Strategy Iteration using Non-Deterministic Strategies for Solving Parity Games

Michael Luttenberger

Institut für Informatik, Technische Universität München, 85748 Garching, Germany
luttenbe@model.in.tum.de

Abstract. This article extends the idea of solving parity games by strategy iteration to non-deterministic strategies: In a non-deterministic strategy a player restricts himself to some non-empty subset of possible actions at a given node, instead of limiting himself to exactly one action.

We show that a strategy-improvement algorithm by by Björklund, Sandberg, and Vorobyov [3] can easily be adapted to the more general setting of non-deterministic strategies. Further, we show that applying the heuristic of "all profitable switches" (cf. [1]) leads to choosing a "locally optimal" successor strategy in the setting of non-deterministic strategies, thereby obtaining an easy proof of an algorithm by Schewe [13].

In contrast to [3], we present our algorithm directly for parity games which allows us to compare it to the algorithm by Jurdzinski and Vöge [15]: We show that the valuations used in both algorithm coincide on parity game arenas in which one player can "surrender". Thus, our algorithm can also be seen as a generalization of the one by Jurdzinski and Vöge to non-deterministic strategies.

Finally, using non-deterministic strategies allows us to show that the number of improvement steps is bound from above by $O(1.724^n)$. For strategy-improvement algorithms, this bound was previously only known to be attainable by using randomization (cf. [1]).

1 Introduction

A parity game arena consists of a directed graph $G = (V, E)$ where every vertex belongs to exactly one of two players, called player 0 and player 1. Every vertex is colored by some natural number in $\{0, 1, \ldots, d-1\}$. Starting from some initial vertex v_0, a play of both players is an infinite path in G where the owner of the current node determines the next vertex. In a parity game, the winner of such an infinite play is then defined by the parity of the maximal color which appears infinitely often along the given play.

As shown by Mostowski [11], and independently by Emerson and Jutla [4], there exists a partition of V in two sets W_0 and W_1 such that player i has a memoryless strategy, i.e. a map $\sigma_i : V_i \to V$ which maps every vertex v controlled by player i to some successor v, so that player i wins any play starting from some $w \in W_i$ by using σ_i to determine his moves.

Interest in parity games arises as determining the winning set W_0 is equivalent to deciding whether a given μ-calculus formula holds w.r.t. to a given Kripke structure, i.e. determining W_0 is equivalent to the model checking problem of μ-calculus. Further interest is sparked as it is known that solving parity games is in UP$\cap$co-UP [8], but no polynomial time algorithm has been found yet.

In this article we consider an approach for calculating the winning sets which is known as strategy iteration or strategy improvement, and can be described as follows in the setting of games: In a first step, a way for valuating the strategies of player 0 is fixed, thereby inducing a partial order on the strategies of player 0. Then, one chooses an initial strategy $\sigma : V_0 \to V$ for player 0. Iteratively (i) the current strategy is valuated, (ii) by means of this valuation possible improvements of the current strategy are determined, i.e. pairs (u, v) such that $\sigma[u \mapsto v]$ is a strategy having a better valuation than σ, (iii) a subset of the possible improvements is selected and implemented yielding a better strategy $\sigma' : V_0 \to V$. These steps are repeated until no improvements can be found anymore.

Although this approach usually (using no randomization [1]) allows only to give a bound exponential in $|V_0|$ on the number of iterations needed till termination, there is no family of games known for which this approach leads to a super-polynomial number of improvement steps. It is thus also used in practice e.g. in compilers [14].

In particular, this approach has been successfully applied in several different scenarios like Markov decision processes [6], stochastic games [5], or discounted payoff games [12]. Using reductions, these algorithms can also be used for solving parity games. In 2000 Jurdzinski and Vöge [15] presented the first strategy-improvement algorithm for parity games which directly works on the given parity game without requiring any reductions to some intermediate representation. Although the algorithm by Jurdzinski and Vöge did not lead to a better upper bound on the complexity of deciding the winner of a parity game with n nodes and d colors (the algorithm in [15] has a complexity of $O((n/d)^d)$ whereas the upper bound of $O((n/d)^{d/2})$ was already known at that time [9]), it sparked a lot of interest as the strategy-improvement process w.r.t. parity games is directly observable and not obfuscated by some reduction.

