# Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games

Hugo Gimbert, Wieslaw Zielonka

# Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games

Hugo Gimbert* Wiesław Zielonka†

December 3, 2009

**Abstract**

We consider two-players zero-sum perfect information stochastic games with finitely many states and actions and examine the problem of existence of pure stationary optimal strategies. We show that the existence of such strategies for one-player games (Markov decision processes) implies the existence of such strategies for two-player games. The result is general and holds for any payoff mapping.

## 1 Introduction

Given a perfect-information zero-sum stochastic game with finite set of states and actions, the existence of pure and stationary optimal strategies is a very desirable property.

For example, since there are finitely many such strategies, computability of the values of a stochastic game is often a direct corollary of the existence of pure and stationary optimal strategies.

Of course not in every game both players have pure stationary optimal strategies, this depends on the transition rules of the game (the arena) and on the way players' payoffs are computed (the payoff function). Actually, for various payoff functions like the mean-payoff function, the discounted payoff function and also parity games, players have pure and stationary optimal strategies whatever is the arena they are playing in.

*CNRS, LaBRI, gimbert@labri.fr.

†Université Paris 7, LIAFA, zielonka@liafa.jussieu.fr.

We provide a result which is very useful for establishing existence of pure and stationary optimal strategies: if for some fixed payoff function $f$, players have pure stationary optimal strategies in every *one-player* stochastic game then this is also the case for zero-sum *two-player* stochastic games with perfect information.

In fact we prove a more general result. We show that the existence of pure stationary optimal strategies for one-player games implies the existence of such strategies for two-player games for each class of games satisfying certain closure properties. Consequently, the same result holds if one only considers deterministic arenas or only arenas without cycles except self-loops.

# 2 Stochastic Games with Perfect Information

## 2.1 Games and Arenas

Two players Max and Min are playing an infinite game on an arena. An arena is a tuple $\mathcal{A} = (S, S_{\text{Max}}, S_{\text{Min}}, A, (A(s))_{s \in S}, p)$, where the set of states $S$ is partitioned in two sets: the set $S_{\text{Max}}$ of states controlled by player Max and the set $S_{\text{Min}}$ of states controlled by Min.

For each state $s \in S$ there a non-empty set $A(s)$ of actions available at $s$, $A = \bigcup_{s \in S} A(s)$. The game is played by stages. If at some stage $i$ the game is in state $s_i \in S$ then the player controlling $s_i$ chooses an action from $A(s_i)$ and at the next stage the game enters a new state $s_{i+1}$ with probability specified by the transition mapping $p$.

Let $\mathcal{D}(S)$ be the set of probability distributions over $S$, i.e. the set of mappings $\delta : S \to [0, 1]$ such that $\sum_s \delta(s) = 1$. Transition mapping $p$ maps each pair $(s, a)$, where $s \in S$ and $a \in A(s)$, to an element of $\mathcal{D}(S)$ and $p(s, a)(t)$ gives the probability that at the next stage the game comes to state $t$ if action $a$ is executed at state $s$. To simplify the notation we shall write $p(s, a, t)$ rather than $p(s, a)(t)$.

Throughout the paper we assume that all arenas are finite, i.e. the sets of states and actions are finite (and the term arena means always finite arena).

An arena is said to be a *one-player arena* for player Max if, for every state $s$ of Min, the set $A(s)$ is a singleton. One-player arenas for player Min are defined similarly. Let us note that in any game it is essentially irrelevant who controls the states with only one available action (or if there is somebody controlling such states). Thus we can as well assume that in a one-player

arena all states are controlled by one-player, i.e. such arenas are just Markov decision processes.

## 2.2 Payoffs

A finite (resp. infinite) history in the arena $\mathcal{A}$ is a non-empty finite (resp. infinite) sequence of states and actions in $(SA)^*S$ (resp. in $(SA)^\omega$).

After an infinite history player Max receives a payoff from player Min. The goal of Max is to maximize this payoff, player Min tries to minimize it.

The payoff can be computed in various ways. For example in a mean-payoff games each action is labeled with a real number called the immediate reward and after an infinite history the payoff of player Max is the mean value of the sequence of immediate rewards. In parity games, each action is labeled with an integer called a priority and player Max receives payoff 0 or 1 depending on the parity of the highest priority seen infinitely often during the history. In both examples, the way payoffs are computed is independent from the transitions rules of the game (the arena). In general also, we want to keep these two aspects of the game independent, therefore the function computing payoff should be defined independently of a specific state or action set. The way to do that is by mean of colors.

From now on we fix a set of colors $C$ and we slightly extend the definition of arenas: each arena $\mathcal{A}$ comes with a mapping $\mathrm{col} : A \to C$ which maps actions to colors. Thus from this moment onward an arena is a tuple $(S, S_{\mathrm{Max}}, S_{\mathrm{Min}}, A, (A(s))_{s\in S}, p, \mathrm{col})$.

For the sake of simplicity we assume that the set of colors $C$ is finite (but we show later that this restriction is not essential).

In this setting a payoff mapping is a function $f : C^\omega \to \mathbb{R}$ from infinite sequences of colors to real numbers.

After an infinite history $s_0a_0s_1a_1\cdots$ whose actions are labeled by $c_0 = \mathrm{col}(a_0), c_1 = \mathrm{col}(a_1), \ldots$ player Max receives payoff $f(c_0c_1\cdots)$ from player Min.

A game is a couple $(\mathcal{A}, f)$ made of an arena and a payoff mapping.

## 2.3 Strategies

Playing a game the players use strategies. A *strategy* for player Max is a mapping $\sigma : (SA)^*S_{\mathrm{Max}} \to \mathcal{D}(A)$ such that for every finite history $s_0a_0s_1a_1\ldots s_n$

with $s_n \in S_{\text{Max}}$, the support of $\sigma(s_0 a_0 s_1 \dots s_n)$ is a subset of the set of actions available in $s_n$, i.e. for all $a \in A$, if $\sigma(s_0 \cdots s_n)(a) > 0$ then $a \in A(s_n)$. Strategies for player Min are defined similarly and denoted $\tau$.

As explained in the introduction, certain types of strategies are of particular interest. A strategy is *pure* if it chooses actions in a deterministic way, and it is *stationary* if it does not have any memory, i.e. it depends only on the current state, and not on the preceding game history. Formally:

**Definition 1.** *A strategy $\sigma$ of player $i \in \{\text{Min}, \text{Max}\}$ is said to be:*

- pure *if,* $\forall p \in (SA)^* S_i$, *if* $\sigma(p)(a) > 0$ *then* $\sigma(p)(a) = 1$,
- stationary *if,* $\forall p \in (SA)^*$, $\forall t \in S_i$, $\sigma(pt) = \sigma(t)$.

