategy $\hat{\sigma}$.*

4 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, we give examples in Section 6. Necessary and sufficient conditions for positionality are known for deterministic games [GZ05]. However the shift-invariant and submixing conditions are general enough to recover several known classical results, and to provide several new examples of games with deterministic stationary optimal strategies. Before we proceed with the proof we remark:

Remark 4.1. *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.*

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

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

If $N(\mathbf{G}) = 0$ there is no choice for maximizer, hence he has a deterministic and stationary optimal strategy. Suppose that the theorem has been proved for all games $\mathbf{G}'$ with $N(\mathbf{G}') < N(\mathbf{G})$, we prove the theorem for the game $\mathbf{G}$. Since $N(\mathbf{G}) > 0$ there must be a state $\tilde{s} \in \mathbf{S}$ such that Player 1 has at least two actions in $\tilde{s}$, *i.e.* $\mathbf{A}(\tilde{s})$ has at least two elements. Partition $\mathbf{A}(\tilde{s})$ into two non-empty sets: $\mathbf{A}_1$ and $\mathbf{A}_2$. Define the games $\mathbf{G}_1$ and $\mathbf{G}_2$ to be the games that one obtains from $\mathbf{G}$ by restricting the actions in the state $\tilde{s}$ to $\mathbf{A}_1$ and $\mathbf{A}_2$ respectively. The proof of the theorem hinges on the following inequality, that says that the value of $\tilde{s}$ in the original game cannot be larger than that of the restricted games:

$$\mathrm{val}(\mathbf{G})(\tilde{s}) \leq \max\{\mathrm{val}(\mathbf{G}_1)(\tilde{s}), \mathrm{val}(\mathbf{G}_2)(\tilde{s})\}. \tag{17}$$

We will witness this inequality with a strategy for Player 2, which will merge two ϵ-subgame-perfect strategies in the respective smaller games. Then we analyse the

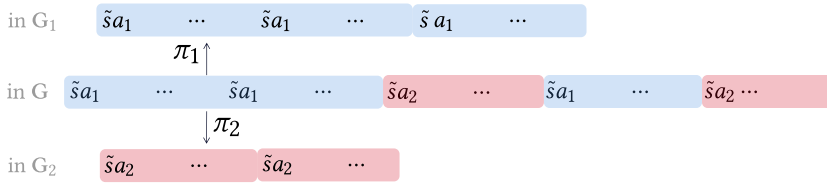
two possible outcomes: (a) after some date the play remains only in game $\mathbf{G}_1$ (or only in game $\mathbf{G}_2$), (b) the play switches between the two smaller games. For the latter case we use the submixing property which in this case states that one cannot get a better payoff by switching between the two smaller games. We start by defining the merge strategy.

4.1 The Merge Strategy

Recall the notation $X^\infty = X^\omega \cup X^*$, the set of (in)finite sequences with alphabet X. We define two projection maps π_1, π_2,

$$\pi_i \ : \ \tilde{s}(\mathbf{AS})^\infty \longrightarrow \tilde{s}(\mathbf{AS})^\infty,$$

that project (in)finite plays of the game $\mathbf{G}$ to plays in the restricted games by deleting factors. For example:



Let us formally define these two projections for finite plays. Let $h = s_0 a_0 s_1 a_1 \cdots s_n$ be a finite play and define

$$i(1) < i(2) < \ldots < i(k) \stackrel{\text{def}}{=} \{ i \ : \ s_i = \tilde{s} \},$$

the increasing sequence of dates where the play reaches $\tilde{s}$. For $1 \leq \ell < k$ let h_ℓ be the ℓ-th factor of h defined by

$$h_\ell \stackrel{\text{def}}{=} s_{i(\ell)} a_{i(\ell)} \cdots a_{i(\ell+1)-1}$$

and

$$h_k \stackrel{\text{def}}{=} s_{i(k)} a_{i(k)} \cdots a_{n-1} s_n.$$

Then the projections for $j \in \{1, 2\}$, are

$$\pi_j(h) \stackrel{\text{def}}{=} \text{concatenation of } (h_\ell)_{1 \leq \ell \leq k \ \wedge \ a_{i(\ell)} \in A_j}.$$

This definition can be extended to infinite plays in a natural way: for $h \in \tilde{s}(\mathbf{AS})^\omega$ let $\pi_j(h)$ be the limit of the sequence

$$\left(\pi_j(h_{<n})\right)_{n \in \mathbb{N}},$$

where $h_{<n}$ is the prefix of h of length $2n+1$. The projections have a couple of properties which we name here so that we are able to refer to them later.

