Pure and Stationary Optimal Strategies in Perfect-Information Stochastic Games with Global Preferences

Hugo Gimbert, Wieslaw Zielonka

Optimal deterministic stationary strategies in perfect-information stochastic games with global preferences

Hugo Gimbert[*] Wiesław Zielonka[†]

November 25, 2016

Abstract

We examine the problem of the existence of optimal deterministic stationary strategies in two-players antagonistic (zero-sum) perfect information stochastic games with finitely many states and actions. We show that the existence of such strategies follows from the existence of optimal deterministic stationary strategies for some derived one-player games. Thus we reduce the problem from two-player to one-player games (Markov decision problems), where usually it is much easier to tackle. The reduction is very general, it holds not only for all possible payoff mappings but also in more a general situations where players' preferences are not expressed by payoffs.

1 Introduction

Given a perfect-information zero-sum stochastic game with a finite set of states and actions, the existence of deterministic and stationary optimal strategies is a useful property. The existence of such simple strategies has been well-studied for several examples of games and Markov decision processes.

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 deterministic and stationary optimal strategies.

Of course not in every game both players have optimal deterministic stationary 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 deterministic and stationary optimal strategies whatever is the arena they are playing in.

We provide a result which is very useful for establishing existence of deterministic and stationary optimal strategies: if for some fixed payoff function f, players have optimal deterministic stationary 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 optimal deterministic stationary strategies for one player games implies the existence of such

[*]CNRS, LaBRI, hugo.gimbert@cnrs.fr.

[†]Université Paris 7, LIAFA, zielonka@liafa.univ-paris-diderot.fr.

strategies for two-player games for each class of games satisfying certain closure properties. These closure properties are satisfied by the class of all games. We prove that this result holds also for some subclasses of perfect information games, for example deterministic arenas or arenas without cycles except self-loops.

The reduction is very general, it holds not only for all possible payoff mappings but also in more a general situations where players' preferences are not expressed by payoffs but rather by preference orders on the set of probability measures over plays.

2 Stochastic Perfect Information Games

Notation. $\mathbb{N} = \{1, 2, 3, \ldots\}$, $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$. For a set X, $|X|$ is the cadinality of X, $\mathcal{P}(X)$ is the power set of X. $\mathcal{M}(X, \mathcal{F}(X))$ will stand for the set of probability measures on a measurable space $(X, \mathcal{F}(X))$, where $\mathcal{F}(X)$ is a σ-algebra of subsets of X.

If X is finite or countably infinite then we always assume that $\mathcal{F}(X) = \mathcal{P}(X)$ and to simplify the notation we write $\mathcal{M}(X)$ rather than $\mathcal{M}(X, \mathcal{P}(X))$ to denote the set of all probability measures over X. Moreover for $\sigma \in \mathcal{M}(X)$, $\mathrm{supp}(\sigma) = \{x \in X \mid \sigma(x) > 0\}$ will denote the support of measure σ.

2.1 Arenas and games

We consider games that two players, Max and Min, play on an arena

$$\mathbb{A} = (\mathbf{S}, \mathbf{A}, \mathbf{T}, p, \mathbf{P}, \mathbf{R}, \rho)$$

consisting of the following ingredients:

- $\mathbf{S}$ and $\mathbf{A}$ are finite nonempty sets of, respectively, states and actions,
- $\mathbf{T} \subseteq \mathbf{S} \times \mathbf{A} \times \mathbf{S}$ is the set of transitions
- $p : \mathbf{T} \to (0, 1]$ is the transition probability function which assigns for each transition $(s, a, s') \in \mathbf{T}$ a positive probability $p(s, a, s')$ of transition from s to s' if a is executed at s. We extend p to all elements of $\mathbf{S} \times \mathbf{A} \times \mathbf{S}$ by setting $p(s, a, s') = 0$ if $(s, a, s') \notin \mathbf{T}$.
- $\mathbf{P} : \mathbf{S} \to \{\mathrm{Min}, \mathrm{Max}\}$ is a mapping associating with each state $s \in \mathbf{S}$ the player $\mathbf{P}(s)$ controling s.
- finally, $\mathbf{R}$ a set of rewards and $\rho : \mathbf{T} \to \mathbf{R}$ is a reward mapping assigning to each transition (s, a, s') a reward $\rho(s, a, s')$.

For each $s \in \mathbf{S}$, we define $\mathbf{A_T}(s) := \{a \in \mathbf{A} \mid \exists s' \in \mathbf{S}, (s, a, s') \in \mathbf{T}\}$ to be the set of actions *available* at s. We assume that all $\mathbf{A_T}(s)$ are nonempty and, for each action $a \in \mathbf{A_T}(s)$, $\sum_{s' \in \mathbf{S}} p(s, a, s') = 1$.

A infinite game is played by players Max and Min on $\mathbb{A}$, at each stage player $\mathbf{P}(s)$ controlling the current state s chooses an available action $a \in \mathbf{A_T}(s)$ which results in a transition to a state s' (which can be equal to s) with probability $p(s, a, s')$. Let us note that the set of rewards can be uncountable, for example $\mathbf{R}$ can be equal to the set $\mathbb{R}$ of real numbers, however since the set of transition is finite each arena contains only finitely many rewards.

In the case where all transition probabilities are equal either to 0 or to 1 the arena is said to be *deterministic*. In a deterministic arena, given a state s and an action $a \in \mathbf{A_T}(s)$ available at s there is a unique state t such that $(s, a, t) \in \mathbf{T}$.

2.2 Strategies

A *finite history* of length $|h| = n$ in the arena $\mathbb{A}$ is a finite sequence $h = s_1 a_1 s_2 \dots s_{n-1} a_{n-1} s_n$ alternating states and actions such that h starts and ends in a state, contains n states and for every $1 \leq i < n$ the triplet (s_i, a_i, s_{i+1}) is a transition i.e. $(s_i, a_i, s_{i+1}) \in \mathbf{T}$.

The set of finite histories of length n is denoted $\mathcal{H}^n(\mathbb{A})$ and $\mathcal{H}(\mathbb{A}) := \cup_{i=1}^{\infty} \mathcal{H}^n$ is the set of all finite histories. When $\mathbb{A}$ is clear from the context we simply write $\mathcal{H}$ instead of $\mathcal{H}(\mathbb{A})$.

Let $\mathcal{H}_{\text{Max}}$ be the subset of $\mathcal{H}$ consisting of finite histories with the last state controlled by player Max.

A strategy of player Max is a mapping $\sigma : \mathcal{H}_{\text{Max}} \to \mathcal{M}(\mathbf{A})$ which assigns to each $h \in \mathcal{H}_{\text{Max}}$ a probability measure over actions. We write $\sigma[h](a)$ for the probability that σ assigns to action a for $h \in \mathcal{H}_{\text{Max}}$. We assume that only actions available in the last state of h can be chosen, i.e. $\text{supp}(\sigma(h)) \subseteq \mathbf{A}_{\mathbf{T}}(s)$ where s is the last state of h.

Strategy σ is said to be *deterministic* if, for each $h \in \mathcal{H}_{\text{Max}}$ there is an action $a \in \mathbf{A}$ such that $\sigma[h](a) = 1$. We can identify the deterministic strategies of player Max with the mappings $\sigma : \mathcal{H}_{\text{Max}} \to \mathbf{A}$ such that for $h \in \mathcal{H}_{\text{Max}}$, $\sigma(h)$ is the action selected by σ if the current finite history is h.

A strategy σ is *stationary* if, for all $h \in \mathcal{H}_{\text{Max}}$, $\sigma[h](a) = \sigma[s](a)$, where s is the last state of h.

Deterministic stationary strategies of Max can be seen as mappings from the set $\mathbf{P}^{-1}(\text{Max})$ of states controlled by Max to the set of actions such that $\sigma[s] \in \mathbf{A}_{\mathbf{T}}(s)$ for $s \in \mathbf{P}^{-1}(\text{Max})$.

Strategies for player Min (deterministic, stationary or general) are defined *mutatis mutandis*.

$\Sigma(\mathbb{A})$ and $\mathcal{T}(\mathbb{A})$ will stand for the sets of strategies of Max and Min respectively, and we use σ and τ (with subscripts or superscripts if necessary) to denote the elements of $\Sigma(\mathbb{A})$ and $\mathcal{T}(\mathbb{A})$ respectively.

A strategy profile $(\sigma, \tau) \in \Sigma(\mathbb{A}) \times \mathcal{T}(\mathbb{A})$ (consisting of strategies for each player) defines a mapping $(\sigma \cup \tau) : \mathcal{H} \to \mathcal{M}(\mathbf{A})$,

$$(\sigma \cup \tau)[h](a) = \begin{cases} \sigma[h](a) & \text{if the last state of } h \text{ is controlled by player Max,} \\ \tau[h](a) & \text{if the last state of } h \text{ is controlled by player Min.} \end{cases}$$

The set of infinite histories $\mathcal{H}^\infty(\mathbb{A})$ consists of infinite sequences $h = s_1 a_1 s_2 a_2 \dots$ alternating states and actions such that $(s_i, a_i, s_{i+1}) \in \mathbf{T}$ for every $i \geq 1$. Again we write $\mathcal{H}^\infty$ instead of $\mathcal{H}^\infty(\mathbb{A})$ when $\mathbb{A}$ is clear from the context.

For a finite history $h = s_1 a_1 s_2 \dots s_n$ by h^+ we denote the cylinder generated by h which consists of all infinite histories having prefix h. An initial state $s \in \mathbf{S}$ and a strategy profile (σ, τ) determine a probability measure $\mathbb{P}^{\sigma,\tau}_{\mathbb{A},s} \in \mathcal{M}(\mathcal{H}^\infty, \mathcal{F}(\mathcal{H}^\infty))$, where $\mathcal{F}(\mathcal{H}^\infty)$ is the σ-algebra generated by the set of all cylinders. When it is clear from the context, we remove the arena from this notation and simply write $\mathbb{P}^{\sigma,\tau}_s$.

Given an initial state s_1 and strategies σ, τ of players Max and Min we define the probability $\mathbb{P}^{\sigma,\tau}_s$ of h^+:

$$\begin{aligned} \mathbb{P}^{\sigma,\tau}_s(h^+) = I_{s=s_1} \cdot (\sigma \cup \tau)[s_1](a_1) \cdot p(s_1, a_1, s_2) \cdot \\ (\sigma \cup \tau)[s_1 a_1 s_2](a_2) \cdot p(s_2, a_2, s_3) \cdot \dots \cdot \\ (\sigma \cup \tau)[s_1 a_1 s_2 \dots s_{n-1}](a_{n-1}) \cdot p(s_{n-1}, a_{n-1}, s_n) \end{aligned} \quad (1)$$

where $I_{s=s_1}$ is the indicator function equal to 1 is $s = s_1$ and 0 otherwise.