In this article, we extend strategy iteration to *non-deterministic strategies*: In a non-deterministic strategy a player is not required to fix a single successor for any vertex controlled by him instead he restricts himself to some non-empty subset of all possible successors. Using non-deterministic strategies seems to be more natural, as it allows a player to only “disable” those moves along which the valuation of the current strategy decreases. Our algorithm is an extension of an algorithm by Björklund, Sandberg, and Vorobyov [3] proposed in 2004. In particular, we borrow their idea of giving one of the two players the option to give up and “escape” an infinite play he would lose by introducing a sink. In contrast to the original algorithm in [3] we present this extended algorithm directly for parity games in order to be able to compare this algorithm directly with the one by Jurdzinski and Vöge, and also in the hope that this might lead to better insights regarding the strategy improvement process.

Strategy iteration, as described above, chooses in step (iii) some subset of possible changes in order to obtain the next (deterministic) strategy. A natural question is how to choose this set of changes. Obviously, one would like to choose these sets in such a way that the total number of improvements steps is as small as possible – we call this “globally optimal”. As no efficient algorithm for determining these sets is known, usually heuristics are used instead. One heuristic applied quite often in the case of a binary arena is called "all profitable switches" [1]: In a binary arena, given a strategy $\sigma : V_0 \to V$ we can refer to the successors of $v \in V_0$ by $\sigma(v)$ and $\overline{\sigma(v)}$. A strategy improvement step then amounts to deciding for every node $v \in V_0$ whether to switch from $\sigma(v)$ to $\overline{\sigma(v)}$, or not. "All possible switches" refers then to the heuristic of switching to $\overline{\sigma(v)}$ of every $v \in V$ if this switch is an improvement w.r.t. the used valuation. Transferring this heuristic to the setting of non-deterministic strategies the heuristic becomes simply to choose the set of all possible improvements of the given strategy as the new strategy considered in the next step. We show that this simple heuristic leads to the "locally optimal" improvement, i.e. the strategy which is at least as good as any other strategy obtainable by implementing a subset of the possible improvements. By applying this heuristic in every step we obtain a new, in our opinion more natural and accessible, presentation of the algorithm by Schewe proposed in [13]: There only valuations (referred to as "estimations" there), and deterministic strategies are considered, whereas the strategy improvement process itself, and the connection to [3] are obfuscated. Further, the algorithm in [13] does not work directly on parity games, and requires some unnecessary restrictions on the graph structure of the arena, e.g. only bipartite arenas are considered.

We then compare our algorithm using non-deterministic strategies to the one by Jurdzinski and Vöge [15]. This is not possible w.r.t. the algorithm in [3] or [13] as these do not work directly on parity games. Here, we can show that the valuation used in our algorithm, resp. in [15] coincide, which readily allows us to conclude that the locally optimal improvement obtained by our algorithm is always at least as good as any locally improvement obtainable by [15].

We obtain an upper bound of $O(|V|^2 \cdot |E| \cdot (\frac{|V|}{d} + 1)^d)$ for our algorithm which is the same as the one obtainable when using deterministic strategies [3]. So using non-deterministic strategies comes "for free". Of course, w.r.t. to the sub-exponential bound of $|V|^{O(\sqrt{|V|})}$ obtainable for the algorithm by Jurdzinski, Paterson and Zwick [7], our algorithm is not competitive. Still, we think that our algorithm is interesting as strategy-iteration in practice only requires a polynomial number of improvement steps in general, as already mentioned above. In particular, we can show that the number of improvement steps done by our algorithm when using the "all profitable switches"-heuristic, and thus by the one by Schewe [13], is bounded by $O(1.724^{|V_0|})$, whereas the best known upper bound for strategy iteration when using only deterministic strategies and no randomization in the improvement selection is $O(2^{|V_0|}/|V_0|)$ [1]. In particular, the bound of $O(1.724^{|V_0|})$ was previously known to be obtainable only be choosing the improvements randomly [1].