Remark 4.2. *For all infinite plays h, if $\pi_1(h)$ (respectively $\pi_2(h)$) is finite then*

$$h \text{ has a suffix that is an infinite play in } \mathbf{G}_2 \text{ (respectively } \mathbf{G}_1) \tag{18}$$

starting in state $\tilde{s}$. If both $\pi_1(h)$ and $\pi_2(h)$ are infinite then both plays reach $\tilde{s}$ infinitely often and

$$h \text{ is a shuffle of } \pi_1(h) \text{ and } \pi_2(h). \tag{19}$$

Fix $\epsilon > 0$, and $\tau_1^\#$ and $\tau_2^\#$ to be ϵ-subgame strategies for Player 2 in the game $\mathbf{G}_1$ and $\mathbf{G}_2$ respectively. Their existence is guaranteed by Theorem 1.2. We construct the *merge* strategy $\tau^\#$, using the two strategies above, to be the strategy that switches between $\tau_1^\#$ and $\tau_2^\#$ depending on the action chosen at most recent visit to $\tilde{s}$. Formally it is defined for all finite plays starting in $\tilde{s}$ as follows. Let $h = \tilde{s} \cdots t$ be such a play and define $\text{last}(h) \in \mathbf{A}$ to be the action that is played after the last visit of h to $\tilde{s}$, then

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

4.2 Proof of (17)

We prove that the merge strategy witnesses (17); meaning that when Player 2 utilises this strategy, no matter what the choice of his opponent, when the game is started in $\tilde{s}$ the payoff will not be larger than the maximum of the values of $\tilde{s}$ in the smaller games plus ϵ. The following lemmas prove this proposition conditioned on whether the play stays in one of the smaller games indefinitely or it switches between them. Define the following abbreviations:

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

Lemma 4.3. *For all strategies σ,*

$$\mathbb{P}_{\tilde{s}}^{\sigma,\tau^\#}\left(f(\Pi_1) \le \text{val}(\mathbf{G}_1)(s) + \epsilon \mid \Pi_1 \text{ is infinite and reaches } \tilde{s} \text{ infinitely often}\right) = 1. \tag{20}$$

Proof. We prove first that for every strategy σ in $\mathbf{G}$, there is a strategy σ_1 in $\mathbf{G}_1$ such that for every measurable event $\mathcal{E}_1$ in the game $\mathbf{G}_1$,

$$\mathbb{P}_{\tilde{s}}^{\sigma_1,\tau_1^\#}(\mathcal{E}_1) \ge \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\#}\left(\Pi_1 \text{ is infinite and } \Pi_1 \in \mathcal{E}_1\right). \tag{21}$$

Denote by $\preceq$ (respectively $\prec$) the prefix relation (respectively strict prefix) over words in $(\mathbf{SA})^\infty$. The strategy σ_1 is defined as:

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

if $\mathbb{P}_{\tilde{s}}^{\sigma,\tau^\#}(h \prec \Pi_0) > 0$ and otherwise $\sigma_1(h)$ is chosen arbitrarily. Let $\mathfrak{E}$ be the set of measurable events $\mathcal{E}_1$ in $\mathbf{G}_1$ for which (21) holds. First, $\mathfrak{E}$ contains all cylinders $h(\mathbf{SA})^\omega$ of $\mathbf{G}_1$ because:

$$\mathbb{P}_{\tilde{s}}^{\sigma_1,\tau_1^\#}(h) \ge \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\#}(h \preceq \Pi_1) \ge \mathbb{P}_{\tilde{s}}^{\sigma,\tau^\#}\left(\Pi_1 \text{ is infinite and } \Pi_1 \in h(\mathbf{SA})^\omega\right),$$

where the first inequality can be proved by induction on the length of h, using the definition of σ_1 and where the second inequality is by definition of prefixes. 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]), which implies that $\mathfrak{E}$ contains the sigma-field that is generated by cylinders, which by definition is the set of all measurable events in the game $\mathbf{G}_1$. This completes the proof of (21).