By the Ionescu Tulcea theorem [Nev71] there exists a unique probability measure $\mathbb{P}_s^{\sigma,\tau} \in \mathcal{M}(\mathcal{H}^\infty, \mathcal{F}(\mathcal{H}^\infty))$ satisfying (1). Moreover the support of this probability measure is the set $\mathcal{H}^\infty(\mathbb{A}, s, \sigma, \tau)$ of infinite histories whose every finite prefix has positive probability i.e.

$$\mathcal{H}^\infty(\mathbb{A}, s, \sigma, \tau) = \{s_1 a_1 s_2 \cdots \in \mathcal{H}^\infty(\mathbb{A}) \mid \forall n, (\sigma \cup \tau)[s_1 a_1 s_2 \cdots s_n](a_n) > 0\} \ .$$

2.3 Players preferences

We extend the reward mapping ρ to finite and infinite histories: for $h = s_1 a_1 s_2 a_2 s_3 a_3 s_4 \ldots \in \mathcal{H}^{\leq\infty}$, we set $\rho(h) = \rho(s_1, a_1, s_2)\rho(s_2, a_2, s_3)\rho(s_3, a_3, s_4)\ldots$.

We assume that the set $\mathbf{R}^\infty$ of infinite reward sequences is endowed with the product σ-algebra $\mathcal{F}(\mathbf{R}^\infty) := \otimes_{i=1}^\infty \mathcal{P}(\mathbf{R})$ Then the mapping ρ defined above is a measurable mapping from $(\mathcal{H}^\infty(\mathbb{A}), \mathcal{F}(\mathcal{H}^\infty(\mathbb{A}))$ to $(\mathbf{R}^\infty, \mathcal{F}(\mathbf{R}^\infty))$.

This implies that, for each probability measure $\mathbb{P} \in \mathcal{M}(\mathcal{H}^\infty, \mathcal{F}(\mathcal{H}^\infty))$ with support $\mathrm{supp}(\mathbb{P})$ included in $\mathcal{H}^\infty(\mathbb{A})$, the mapping ρ induces a probability measure $\rho\mathbb{P} \in \mathcal{M}(\mathbf{R}^\infty, \mathcal{F}(\mathbf{R}^\infty))$ such that, for $U \in \mathcal{F}(\mathbf{R}^\infty)$,

$$\rho\mathbb{P}(U) := \mathbb{P}(\rho^{-1}(U)) \ .$$

We assume that players are interested only in infinite sequence of rewards obtained during the play. This leads to the following definition.

Definition 1 (Outcomes and preference relations)**.** *Fix an arena $\mathbb{A}$. For a given strategy profile (σ, τ) and an initial state s the* outcome *of the game is the probability measure*

$$\mathbb{O}(\mathbb{A}, s, \sigma, \tau) = \rho\mathbb{P}_{\mathbb{A},s}^{\sigma,\tau} \in \mathcal{M}(\mathbf{R}^\infty, \mathcal{F}(\mathbf{R}^\infty))$$

Given two set of strategies $\Sigma \subseteq \Sigma(\mathbb{A})$ and $\mathcal{T} \subseteq \mathcal{T}(\mathbb{A})$ We set

$$\mathcal{O}_s(\mathbb{A}, \Sigma, \mathcal{T}) := \{\mathbb{O}(\mathbb{A}, s, \sigma, \tau) \mid (\sigma, \tau) \in \Sigma \times \mathcal{T}\}$$

to be the set of outcomes in $\mathbb{A}$ starting at s and using strategies from Σ and $\mathcal{T}$ and

$$\mathcal{O}(\mathbb{A}, \Sigma, \mathcal{T}) = \bigcup_{s \in \mathbf{S}} \mathcal{O}_s(\mathbb{A}, \Sigma, \mathcal{T})$$

the set of all possible outcomes using strategies from Σ and $\mathcal{T}$.

A preference relation in $\mathbb{A}$ with strategies in Σ and $\mathcal{T}$ *is a reflexive and transitive binary relation $\preceq$ over $\mathcal{O}(\mathbb{A}, \Sigma, \mathcal{T})$*

A preference relation $\preceq$ is *total* if for all outcomes $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{O}(\mathbb{A}, \Sigma, \mathcal{T})$, either $\mathbb{P}_1 \preceq \mathbb{P}_2$ or $\mathbb{P}_2 \preceq \mathbb{P}_1$. Usually naturally arising preference relations are total but this assumption is not necessary to formulate and prove our main result.

2.4 Games and optimal strategies

A *game* is a tuple $\Gamma = (\mathbb{A}, \Sigma, \mathcal{T}, \preceq)$ composed of an arena, some set of strategies for the players in $\mathbb{A}$ and a preference relation $\preceq$ in $\mathbb{A}$ with strategies in Σ and $\mathcal{T}$. The aim of Max is to maximize the obtained outcome with respect to $\preceq$. We consider only zero-sum games where the preference relation of player Min is the inverse of $\preceq$.

A strategy $\sigma \in \Sigma$ is a *best response* to a strategy $\tau \in \mathcal{T}$ in Γ if for each state s and each strategy $\sigma' \in \Sigma$ of Max,

$$\mathbb{O}(\mathbb{A}, s, \sigma', \tau) \preceq \mathbb{O}(\mathbb{A}, s, \sigma, \tau) \ .$$

Symmetrically, a best response of player Min to a strategy σ of Max is a strategy τ such that for all strategies $\tau' \in \mathcal{T}$ of Min,

$$\mathbb{O}(\mathbb{A}, s, \sigma, \tau) \preceq \mathbb{O}(\mathbb{A}, s, \sigma, \tau') \ .$$

A pair of strategies $\sigma^\sharp \in \Sigma, \tau^\sharp \in \mathcal{T}$ is *optimal* if $\sigma^\sharp$ is a best response to $\tau^\sharp$ and $\tau^\sharp$ a best response to $\sigma^\sharp$, i.e. if, for each state s and all strategies $\sigma \in \Sigma$, $\tau \in \mathcal{T}$,

$$\mathbb{O}(\mathbb{A}, s, \sigma, \tau^\sharp) \preceq \mathbb{O}(\mathbb{A}, s, \sigma^\sharp, \tau^\sharp) \preceq \mathbb{O}(\mathbb{A}, s, \sigma^\sharp, \tau) \ . \tag{2}$$

It is an elementary exercise to check that if $(\sigma_1^\sharp, \tau_1^\sharp)$ and $(\sigma_2^\sharp, \tau_2^\sharp)$ are pairs of optimal strategies then $(\sigma_1^\sharp, \tau_2^\sharp)$ and $(\sigma_2^\sharp, \tau_1^\sharp)$ also are. As a consequence we say that a single strategy $\sigma^\sharp \in \Sigma$ itself is optimal whenever it belongs to some pair of optimal strategies $(\sigma^\sharp, \tau^\sharp)$, and similarly for $\tau^\sharp \in \mathcal{T}$.

2.5 Deterministic games

For some applications, it is natural to require both arenas and strategies to be deterministic. In this case, it is enough to express the preferences of the players between infinite sequences of rewards because the probability measures defined by strategy profiles are Dirac measures over $\mathbf{R}^\infty$. For this purpose we define *deterministic preference relations* as follows.

Definition 2. *Fix a deterministic arena $\mathbb{A}$. For a given profile (σ, τ) of deterministic strategies and an initial state s there is a unique infinite history $h(s, \sigma, \tau) = s_1 a_1 s_2 a_2 \ldots \in \mathcal{H}^\infty$ such that $s = s_1$ and for every n,*

$$(\sigma \cup \tau)(s_1 a_1 \ldots a_{n-1} s_n) = a_n \ \textit{and}$$
$$(s_n, a_n, s_{n+1}) \in \mathbf{T} \ ,$$

and $\rho(h(s, \sigma, \tau)) \in \mathbf{R}^\infty$ is called the deterministic outcome of the game. A deterministic preference relation in $\mathbb{A}$ *is a reflexive and transitive binary relation $\preceq$ over the set of deterministic outcomes in $\mathbb{A}$.*

A *deterministic game* is a tuple $\Gamma = (\mathbb{A}, \Sigma, \mathcal{T}, \preceq)$ composed of a deterministic arena, the sets Σ and $\mathcal{T}$ of deterministic strategies in $\mathbb{A}$ and a deterministic preference relation $\preceq$ of player Max. In a deterministic game strategies $\sigma^\sharp, \tau^\sharp$ are optimal if for each state s and all deterministic strategies $\sigma \in \Sigma$, $\tau \in \mathcal{T}$,

$$\rho(h(s, \sigma, \tau^\sharp)) \preceq \rho(h(s, \sigma^\sharp, \tau^\sharp)) \preceq \rho(h(s, \sigma^\sharp, \tau)). \tag{3}$$

3 Examples

Specifying players' preferences by means of preference relations over measures allows us to cover a wide range of optimality criteria. We illustrate this flexibility with four examples.

In most application the preferences are rather defined by means of a payoff mapping.

A payoff mapping is a measurable mapping f from the set $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ of infinite reward sequences to the set $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ of real numbers equipped with the σ-algebra of Borel sets.

For each outcome $\rho\mathbb{P}_s^{\sigma,\tau}$ we write $\rho\mathbb{P}_s^{\sigma,\tau}(f)$ for the expectation of f,

$$\rho\mathbb{P}_s^{\sigma,\tau}(f) := \int_{\mathbf{R}^\infty} f d\mathbb{P}$$

where $\mathbb{P} := \rho\mathbb{P}_s^{\sigma,\tau}$. We assume that f is integrable for all outcome measures $\rho\mathbb{P}_s^{\sigma,\tau}$.

We say that a preference relation $\preceq$ is induced by a payoff mapping f if for any two outcomes $\mathbb{P}_1$ and $\mathbb{P}_2$, $\mathbb{P}_1 \preceq \mathbb{P}_2$ iff $\mathbb{P}_1(f) \leq \mathbb{P}_2(f)$.

Mean-payoff games and parity games are two well-known examples of games with preferences induced by payoff mappings.

Example 1 (Mean-payoff games)**.** A mean-payoff game is a game played on arenas equipped with a reward mapping $\rho : \mathbf{T} \to \mathbb{R}$ and the payoff of an infinite reward sequence $r_1 r_2 r_3 \dots$ is given by $\limsup \frac{\sum_{i=1}^n r_i}{n}$.