A pure stationary strategy of player $i \in \{\text{Max}, \text{Min}\}$ can be seen as a mapping $\sigma : S_i \to A$, where for each $s \in S_i$, $\sigma(s) \in A(s)$ is the action taken by player $i$ at state $s$ according to $\sigma$.

For any finite history $p \in (SA)^* S$ and action $a \in A$ we define the cones $\mathcal{O}(p)$ and $\mathcal{O}(pa)$ as the sets consisting of all infinite histories with prefix $p$ and $pa$ respectively.

In the sequel we assume that the set of infinite histories $(SA)^\omega$ is equipped with the $\sigma$-field $\mathcal{H}_\infty$ generated by the collection of all cones $\mathcal{O}(p)$ and $\mathcal{O}(pa)$. Elements of $\mathcal{H}_\infty$ are called *events*. Moreover, when there is no risk of confusion, the events $\mathcal{O}(p)$ and $\mathcal{O}(pa)$ will be denoted simply $p$ and $pa$.

Suppose that players Max and Min are playing accordingly to strategies $\sigma$ and $\tau$. Then after a finite history $s_0 a_1 \dots s_n$ the probability of choosing an actions $a_{n+1}$ is either $\sigma(s_0 a_1 \dots s_n)(a_{n+1})$ or $\tau(s_0 a_1 \dots s_n)(a_{n+1})$ depending on whether $s_n$ belongs to $S_{\text{Max}}$ or to $S_{\text{Min}}$. Fixing the initial state $s \in S$ these probabilities and the transition probability $p$ yield the following probabilities

$$\mathbb{P}_s^{\sigma,\tau}(s_0) = \begin{cases} 1 & \text{if } s_0 = s \\ 0 & \text{if } s_0 \neq s \end{cases} \tag{1}$$

is the probability of the cone $\mathcal{O}(s_0)$,

$$\mathbb{P}_s^{\sigma,\tau}(s_0 a_1 \dots s_n a_{n+1} \mid s_0 a_1 \dots s_n) = \begin{cases} \sigma(s_0 a_1 \dots s_n)(a_{n+1}) & \text{if } s_n \in S_{\text{Max}} \\ \tau(s_0 a_1 \dots s_n)(a_{n+1}) & \text{if } s_n \in S_{\text{Min}} \end{cases} \tag{2}$$

is the conditional probability of $\mathcal{O}(s_0 a_1 \dots s_n a_{n+1})$ given $\mathcal{O}(s_0 a_1 \dots s_n)$ and

$$\mathbb{P}_s^{\sigma,\tau}(s_0 a_1 \dots s_n a_{n+1} s_{n+1} \mid s_0 a_1 \dots s_n a_{n+1}) = p(s_n, a_{n+1}, s_{n+1}) \tag{3}$$

is the conditional probability of the cone $\mathcal{O}(s_0 a_1 \dots s_n a_{n+1} s_{n+1})$ given the cone $\mathcal{O}(s_0 a_1 \dots s_n a_{n+1})$.

Ionescu Tulcea's theorem [BS78] implies that there exists a unique probability measure $\mathbb{P}_s^{\sigma,\tau}$ on $\mathcal{H}_\infty$ satisfying (1), (2) and (3).

## 2.4 Game value and optimal strategies

For each finite sequence $c_1 \dots c_k$ of colors, let $\mathcal{O}(c_1 \dots c_k)$ be the cone consisting of all elements of $C^\omega$ with prefix $c_1 \dots c_k$. We equip the set $C^\omega$ with the $\sigma$-field $\mathcal{F}$ generated by all cones $\mathcal{O}(c_1 \dots c_k)$.

We extend the coloring mapping col to a mapping from infinite histories to $C^\omega$ by setting $\mathrm{col}(s_0 a_1 s_1 a_2 \dots) = \mathrm{col}(a_1)\,\mathrm{col}(a_2)\dots$. Note that col defined above is a measurable mapping from $((SA)^\omega, \mathcal{H}_\infty)$ into $(C^\omega, \mathcal{F})$.

In the sequel we assume that the payoff mapping $f$ is a measurable bounded (it is sufficient to assume that $f$ is bounded either from below or from above) mapping from $(C, \mathcal{F})$ into the set $\mathbb{R}$ of real numbers equipped with the Borel $\sigma$-field. Under this assumption the composition $f \circ \mathrm{col}$ of col and $f$ becomes a measurable mapping from the set of infinite histories into $\mathbb{R}$.

Given an arena $\mathcal{A} = (S, S_{\mathrm{Max}}, S_{\mathrm{Min}}, A, (A(s))_{s\in S}, p, \mathrm{col})$, an initial state $s$, the strategies $\sigma$ and $\tau$ of players Max and Min, the expected value of $(f \circ \mathrm{col})$ under $\mathbb{P}_s^{\sigma,\tau}$ will be denoted $\mathbb{E}_s^{\sigma,\tau}[f, \mathcal{A}]$.

The aim of player Max is to choose a strategy $\sigma$ that maximizes $\mathbb{E}_s^{\sigma,\tau}[f, \mathcal{A}]$ while player Min wish to minimize the expected payoff.

For each state $s \in S$ the following inequality always holds:

$$\underline{\mathrm{val}}_s = \sup_\sigma \inf_\tau \mathbb{E}_s^{\sigma,\tau}[f, \mathcal{A}] \leq \inf_\tau \sup_\sigma \mathbb{E}_s^{\sigma,\tau}[f, \mathcal{A}] = \overline{\mathrm{val}}_s$$

and if $\underline{\mathrm{val}}_s = \overline{\mathrm{val}}_s$ then this quantity is called the value of the game $(\mathcal{A}, f)$ for the initial state $s$ and noted as $\mathrm{val}_s(f, \mathcal{A})$.

Strategies $\sigma^\sharp$ and $\tau^\sharp$ for players Max and Min are termed *optimal* if for all strategies $\sigma$ and $\tau$ and all states $s \in S$

$$\mathbb{E}_s^{\sigma,\tau^\sharp}[f, \mathcal{A}] \leq \mathbb{E}_s^{\sigma^\sharp,\tau^\sharp}[f, \mathcal{A}] \leq \mathbb{E}_s^{\sigma^\sharp,\tau}[f, \mathcal{A}]\,.$$

If $\sigma^\sharp$ and $\tau^\sharp$ are optimal then, for each state $s$, the game has the value and $\mathrm{val}_s(f, \mathcal{A}) = \mathbb{E}_s^{\sigma^\sharp,\tau^\sharp}[f, \mathcal{A}]$.

Under the hypothesis that $f$ is measurable and bounded, Martin's theorem [Mar98] guarantees that every state has a value. Notice however that

Martin's theorem does not guarantee the existence of optimal strategies and even less the existence of pure stationary optimal strategies. The existence of such strategies is the main subject of this paper.