Next we prove that

$$\mathbb{P}_{\tilde{s}}^{\sigma_1,\tau_1^\sharp}\left(f \leq \liminf_n \operatorname{val}(\mathbf{G}_1)(S_n) + \epsilon\right) = 1. \tag{22}$$

To this end, we apply Lévy's 0-1 law (see *e.g.* [Wil91, Theorem 14.4]) to the sequence of random variables:

$$\mathbb{E}_{\tilde{s}}^{\sigma_1,\tau_1^\sharp}[f \mid S_0, A_0, \dots, S_n], \qquad n \in \mathbb{N},$$

which says that the latter sequence converges point-wise to the random variable $f(S_0A_0\cdots)$. Next observe that due to the fact that $\tau_1^\sharp$ is ϵ-subgame-perfect (and that f is shift-invariant), the random variables in the sequence are upper bounded, that is for all $n \in \mathbb{N}$,

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

These two facts together imply (22). A consequence of which is that

$$\mathbb{P}_{\tilde{s}}^{\sigma_1,\tau_1^\sharp}(f > \operatorname{val}(\mathbf{G}_1)(\tilde{s}) + \epsilon \text{ and } \tilde{s} \text{ is reached infinitely often}) = 0.$$

The lemma follows by applying (21) to the event being measured above. □

The proof above is symmetric, and in fact it can be used to deduce the following statement as well: for all strategies σ,

$$\mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}(f(\Pi_2) \leq \operatorname{val}(\mathbf{G}_2)(s) + \epsilon \mid \Pi_2 \text{ is infinite and reaches } \tilde{s} \text{ infinitely often}) = 1. \tag{23}$$

We join these two facts to show a corollary, namely that:

Lemma 4.4. *For all strategies σ,*

$$\mathbb{P}_{\tilde{s}}^{\sigma,\tau^\sharp}(f \leq \max\{\operatorname{val}(\mathbf{G}_1)(\tilde{s}), \operatorname{val}(\mathbf{G}_1)(\tilde{s})\} + \epsilon \mid \tilde{s} \textit{ is reached infinitely often}) = 1. \tag{24}$$

Proof. This equation is proved by case inspection. Assuming $\tilde{s}$ is reached infinitely often, then the projection Π_1 is either finite or is infinite and reaches $\tilde{s}$ infinitely often. If Π_1 is finite then according to (18), the projection Π_2 is a suffix of the play Π. Thus we can apply (20) and conclude that (24) holds since f is shift-invariant. The case where Π_2 is finite is symmetrical. In the remaining case, both projections Π_1 and Π_2 are infinite. Then according to property (19), both Π_1 and Π_2 reach $\tilde{s}$ infinitely often, and Π is a shuffle of Π_1 and Π_2. In this case both (20) and (23) hold and the statement of the lemma follows because f is submixing. □

For any strategy σ, define $\mathfrak{C}^1_\sigma$ to be the set of finite plays h in $\mathbf{G}_2$ ending in the state $\tilde{s}$, and such that $\sigma(h)$ wants to play an action in $\mathbf{A}_1$; *i.e.* prefixes where a switch to $\mathbf{G}_1$ is about to occur. The event that is generated by the cylinders of prefixes in $\mathfrak{C}^1_\sigma$ (respectively its complement) abbreviates as

$$\text{leaves } \mathbf{G}_2 \qquad \text{(respectively stays in } \mathbf{G}_2\text{)}.$$

One can define symmetrically the events for $\mathbf{G}_1$.

Lemma 4.5. *For all strategies σ and τ such that the restriction of τ to $\mathbf{G}_2$ (called τ_2) is ϵ-subgame perfect in that game, we have*

$$\mathbb{E}^{\sigma,\tau}_{\tilde{s}}\left[f \mid \textit{stays in } \mathbf{G}_2\right] \leq \mathrm{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon. \tag{25}$$

Proof. In case $\mathbb{P}^{\sigma,\tau}_{\tilde{s}}(\text{stays in } \mathbf{G}_2) = 0$ the left handside of (25) is undefined and there is nothing to prove. For the rest of the proof, we assume $\mathbb{P}^{\sigma,\tau}_{\tilde{s}}(\text{stays in } \mathbf{G}_2) > 0$.