Example 2 (Parity games)**.** The class of games with many applications in computer science and logic is the class of parity games [GTW02]. These games are played on arenas endowed with a priority mapping $\beta : \mathbf{S} \to \mathbb{Z}_+$ and the payoff for an infinite history is either 1 or 0 depending on whether $\limsup_i \beta(s_i)$, the maximal priority visited infinitely often, is odd or even, where s_i is the state visited at stage i. Again the aim of players Max and Min is, respectively, to maximize/minimize the probability

$$\mathbb{P}_s^{\sigma,\tau}(\limsup_n \beta(s_n) \text{ is even}) \ .$$

The next example is a variant of parity games which is positional in deterministic arena but in general not in stochastic arenas.

Example 3 (Simple parity games)**.** A simple parity game is played in a parity game arena, the aim of players Max and Min is, respectively, to maximize/minimize the probability

$$\mathbb{P}_s^{\sigma,\tau}(\sup_n \beta(s_n) \text{ is even}) \ .$$

We continue with two examples where the preference relation is not induced by a payoff over infinite reward sequences. The first is a well-known variant of mean-payoff games of Example 1.

Example 4 (Mean-payoff games)**.** The arena is equipped with a real valued reward mapping $\rho : \mathbf{T} \to \mathbb{R}$ exactly like in Example 1. Let $f_n = \frac{\sum_{i=1}^n \rho(s_i, a_i, s_{i+1})}{n}$ be the mean-payoff over first n periods. For outcomes $\mathbb{P}_1$ and $\mathbb{P}_2$ we set $\mathbb{P}_1 \preceq \mathbb{P}_2$ if $\limsup_n \mathbb{P}_1(f_n) \leq \limsup_n \mathbb{P}_2(f_n)$.

The next example is variant of the overtaking optimality criterion:

Example 5 (Overtaking)**.** Let $\rho : \mathbf{T} \to \mathbb{R}$ be a real valued reward mapping. For outcomes $\mathbb{P}_1$ and $\mathbb{P}_2$ we set $\mathbb{P}_1 \preceq \mathbb{P}_2$ if there exists n such that for all $k \geq n$, $\mathbb{P}_1(\sum_{i=1}^k \rho(s_i, a_i, s_{i+1})) \leq \mathbb{P}_2(\sum_{i=1}^k \rho(s_i, a_i, s_{i+1}))$. This preference relation is not total.

In the first three examples we associated with each outcome a real number which allows us to order the outcomes. These real numbers can be used also to quantify how much one outcome is better than another one for a given player but the question "how much better" is irrelevant when we are interested in optimal strategies. Moreover, as Example 5 shows, for some preference relations it is difficult to define a corresponding payoff mapping.

4 From one-player to two-player games: the case of deterministic games and perfect-information stochastic games

Our main result relies on three notions: one player games, subarenas and coverings. Before stating our theorem in its full generality, we provide in this section a weaker form of Theorem 9 which does not make use of the notions of subarenas and coverings and is enough to cover most applications.

4.1 One-player arenas

We say that arena $\mathbb{A} = (\mathbf{S}, \mathbf{A}, \mathbf{T}, p, \mathbf{P}, \mathbf{R}, \rho)$ is a *one-player arena* controlled by Max if each state s controlled by player Min has only one available action. But for states with one available action it is essentially irrelevant which player controls them and we can as well assume that all states of such an arena $\mathbb{A}$ are controlled by Max. Therefore games on one-player arenas controlled by player Max are nothing else but Markov decision processes where the unique player Max wants to maximize (relative to $\preceq$) the resulting outcome.

One-player arenas controlled by player Min are defined in a symmetric way and they can be identified with Markov decision processes where the aim of the unique player Min is to minimize the outcome relative to $\preceq$.

A one-player game is a game on a one-player arena $\mathbb{A}$ controlled either by Max or by Min.

4.2 Specializations of the one-to-two theorem

Our main result reduces the problem of the existence of optimal deterministic stationary strategies in two-player games to the same problem for one-player games.

For the reader interested only in using our theorem for a particular class of games, it may be sufficient to make use of one of the two following weak forms of our result, which adresses specifically the cases of deterministic games and of perfect-information stochastic games.

The general statement and its proof may be found in the next section.

We start with a specialization of our theorem to the class of deterministic games.

Theorem 3 (One-to-two theorem for deterministic games). *Let $\mathbf{R}$ be a set of rewards and $\preceq$ a reflexive and transitive relation on $\mathbf{R}^\infty$. A deterministic game $G = (\mathbb{A}_G, \preceq_G)$ with rewards in $\mathbf{R}$ is said to be* compatible *with $\preceq$ if $\preceq_G$ is the restriction of $\preceq$ to the set of deterministic outcomes of G.*

Assume that optimal deterministic stationary strategies exist in every deterministic one-player games compatible with $\preceq$. Then optimal deterministic stationary strategies exist in every deterministic two-player games compatible with $\preceq$.

Remark that we do not require the transitive relation $\preceq$ to be total, in other words $\preceq$ is simply assumed to be a partial preorder. This degree of generality is natural because in this paper all arenas are assumed to be finite, hence any infinite history arising in a game goes through finitely many different transitions and generates a sequence of rewards that takes finitely many different values in $\mathbf{R}$. Thus it is useless to define $u \preceq v$ when either u or v takes infinitely many different letters from $\mathbf{R}$.

When applied to the class of all perfect-information two-player games, our theorem specializes as follows.

Theorem 4 (One-to-two theorem for games with perfect-information)**.** *Let* $\mathbf{R}$ *be a set of rewards and* $\preceq$ *a reflexive and transitive relation on the set* $\mathcal{M}(\mathcal{H}^\infty, \mathcal{F}(\mathcal{H}^\infty))$ *of probability measures on* $\mathcal{H}^\infty$ *equipped with the* σ*-algebra* $\mathcal{F}(\mathcal{H}^\infty)$ *generated by the set of all cylinders.*

A game $G = (\mathbb{A}_G, \preceq_G)$ *with rewards in* $\mathbf{R}$ *is said to be* compatible *with* $\preceq$ *if* $\preceq_G$ *is the restriction of* $\preceq$ *to the set of outcomes of* G*.*

Assume that optimal deterministic stationary strategies exist in every deterministic one-player games compatible with $\preceq$*. Then optimal deterministic stationary strategies exist in every deterministic two-player games compatible with* $\preceq$*.*

Again remark that $\preceq$ is not assumed to be total, and actually we only need to compare probability measures that arise as outcomes in an arena $\mathbb{A}$ with rewards $\mathbf{R}$ (i.e. elements of $\mathcal{O}(\mathbb{A})$). In particular, it is enough to define $\preceq$ for probability measures whose support is included in the set of histories that take finitely many different values in $\mathbf{R}$.

Although the formulations of the two specializations of our theorem are quite close, they apply to different situations. The first one adresses deterministic games, where players are restricted to deterministic strategies and the preference relations are defined only on infinite sequences of rewards, while the second one adresses the full class of stochastic games where players use behavioural strategies and their preferences are expressed with respect to probability measures on infinite reward sequences. As a matter of fact there is no way, at least to our knowledge, to deduce the second theorem from the first one nor the contrary.

4.3 Examples revisited

Bierth [Bie87] shows that one-player games of Example 1 have optimal deterministic stationary strategies. A more readable proof of this fact is given in [Ney04].

One-player parity games of Example 2 have also optimal deterministic stationary strategies [dA97], thus the same holds for two-player parity games (the latter fact was proved by several authors but we can see now that this is a consequence of the result for one-player games).

Unlike parity games, one-player *simple* parity games are *not* positional, however they become positional when played on deterministic arenas. This is discussed in the next subsection.

One-player games of Example 4 have optimal deterministic stationary strategies [Der62]. Theorem 4 allows us to deduce that the same holds for two-player games with perfect information.

One-player games of Example 5 do not have optimal deterministic stationary strategies. For example if the game is made of two simple cycles of length 4 on the initial state, labelled with $0, 1, 1, 0$ on one hand and $1, 0, 0, 1$ on the other hand then neither cycle dominates the other.

4.4 Application to deterministic games

A deterministic game is a game played on a deterministic arena with the additional constraint that both players can use only deterministic strategies.

Even if deterministic games are usually much simpler to analyze than their stochastic counterparts, this does not mean that they are always easy. They are also interesting by themselves for at least two reasons. First of all deterministic games prevail in computer science applications related to automata theory, logic and verification [GTW02]. The second reason is that, concerning our main problem – the existence of optimal deterministic stationary strategies, deterministic games can differ from their stochastic counterparts.

It turns out that for the payoff mapping f associated to simple parity games, all deterministic games with payoff f have optimal deterministic stationary strategies but the same is not true for simple stochastic parity games. A suitable example was shown to us by Florian Horn, see Figure 1.

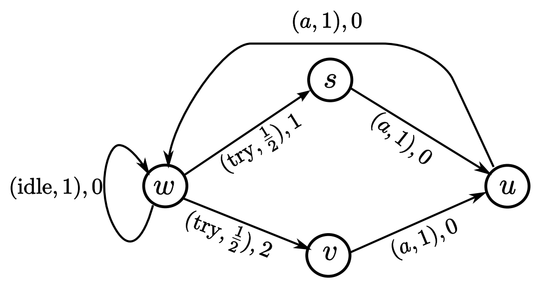

Figure 1: A simple parity game with all states controlled by player Max. Each transition t is labelled by a triple $((b,p),r)$ where b is an action, r the reward and p is the transition probability of t. There is only one randomized action **try**, its execution yields with probability 1/2 either the reward 1 or 2.

In the game in Figure 1 only the state w has two available actions **try** and **idle**. The payoff of Max is determined in the following way. Let $r = r_0r_1r_2\cdots$ be the infinite sequence of rewards, player Max obtains payoff 1 if this sequence contains at least one occurence of reward 1 and no occurence of reward 2, otherwise Max gets payoff 0. Clearly the optimal strategy for player Max is to execute the action try exactly once. If the transition with reward 1 is taken then returning to w he should always execute **idle**, if the transition with reward 2 is taken then he can do anything, his payoff will be 0. Thus his maximal expected payoff is 1/2. This strategy is not stationary since player Max should remember if it is the first visit to w or not.

On the other hand it is easy to see that all stationary strategies (randomized or not) yield the expected payoff 0.

4.5 A remark on one-player stochastic games with sub-mixing payoff functions

In light of Theorem 4 it is important to be able to decide if one-player games have optimal deterministic stationary strategies. A convenient way to tackle this problem was discovered by the first author [Gim07].

Let $f : \mathbf{A}^\infty \to \mathbb{R}$ be a bounded (either from below or from above) and measurable.