Organization: Section 2 summarizes the standard definitions and results regarding parity games. In Section 3 we extend parity games by allowing player 0 to terminate infinite plays in order to escape an infinite play he would lose. This idea was first stated in [3]. We combine this with a generalization of the path profiles used in [15] in order to get an algorithm working directly on parity games. Section 4 summarizes our strategy improvement algorithm using non-deterministic strategies. Section 5 then compares the algorithm presented in this article with the one by Jurdzinski and Vöge.

2 Preliminaries

In this section we repeat the standard definitions and notations regarding parity games.

An *arena* $\mathcal{A}$ is given by (V, E, o), if (V, E) is a finite, directed graph, where $o : V \to \{0, 1\}$ assigns each node an owner. We denote by $V_i := o^{-1}(i)$ the set of all nodes belonging to player $i \in \{0, 1\}$, and write E_i for $E \cap V_i \times V$. Given some subset $V' \subseteq V$ we write $\mathcal{A}|_{V'}$ for the restriction of the arena $\mathcal{A}$ to the nodes V'. A play $\pi \in V^{\mathbb{N}} \cup V^*$ in $\mathcal{A}$ is any maximal path in $\mathcal{A}$ where we assume that player i determines the move $(\pi(i), \pi(i + 1))$, if $\pi(i) \in V_i$. For (V, E) a directed graph, and $s \in V$ a node we write sE for the set of successors of s.

For $\mathcal{A} = (V, E, o)$ an arena, a (memoryless) *strategy of player i (short: i-strategy)* ($i \in \{0, 1\}$) is any subset $\sigma \subseteq E_i$ satisfying $\forall s \in V_i : |sE| > 0 \Rightarrow |s\sigma| > 0$, i.e. a strategy does not introduce any new dead ends. σ is *deterministic*, if $|s\sigma| \leq 1$ for all $s \in V_i$. We write E_σ for $E_\sigma = E_{1-i} \cup \sigma$, and $\mathcal{A}|_\sigma$ for (V, E_σ, o).

We assume that the reader is familiar with the concept of attractors. For convenience, a definition can be found in the appendix.

A *parity game arena* $\mathcal{A}$ is given by (V, E, o, c) where (V, E, o) is an arena with $vE \neq \emptyset$ for all $v \in V$, and $c : V \to \{0, 1, \ldots, d - 1\}$ assigns each node a color. The winner of a play π in a parity game arena is given by $\limsup_{i \in \mathbb{N}} c(\pi(i)) \pmod 2$. Given a node s, a strategy $\sigma \in E_i$ is a winning strategy for s of player i, if he wins any play in $\mathcal{A}|_\sigma$ starting from s. Player i wins a node s, if he has a winning strategy for it. W_i denotes the set of nodes won by player i. As we assume that every node has at least one successor, there are only infinite plays in a parity game arena. Wlog., we further assume that $c^{-1}(k) \neq \emptyset$ for all $k \in \{0, 1, \ldots, d - 1\}$ as we may otherwise reduce d. A cycle $s_0 s_1 \ldots s_{n-1}$ (with $s_{i+1 \pmod n} \in s_i E$) in a parity game arena $\mathcal{A}$ is called *i-dominated*, if the parity of its highest color is i. Player i wins the node s using strategy $\sigma \subseteq V_i \times V$, iff every cycle reachable from s in $\mathcal{A}|_\sigma$ is i-dominated.

Theorem 1. *[11, 4] For any a parity game arena $\mathcal{A}$ we have $W_0 \cup W_1 = V$. Player i possesses a deterministic strategy $\sigma_i^* : V_i \to V$ with which he win every node $s \in W_i$.*

3 Escape Arenas

In this section we extend parity games by allowing player 0 to *escape* an infinite play which he would loose w.r.t. the parity game winning condition:

Let $\mathcal{A} = (V, E, o)$ be a parity game arena. We obtain the arena $\mathcal{A}_\perp = (V_\perp, E_\perp, o_\perp)$ from $\mathcal{A}$ by introducing a sink $\perp$ $V_\perp := V \uplus \{\perp\}$ where only player 0 can choose to play to $\perp$ ($E_\perp := E \cup V_0 \times \{\perp\}$). The sink $\perp$ itself has no out-going edges, and we assume that player 0 controls $\perp$ ($o_\perp := o \cup \{(\perp, 0)\}$ although this is of no real importance. Although, this construction was first proposed in [3] we refer to $\mathcal{A}_\perp$ as *escape arena* in the style of [13]. As $\mathcal{A}_\perp$ itself is no parity game arena anymore, we have to define the winner of such a finite play as well. For this we extend the definition of *color profile*, which was first stated in [2], to finite plays:

For a given escape arena $\mathcal{A}_\perp$ using d colors $\{0, 1, \ldots, d - 1\}$, we define the set $\mathcal{P}$ of *color profiles* by $\mathcal{P} := \mathbb{Z}^d \cup \{-\infty, \infty\}$ where $\mathbb{Z}^d$ is the set of d-dimensional integer vectors. We write ø for the zero-profile $(0, 0, \ldots, 0) \in \mathbb{Z}^d$, and use standard addition on $\mathbb{Z}^d$ for two profiles $\wp, \wp' \in \mathbb{Z}^d$. The idea of a profile $\wp \in \mathcal{P}$ is to count how often a given color appears a long a finite play, whereas $-\infty$, reps. ∞ correspond to infinite plays won by player 1, resp. player 0. More precisely, for a finite sequence $\pi = s_0 s_1 \ldots s_l$ of vertices, the *value* $\wp(\pi)$ of π is the profile which counts how often a color $k \in \{0, 1, \ldots, d-1\}$ appears in $c(s_0)c(s_1)\ldots c(s_l)$. For an infinite sequence $\pi = s_0 s_1 \ldots$, its *value* $\wp(\pi)$ is defined to be ∞, if π is won by player 0 w.r.t. the parity game winning condition; otherwise $\wp(\pi) := -\infty$. Finally, we introduce a total order $\prec$ on $\mathcal{P}$ which tries to capture the notion of when one of two given plays is better than the other for player 0: For this we set (i) $-\infty$ to be the bottom element of $\prec$, (ii) ∞ to be the top element of $\prec$, and (iii) for all $\wp, \wp' \in \mathcal{P} \setminus \{-\infty, \infty\}$ we set:

$$\wp \prec \wp' :\Leftrightarrow \exists k \in \{0, 1, \ldots, d-1\} : k = \max\{k \in \{0, 1, \ldots, d-1\} \mid \wp_k \neq \wp'_k\} \\ \wedge (k \equiv_2 0 \wedge \wp_k < \wp'_k \vee k \equiv_2 1 \wedge \wp_k > \wp'_k) \, .$$

Informally, the definition of $\prec$ says that player 0 hates to loose in an infinite play, whereas he likes it the most to win an infinite play. So, whenever he can, he will try to escape an infinite play he cannot win, therefore resulting in a finite play to $\bot$: here, given two finite plays π_1, π_2 ending in $\bot$, player 0 looks for the highest color c which does not appear equally often along both plays. If c is even, he prefers that play in which it appears more often; if it is odd, he prefers the one in which it appears less often. In particular, player 0 dislikes visiting odd-dominated cycle, while he likes visiting even-dominated ones:

Lemma 1. *Assume that $\chi = s_0 s_1 \ldots s_n$ is a non-empty cycle in the parity game arena $\mathcal{A}$, i.e. $s_0 \in s_n E$ and $n \geq 0$. χ is 0-dominated, i.e. the highest color in χ is even if and only if $\wp(\chi) \succ$ ø. χ is 1-dominated if and only if $\wp(\chi) \prec$ ø.*

Now, for a given parity game arena $\mathcal{A}$ let σ_0^*, σ_1^* be the optimal winning strategies of player 0, resp. 1. Further, let W_0, W_1 be the corresponding winning sets. Obviously, both players can still use these strategies in $\mathcal{A}_\bot$, too, as we only added additional edges. Especially, player 0 can still use σ_0^* to win W_0 in $\mathcal{A}_\bot$ as only he has the option to move to $\bot$. In the case of player 1, by applying σ_1^* any cycle in $\mathcal{A}_\bot|_{\sigma_1^*}$ reachable from a vertex $v \in W_1$ has to be odd-dominated. Hence, player 0 prefers to play in an acyclic path from v to $\bot$ in $\mathcal{A}_\bot|_{\sigma_1}$ when starting from a vertex in W_1.