# 3 From one-player to two-player games: a transfer theorem

Our main theorem states that the existence of pure stationary optimal strategies in one-player games implies the same property for two-player games.

**Theorem 2.** *Let $f : C^\omega \to \mathbb{R}$ be a measurable and bounded payoff function.*

*Suppose that for every* one-player *arena $\mathcal{A}$ colored by $C$, the player controlling $\mathcal{A}$ has a pure stationary optimal strategy for the game $(\mathcal{A}, f)$.*

*Then for every two-player arena $\mathcal{A}$ labelled by $C$ both players have pure stationary optimal strategies in the game $(\mathcal{A}, f)$.*

Actually the result can be refined to sub-classes of arenas with good closure properties.

**Subarenas.** An arena $\mathcal{A}'$ is a subarena of an arena $\mathcal{A}$ if both arenas are identical except that for every state $s$ the set of actions available in $s$ in $\mathcal{A}'$ is a subset (not necessarily proper) of the set of actions available in $s$ in $\mathcal{A}$.

**Merged union of two arenas.** Intuitively, the *merged union* of two arenas $\mathcal{A}_L$ and $\mathcal{A}_R$ colored by $C$ consists in merging a state of $\mathcal{A}_L$ with a state of $\mathcal{A}_R$.

Formally, let $\mathcal{A}_L = (S_L, S_{L,\mathrm{Max}}, S_{L,\mathrm{Min}}, A_L, (A_L(s))_{s\in S_L}, p_L, \mathrm{col}_L)$ and $\mathcal{A}_R = (S_R, S_{R,\mathrm{Max}}, S_{R,\mathrm{Min}}, A_R, (A_R(s))_{s\in S_R}, p_R, \mathrm{col}_R)$ be two arenas called respectively the left and the right arena. We assume that:

1. there exists a unique state called the pivot state and denoted $\pi$ such that $S_L \cap S_R = \{\pi\}$, and this pivot state is controlled by the same player in both arenas $\mathcal{A}_L$ and $\mathcal{A}_R$,
2. the sets $A_L(\pi)$ and $A_R(\pi)$ of actions available in the pivot state in arenas $\mathcal{A}_L$ and $\mathcal{A}_R$ are disjoint: $A_L(\pi) \cap A_R(\pi) = \emptyset$,
3. the coloring mappings are compatible i.e. for every action $a \in A_L \cap A_R$, $\mathrm{col}_L(a) = \mathrm{col}_R(a)$.

When making use of the merged union in proofs, we start with renaming states and actions so that these three constraints are met.

The *merged union* of arenas $\mathcal{A}_L$ and $\mathcal{A}_R$ is the arena $\mathcal{A}_{LR} = (S_{LR}, S_{LR,\mathrm{Max}}, S_{LR,\mathrm{Min}}, A_L \cup A_R, (A_{LR}(s))_{s \in S_{LR}}, p_{LR}, \mathrm{col}_{LR})$ defined as follows.

- The set of states $S_{LR}$ is the union of $S_L$ and $S_R$, and the controller of a state is the same in $\mathcal{A}_{LR}$ as in $\mathcal{A}_L$ or $\mathcal{A}_R$. By hypothesis, there is a unique player controlling the pivot state in both arenas, suppose from now on that this player is Max.
- In arena $\mathcal{A}_{LR}$, an action is available in the pivot state if it is available either in arena $\mathcal{A}_L$ or in arena $\mathcal{A}_R$, i.e. $A_{LR}(\pi) = A_L(\pi) \cup A_R(\pi)$. By hypothesis this union is disjoint. The actions available in all other states are the same as in $\mathcal{A}_L$ or $\mathcal{A}_R$.
- Transition probabilities in $\mathcal{A}_{LR}$ are inherited directly from $\mathcal{A}_L$ and $\mathcal{A}_R$, which makes sense because in the only common state of $\mathcal{A}_L$ and $\mathcal{A}_R$ the sets of available actions are disjoint.
- The coloring mapping $\mathrm{col}_{LR}$ is the extension of the two compatible mappings $\mathrm{col}_L : A_L \to C$ and $\mathrm{col}_R : A_R \to C$ to $A_L \cup A_R$.

The merged union merging two states controlled by player Min is defined mutatis mutandis.

We can now restate Theorem 2 with weaker hypotheses.

**Theorem 3.** *Let $\mathcal{C}$ be a collection of game arenas colored by $C$ closed under renaming states and actions, taking subarenas and taking merged unions. Let $f : C^\omega \to \mathbb{R}$ be a bounded measurable payoff function. Suppose that for every one-player arena $\mathcal{A}$ in $\mathcal{C}$, the player controlling $\mathcal{A}$ has a pure stationary optimal strategy for the game $(\mathcal{A}, f)$. Then, for every arena $\mathcal{A}$ in $\mathcal{C}$, both players have pure stationary optimal strategies for the game $(\mathcal{A}, f)$.*

Of course Theorem 2 is a special case of Theorem 3 since the collection of all arenas with a fixed coloring is obviously closed under taking subarenas and merged union.

*Proof.* The proof is carried out by induction on the size of arenas where the size $N(\mathcal{A})$ of an arena $\mathcal{A} = (S, S_{\mathrm{Max}}, S_{\mathrm{Min}}, A, (A(s))_{s \in S}, p, \mathrm{col})$ is defined as $N(\mathcal{A}) = \sum_{s \in S}(|A(s)| - 1)$.

The base case is immediate: if $N(\mathcal{A}) = 0$ then only one action is available at each state, hence there is only one strategy for each player and this strategy is pure stationary and optimal.

Now we prove the inductive step. Suppose that for some $n \in \mathbb{N}$, for every arena $\mathcal{B} \in \mathcal{C}$ of size $N(\mathcal{B}) \leq n$, both players have pure stationary optimal strategies in the game $(\mathcal{B}, f)$. Let $\mathcal{A}$ be an arena of size $n+1$.

Let $\pi$ be a state with at least two available actions, $|A(\pi)| \geq 2$. We fix such a state $\pi$ for the rest of the proof and we will call it the *pivot state*. We also assume that

$$\text{the pivot state } \pi \text{ is controlled by player Max} \ , \tag{4}$$

the symmetric case when $\pi$ is controlled by player Min will be discussed briefly later.

Now we build two subarenas of $\mathcal{A}$ to which we apply the inductive hypothesis.

Let us partition the set $A(\pi)$ of actions available in the pivot state into two non-empty sets $A_L(\pi)$ and $A_R(\pi)$. Restricting the actions available at $\pi$ to $A_L(\pi)$ we get the arena $\mathcal{A}_L$, similarly restricting the actions available at $\pi$ to $A_R(\pi)$ we get the arena $\mathcal{A}_R$. For all states other than the pivot state both arenas keep the same sets of available action as in $\mathcal{A}$.