We say that f is prefix-independent f does not depend on the finite initial segment, i.e. $f(a_1a_2a_3\ldots) = f(a_2a_3\ldots)$ for each infinite sequence $a_1a_2a_3\ldots \in \mathbf{A}^\infty$. And f is said to be sub-mixing if

$$f(a_1a_2a_3\cdots) \le \max\{f(a_{i_1}a_{i_2}\cdots), f(a_{j_1}a_{j_2}\cdots)\},$$

for each infinite sequence $a_1a_2a_3\ldots \in \mathbf{A}^\infty$ of actions and each partition of $\mathbb{N}$ onto two infinite sets $I = \{i_1 < i_2 < i_3 < \ldots\}$, $J = \{j_1 < j_2 < j_3 < \ldots\}$.

In [Gim07] it is proved that if f is prefix independent and sub-mixing then for each one-player game $\Gamma = (\mathbb{A}, \preceq)$ with the preference relation induced by f and such that $\mathbb{A}$ is controlled by player Max, this player has an optimal deterministic stationary strategy.

In order to prove that one-player games controlled by player Min have optimal deterministic stationary strategies it suffices to verify that $-f$ is sub-mixing. Thus verifying that f and $-f$ are prefix independent and sub-mixing can be used to prove that two-player games with the preference relation induced by f have

Let us note that the payoff f of the parity game of Example 2 is prefix independent and sub-mixing and the same holds for $-f$.

For the mean-payoff games of Example 1 the payoff f is sub-mixing, but $-f$ is not (replacing lim sup by lim inf we obtain the payoff mapping which is not sub-mixing). However, for if players use deterministic stationary strategies then we a get a finite state Markov chain and for the resulting outcome $\mathbb{P}$ we have $\limsup \frac{\sum_{i=1}^{n} \rho(s_i,a_i,s_{i+1})}{n} = \lim \frac{\sum_{i=1}^{n} \rho(s_i,a_i,s_{i+1})}{n}$ almost surely, where $s_i, a_i, s_{i+1})$ is the transition at stage i. This can be used to prove that one-player games controlled by player Min have optimal deterministic stationary strategies we can apply Theorem 4 once again.

5 From one- to two-player games: the general theorem

Our main result relies on the notions of subarenas and splits of an arena on a state.

5.1 Subarenas and subgames

An arena $\mathbb{A}' = (\mathbf{S}, \mathbf{A}, \mathbf{T}', p', \mathbf{P}, \mathbf{R}, \rho')$ is a subarena of another arena
$\mathbb{A} = (\mathbf{S}, \mathbf{A}, \mathbf{T}, p, \mathbf{P}, \mathbf{R}, \rho)$ if $\mathbf{T}' \subseteq \mathbf{T}$ and p' and ρ' are the restrictions of p and ρ to $\mathbf{T}'$.

Clearly the set $\mathbf{T}'$ of transition of subarena $\mathbb{A}'$ satisfies the following conditions:

(a) if $(s, a, s') \in \mathbf{T}'$ and $(s, a, s'') \in \mathbf{T}$ then $(s, a, s'') \in \mathbf{T}'$, i.e. for each state s and each action a either we keep all transitions $(s, a, \cdot)$ in $\mathbb{A}'$ or we remove them all,

(b) for each $s \in \mathbf{S}$ there exist $a \in \mathbf{A}$ and $s' \in \mathbf{S}$ such that $(s, a, s') \in \mathbf{T}'$.

By definition of an arena, both conditions (a) and (b) are necessary for $\mathbb{A}'$ to be an arena: (a) is necessary for the sum of transition probabilities induced by a pair of states and actions to be 1 and (b) because in each state there should be at least one available action. They are also sufficient: if $\mathbf{T}' \subset \mathbf{T}$ satisfies (a) and (b) then there exists a unique subarena $\mathbb{A}'$ of $\mathbb{A}$ having $\mathbf{T}'$ as the set of transitions.

If $\mathbb{A}'$ is a subarena of $\mathbb{A}$ then all finite histories consistent with $\mathbb{A}'$ are consistent with $\mathbb{A}$.

We say that a strategy σ in $\mathbb{A}$ is compatible with $\mathbb{A}'$ if its restriction to $\mathcal{H}(\mathbb{A}')$ is a strategy in $\mathbb{A}'$. By definition of a strategy, this requires that σ restricted to

$\mathcal{H}(\mathbb{A}')$ only puts positive probability on actions available in $\mathbb{A}'$, i.e. formally for every $s_1a_2s_2\cdots s_n \in \mathcal{H}(\mathbb{A}')$ and $a \in \mathbf{A}$,

$$\sigma(s_1a_2s_2\cdots s_n)(a) > 0 \implies \exists t \in \mathbf{S}, (s_n, a, t) \in \mathbf{T}' \enspace .$$

This condition is also sufficient for σ to be compatible with $\mathbb{A}'$.

Conversely, every strategy σ' in $\mathbb{A}'$ defined on $\mathcal{H}(\mathbb{A}')$ can be extended in an arbitrary way to a strategy in $\mathbb{A}$ defined on $\mathcal{H}(\mathbb{A})$. This implies the inclusion $\mathcal{O}(\mathbb{A}') \subseteq \mathcal{O}(\mathbb{A})$ of the sets of outcomes. Therefore a preference $\preceq$ relation defined on outcomes of $\mathbb{A}$ is also a preference on outcomes of $\mathbb{A}'$.

Definition 5 (Subgame induced by a subarena)**.** *Let $\mathbb{A}'$ be a subarena of $\mathbb{A}$ and $G = (\mathbb{A}, \Sigma, \mathcal{T}, \preceq)$ a game played in $\mathbb{A}$. The subgame of G induced by $\mathbb{A}'$ is the game $G' = (\mathbb{A}', \Sigma', \mathcal{T}', \preceq')$ where:*

i) Σ' is the set of strategies obtained by restricting to $\mathcal{H}(\mathbb{A}')$ the strategies in Σ compatible with $\mathbb{A}'$,

ii) $\mathcal{T}'$ is the set of strategies obtained by restricting to $\mathcal{H}(\mathbb{A}')$ the strategies in $\mathcal{T}$ compatible with $\mathbb{A}'$,

iii) $\preceq'$ is the restriction of $\preceq$ to $\mathcal{O}(\mathbb{A}', \Sigma', \mathcal{T}')$.

5.2 Split of an arena

Intuitively, the split of an arena $\mathbb{A} = (\mathbf{S}, \mathbf{A}, \mathbf{T}, p, \mathbf{P}, \mathbf{R}, \rho)$ on a state $\omega \in \mathbf{S}$ consists in adding in the states of the arena an extra information: the last action chosen in the state ω by the player controlling ω. This memory is reset each time ω is reached. The state ω is called the *separation state*. An example is provided on Fig. 2.

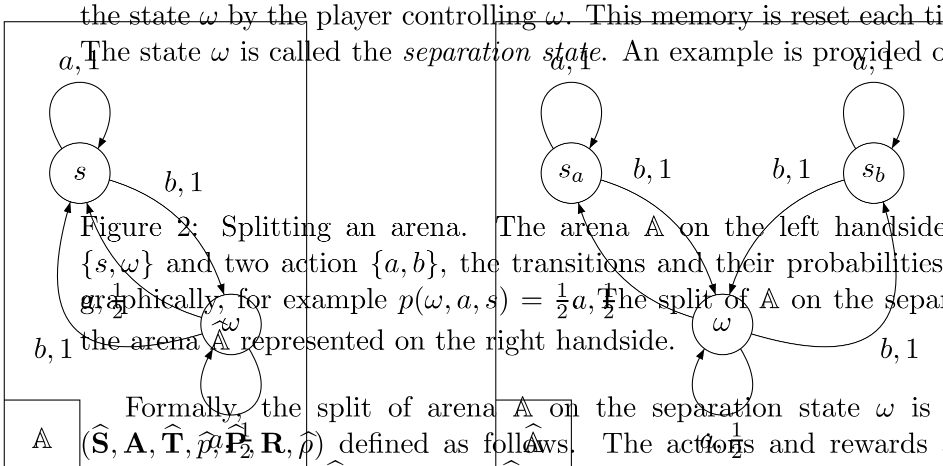

Figure 2: Splitting an arena. The arena $\mathbb{A}$ on the left handside has two states $\{s, \omega\}$ and two action $\{a, b\}$, the transitions and their probabilities are represented graphically, for example $p(\omega, a, s) = \frac{1}{2}a$. The split of $\mathbb{A}$ on the separation state ω is the arena $\widehat{\mathbb{A}}$ represented on the right handside.

Formally, the split of arena $\mathbb{A}$ on the separation state ω is the arena $\widehat{\mathbb{A}} = (\widehat{\mathbf{S}}, \mathbf{A}, \widehat{\mathbf{T}}, \widehat{p}, \widehat{\mathbf{P}}, \mathbf{R}, \widehat{\rho})$ defined as follows. The actions and rewards are the same in both arenas $\mathbb{A}$ and $\widehat{\mathbb{A}}$. A state of $\widehat{\mathbb{A}}$ is either the separation state ω or a pair made of another state and an action available in the separation state, thus

$$\widehat{\mathbf{S}} = \{\omega\} \cup (\mathbf{S} \setminus \{\omega\}) \times \mathbf{A}_{\mathbf{T}}(\omega) \enspace ,$$

where $\mathbf{A}_{\mathbf{T}}(\omega)$ denotes the set of actions available in state ω. For every state $s \in \mathbf{S}\setminus\{\omega\}$ and action $x \in \mathbf{A}_{\mathbf{T}}(\omega)$ we denote $s_x = (s, x)$. To make the notation uniform we set $\omega_x = \omega$ i.e. ω_x is just an alias name for ω. There is a natural mapping

$$\pi : \widehat{\mathbf{S}} \to \mathbf{S}$$

which sends s_x to s. The rewards and the controlling player are the same in states s and s_x: we set $\widehat{\rho} = \rho \circ \pi$ and $\widehat{\mathbf{P}} = \mathbf{P} \circ \pi$. The transitions of $\widehat{\mathbb{A}}$ and their probabilities are inherited from $\mathbb{A}$ so that the second component of a state in $\widehat{\mathbb{A}}$ keeps track of the last action chosen in the separation state:

$$\widehat{\mathbf{T}} = \{(\omega, x, t_x) \mid (\omega, x, t) \in \mathbf{T}\} \cup \{(s_x, a, t_x) \mid (s, a, t) \in \mathbf{T} \wedge s \neq \omega\} \ ,$$
$$\text{for every } (s_x, a, t_y) \in \widehat{\mathbf{T}} \text{ we set } \widehat{p}(s_x, a, t_y) = p(s, a, t) \ .$$

The separation state ω separates the states of the split $\widehat{\mathbb{A}}$ in the following sense. For every action $x \in \mathbf{A_T}(\omega)$ we denote

$$\widehat{\mathbf{S}}_x = (\mathbf{S} \setminus \{\omega\}) \times \{x\} \ .$$

Then $\widehat{\mathbf{S}}$ is partitioned in $\{\omega\}$ on one hand and the sets $(\widehat{\mathbf{S}}_x)_{x \in \mathbf{A_T}(\omega)}$ on the other hand. By construction of the split:

Proposition 6 (Separation property)**.** *Every finite history of $\widehat{\mathbb{A}}$ starting in $\widehat{\mathbf{S}}_x$ and ending in $\widehat{\mathbf{S}}_y$ with $x \neq y$ passes through ω.*

5.3 Projecting histories and lifting strategies

Projecting histories $\quad \mathcal{H}(\widehat{\mathbb{A}}) \xrightarrow{\pi} \mathcal{H}(\mathbb{A})$, $\quad s_1 a_1 s_2 \cdots s_n \mapsto \pi(s_1) a_1 \pi(s_2) \cdots \pi(s_n)$

Lifting strategies $\quad \Sigma(\widehat{\mathbb{A}}) \xleftarrow{\Pi} \Sigma(\mathbb{A})$, $\quad \sigma \circ \pi \leftarrow \sigma$

Fix $x \in \mathbf{A_T}(\omega)$.