We prove (25) by contradiction. Suppose *a contrario* that there exists $\epsilon' > 0$ and strategies σ, τ as in the statement of the lemma such that

$$\mathbb{E}^{\sigma,\tau}_{\tilde{s}}\left[f \mid \text{stays in } \mathbf{G}_2\right] \geq \mathrm{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon + \epsilon'.$$

We define the event $\mathcal{Z}$ and quantity x_0 as:

$$\mathcal{Z} \stackrel{\text{def}}{=} \{\text{stays in } \mathbf{G}_2 \text{ and } f \geq \underbrace{\mathrm{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon + \epsilon'}_{x_0}\}.$$

Then $\mathbb{P}^{\sigma,\tau}_{\tilde{s}}(\mathcal{Z}) > 0$. We apply Lévy's 0-1 law (see *e.g.* [Wil91, Theorem 14.4]) to the sequence of random variables:

$$\mathbb{E}^{\sigma,\tau}_{\tilde{s}}\left[\mathbb{1}_{\mathcal{Z}} \mid S_0, A_0, \ldots, S_n\right], \qquad n \in \mathbb{N},$$

which says that the latter sequence converges point-wise to the random variable $\mathbb{1}_{\mathcal{Z}}$. Then

$$\mathbb{P}^{\sigma,\tau}_{\tilde{s}}\left(\mathbb{E}^{\sigma,\tau}_{\tilde{s}}\left[\mathbb{1}_{\mathcal{Z}} \mid S_0, A_0, \ldots, S_n\right] \longrightarrow_n 1\right) = \mathbb{P}^{\sigma,\tau}_{\tilde{s}}(\mathcal{Z}) > 0.$$

Set $\epsilon'' > 0$, whose exact value shall be determined later. Then there exists a finite play h in $\mathbf{G}_2$ consistent with σ and τ and such that

$$\mathbb{P}^{\sigma,\tau}_{\tilde{s}}(\mathcal{Z} \mid h) = \mathbb{E}^{\sigma,\tau}_{\tilde{s}}\left[\mathbb{1}_{\mathcal{Z}} \mid h\right] \geq 1 - \epsilon''.$$

By definition of the event $\mathcal{Z}$, we have

$$\mathbb{E}^{\sigma,\tau}_{\tilde{s}}\left[f \cdot \mathbb{1}_{\mathcal{Z}} \mid h\right] \geq (1 - \epsilon'') \cdot x_0 = \mathrm{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon + \epsilon' - \epsilon'' x_0. \tag{26}$$

Modify the strategy σ into σ' such that as soon as a prefix in $\mathfrak{C}^1_\sigma$ is reached, then instead of playing an action in $\mathbf{A}_1$, the strategy σ' switches to an arbitrary strategy σ_2 in $\mathbf{G}_2$. Let M denote the maximum absolute value of f (which is bounded, by definition of payoff functions). Since σ and σ' coincide as long as the play stays in $\mathbf{G}_2$, they

coincide on every prefix of every infinite play in $\mathcal{Z}$ thus (26) holds when replacing σ by σ' and with the same argument, hence

$$\mathbb{P}_{\tilde{s}}^{\sigma',\tau}(\mathcal{Z} \mid h) \geq 1 - \epsilon'', \text{ and}$$
$$\mathbb{E}_{\tilde{s}}^{\sigma',\tau}\left[f \cdot (1 - \mathbb{1}_{\mathcal{Z}}) \mid h\right] \geq -\epsilon'' M.$$

Summing this inequality with (26) (where σ is replaced by σ') we get

$$\mathbb{E}_{\tilde{s}}^{\sigma',\tau}[f \mid h] \geq \mathrm{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon + \epsilon' - \epsilon''(x_0 + M).$$

In particular, such a finite play h consistent with σ' and τ exists when choosing

$$\epsilon'' \stackrel{\text{def}}{=} \frac{1}{2}\epsilon'/(x_0 + M),$$

which is positive due to the definition of M. It now follows that

$$\mathbb{E}_{\tilde{s}}^{\sigma',\tau}[f \mid h] \geq \mathrm{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon + \epsilon'/2.$$

From this inequality, and since σ' is a strategy in $\mathbf{G}_2$, by analysing extensions of h to infinite plays that reach $\tilde{s}$ again and those that do not, we can reach a contradiction to the ϵ-subgame perfection of τ in $\mathbf{G}_2$. This terminates the proof of the lemma. □

We are now ready to prove (17) by showing that for all strategies σ,

$$\mathbb{E}_{\tilde{s}}^{\sigma,\tau^\#}[f] \leq \max\{\mathrm{val}(\mathbf{G}_1)(\tilde{s}), \mathrm{val}(\mathbf{G}_2)(\tilde{s})\} + \epsilon, \tag{27}$$

this suffices since it holds for any $\epsilon > 0$.