The sizes of arenas $\mathcal{A}_L$ and $\mathcal{A}_R$ are smaller than the size of $\mathcal{A}$, thus, by induction hypothesis, players Max and Min have pure stationary optimal strategies in games $(\mathcal{A}_L, f)$ and $(\mathcal{A}_R, f)$.

We denote $\sigma_L^\sharp$ and $\tau_L^\sharp$ pure stationary optimal strategies of Max and Min in $(\mathcal{A}_L, f)$, and $\sigma_R^\sharp$ and $\tau_R^\sharp$ their pure stationary optimal strategies in $(\mathcal{A}_R, f)$.

We are going to prove that either $\sigma_L^\sharp$ or $\sigma_R^\sharp$ is optimal not only in $(\mathcal{A}_L, f)$ or $(\mathcal{A}_R, f)$ but also in the game $(\mathcal{A}, f)$ as well.

However how can we tell $\sigma_L^\sharp$ and $\sigma_R^\sharp$ apart and determine which one is optimal in $(\mathcal{A}, f)$?

For that we need to build the merged union of arenas $\mathcal{A}_L$ and $\mathcal{A}_R$. To begin with we rename the states of arenas $\mathcal{A}_L$ and $\mathcal{A}_R$ different from pivot state $\pi$ in such a way that $\pi$ becomes the only common state of both arenas. In other words we make two copies of each state $t \in S \setminus \{\pi\}$ and obtain a left copy $t_L$ and a distinct right copy $t_R$.

It is convenient to define explicitly two corresponding renaming mappings:

$$\varphi_L : S \to S_L \quad \text{and} \quad \varphi_R : S \to S_R \tag{5}$$

such that $\varphi_L(t) = t_L$ and $\varphi_R(t) = t_R$, for $t \in S \setminus \{\pi\}$, and $\varphi_L(\pi) = \varphi_R(\pi) = \pi$ and the converse mapping

$$\varphi : S_L \cup S_R \to S \tag{6}$$

such that $\varphi(t_L) = \varphi(t_R) = t$ for $t \in S \setminus \{\pi\}$ and $\varphi(\pi) = \pi$.

We call $\mathcal{B}_L$ and $\mathcal{B}_R$ the two arenas obtained from $\mathcal{A}_L$ and $\mathcal{A}_R$ after the renaming. Thus arenas $\mathcal{B}_L$ and $\mathcal{B}_R$ share the same set of actions and their only common state is the pivot state.

By design, arenas $\mathcal{B}_L$ and $\mathcal{B}_R$ meet the three constraints in the definition of the merged union. Let $\mathcal{A}_{LR}$ be the merged union of $\mathcal{B}_L$ and $\mathcal{B}_R$.

Let us describe $\mathcal{A}_{LR}$ in detail. The set of states of $\mathcal{A}_{LR}$ is

$$S_{LR} = S_L \cup S_R.$$

We call the states of $S_L$ the left states of $\mathcal{A}_{LR}$ while $S_R$ are the right states of $\mathcal{A}_{LR}$ (in particular $\pi$ is the only state that is both left and right).

For each state $s \in S_{LR}$, $s$ and $\varphi(s)$ have the same sets of available actions, $A_{LR}(s) = A(\varphi(s)))$.

The transition probabilities $p_{LR}$ in $\mathcal{A}_{LR}$ are inherited from $\mathcal{A}$ with the proviso that at pivot state $\pi$ left actions from $A_L(\pi)$ lead to left states and right actions from $A_R(\pi)$ lead to right states:

$$\text{for } s \in S \setminus \{\pi\}, t \in S \text{ and } a \in A(s), \qquad p_{LR}(s_L, a, \varphi_L(t)) = p_{LR}(s_R, a, \varphi_R(t)) = p(s, a, t)$$

and

$$\begin{aligned} \text{for } a \in A_L(\pi), \quad & p_{LR}(\pi, a, \varphi_L(t)) = p(\pi, a, t) \\ \text{for } a \in A_R(\pi), \quad & p_{LR}(\pi, a, \varphi_R(t)) = p(\pi, a, t). \end{aligned}$$

Of course all actions retain their colors.

It is important to note the arenas $\mathcal{A}$ and $\mathcal{A}_{LR}$ are very closely related. $\mathcal{A}_{LR}$ was obtained from $\mathcal{A}$ by splitting each state $s$ different from the pivot into two states $s_L$ and $s_R$. All three states have the same set of available actions $A(s) = A_{LR}(s_L) = A_{LR}(s_R)$ however actions executed at $s_L$ lead to corresponding left copies of target states while the same actions executed in $s_R$ lead to the corresponding right copies of target states. The only junction between the left and the right part of $\mathcal{A}_{LR}$ takes place in the pivot state

where the left actions $A_L(\pi)$ have targets in left states while right actions have targets in right states.

Let $s \in S$ be a state of $\mathcal{A}$ and let $\varphi_L(s)$ be the corresponding "left" state in $\mathcal{A}_{LR}$. Let $H_s(\mathcal{A})$ be the set of all (finite and infinite) histories in $\mathcal{A}$ starting at $s$. Let $H_{\varphi_L(s)}(\mathcal{A}_{LR})$ be the set of all histories in $\mathcal{A}_{LR}$ starting at $\varphi_L(s)$.

There is a natural bijective correspondence between $H_s(\mathcal{A})$ and $H_{\varphi_L(s)}(\mathcal{A}_{LR})$. For each history $h = s_0 a_1 s_1 a_2 \ldots \in H_{\varphi(s_L)}(\mathcal{A}_{LR})$ the corresponding history in arena $\mathcal{A}$ has the form

$$\varphi(h) = \varphi(s_0) a_1 \varphi(s_1) a_2 \ldots, \tag{7}$$

where $\varphi$ is given by (6). We call $\varphi(h)$ the projection of $h$. Conversely, for each history $h = s_0 a_1 s_1 a_2 \ldots \in H_{s_0}(\mathcal{A})$ there is a unique corresponding history $h'$ in $\mathcal{A}_{LR}$ starting at $\varphi_L(s_0)$, $h' = s'_0 a_1 s'_1 a_2 \ldots$, where

$$s'_k = \begin{cases} \varphi_R(s_k) & \text{if there is } i < k \text{ such that } s_i = \pi,\ a_{i+1} \in A_R(\pi), \\ & \text{and } s_j \neq \pi \text{ for } i < j < k \\ \varphi_L(s_k) & \text{otherwise.} \end{cases}$$

We shall write $\varphi_L(h)$ to denote such a history $h'$ and call it the left-lifting of $h$. Obviously, the projection and the left lifting are mutual inverses.