Lifting histories $\quad \mathcal{H}(\mathbb{A}) \xrightarrow{\phi_x} \mathcal{H}(\widehat{\mathbb{A}})$

$\phi_x(s) = s_x$

$\pi \circ \phi_x$ is the identity.

Projecting strategies $\quad \Sigma(\widehat{\mathbb{A}}) \xrightarrow{\Phi_x} \Sigma(\mathbb{A})$, $\quad \sigma \mapsto \sigma \circ \phi_x$

Figure 3: Projections and liftings of histories and strategies and their characteristic properties. Strategies of Min in $\mathcal{T}(\mathbb{A})$ and $\mathcal{T}(\widehat{\mathbb{A}})$ are projected and lifted in a symmetric way.

Let $\widehat{\mathbb{A}}$ be the split of an arena $\mathbb{A}$ on some separation state ω.

There is a natural projection π of finite histories in the split $\widehat{\mathbb{A}}$ to finite histories in the original arena $\widehat{\mathbb{A}}$ and conversely a natural lifting Π of strategies in the original arena to strategies in the split. These mappings are represented on Fig. 3.

The projection $\pi : \mathcal{H}(\widehat{\mathbb{A}}) \to \mathcal{H}(\mathbb{A})$ is the natural extension to histories of the mapping $\pi : \widehat{\mathbf{S}} \to \mathbf{S}$ which maps ω to ω and forgets the second component of other states ($\forall s_x \in \widehat{\mathbf{S}}, \pi(s_x) = s$). For every finite history $h = s_1 a_1 s_2 \cdots s_n \in \mathcal{H}(\widehat{\mathbb{A}})$ in $\widehat{\mathbb{A}}$, we set $\pi(h) = \pi(s_1) a_1 \pi(s_2) \cdots \pi(s_n)$. That $\pi(h)$ is a finite history in $\mathcal{H}(\mathbb{A})$ is immediate from the definition of $\widehat{\mathbf{T}}$.

The mapping $\Pi : \Sigma(\mathbb{A}) \to \Sigma(\widehat{\mathbb{A}})$ transforms a strategy $\sigma \in \Sigma(\mathbb{A})$ to the strategy $\Pi(\sigma) = \sigma \circ \pi$ called the *lifting* of σ from $\mathbb{A}$ to $\widehat{\mathbb{A}}$. Obviously $\sigma \circ \pi$ is a strategy in $\widehat{\mathbb{A}}$ because for every state $s_x \in \widehat{\mathbf{S}}$, the same actions are available in state s_x of the arena $\widehat{\mathbb{A}}$ and in the state s of the arena $\mathbb{A}$.

5.4 Lifting histories and projecting strategies

There is a canonical way to lift histories from the original arena to the splitted arena, provided the initial state is fixed.

Proposition 7. *Fix an action $x \in \mathbf{A_T}(\omega)$. For every state s of $\mathbb{A}$ and history $h \in \mathcal{H}(\mathbb{A}, s)$ starting in s there exists a unique history $\phi_x(h) \in \mathcal{H}(\widehat{\mathbb{A}}, s_x)$ such that $\pi(\phi_x(h)) = h$. F every state $s \in \mathbf{S}$, $\phi_x(s) = s_x$ and in particular $\phi_x(\omega) = \omega$. Moreover, for every for every state $s \in \mathbf{S}$ and strategies $\widehat{\sigma} \in \Sigma(\widehat{\mathbb{A}})$ and $\widehat{\tau} \in \mathcal{T}(\widehat{\mathbb{A}})$,*

$$\pi \circ \phi_x \textit{ is the identity on } \mathcal{H}(\mathbb{A}) \tag{4}$$
$$\phi_x \circ \pi \textit{ is the identity on } \mathcal{H}(\mathbb{A}, s_x, \widehat{\sigma}, \widehat{\tau}) \ . \tag{5}$$

Proof. The mapping ϕ_x is defined inductively. Initially, ϕ_x maps s to s_x. Assume $\phi_x(h)$ is uniquely defined for some $h \in \mathcal{H}(\mathbb{A})$ and let $a \in \mathbf{A}$ and $s \in \mathbf{S}$ such that $has \in \mathcal{H}(\mathbb{A})$. Let t be the last state of h. If $t = \omega$ then there is a unique $y \in \mathbf{A_T}(\omega)$ such that (ω, a, s_y) is a transition in $\widehat{\mathbb{A}}$: necessarily $y = a$. If $t \neq \omega$ then by induction, according to property i) there exists $b \in \mathbf{A_T}(\omega)$ such that t_b is the last state of $\phi_x(h)$. There is a unique $y \in \mathbf{A_T}(\omega)$ such that (t_b, a, s_y) is a transition in $\widehat{\mathbb{A}}$: necessarily $y = b$. Equations (4) holds by definition of ϕ_x. Equation (5) follows from an easy induction. □

Beware that the lifting of a finite history $s_1 a_1 s_2 \cdots s_n$ in $\mathbb{A}$ is in general different from $\phi_x(s_1) a_1 \phi_x(s_2) \cdots \phi_x(s_n)$ which actually may not even be a finite history in $\widehat{\mathbb{A}}$. For example, in the arena of Fig. 2, the lifting of $h = s, a, \omega, b, s, b, s, b, s$ with respect to a is $\phi_a(h) = s_a, a, \omega, b, s_b, b, s_b, b, s_b$ and not $s_a, a, \omega, b, s_a, b, s_a, b, s_a$.

The lifting of histories from $\mathbb{A}$ to $\widehat{\mathbb{A}}$ can be used to project strategies from $\widehat{\mathbb{A}}$ to $\mathbb{A}$. For this purpose for every action $x \in \mathbf{A_T}(\omega)$ we define $\Phi_x : \Sigma(\widehat{\mathbb{A}}) \to \Sigma(\mathbb{A})$ as

$$\Phi_x(\sigma) = \sigma \circ \phi_x \ .$$

$\Phi_x(\sigma)$ is a strategy in the arena $\mathbb{A}$ because for every state $s_x \in \widehat{\mathbf{S}}$, the same actions are available in state s_x of the arena $\widehat{\mathbb{A}}$ and in the state s of the arena $\mathbb{A}$.

6 The general one-to-two theorem

The general formulation of our theorem concerns classes of games that have three properties: every subgame or split of a game of the class should also be in the class. Moreover in every game in the class the set of strategies available for the players should contain all the deterministic stationary strategies.

Definition 8. *Let $\mathcal{G} = (\mathbb{A}_i, \Sigma_i, \mathcal{T}_i, \preceq_i)_{i \in I}$ be a collection of games. We say that $\mathcal{G}$:*

- *is* subgame-closed *if every subgame of a game in $\mathcal{G}$ is also in $\mathcal{G}$.*
- *is* split-closed *if for every game $G = (\mathbb{A}, \Sigma, \mathcal{T}, \preceq)$ in $\mathcal{G}$, every state ω of $\mathbb{A}$ there is a game $\widehat{G} = (\widehat{\mathbb{A}}, \widehat{\Sigma}, \widehat{\mathcal{T}}, \preceq)$ in $\mathcal{G}$, such that $\widehat{\mathbb{A}}$ is the split of $\mathbb{A}$ on ω and for every action x available in ω,*

$$\begin{aligned} &\Pi(\Sigma) \subseteq \widehat{\Sigma} \text{ and } \Pi(\mathcal{T}) \subseteq \widehat{\mathcal{T}} \ , \\ &\Phi_x(\widehat{\Sigma}) \subseteq \Sigma \text{ and } \Phi_x(\widehat{\mathcal{T}}) \subseteq \mathcal{T} \ . \end{aligned} \tag{6}$$

- contains all deterministic stationary strategies *if in every game $G = (\mathbb{A}, \Sigma, \mathcal{T}, \preceq$ $)$ in $\mathcal{G}$, every deterministic stationary strategy in the arena $\mathbb{A}$ is contained in $\Sigma \cup \mathcal{T}$.*

The following theorem reduces the problem of the existence of optimal deterministic stationary strategies from two-player games to one-player games.

Theorem 9 (The one-to-two theorem)**.** *Let $\mathcal{G}$ be a subgame-closed and split-closed collection of games, which contains all deterministic stationary strategies. If optimal deterministic stationary strategies exist in every one-player game of $\mathcal{G}$ then optimal deterministic stationary strategies exist in every two-player game of $\mathcal{G}$.*

In Section 4.2 are presented two specializations of Theorem 9 to deterministic games (Theorem 3) and games with perfect-information (Theorem 4). In both cases a transitive relation $\preceq$ is fixed, on either $\mathbf{R}^\infty$ for deterministic games or $\mathcal{M}(\mathcal{H}^\infty, \mathcal{F}(\mathcal{H}^\infty))$ for games with perfect-information. Both results follow from an application of Theorem 9 to the class of all games compatible with $\preceq$ (all deterministic games for Theorem 3 and all games with perfect-information for Theorem 4).

7 From one-player games to two-player games – the proof

This section is devoted to the proof of Theorem 9. We start with some trivial though crucial properties of projections and liftings.