There are three types of infinite plays, depending on whether their projections Π_1 and Π_2 are finite of infinite. Take first the infinite plays whose projection Π_1 is finite. Such plays have a prefix h that ends up in state $\tilde{s}$ and after which the play remains in $\mathbf{G}_2$. This suffix that is played in $\mathbf{G}_2$ is consistent with the strategies $\sigma[h]$ and $\tau^\#[h]$. The latter is by definition equal to $\tau_2^\#[\pi_2(h)]$. Since $\tau_2^\#$ is ϵ-subgame-perfect, $\tau_2^\#[\pi_2(h)]$ is ϵ-subgame-perfectas well. Consequently, we can apply Lemma 4.5 to strategies $\sigma[h]$ and $\tau^\#[h]$ and deduce that

$$\mathbb{E}_{\tilde{s}}^{\sigma[h],\tau^\#[h]}[f \mid \text{stays in } \mathbf{G}_2] \leq \mathrm{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon,$$

and as a consequence

$$\mathbb{E}_{\tilde{s}}^{\sigma,\tau^\#}[f \mid \text{stays in } \mathbf{G}_2 \text{ after } h \mid h] \leq \mathrm{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon.$$

Conditioning over all such finite plays h we cover all cases where Π_1 is finite hence:

$$\mathbb{E}_{\tilde{s}}^{\sigma,\tau^\#}[f \mid \Pi_1 \text{ is finite}] \leq \mathrm{val}(\mathbf{G}_2)(\tilde{s}) + \epsilon.$$

Symmetrically,

$$\mathbb{E}_{\tilde{s}}^{\sigma,\tau^\#}[f \mid \Pi_2 \text{ is finite}] \leq \mathrm{val}(\mathbf{G}_1)(\tilde{s}) + \epsilon.$$

The last possibility is when both Π_1 and Π_2 are infinite, in such plays there is a visit to $\tilde{s}$ infinitely many times thus according to Lemma 4.4,

$$\mathbb{E}_{\tilde{s}}^{\sigma,\tau^\#}[f \mid \text{both } \Pi_1 \text{ and } \Pi_2 \text{ are infinite}] \leq \max\{\text{val}(\mathbf{G}_1)(\tilde{s}), \text{val}(\mathbf{G}_2)(\tilde{s})\} + \epsilon.$$

The last three inequalities cover all the cases and we therefore have (27).

4.3 Proof of Theorem 1.1

Without loss of generality assume that

$$\max\{\text{val}(\mathbf{G}_1)(\tilde{s}), \text{val}(\mathbf{G}_2)(\tilde{s})\} = \text{val}(\mathbf{G}_1)(\tilde{s}). \tag{28}$$

By (27) and since Player 1 has more choice in $\mathbf{G}$ than he does in $\mathbf{G}_1$, we have

$$\text{val}(\mathbf{G})(\tilde{s}) = \text{val}(\mathbf{G}_1)(\tilde{s}).$$

To finish the induction step and the proof of Theorem 1.1, we have to show that the equality above holds for all other states $s \in \mathbf{S}$, as well. Recall that the merge strategy was defined only for plays that start in state $\tilde{s}$; we enlarge this definition, profiting from the assumption (28) to the following. For all finite plays h that end in state t,

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

We will prove that the values of every state in $\mathbf{G}$ are equal to those in $\mathbf{G}_1$, by witnessing it with $\tau^\#$; that is for every $s \in \mathbf{S}$

$$\text{val}(\mathbf{G})(s) = \text{val}(\mathbf{G}_1)(s). \tag{29}$$

We show this by demonstrating that $\tau^\#$ guarantees a payoff smaller than $\text{val}(\mathbf{G}_1)(s) + \epsilon$ for every state s. Fix σ a strategy for Player 1, and define σ' to be the strategy that plays like σ until state $\tilde{s}$ is reached in which case it switches definitively to the strategy $\sigma_1^\#$, that is optimal in the game $\mathbf{G}_1$, and whose existence is guaranteed by the induction hypothesis. The plays consistent with σ' and $\tau^\#$ are plays in the game $\mathbf{G}_1$, so we can write for all s:

$$\mathbb{E}_s^{\sigma',\tau^\#}[f] = \mathbb{E}_s^{\sigma',\tau_1^\#}[f] \leq \text{val}(\mathbf{G}_1)(s) + \epsilon, \tag{30}$$