The symmetric right lifting where the corresponding lifted history starts at $\varphi_R(s_1)$ is defined in a similar way and defines a bijection between $H_s(\mathcal{A})$ and $H_{\varphi_R(s)}(\mathcal{A}_{LR})$. Again the projection $\varphi$ defined by (7) is the inverse of $\varphi_R$.

Not only the histories but also the strategies in $\mathcal{A}$ and $\mathcal{A}_{LR}$ are closely related.

Let $\sigma$ be a strategy of player Max on $\mathcal{A}$. This strategy can be lifted up to $\mathcal{A}_{LR}$ using $\varphi$. The image $\varphi(\sigma)$ of $\sigma$ under $\varphi$ is the strategy on $\mathcal{A}_{LR}$ defined in the following way. For each finite history $h$ in $\mathcal{A}_{LR}$ terminating at a state controlled by player Max and for each action available after $h$ we set

$$\varphi(\sigma)(h)(a) = \sigma(\varphi(h))(a) \ .$$

We call the strategy $\varphi(\sigma)$ the lifting of $\sigma$ (note that to lift up a strategy we use the projection $\varphi$ of histories). The definition (3) is very natural, to determine the probability of execution $a$ after a history $h$ we project this history on $\mathcal{A}$ and apply the strategy $\sigma$. Let us note that if $\sigma$ is pure stationary then $\varphi(\sigma)$ is also pure stationary. The lifting of strategies of player Min from $\mathcal{A}$ to $\mathcal{A}_{LR}$ is defined similarly.

We need also an inverse operation which transforms strategies defined on $\mathcal{A}_{LR}$ into strategies on $\mathcal{A}$. In fact we have two such natural transformations, depending whether histories in $\mathcal{A}$ are lifted up to $\mathcal{A}_{LR}$ using $\varphi_R$ or $\varphi_L$.

Let $\tau_{LR}$ be a strategy of player Min on $\mathcal{A}_{LR}$. Then $\varphi_L(\tau_{LR})$, *the left projection* of $\tau_{LR}$, is a strategy on $\mathcal{A}$ defined in the following way. Let $h$ be a finite history in $\mathcal{A}$ ending at a state controlled by player Min and let $a$ be an action available after $h$. Then we set

$$\varphi_L(\tau_{LR})(h)(a) = \tau_{LR}(\varphi_L(h))(a) \ . \tag{8}$$

(Note the use of the left lifting of histories to define the left projection of strategies.)

Thus to obtain the probability of executing $a$ after a history $h$ in $\mathcal{A}$ we use the left lifting to lift up this history to arena $\mathcal{A}_{LR}$, and we take the probability of executing $a$ after the lifted history $\varphi_L(h)$ given by by $\tau_{LR}$.

Let us stress that the left projection of a pure stationary strategy is always pure but not necessarily stationary. This is due to the fact that if $\tau_{LR}$ is pure stationary on $\mathcal{A}_{LR}$ then still $\tau_{LR}$ can choose different actions on the left and on the right copies of a state $s$, i.e. we can have $\tau_{LR}(s_L) \neq \tau_{LR}(s_R)$ for $s \in S_{\mathrm{Min}}$. One can prove however that if $\tau_{LR}$ is pure stationary then $\varphi_L(\tau_{LR})$ can be implemented using a finite memory (more exactly a two-valued memory is sufficient). We will return to this point later.

There is of course a corresponding notion of the right projection $\varphi_R(\tau_{LR})$ where the right lifting is used to lift up histories from $\mathcal{A}$ to $\mathcal{A}_{LR}$.

Similarly, it should be obvious how to define the left and the right projections of strategies of player Max from $\mathcal{A}_{LR}$ to $\mathcal{A}$.

We end by listing some easy but crucial properties of lifting and projection.

**Lemma 4.** *Let $s \in S$ be a state of $\mathcal{A}$, let $\sigma$ be a strategy of player* Max *on $\mathcal{A}$ and let $\tau_{LR}$ be a strategy of player* Min *on $\mathcal{A}_{LR}$. Then*

$$\mathbb{E}_s^{\sigma,\varphi_L(\tau_{LR})}[f, \mathcal{A}] = \mathbb{E}_{\varphi_L(s)}^{\varphi(\sigma),\tau_{LR}}[f, \mathcal{A}_{LR}] \, . \tag{9}$$

*Proof.* Let $h = s_0 a_1 s_1 a_2 \dots s_n$ be a finite history in $\mathcal{A}$ starting at state $s = s_0$. Let $h' = \varphi_L(h) = s'_0 a_1 s'_1 a_2 \dots s'_n$ be the corresponding history in $\mathcal{A}_{LR}$ obtained by the left lifting of $h$. Thus $s_n$ and $s'_n$ have the same set of available actions and $s'_n$ is either the left or the right copy of $s_n$.

Suppose that $s_n$ and $s'_n$ are controlled by player Max. From the definition of the strategy $\varphi(\sigma)$ it follows that the probability of executing an action $a$ after $h$ given by strategy $\sigma$ is the same as the probability of executing $a$ after $h'$ by strategy $\varphi(\sigma)$.

Similarly, if $s_n$ is controlled by player Min then, by definition of $\varphi_L(\tau_{LR})$ the probability of executing $a$ after $h$ if player Min uses strategy $\varphi_L(\tau_{LR})$ id the same as the probability of executing $a$ after strategy $h'$ if player Min uses $\tau_{LR}$. Given action $a$ available in $s_n$ (and thus in $s'_n$) the distribution probability of the states reached on the next stage is given by the transition probabilities.

All allows to show by a trivial induction on the length of histories that

$$\mathbb{P}_s^{\sigma,\varphi_L(\tau_{LR})}(h) = \mathbb{P}_{\varphi_L(s)}^{\varphi(\sigma),\tau_{LR}}(\varphi_L(h)),$$

i.e. the probabilities of the cones generated by $h$ and $\varphi_L(h)$ are the same. This implies that the mapping sending infinite histories $h$ in $\mathcal{A}$ to infinite histories $\varphi_L(h)$ in $\mathcal{A}_{LR}$ preserves probabilities of all events. The same mapping preserves also color sequences of infinite plays, thus it preserves payoffs, which proves (9).