Let therefore be $\overline{\wp}$ the $\prec$-maximal value of any acyclic path terminating in $\bot$ in $\mathcal{A}_\bot$. $\overline{\wp}$ is the best player 0 can hope to achieve starting from a node $v \in W_1$ when player 1 plays optimal. We therefore define: player 0 wins a play π, if $\wp(\pi) \succ \overline{\wp}$, otherwise player 1 wins the play. Player i wins a node $s \in V$, if he has a strategy $\sigma \subseteq E_i$ with which he wins any play starting from s in $\mathcal{A}|_\sigma$. As already sketched, this leads then to the following theorem.

Theorem 2. *Player i wins the node s in $\mathcal{A}$ iff he wins it in $\mathcal{A}_\bot$.*

4 Strategy Improvement

We now turn to the problem of finding optimal winning strategies by iteratively valuating the strategy, and determining from this valuation possible better strategies. The following section can be seen as the generalization of the algorithm in [3] to non-deterministic strategies and explicitly stated in the setting of parity games. In fact, we will only consider a special class of strategies for player 0, i.e. such strategies which do not introduce any 1-dominated cycles. The strategy improvement process will assure that no 1-dominated cycles are created. If there are any 1-dominated cycles in $\mathcal{A}_\perp|_{V_1}$, then player 1 wins all the nodes in the 1-attractor to these cycles. We may, thus, identify the nodes trivially won by player 1 in a preprocessing step, and remove them.

Assumption 1. *The arena $\mathcal{A}_\perp|_{V_1}$ has no 1-dominated cycles.*

Definition 1. *We call a strategy $\sigma \subseteq E_0$ of player 0* reasonable, *if there are no 1-dominated cycles in $\mathcal{A}_\perp|_\sigma$.*

Remark 1. **(a)** By our assumption above the strategy $\sigma_\perp := V_0 \times \{\perp\}$ is reasonable, as every 1-dominated cycle in $\mathcal{A}$ consists of at least one node controlled by player 0. **(b)** Let σ be any strategy of player 0, and W_σ the set of nodes won by σ. Then, the strategy $\sigma' = \sigma \cap (W_\sigma \times W_\sigma) \cup \{(s,\perp) \mid s \in V_0 \setminus W_\sigma\}$ is reasonable with $W_\sigma = W_{\sigma'}$.

We may thus assume that player 0 uses only reasonable strategies.

Definition 2. *Let σ be some reasonable strategy of player 0. Its valuation $\mathcal{V}_\sigma : V \cup \{\perp\} \to \mathcal{P}$ maps every node s on the $\prec$-minimal value $\mathcal{V}_\sigma(s)$ which player 1 can* guarantee *to achieve in any play starting from s in $\mathcal{A}_\perp|_\sigma$ by using some memoryless strategy:*

$$\mathcal{V}_\sigma(s) := \overset{\prec}{\min_{\tau\subseteq E_1 \textit{ strategy}}} \overset{\prec}{\max}\{\wp(\pi) \mid \pi \textit{ is a play in } \mathcal{A}_\perp|_{\sigma,\tau} \wedge \pi(0) = s\},$$

where we set $\mathcal{V}_\sigma(\perp) :=$ ø.

Remark 2. **(a)** We will show later that, if we start from the reasonable strategy $\sigma_\perp := V_0 \times \{\perp\}$, then our strategy-improvement algorithm will only generate reasonable strategies. (Note, if $\mathcal{A}_\perp|_{\sigma_\perp}$ had 1-dominated cycles, then these would need to exist solely in $\mathcal{A}|_{V_1}$ – but we have assumed above that we removed those in a preprocessing step.) **(b)** As shown above, for all $s \in W_1$ player 1 can use his optimal winning strategy σ_1^* from the parity game to guarantee $\mathcal{V}_\sigma(s) \preceq \overline{\wp} \prec \infty$.