since $\tau_1^\#$ is ϵ-optimal in $\mathbf{G}_1$. Let h be a finite play that is consistent with σ and $\tau^\#$ that reaches $\tilde{s}$ in the last step (and not before). The strategy $\tau^\#[h]$ is ϵ-optimal from state $\tilde{s}$ for the following reason. By definition of the merge strategy above, $\tau^\#[h]$ is the strategy that one obtains by merging the strategy $\tau_1^\#[h]$ and $\tau_2^\#$, both of which are ϵ-subgame-perfect in the respective games. Since (27) was proved for any merge of two ϵ-subgame-perfect strategies, we can apply it to the strategy $\tau^\#[h]$, to conclude that the latter is ϵ-optimal from state $\tilde{s}$. As a consequence we can write:

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

Since the strategies σ and σ' coincide on those plays that never reach $\tilde{s}$, the inequality above implies that for all s,

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

By using (30) now we have that for all s,

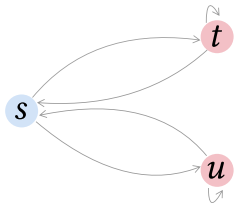
$$\mathbb{E}_s^{\sigma,\tau^\sharp}[f] \leq \mathrm{val}(\mathbf{G}_1)(s) + 2\epsilon. \tag{31}$$

Since this holds for any $\epsilon > 0$, (29) has been proved, and with it the induction step and Theorem 1.1.

Remarks about the merge strategy. We observe a byproduct of the proof above, namely that (31) has yielded 2ϵ-optimality of the merge strategy:

Observation 4.6. *The merge strategy $\tau^\sharp$ constructed with ϵ-subgame-perfect pieces is 2ϵ-optimal in the game* $\mathbf{G}$.

After this observation, since the merge strategy is obtained by merging two ϵ-subgame-perfect strategies, a natural question to ask is whether $\tau^\sharp$ is 2ϵ-subgame-perfect in the $\mathbf{G}$? 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 $\mathbf{G}_1$ we remove the action $s \longrightarrow t$. In particular in the game $\mathbf{G}_1$ the positional strategy $\tau_1^\sharp$ which chooses $u \longrightarrow s$ and $t \longrightarrow s$ is subgame-perfect. We can therefore use it to construct a merge strategy $\tau^\sharp$. However this merge strategy is not 2ϵ-subgame-perfect, since in case Player 1 uses the suboptimal action $s \longrightarrow u$, his opponent does not profit by taking the self-loop forever.

5 Finite Memory Transfer Theorem

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

Theorem 1.3. *For every payoff function f that is both shift-invariant and submixing, if the minimizer has optimal strategies with finite memory in one player games equipped with f, then the minimizer has the same for two-player games equipped with f.*

We prove a slightly stronger theorem and derive Theorem 1.3 as a corollary. An arena is said to be *controlled by Player 2* if in every state that belongs to his opponent there is only one action available. (In other words these arenas are one-player games, or Markov decision processes).

Theorem 5.1. *Let f be a shift-invariant and submixing payoff function.*

If for all $\epsilon > 0$, Player 2 has an ϵ-optimal strategy with finite memory in every game controlled by himself, then in every (two-player) game he has an ϵ-subgame-perfect strategy that has finite memory.

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

Proof. The proof is by induction on actions of $\mathbf{G}$, by defining the smaller games $\mathbf{G}_1$ and $\mathbf{G}_2$ as in the proof of the main theorem in the previous section. The base of the induction follows from the assumption of the theorem, the induction hypothesis says that there are two ϵ-subgame perfect strategies $\tau_1^\sharp$ and $\tau_1^\sharp$ with finite memory, given by the transducers:

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

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

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

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

According to Observation 4.6, $\tau^\sharp$ is 2ϵ-optimal. According to Proposition 3.19, there exists a 4ϵ-subgame-perfect strategy in $\mathbf{G}$. Since this holds for any $\epsilon > 0$, it concludes the induction step.

The second part of the theorem follows similarly because of Remark 3.20. □

On the size of the memory How large is the memory $\mathcal{M}_\mathbf{G}$ needed by Player 2 to play optimally in some $\mathbf{G} = (\mathcal{A}, f)$? Every deterministic and stationary strategy σ for Player 1 in $\mathbf{G}$ induces a game $\mathbf{G}_\sigma$ that is controlled by Player 2. Let $\mathfrak{M}$ be the maximal memory size required by Player 2 to play optimally in the games G_σ. According to the proof of the theorem above, the memory $\mathcal{M}_\mathbf{G}$ needed by Player 2 to play optimally in $\mathbf{G}$ is of size $2 \cdot |\mathcal{M}_{\mathbf{G}_1}| \cdot |\mathcal{M}_{\mathbf{G}_2}|$. By induction we derive the following bound:

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

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.