□

Let us return to the proof of Theorem 3. We define a pure stationary strategy $\tau_{LR}$ of player Min on arena $\mathcal{A}_{LR}$ as the union $\tau_L^\sharp \cup \tau_R^\sharp$ of strategies $\tau_L^\sharp$ and $\tau_R^\sharp$, remember that $\tau_L^\sharp$ and $\tau_R^\sharp$ are two pure and stationary strategies that are optimal respectively in the games $(\mathcal{A}_L, f)$ and $(\mathcal{A}_R, f)$. Thus strategy $\tau_{LR}$ consists in playing according to $\tau_L^\sharp$ on the left part of $\mathcal{A}_{LR}$ and according to $\tau_R^\sharp$ on the right part of $\mathcal{A}_{LR}$. Formally, for every state $t \in S_{\text{Min}}$ and its two copies $t_L = \varphi_L(t)$ and $t_R = \varphi_R(t)$ in $\mathcal{A}_{LR}$,

$$\tau_{LR}(t_L) = \tau_L^\sharp(t) \text{ and } \tau_{LR}(t_R) = \tau_R^\sharp(t) \ .$$

According to hypothesis 4, the pivot state $\pi$ is controlled by player Max hence there is no need to define $\tau_{LR}(\pi)$.

Having fixed the strategy $\tau_{LR}$ of player Min in the game $(\mathcal{A}_{LR}, f)$ we consider a one-player game on $\mathcal{A}_{LR}$ where player Min is obliged to play according to $\tau_{LR}$ while player Max keeps all his moves unrestrained. This amounts to examine a game on a subarena $\mathcal{A}_{LR}[\tau_{LR}]$ of $\mathcal{A}_{LR}$ such that for each state controlled by player Min there is only one available action, the one provided by strategy $\tau_{LR}$. Player Max conserves in $\mathcal{A}_{LR}[\tau_{LR}]$ all actions

available to him in $\mathcal{A}_{LR}$. Thus $\mathcal{A}_{LR}[\tau_{LR}]$ is a one-player arena controlled by player Max. Therefore, by hypothesis, player Max has a pure stationary optimal strategy $\sigma_{LR}$ in the game $(\mathcal{A}_{LR}[\tau_{LR}], f)$. (Thus, intuitively, $\sigma_{LR}$ is the best response of player Max to strategy $\tau_{LR}$ of player Min when the game $(\mathcal{A}_{LR}, f)$ is played).

Since the pivot state $\pi$ is controlled by player Max, $\sigma_{LR}$ is defined on $\pi$ and $\sigma_{LR}(\pi) \in A(\pi) = A_L(\pi) \cup A_R(\pi)$. Without loss of generality we can assume that

$$\sigma_{LR}(\pi) \in A_L(\pi) \tag{10}$$

(if this is not the case then it suffices to switch left and right).

We claim that (10) implies that $\sigma_L^\sharp$ is an optimal strategy of player Max in the game $(\mathcal{A}, f)$. It turns out that the corresponding optimal strategy of player Min in $(\mathcal{A}, f)$ is the left projection $\varphi_L(\tau_{LR})$ of the strategy $\tau_{LR}$.

To prove the optimality of strategies $\sigma_L^\sharp$ and $\varphi_L(\tau_{LR})$ we shall show that, for each state $s \in S$ and all strategies $\sigma$ and $\tau$ of players Max and Min in the game $(\mathcal{A}, f)$,

$$\mathbb{E}_s^{\sigma,\varphi_L(\tau_{LR})}[f, \mathcal{A}] \leq \mathrm{val}_s(f, \mathcal{A}_L) \leq \mathbb{E}_s^{\sigma_L^\sharp,\tau}[f, \mathcal{A}]\,. \tag{11}$$

In other words, using $\sigma_L^\sharp$ in $(\mathcal{A}, f)$, player Max secures the (expected) payoff of at least $\mathrm{val}_s(f, \mathcal{A}_L)$ while player Min using $\varphi_L(\tau_{LR})$ guarantees that the payoff will not exceed $\mathrm{val}_s(f, \mathcal{A}_L)$. This implies immediately that $\sigma_L^\sharp$ and $\varphi_L(\tau_{LR})$ are optimal in $(\mathcal{A}, f)$ and that the value $\mathrm{val}_s(f, \mathcal{A})$ is equal to $\mathrm{val}_s(f, \mathcal{A}_L)$.

Let us note also that since we have assumed that $\sigma_L^\sharp$ and $\tau_L^\sharp$ are optimal in the game $(\mathcal{A}_L, f)$, the value $\mathrm{val}_s(f, \mathcal{A}_L)$ exists and is equal to $\mathbb{E}_s^{\sigma_L^\sharp,\tau_L^\sharp}[f, \mathcal{A}_L]$.

We begin by proving the right-hand side of (11).

If $\tau$ is any strategy of player Min on arena $\mathcal{A}$ then restricting $\tau$ to histories which are also valid in subarena $\mathcal{A}_L$ we obtain a strategy of player Min in the game $(\mathcal{A}_L, f)$, to avoid clutter we note this restricted strategy using the same symbol $\tau$. By optimality of $\sigma_L^\sharp$ in $(\mathcal{A}_L, f)$ we get $\mathbb{E}_s^{\sigma_L^\sharp,\tau}[f, \mathcal{A}] = \mathbb{E}_s^{\sigma_L^\sharp,\tau}[f, \mathcal{A}_L] \geq \mathrm{val}_s(f, \mathcal{A}_L)$. This terminates the proof of the right-hand side of (11).

The proof of the left-hand side inequality of (11) is less straightforward and goes through several stages.

Take any strategy $\sigma$ for player Max defined on arena $\mathcal{A}$ and lift it up to the strategy $\varphi(\sigma)$ on $\mathcal{A}_{LR}$. By Lemma 4

$$\mathbb{E}_s^{\sigma,\varphi_L(\tau_{LR})}[f, \mathcal{A}] = \mathbb{E}_s^{\varphi(\sigma),\tau_{LR}}[f, \mathcal{A}_{LR}] \quad .$$

However, by definition, $\sigma_{LR}$ is the best response strategy of player Max if player Min uses $\tau_{LR}$, implying

$$\mathbb{E}_s^{\varphi(\sigma),\tau_{LR}}[f,\mathcal{A}_{LR}] \leq \mathbb{E}_s^{\sigma_{LR},\tau_{LR}}[f,\mathcal{A}_{LR}] \quad .$$

If player Max uses the strategy $\sigma_{LR}$ then (10) implies that for plays starting in the left part of $\mathcal{A}_{LR}$ we never come to the right part of $\mathcal{A}_{LR}$. But on the left part of $\mathcal{A}_{LR}$ the strategy $\tau_{LR}$ is nothing else but $\tau_L^\sharp$, the optimal strategy of Min in the game $(\mathcal{A}_L, f)$. (10) implies also that strategy $\sigma_{LR}$ when restricted to the left states of $\mathcal{A}_{LR}$ is a valid pure stationary strategy of player Max an $\mathcal{A}_L$, that we denote $\sigma_L$. Thus

$$\mathbb{E}_s^{\sigma_{LR},\tau_{LR}}[f,\mathcal{A}_{LR}] = \mathbb{E}_s^{\sigma_L,\tau_L^\sharp}[f,\mathcal{A}_L] \quad ,$$

and optimality of $\tau_L^\sharp$ implies

$$\mathbb{E}_s^{\sigma_L,\tau_L^\sharp}[f,\mathcal{A}_L] \leq \mathrm{val}_s(f,\mathcal{A}_L) \quad .$$

This terminates the proof of the left-hand side inequality in (11).