Proposition 10 (Properties of liftings and projections)**.** *Let $G = (\mathbb{A}, \Sigma, \mathcal{T}, \preceq)$ and $\widehat{G} = (\widehat{\mathbb{A}}, \widehat{\Sigma}, \widehat{\mathcal{T}}, \preceq)$ be two games on arenas $\mathbb{A}$ and $\widehat{\mathbb{A}}$ such that $\widehat{\mathbb{A}}$ is the split of $\mathbb{A}$ on some separating state ω. Let $x \in \mathbf{A_T}(\omega)$ an action available in ω, let $\sigma \in \Sigma$ and $\tau \in \mathcal{T}$ some strategies in G and $\widehat{\sigma} \in \widehat{\Sigma}$ and $\widehat{\tau} \in \widehat{\mathcal{T}}$ some strategies in $\widehat{G}$.*

Lifting and projection preserve outcomes: for every state $s \in \mathbf{S}$,

$$\mathbb{O}(\mathbb{A}, s, \sigma, \tau) = \mathbb{O}(\widehat{\mathbb{A}}, s_x, \Pi(\sigma), \Pi(\tau))$$
$$\mathbb{O}(\widehat{\mathbb{A}}, s_x, \widehat{\sigma}, \widehat{\tau}) = \mathbb{O}(\mathbb{A}, s, \Phi_x(\widehat{\sigma}), \Phi_x(\widehat{\tau})) \ .$$

Assume moreover that condition (6) of Definition 8 holds, i.e.

$$\Pi(\Sigma) \subseteq \widehat{\Sigma} \text{ and } \Pi(\mathcal{T}) \subseteq \widehat{\mathcal{T}} \ ,$$
$$\Phi_x(\widehat{\Sigma}) \subseteq \Sigma \text{ and } \Phi_x(\widehat{\mathcal{T}}) \subseteq \mathcal{T} \ .$$

If $\widehat{\sigma}$ and $\widehat{\tau}$ are optimal strategies in $\widehat{G}$ from state s_x then $\Phi_x(\widehat{\sigma})$ and $\Phi_x(\widehat{\tau})$ are optimal in G from state s. If $\widehat{\sigma}$ is deterministic and stationary and $x = \widehat{\sigma}(\omega)$ then also $\Phi_x(\widehat{\sigma})$ is deterministic and stationary.

Proof. We prove the first part of the proposition. Define

$$\pi_\infty : \mathcal{H}^\infty(\widehat{\mathbb{A}}, s_x, \widehat{\sigma}, \widehat{\tau}) \to \mathcal{H}^\infty(\mathbb{A}, s, \Phi_x(\widehat{\sigma}), \Phi_x(\widehat{\tau}))$$

as $\pi^\infty(s_1 a_1 s_2 \cdots) = \pi(s_1) a_1 \pi(s_2) \cdots$. Equivalently, $\pi^\infty(h) = \lim_n \pi(h_n)$ where h_n is the prefix of h of length $2n+1$. First we show

$$\mathbb{P}_{\mathbb{A},s}^{\Phi(\widehat{\sigma}),\Phi(\widehat{\tau})} = \pi_\infty \mathbb{P}_{\widehat{\mathbb{A}},s_x}^{\widehat{\sigma},\widehat{\tau}} \tag{7}$$

Let $h \in \mathcal{H}(\mathbb{A}, s, \Phi_x(\widehat{\sigma}), \Phi_x(\widehat{\tau}))$, remember that $h^+ \subseteq \mathcal{H}^\infty(\mathbb{A}, s)$ denotes the cylinder of infinite histories starting with h. According to Proposition 7,

$$\pi_\infty^{-1}(h^+) = (\phi_x(h))^+ \ .$$

By definition of $\mathbb{P}_{\mathbb{A},s}^{\Phi(\widehat{\sigma}),\Phi(\widehat{\tau})}$ and since $\pi \circ \phi_x$ is the identity on $\mathcal{H}(\mathbb{A})$, an easy induction shows that $\mathbb{P}_{\mathbb{A},s}^{\Phi(\widehat{\sigma}),\Phi(\widehat{\tau})}(h^+) = \mathbb{P}_{\widehat{\mathbb{A}},s_x}^{\widehat{\sigma},\widehat{\tau}}((\phi_x(h))^+)$. Finally $\mathbb{P}_{\widehat{\mathbb{A}},s_x}^{\widehat{\sigma},\widehat{\tau}}(\pi_\infty^{-1}(h^+)) = \mathbb{P}_{\mathbb{A},s}^{\Phi(\widehat{\sigma}),\Phi(\widehat{\tau})}(h^+)$. Since this holds for every cylinder h^+ with $h \in \mathcal{H}(\mathbb{A}, s, \Phi_x(\widehat{\sigma}), \Phi_x(\widehat{\tau}))$ we get (7).

The projection π_∞ does not change the sequence of rewards thus $\rho \circ \pi_\infty = \rho$ and

$$\mathbb{O}(\mathbb{A}, s, \Phi_x(\widehat{\sigma}), \Phi_x(\widehat{\tau})) = \rho \mathbb{P}_{\mathbb{A},s}^{\Phi_x(\widehat{\sigma}),\Phi_x(\widehat{\tau})} = \rho \pi_\infty \mathbb{P}_{\widehat{\mathbb{A}},s_x}^{\widehat{\sigma},\widehat{\tau}} = \rho \mathbb{P}_{\widehat{\mathbb{A}},s_x}^{\widehat{\sigma},\widehat{\tau}} = \mathbb{O}(\widehat{\mathbb{A}}, s_x, \widehat{\sigma}, \widehat{\tau})$$

which proves $\mathbb{O}(\widehat{\mathbb{A}}, s_x, \widehat{\sigma}, \widehat{\tau}) = \mathbb{O}(\mathbb{A}, s, \Phi_x(\widehat{\sigma}), \Phi_x(\widehat{\tau}))$. We can apply this equality to $\widehat{\sigma} = \Pi(\sigma)$ and $\widehat{\tau} = \Pi(\tau)$ and since $\Phi_x \circ \Pi$ is the identity (Proposition 7) this implies $\mathbb{O}(\widehat{\mathbb{A}}, s_x, \Pi(\sigma), \Pi(\tau)) = \mathbb{O}(\mathbb{A}, s, \sigma, \tau)$ which terminates the proof of the first part of the proposition.

We prove the second part of the proposition. Assume $\widehat{\sigma}$ and $\widehat{\tau}$ are optimal in $\widehat{\mathbb{A}}$ from state s_x. Let $\sigma \in \Sigma$ and $\tau \in \mathcal{T}$ be any strategies in G. We prove that

$$\mathbb{O}(\mathbb{A}, s, \sigma, \Phi_x(\widehat{\tau})) \preceq \mathbb{O}(\mathbb{A}, s, \Phi_x(\widehat{\sigma}), \Phi_x(\widehat{\tau})) \preceq \mathbb{O}(\mathbb{A}, s, \Phi_x(\widehat{\sigma}), \tau) \ . \tag{8}$$

By symmetry, it is enough to prove the left inequality:

$$\begin{aligned}
\mathbb{O}(\mathbb{A}, s, \sigma, \Phi_x(\widehat{\tau})) &= \mathbb{O}(\mathbb{A}, s, \Phi_x(\Pi(\sigma)), \Phi_x(\widehat{\tau})) \\
&= \mathbb{O}(\widehat{\mathbb{A}}, s_x, \Pi(\sigma), \widehat{\tau})) \\
&\preceq \mathbb{O}(\widehat{\mathbb{A}}, s_x, \widehat{\sigma}, \widehat{\tau})) \\
&= \mathbb{O}(\mathbb{A}, s, \Phi_x(\widehat{\sigma}), \Phi_x(\widehat{\tau})) \ .
\end{aligned}$$

The first equality holds because $\pi \circ \phi_x$ is the identity on $\mathcal{H}(\mathbb{A})$ thus $\Phi_x \circ \Pi$ is the identity on Σ. The second equality was established in the first part of the proposition. The inequality holds because $(\widehat{\sigma}, \widehat{\tau})$ are optimal in $\widehat{\mathbb{A}}$ and $\Pi(\sigma)$ is a strategy in $\widehat{G}$ since $\Pi(\Sigma) \subseteq \widehat{\Sigma}$ by hypothesis. The last equality was established in the first part.

Since by hypothesis $\Phi_x(\widehat{\Sigma}) \subseteq \Sigma$ and $\Phi_x(\widehat{\mathcal{T}}) \subseteq \mathcal{T}$ then $\Phi_x(\widehat{\sigma})$ and $\Phi_x(\widehat{\tau}))$ are strategies in $\widehat{G}$ and according to (8) they are optimal in $\widehat{G}$.

Assume now that $\widehat{\sigma}$ is deterministic and stationary and $x = \widehat{\sigma}(\omega)$. On $\mathcal{H}(\mathbb{A}_x)$ the lifting ϕ_x is very simple, it sends every finite history $s_1 a_1 s_2 \cdots s_n \in \mathcal{H}(\mathbb{A}_x)$ to $\phi_x(s_1) a_1 \phi_x(s_2) \cdots \phi_x(s_n)$. As a consequence, a simple induction shows that every history consistent with $\Phi(\widehat{\sigma})$ is a history of $\mathbb{A}_x$ and $\Phi(\widehat{\sigma})(s_1 a_1 s_2 \cdots s_n) = \widehat{\sigma}(\phi(s_n))$. Thus $\Phi(\widehat{\sigma})$ is deterministic stationary. $\square$

Remark that the last property stated in Proposition 10 is specific to the case where $x = \widehat{\sigma}(\omega)$, because in general projection turns deterministic strategies into deterministic strategies but does not preserve stationarity. For example in the game of Fig. 2, the stationary strategy σ in $\widehat{\mathbb{A}}$ which plays b in s_a and a in s_b is not stationary anymore once projected to $\mathbb{A}$: when playing with $\Phi_\omega(\sigma)$, it is necessary to remember which action was taken on the last visit to ω.

The proof of Theorem 9 is performed by induction on the size of the games in $\mathcal{G}$. The size of an arena $\mathbb{A} = (\mathbf{S}, \mathbf{A}, \mathbf{T}, p, \mathbf{P}, \mathbf{R}, \rho)$ is defined as

$$\text{size}(\mathbb{A}) = \sum_{s \in \mathbf{S}} (|\mathbf{A_T}(s)| - 1).$$

Note that $\text{size}(\mathbb{A}) \geq 0$ since each state has at least one available action. By extension, the size of a game is the size of its arena.