6 Applications

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

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

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

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 $(r_0, \lambda_0)(r_1, \lambda_1) \cdots \in (\mathbb{R} \times [0, 1))^\omega$ of daily payoffs and discount factors is associated the payoff:

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

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

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

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:

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

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

$$\limsup_n \sum_{0 \leq i \leq n} c_i = +\infty \tag{32}$$

$$\limsup_n \sum_{0 \leq i \leq n} c_i = -\infty \tag{33}$$

$$f_{\mathrm{mean}}(c_0 c_1 \ldots) > 0 \tag{34}$$

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

$$\forall 1 \leq i \leq k, f^i_{\text{mean}}\left(r_0^{(i)} r_1^{(i)} \dots\right) > 0. \tag{35}$$

6.1 Unification of Classical Results

The existence of deterministic stationary optimal strategies in Markov decision processes with parity [CY90], limsup, liminf [MS96], mean-payoff [LL69, NS03, Bie87, VTRF83] or discounted payoff functions [Sha53] is well-known. Theorem 1.1 provides a unified proof of these five results, as a corollary of the following proposition.

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

The proof of this proposition is an elementary exercise.

Corollary 6.2. *In every two-player stochastic game equipped with the parity, limsup, liminf, mean or discounted payoff function, Player 1 has a deterministic and stationary strategy which is optimal. The same is true for Player 2 for the parity, limsup and liminf payoff.*

Proof. Except for the discounted payoff function, this is a direct consequence of Proposition 6.1, Theorem 1.1, and Remark 4.1. The case of the discounted payoff function can be reduced to the case of the mean-payoff function, interpreting discount factors as stopping probabilities as was done in the seminal paper of Shapley [Sha53]. Details can be found in [Gim07, Gim06]. □

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

The existence of deterministic and stationary optimal strategies in mean-payoff games has attracted much attention. The first proof was given by Gilette [Gil57] and based on a variant of Hardy and Littlewood theorem. Later on, Ligget and Lippman found the variant to be wrong and proposed an alternative proof based on the existence of Blackwell optimal strategies plus a uniform boundedness result of Brown [LL69]. For one-player games, Bierth [Bie87] gave a proof using martingales and elementary linear algebra while [VTRF83] provided a proof based on linear programming and a modern proof can be found in [NS03] based on a reduction to discounted games and the use 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].

6.2 Variants of Mean-Payoff Games

The positive average condition defined by (34) is a variant of mean-payoff games which may be more suitable to model quality of service constraints or decision makers with a loss aversion.

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

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

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

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

6.3 New Examples of Positional Games

Although the generalized mean-payoff condition defined by (35) is not submixing a variant is. *Optimistic generalized mean-payoff games* are defined similarly except the winning condition is

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

It is a basic exercise to show that this winning condition is submixing. More generally, if $f_1, \ldots, f_n$ are submixing payoff mappings then $\max\{f_1, \ldots, f_n\}$ is submixing as well. As a consequence, those games are half-positional. It is a simple exercise to observe that optimistic generalized mean-payoff games are half-positional but not positional, however we can infer from Theorem (35)

We give a final example of a payoff function that is shift-invariant, submixing, and inverse-submixing (hence positional for both players in two-players games): 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:

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

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

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

and the payoff

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

Other examples can be found in [Gim07, Kop09, Gim06], and in the papers cited in the introduction.

Acknowledgments

We are very grateful to Pierre Vandenhove for finding an error in the previous version of the proof of Lemma 4.5. This work was supported by the ANR projet "Stoch-MC" and the LaBEX "CPU".

References

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

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

[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+20] Patricia Bouyer, Stéphane Le Roux, Youssouf Oualhadj, Mickael Randour, and Pierre Vandenhove. 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. Chatterejee, M. Jurdziński, and T.A. Henzinger. Quantitative stochastic parity games. In *SODA*, 2003.

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

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

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

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

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

[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 *FOSSACS 2004*, volume 2987 of *Lecture Notes in Computer Science*, pages 499–513. Springer, 2004.

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