Thus $(\sigma_L^\sharp, \varphi_L(\tau_{LR}))$ is a pair of optimal strategies in the game $(\mathcal{A}, f)$. The problem is that while $\sigma_L^\sharp$ is pure stationary, $\varphi_L(\tau_{LR})$ may be non-stationary.

To end the proof, we use the inherent symmetry of the problem.

Take a state $s$ of $\mathcal{A}$ controlled by player Min such that $|A(s)| > 1$ and make it the pivot.

Proceeding exactly as previously we will find a pair of optimal strategies $(\sigma^\sharp, \tau^\sharp)$ in the game $(\mathcal{A}, f)$ but now this is the strategy $\tau^\sharp$ of player Min that will be pure stationary and finally $(\sigma_L^\sharp, \tau^\sharp)$ will be the required pair of pure stationary strategies optimal in $(\mathcal{A}, f)$. □

## 3.1 Remarks

In the proof of Theorem 3 we mentioned that if $\tau_{LR}$ is a pure stationary strategy of player Min on $\mathcal{A}_{LR}$ then its projection $\varphi_L(\tau_{LR})$ on $\mathcal{A}$ is a finite memory strategy. This fact is not used in the proof of Theorem 3 but may be of some interest if one is interested in finite memory strategies. For this reason we provide a short proof of this fact (but the reader can go directly to the next section since the material exposed here is not used elsewhere in this paper).

Strategy $\tau_{LR}$ is played on an arena where all states different from the pivot are duplicated. We construct a strategy $\tau^\sharp$ on $\mathcal{A}$ that will simulate the strategy $\tau_{LR}$. To make a simulation of $\tau_{LR}$ on $\mathcal{A}$ possible we provide $\tau^\sharp$ with a finite (two state) memory $\mathcal{M}$ used to remember the last move of player Max in the pivot state $\pi$. The memory used by $\tau^\sharp$ can store one of two values, either $L$ or $R$, and initially it is set to $L$,

$$\text{initialization:} \quad \mathcal{M} := L \tag{12}$$

(the initialization should match (10), if $\sigma_{LR}(\pi) \in A_R(\pi)$ then the memory should be initialized to $R$).

Now the strategy $\tau^\sharp$ is defined in the following way: for each state $s \in S_L$ of $\mathcal{A}$ controlled by Min

$$\tau^\sharp(s) = \begin{cases} \tau^\sharp_L(s) & \text{if } \mathcal{M} = L, \\ \tau^\sharp_R(s) & \text{if } \mathcal{M} = R. \end{cases} \tag{13}$$

Thus player Min plays either according to his optimal strategy in the game $(\mathcal{A}_L, f)$ or according to his optimal strategy in $(\mathcal{A}_R, f)$ depending on the current state of the memory.

To complete the description of the strategy $\tau^\sharp$ it remains to explain how memory $\mathcal{M}$ evolves in time. Each time the play crosses the pivot state $\pi$ the memory is set to $L$ if the action chosen by player Max belongs to $A_L(\pi)$ and it is set to $R$ if the chosen action belongs to $A_R(\pi)$. The memory is never modified when other states are visited.

Thus, intuitively, the purpose of $\mathcal{M}$ is to remember the last move of player Max at the pivot state, if the last action played at $\pi$ was in $A_L(\pi)$ then from this moment onward (until the next visit to $\pi$) the game resembles the game on arena $\mathcal{A}_L$ and player Min tries to respond by playing his optimal strategy $\tau^\sharp_L$ on $\mathcal{A}_L$. Otherwise, if the last action played at $\pi$ was in $A_R(\pi)$ then from this moment onward (until the next visit to $\pi$) the game resembles the game on arena $\mathcal{A}_R$ and player Min tries to respond by playing his optimal strategy $\tau^\sharp_R$ on $\mathcal{A}_R$.

It is easy to see that $\tau^\sharp$ and $\varphi_L(\tau_{LR})$ are in fact equal (for pure stationary $\tau_{LR}$), i.e. after each finite history ending at a state controlled by Min both strategies choose the same action to execute.

# 4 Comments

## 4.1 Deterministic arenas

The reader may wonder what is the use of refining Theorem 2 into Theorem 3.

The main reason is that Theorem 3 can be used to address both stochastic games as well as *deterministic* games, i.e. games played on arenas where transition probabilities are either 0 or 1. This is useful because deterministic games are of independent interest (in particular deterministic games prevail in some applications in computer science [GTW02]).

Moreover, the class of payoff mappings that admit pure stationary optimal strategies for the class of *deterministic* games strictly contains the class of payoff mappings that admit pure and stationary optimal strategies for prefect-information *stochastic* games. This fact can be illustrated by so-called *simple parity games*. In simple parity games actions are colored by non-negative integers and the payoff is either 0 or 1 depending on the parity of the greatest integer visited during an infinite history (payoff is 0 if the greatest color seen during the play is even, otherwise, if the greatest integer visited during the play is odd then the payoff is 1). It is easy to see that for one-player deterministic games the player controlling the arena has a pure stationary optimal strategy. By Theorem 3 this implies that both players have pure stationary optimal strategies for two-players deterministic simple parity games.

On the other hand, Florian Horn [Hor07] provided an example of a stochastic one-player simple parity game such for which no stationary strategy is optimal, see Figure 1.

In the game in Figure 1 states $w, v, t$ have only one available action and this action is deterministic. State $s$ has also one available action but this action is non-deterministic, with probability 0.5 it leads either to $t$ or to $u$. State $u$ is the only state with two available actions, we call them `stay` and `risk`. Action `stay` is deterministic and leads back to $u$. Action `risk` leads with probability 0.5 either to $v$ or to $w$. All actions are colored with integer numbers as shown in the figure and all states are controlled by player Max.

Suppose that the game starts at state $t$. Note that for any stationary strategy the expected payoff of player Max is 0. Indeed, suppose that at $u$ he always takes action `risk` with probability $\alpha$ and action `stay` with probability $1-\alpha$. If $\alpha = 0$ then he always executes `stay` and the sequence $2, 1, 1, 1 \ldots$ of visited colors gives him payoff 0. If $\alpha > 0$ then with probability 1 he visits

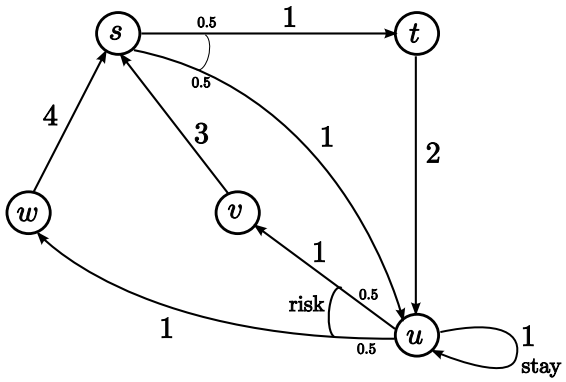


Figure 1: A simple parity game with all states controlled by player Max.

state $w$ and the sequence of visited colors contains 4 which again results in payoff 0.