The core of the inductive step in the proof of Theorem 9 is:

Lemma 11. *Let $G = (\mathbb{A}, \Sigma, \mathcal{T}, \preceq)$ and $\widehat{G} = (\widehat{\mathbb{A}}, \widehat{\Sigma}, \widehat{\mathcal{T}}, \preceq)$ be two games such that $\widehat{\mathbb{A}}$ is the split of $\mathbb{A}$ on some state ω of $\mathbb{A}$ controlled by* Max. *Assume that:*

i) at least two actions are available in state ω in arena $\mathbb{A}$,

ii) there exists optimal deterministic stationary strategies in every subgame G' of $\widehat{G}$ such that such that $\text{size}(G') < \text{size}(G)$, *iii) there exists optimal deterministic stationary strategies in every subgame G' of $\widehat{G}$ which is a one-player game,*

iv) every deterministic stationary strategy in the arena $\widehat{\mathbb{A}}$ is a strategy in the game $\widehat{G}$, i.e. $\widehat{\Sigma} \cup \widehat{\mathcal{T}}$ contains all these strategies.

Then there exists a pair $(\sigma^\sharp, \tau^\sharp)$ of optimal deterministic stationary strategies in $\widehat{G}$.

Proof. We set the usual notations:

$$\mathbb{A} = (\mathbf{S}, \mathbf{A}, \mathbf{T}, p, \mathbf{P}, \mathbf{R}, \rho) \text{ and } \widehat{\mathbb{A}} = (\widehat{\mathbf{S}}, \mathbf{A}, \widehat{\mathbf{T}}, \widehat{p}, \widehat{\mathbf{P}}, \mathbf{R}, \widehat{\rho}) \ .$$

For each action $x \in \mathbf{A_T}(\omega)$ let $\widehat{\mathbb{A}}_x$ be the subarena of $\widehat{\mathbb{A}}$ induced by the set $\widehat{\mathbf{T}}_x$ of transitions:

$$\widehat{\mathbf{T}}_x = \{(\omega, x, s_x) \mid (\omega, x, s) \in \mathbf{T}\} \cup \{(s_x, a, s'_x) \mid (s, a, s') \in \mathbf{T} \text{ and } s \neq \omega\}. \tag{9}$$

For every $x \in \mathbf{A_T}(\omega)$ let $\widehat{G}_x = (\widehat{\mathbb{A}}_x, \Sigma_x, \mathcal{T}_x, \preceq_x)$ be the subgame of $\widehat{G}$ induced by the subarena $\widehat{\mathbb{A}}_x$ of $\widehat{\mathbb{A}}$.

We will make use several times of the following lemma, which relates the outcomes in the games $\widehat{G}$ and $\widehat{G}_x$ when only the action x is played in ω.

Lemma 12. *Let σ and τ be two strategies in $\widehat{G}$ and $x \in \mathbf{A_T}(\omega)$. Assume that σ is deterministic stationary and $\sigma(\omega) = x$. Let σ_x and τ_x be the restrictions of σ and τ to $\mathcal{H}(\widehat{\mathbb{A}}_x)$. Then σ_x and τ_x are strategies in $\widehat{G}_x$. Moreover, for every state $s \in \mathbf{S}$,*

$$\mathbb{O}(\widehat{\mathbb{A}}, s_x, \sigma, \tau) = \mathbb{O}(\widehat{\mathbb{A}}_x, s_x, \sigma_x, \tau_x) \ .$$

Proof. First, remark that σ_x and τ_x are strategies in $\widehat{\mathbb{A}}_x$. For σ_x it is obvious: σ_x is the deterministic stationary strategy in $\widehat{G}_x$ which associates $\sigma(s_x)$ to $s_x \in S_x$ and in particular $\sigma_x(\omega) = \sigma(\omega) = x$. For τ_x it is quite obvious as well, because every state $s_x \in S_x$ controlled by player Min is different from ω thus the same actions are available in s_x in both arenas $\widehat{\mathbb{A}}$ and $\widehat{\mathbb{A}}_x$.

Therefore, σ_x and τ_x are strategies in the subgame $\widehat{G}_x$ by definition of a subgame (Definition 5).

According to the separation property (Proposition 6), since $\sigma(\omega) = x$ every history consistent with σ and starting from S_x stays in $\widehat{\mathbb{A}}_x$, thus $\mathcal{H}(\widehat{\mathbb{A}}, s_x, \sigma, \tau) = \mathcal{H}(\widehat{\mathbb{A}}_x, s_x, \sigma_x, \tau_x)$. Since σ and σ_x on one hand and τ and τ_x on the other hand coincide on $\mathcal{H}(\widehat{\mathbb{A}}_x, s_x, \sigma_x, \tau_x)$ then $\mathbb{P}^{\sigma,\tau}_{\widehat{\mathbb{A}},s_x} = \mathbb{P}^{\sigma_x,\tau_x}_{\widehat{\mathbb{A}}_x,s_x}$. By definition of outcomes, this terminates the proof of Lemma 12. □

Note that, modulo the renaming of states, $\widehat{\mathbb{A}}_x$ is isomorphic to the subarena $\mathbb{A}_x$ of $\mathbb{A}$ obtained by removing from $\mathbf{A_T}(\omega)$ all actions except x. According to hypothesis i) $|\mathbf{A_T}(\omega)| \geq 2$ thus $\text{size}(\widehat{\mathbb{A}}_x) < \text{size}(\mathbb{A})$. According to hypothesis ii), in the game $\widehat{G}_x$ there are optimal deterministic stationary strategies

$$(\sigma^\sharp_x, \tau^\sharp_x) \in \Sigma_x \times \mathcal{T}_x \ .$$

Given these strategies, our aim is to construct a pair $(\sigma^\sharp, \tau^\sharp)$ of optimal deterministic stationary strategies in the game $\widehat{G}$.

Construction of $\tau^\sharp$ Let $\tau^\sharp \in \mathcal{T}(\widehat{\mathbb{A}})$ be the deterministic stationary strategy of player Min in $\widehat{\mathbb{A}}$ defined in the following way: for each $s \in \mathbf{S}$ such that $\mathbf{P}(s) = \text{Min}$ and all $x \in \mathbf{A_T}(\omega)$,

$$\tau^\sharp(s_x) = \tau^\sharp_x(s_x) \ . \tag{10}$$

Construction of $\sigma^\sharp$ The construction of $\sigma^\sharp$ simply consists in selecting one of the optimal deterministic stationary strategies $(\sigma_x)_{x\in\mathbf{A_T}(\omega)}$ in $\widehat{G}_x$ and extend it in a natural way to $\widehat{G}$.

The point is to choose the right action $x \in \mathbf{A_T}(\omega)$. For that we consider the subarena $\widehat{\mathbb{A}}[\tau^\sharp]$ of $\widehat{\mathbb{A}}$ obtained from $\widehat{\mathbb{A}}$ by removing all transitions (s_x, a, s'_x) controlled by Min and such that $a \neq \tau^\sharp(s_x)$. In other words, we restrain the moves of player Min in $\widehat{\mathbb{A}}[\tau^\sharp]$ by forcing him to play actions according to strategy $\tau^\sharp$. Therefore $\widehat{\mathbb{A}}[\tau^\sharp]$ is a subarena of $\widehat{\mathbb{A}}$. We denote $\widehat{G}[\tau^\sharp]$ the subgame of $\widehat{G}$ induced by $\widehat{\mathbb{A}}[\tau^\sharp]$.

Note that only player Max has freedom to choose actions in $\widehat{\mathbb{A}}[\tau^\sharp]$, thus $\widehat{G}[\tau^\sharp]$ is a one-player game controlled by player Max. Therefore, according to hypothesis iii) there are optimal deterministic stationary strategies $(\zeta^\sharp, \tau_0)$ in the game $\widehat{G}[\tau^\sharp]$. Actually τ_0 is the unique trivial strategy of player Min in $\widehat{\mathbb{A}}[\tau^\sharp]$.

Since there is no restriction on actions of player Max in $\widehat{\mathbb{A}}[\tau^\sharp]$ then $\zeta^\sharp$ is also a strategy in arena $\widehat{\mathbb{A}}$ and by definition of a subgame, $\zeta^\sharp$ is a strategy in the game $\widehat{G}$:

$$\zeta^\sharp \in \widehat{\Sigma} \ . \tag{11}$$

Let $e \in \mathbf{A_T}(\omega) = \mathbf{A}_{\widehat{\mathbf{T}}}(\omega)$ be the action chosen by $\zeta^\sharp$ in the state ω,

$$\zeta^\sharp(\omega) = e \ . \tag{12}$$

The stationary and deterministic strategy $\sigma^\sharp : \widehat{\mathbf{S}} \to \mathbf{A}$ is the extension of the optimal deterministic stationary strategy $\sigma^\sharp_e : \widehat{\mathbf{S}}_e \to A$ from $\widehat{\mathbf{S}}_e$ to $\widehat{\mathbf{S}}$ defined by:

$$\forall s_x \in \widehat{\mathbf{S}}, \sigma^\sharp(s_x) = \sigma^\sharp_e(s_e) \ .$$

In particular, since e is the only action available in ω in the arena $\widehat{\mathbb{A}}_e$,

$$\sigma^\sharp(\omega) = e \ . \tag{13}$$

Strategies $(\sigma^\sharp, \tau^\sharp)$ are optimal in $\widehat{G}$. According to hypothesis iv), the strategies strategies $\sigma^\sharp$ and $\tau^\sharp$ are strategies in $\widehat{G}$. We show they are optimal in the game $\widehat{G}$ if the initial state belongs to $\mathbf{S}_e$, i.e. for all strategies $\sigma \in \widehat{\Sigma}, \tau \in \widehat{\mathcal{T}}$ and all $s \in \mathbf{S}$,

$$\mathbb{O}(\widehat{\mathbb{A}}, s_e, \sigma, \tau^\sharp) \preceq \mathbb{O}(\widehat{\mathbb{A}}, s_e, \sigma^\sharp, \tau^\sharp) \preceq \mathbb{O}(\widehat{\mathbb{A}}, s_e, \sigma^\sharp, \tau) \ . \tag{14}$$

We start with the proof of the right-handside of (14). Let τ be any strategy for Min in $\widehat{G}$. Then its restriction τ_e to $\mathcal{H}(\widehat{\mathbb{A}}_e)$ is also strategy in $\widehat{\mathbb{A}}_e$ because for every state $t_e \in S_e$ controlled by player Min, $t_e \neq \omega$ thus exactly the same actions are available in t_e in both arenas $\widehat{\mathbb{A}}$ and $\widehat{\mathbb{A}}_e$. The right inequality in (14) decomposes as:

$$\mathbb{O}(\widehat{\mathbb{A}}, s_e, \sigma^\sharp, \tau^\sharp) = \mathbb{O}(\widehat{\mathbb{A}}_e, s_e, \sigma^\sharp_e, \tau^\sharp_e) \preceq_e \mathbb{O}(\widehat{\mathbb{A}}_e, s_e, \sigma^\sharp_e, \tau_e) = \mathbb{O}(\widehat{\mathbb{A}}, s_e, \sigma^\sharp, \tau) \ .$$

The left and right equalities holds according to Lemma 12. The central inequality holds because $\sigma^\sharp_e$ and $\tau^\sharp_e$ are optimal in $\widehat{G}_e$ and τ_e is a strategy in $\widehat{G}_e$, according to Lemma 12 again. Since $\preceq_e$ is a restriction of $\preceq$ we get the right-handside of (14).

Now we prove the left-handside of (14). Let σ be any strategy for Max in $\widehat{G}$. Then the restriction σ' of σ to $\mathcal{H}(\widehat{\mathbb{A}}[\tau^\sharp])$ is a strategy in $\widehat{\mathbb{A}}[\tau^\sharp]$ because the same actions are available in every state s controlled by Max in both arenas $\widehat{\mathbb{A}}$ and $\widehat{\mathbb{A}}[\tau^\sharp]$. Let $\zeta_e^\sharp$ be the restriction of $\zeta^\sharp$ to the states of $\mathbf{S}_e$. The left inequality in (14) decomposes as:

$$
\begin{aligned}
\mathbb{O}(\widehat{\mathbb{A}}, s_e, \sigma, \tau^\sharp) &= \mathbb{O}(\widehat{\mathbb{A}}[\tau^\sharp], s_e, \sigma', \tau_0) && (15)\\
&\preceq \mathbb{O}(\widehat{\mathbb{A}}[\tau^\sharp], s_e, \zeta^\sharp, \tau_0) && (16)\\
&= \mathbb{O}(\widehat{\mathbb{A}}, s_e, \zeta^\sharp, \tau^\sharp) && (17)\\
&= \mathbb{O}(\widehat{\mathbb{A}}_e, s_e, \zeta_e^\sharp, \tau_e^\sharp) && (18)\\
&\preceq_e \mathbb{O}(\widehat{\mathbb{A}}_e, s_e, \sigma_e^\sharp, \tau_e^\sharp) && (19)\\
&= \mathbb{O}(\widehat{\mathbb{A}}_e, s_e, \sigma^\sharp, \tau^\sharp) \ . && (20)
\end{aligned}
$$

The equality (15) holds because $\mathcal{H}(\widehat{\mathbb{A}}, s_e, \sigma, \tau^\sharp) = \mathcal{H}(\widehat{\mathbb{A}}[\tau^\sharp], s_e, \sigma', \tau_0)$ and for every finite history $h \in \mathcal{H}(\widehat{\mathbb{A}}, s_e, \sigma, \tau^\sharp)$, $\mathbb{P}^{\sigma,\tau^\sharp}_{\widehat{\mathbb{A}},s_e}(h^+) = \mathbb{P}^{\sigma',\tau_0}_{\widehat{\mathbb{A}}[\tau^\sharp],s_e}(h^+)$. The inequality (16) holds because $\zeta^\sharp$ and τ_0 are optimal in $\widehat{G}[\tau^\sharp]$ and by definition of subgames σ' is a strategy in the subgame $\widehat{G}[\tau^\sharp]$ and the preference order in $\widehat{G}[\tau^\sharp]$ is a restriction of $\preceq$. The equality (17) holds for similar reasons than (15) does. The equalities (18) and (20) are consequences of Lemma 12. The inequality (19) holds because $(\sigma_e^\sharp, \tau_e^\sharp)$ is a pair of optimal strategies in $\widehat{G}_e$ and $\zeta_e^\sharp$ is a strategy in $\widehat{G}_e$ according to Lemma 12 again. Since $\preceq_e$ is a restriction of $\preceq$ we get the left inequality in (14).

Finally the two inequalities in (14) do hold and since $\sigma^\sharp$ is deterministic stationary the proof of Lemma 11 is over. □

Proof of Theorem 9. We fix a subgame-closed and split-closed collection $\mathcal{G}$ of games which contains all deterministic stationary strategies. We denote $\mathcal{G} = (G_i)_{i\in I}$ and for every $i \in I$ we denote

$$G_i = (\mathbb{A}_i, \Sigma_i, \mathcal{T}_i, \preceq_i)_{i\in I} \ .$$

The proof of Theorem 9 is carried out by induction on the size of $\mathbb{A}_i$.

The case where $\text{size}(\mathbb{A}_i) = 0$ for some $i \in I$ is trivial. In this case there is a unique action available in each state s of $\mathbb{A}_i$, and each player has a unique strategy which is optimal deterministic stationary.

Let $G = (\mathbb{A}, \Sigma, \mathcal{T}, \preceq)$ a game in $\mathcal{G}$ such that $\text{size}(G) > 0$ and suppose that the theorem holds for all games G_i in $\mathcal{G}$ whose arenas $\mathbb{A}_i$ satisfy $\text{size}(\mathbb{A}_i) < \text{size}(\mathbb{A})$.

In case G is a one-player game, then by hypothesis optimal deterministic stationary strategies exist for both players. In the sequel we assume G is *not* a one-player game.

Therefore there exists a state ω in $\mathbb{A}$ such that $\mathbf{A_T}(\omega) \geq 2$. Assume first that

$$\mathbf{P}(\omega) = \text{Max} \ . \tag{21}$$

Let $\widehat{\mathbb{A}}$ be the split of $\mathbb{A}$ on ω. Then, since $\mathcal{G}$ is split-closed there exists a game $\widehat{G} = (\widehat{\mathbb{A}}, \widehat{\Sigma}, \widehat{\mathcal{T}}, \preceq)$ in $\mathcal{G}$ such that property (6) of Definition 8 holds for every action x available in ω.

All conditions of Lemma 11 are satisfied for the games G and $\widehat{G}$: condition i) is by choice of ω, condition ii) is the inductive hypothesis, and by hypothesis conditions iii) and iv) hold for every game in $\mathcal{G}$, in particular for $\widehat{G}$. As a consequence there exists a pair $(\sigma^\sharp, \tau^\sharp)$ of optimal deterministic stationary strategies in $\widehat{G}$.

Let $e = \sigma^\sharp(\omega)$ be the action played by $\sigma^\sharp$ in the separation state and $\Phi_e(\sigma^\sharp)$ the lifting of $\sigma^\sharp$ from $\widehat{\mathbb{A}}$ to $\mathbb{A}$ with respect to action e. By choice of $\widehat{G}$, condition (6) holds thus $\Phi_x(\widehat{\Sigma}) \subseteq \Sigma$ and $\Phi_e(\sigma^\sharp)$ is a strategy in G. According to the second part of Proposition 10, the strategy $\Phi_e(\sigma^\sharp)$ is both optimal in G and deterministic stationary.

Therefore Max has an optimal deterministic stationary strategy in G.

To find an optimal deterministic stationary for player Min in G it suffices to choose as a separation state a state controlled by player Min with at least two actions available. Such a state exists because G is not a one-player game. By a reasoning symmetric to the one developped previously we can construct another pair of optimal strategies $(\sigma^\star, \tau^\star)$ in G, however now the strategy $\tau^\star$ of player Min will be deterministic stationary.

But in zero-sum games if we have two pairs of optimal strategies $(\sigma^\sharp, \tau^\sharp)$ and $(\sigma^\star, \tau^\star)$ then $(\sigma^\sharp, \tau^\star)$ is also a pair of optimal strategies.

This ends the proof of Theorem 9.

Even if it has no bearing on the proof, we provide an explicit description of the strategy $\Phi_e(\tau^\sharp)$. Let $h \in \mathcal{H}(\mathbb{A})$ be any finite history consistent with $\mathbb{A}$ starting in a state s and ending in a state s controlled by Min and $\widehat{h} = \phi_e(h)$ the lifting of h in $\mathcal{H}(\widehat{\mathbb{A}})$ with respect to e. Then $\widehat{h}$ ends in a state $s_x \in \pi^{-1}(s)$ for some $x \in \mathbf{A_T}(\omega)$. By definition, $\widehat{\pi}^{-1}(\tau^\sharp)[h] = \tau^\sharp[\widehat{h}] = \tau^\sharp[s_x] = \tau^\sharp_x[s_x]$. Informally, for a history ending in a state s player Min chooses one of the actions $\tau^\sharp_x[s_x]$, $x \in \mathbf{A_T}(\omega)$.

The problem is to see which of these actions should be chosen and this depends on the last state of $\widehat{h}$. Thus the question is if we can obtain the last state of $\widehat{h}$ without explicitly calculating the whole lifted history $\widehat{h}$? If h never visits the state ω then, since $\widehat{h}$ begins in a state of $\widehat{\mathbf{S}}_e$ this history can only end in the state $s_e \in \widehat{\mathbf{S}}_e$, i.e $\widehat{\pi}^{-1}(\tau^\sharp)[h] = \tau^\sharp_e(s_e)$. If h visits ω then all depends on the action chosen by Max during the last visit to ω, if this action is $x \in \mathbf{A_T}(\omega)$ then the last state of $\widehat{h}$ is s_x and therefore $\widehat{\pi}^{-1}(\tau^\sharp)[h] = \tau^\sharp_x(s_x)$. Notice that the strategy $\widehat{\pi}^{-1}(\tau^\sharp)$ is not stationary, however we only need a finite amount of memory to implement it, it is sufficient to know all strategies $\tau^\sharp_x$, the last state s of h, whether since the begining of the game the state ω was visited or not and if it was visited then what action $x \in \mathbf{A_T}(\omega)$ was taken during the last visit to ω. $\square$

8 Final remarks

Theorem 9 gives a sufficient condition for the existence of optimal deterministic stationary strategies for a given two-player game.

The examples in Section 3 were given only to illustrate the method, we do not think that reestablishing known results is of particular interest.

Let us note that in recent years, attempting to capture subtle aspects of computer systems behaviour, several new games were proposed, as typical example we can cite [CHJ05] combining parity and mean-payoff games. We hope that Theorem 9 and its deterministic counterpart Theorem 3 will prove useful in the study of such new games.

Finally, note that an improved version of the results is in preparation, for a class of games both more general and simpler, where probabilities are abstracted as non-determinism.

References

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

[CHJ05] Krishnendu Chatterjee, Thomas A. Henzinger, and Marcin Jurdziński. Mean-payoff parity games. In *Proceedings, Twentieth Annual IEEE Symposium on Logic in Computer Science, LICS 2005*, pages 178–187. IEEE Computer Society Press, 2005.

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

[Der62] C. Derman. On sequential decisions and Markov chains. *Managemant Science*, 9:16–24, 1962.

[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.

[Nev71] J. Neveu. *Mathematical Foundations of the Calculus of Probability*. Holden-Day, 1971.

[Ney04] A. Neyman. From Markov chains to stochaqstic 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.