On the other hand player Max has strategies assuring him the expected payoff 0.5. During the first visit at $u$ player Max should play `risk`. If he is lucky then executing `risk` for the first time he reaches $v$ and executes the action labeled 3 which secures him payoff 1 provided that when he returns to state $u$ he will always play `stay` (i.e. he plays `risk` only once).

If he is unlucky when executing `risk` for the first time then he reaches state $w$ and next he is obliged to execute the action labeled 4. What he will do afterward is irrelevant his payoff will be 0.

The non-stationary strategy described above secures for player Max the expected payoff 0.5 and this is the best that he can obtain in this game.

In general deterministic games are simpler than their stochastic counterparts. This does not mean however than they are always trivial. In practice it is often easy to prove the existence of stationary optimal strategies for deterministic *one-player* games, this is certainly easy for example in the case of parity games [GTW02] and in the case of mean-payoff games [EM79] where constructing stationary optimal strategies for one-player games is an easy exercise. However for example the proof of the existence of pure stationary strategies for both players in deterministic parity or mean-payoff games is not so straightforward [BSV04]. Thus Theorem 3 comes as a handy tool even for deterministic games.

Let us notice also that there exist other classes of arenas, closed by merged union and subarenas, which are of interest: for example, arenas with no cycles

except self-loops or arenas with rational transition probabilities.

## 4.2 Perfect information stochastic games

For stochastic games, to our knowledge, Theorem 3 is the key to the simplest proofs of existence of pure and stationary optimal strategies in the two-player games.

For example several proofs of the existence of pure stationary optimal strategies for perfect information stochastic parity games are known. However all these proofs are rather obscure [MM02, CJH04, Zie04].

On the other hand the proof of the existence of pure stationary optimal strategies for *one-player* games based on ideas of [CY90, dA97]) is rather elementary. And now Theorem 3 allows to pass directly from one-player to two-player games. In particular we can see now that the question of the existence of pure stationary optimal strategies has essentially the same difficulty for one-player games as for two-player games.

For mean-payoff games the situation is more complex. Mean-payoff stochastic games come in two distinct flavors. One can take as the expected payoff a limit (either $\limsup$ or $\liminf$) of $\frac{1}{n}\mathbb{E}_s^{\sigma,\tau}\left[\sum_{i=1}^{n} r(a_i)\right]$, where $r(a_i)$ is the real valued reward associated with action $a_i$. Here we calculate the expected mean payoff over $n$ stages and next tend $n$ to infinity. one-player games of this type have pure stationary strategies but these games do not enter in our framework.

What we need are mean-payoff games where the payoff is calculated as $\mathbb{E}_s^{\sigma,\tau}\left[\limsup_n \frac{1}{n}\sum_{i=1}^{n} r(a_i)\right]$, i.e. the mean-value is calculated separately for each infinite history and only then the expectation is applied. It seems that the existence of pure stationary optimal strategies for one-player games of this type was first proved in [Bie87]. A more elegant and more readable proof linking mean-payoff and discounted games is given in [Ney04]. Again Theorem 3 allows now an effortless extension of this result to two player games.

## 4.3 One-player stochastic games and sub-mixing payoff functions

A convenient way of proving the existence of pure and stationary optimal strategies for one-player stochastic games was discovered recently by the first author.

The main result of [Gim07] states that if a payoff function $f$ is *sub-mixing* then for each one-player stochastic game controlled by player Max, this player has a pure stationary optimal strategy.

A payoff function $f$ is sub-mixing if for every sequence $u_0, u_1, u_2, \ldots$ of non-empty finite sequences of colors,

$$f(u_0 u_1 c_2 \cdots) \leq \max\{f(u_0 u_2 \cdots), f(u_1 u_3 \cdots)\} \ .$$

Note that to prove that one-player games controlled by player Min have pure stationary optimal strategies it suffices to verify that $-f$ is sub-mixing. Thus sub-mixity can be used to verify the existence of pure stationary optimal strategies for all one-player games and next Theorem 2 extends the existence of pure stationary optimal strategies to two player games. Let us note that this route can be taken for parity and mean-payoff games since their payoffs are sub-mixing.

# References


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

[BS78] D. Bertsekas and S. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*. Academic Press, 1978.

[BSV04] H. Björklund, S. Sandberg, and S. Vorobyov. Memoryless determinacy of parity and mean payoff games: a simple proof. *Theoretical Computer Science*, 310(1-3):365–378, 2004.

[CJH04] K. Chatterjee, M. Jurdzinski, and T. A. Henzinger. Quantitative stochastic parity games. In *Proc. of SODA'04*, pages 121–130. SIAM, 2004.

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

[dA97] L. de Alfaro. *Formal Verification of Probabilistic Systems*. PhD thesis, Stanford University, december 1997.

[EM79] A. Ehrenfeucht and J. Mycielski. Positional strategies for mean-payoff games. *International Journal of Game Theory*, 8:109–113, 1979.

[Gim07] H. Gimbert. Pure stationary optimal strategies in Markov decision processes. In *STACS 2007, 24th Annual Symposium on Theoretical Aspects of Computer Science*, volume 4393 of *Lecture Notes in Computer Science*, pages 200–211. Springer, 2007.

[GTW02] E. Grädel, W. Thomas, and T. Wilke, editors. *Automata, Logics and Infinite Games*, volume 2500 of *LNCS*. Springer, 2002.

[Hor07] Florian Horn. Personal communication, December, 2007.

[Mar98] D.A. Martin. The determinacy of Blackwell games. *Journal of Symbolic Logic*, 63(4):1565–1581, 1998.

[MM02] A.K. McIver and C.C. Morgan. Games, probability and the quantitative $\mu$-calculus qm$\mu$. In *Proc. of LPAR'02*, pages 292–310. Springer, 2002.

[Ney04] A. Neyman. From Markov chains to stochastic games. In A. Neyman and S. Sorin, editors, *Stochastic Games and Applications*, volume 570 of *NATO Science Series C, Mathematical and Physical Sciences*, pages 9–25. Kluwer Academic Publishers, 2004.

[Zie04] Wieslaw Zielonka. Perfect-information Stochastic Parity Games. In *Proc. of FOSSACS'04*, pages 499–513. Springer, 2004.