By means of the valuation $\mathcal{V}_\sigma$ we can partially order reasonable strategies in the natural way:

Definition 3. *For two (reasonable) strategies σ_a, σ_b of player 0 we write $\sigma_a \preceq \sigma_b$, if $\mathcal{V}_{\sigma_a}(s) \preceq \mathcal{V}_{\sigma_b}(s)$ for all nodes s. We write $\sigma_a \prec \sigma_b$, if there is at least one node s such that $\mathcal{V}_{\sigma_a}(s) \prec \mathcal{V}_{\sigma_b}(s)$. Finally, $\sigma_a \approx \sigma_b$, if $\sigma_a \preceq \sigma_b \wedge \sigma_b \preceq \sigma_a$.*

The following lemma addresses the calculation of $\mathcal{V}_\sigma$ using a straight-forward adaption of the Bellman-Ford algorithm:

Lemma 2. *Let $\sigma \subseteq E_0$ be a reasonable strategy of player 0. We define $\mathcal{V}_\perp : V \cup \{\perp\} \to \mathcal{P}$ by $\mathcal{V}_\perp(\perp) :=$ ø, and $\mathcal{V}_\perp(s) = \infty$ for all $s \in V$, and the operator $F_\sigma : (V \cup \{\perp\} \to \mathcal{P}) \to (V \cup \{\perp\} \to \mathcal{P})$ by*

$$\begin{aligned}
&F_\sigma[\mathcal{V}](\perp) := \text{ø} \\
&F_\sigma[\mathcal{V}](s) \;\; := \wp(s) + \min\nolimits^{\preceq}\{\mathcal{V}(t) \mid (s,t) \in E_1\} \;\; \textit{if } s \in V_1, \\
&F_\sigma[\mathcal{V}](s) \;\; := \wp(s) + \max\nolimits^{\preceq}\{\mathcal{V}(t) \mid (s,t) \in \sigma\} \quad \textit{if } s \in V_0,
\end{aligned}$$

for any $\mathcal{V} : V \cup \{\bot\} \to \mathcal{P}$.

Then, the valuation $\mathcal{V}_\sigma$ of σ is given as the limit of the sequence $F^i_\sigma[\mathcal{V}_\bot]$ for $i \to \infty$, and this limit is reached after at most $|V|$ iterations.

Remark 3. **(a)** We assume unit cost for adding and comparing color profiles. The time needed for calculating $\mathcal{V}_\sigma$ is then simply given by $O(|V| \cdot |E|)$.**(b)** For every $s \in V$ there has to be at least one edge (s,t) with $\mathcal{V}_\sigma(s) = \wp(s) + \mathcal{V}_\sigma(t)$, as $\mathcal{V}_\sigma = F_\sigma[\mathcal{V}_\sigma]$.

W.r.t. $\mathcal{V}_\sigma$ we can identify possible *improvements* of σ:

Definition 4. *Let $\sigma \subseteq E_0$ be a reasonable strategy of player* 0*. The set I_σ of* improvements*, resp. the set S_σ of* strict improvements *of σ is defined by*

$$I_\sigma := \{(s,t) \in E_0 \mid \mathcal{V}_\sigma(s) \preceq \wp(s)+\mathcal{V}_\sigma(t)\},\ \textit{resp. } S_\sigma := \{(s,t) \in E_0 \mid \mathcal{V}_\sigma(s) \prec \wp(s)+\mathcal{V}_\sigma(t)\}.$$

We call any strategy $\sigma \subseteq E_0$ a direct improvement *of σ, if $\sigma \subseteq I_\sigma$.*

Fact 1. *Let σ' be a direct improvement of σ. Then along every edge (u,v) of $\mathcal{A}_\bot|_{\sigma'}$ we have $\mathcal{V}_\sigma(u) \preceq \wp(u) + \mathcal{V}_\sigma(v)$. In particular, we have for any finite path $s_0 s_1 \dots s_{l+1}$ in $\mathcal{A}_\bot|_{\sigma'}$*

$$\mathcal{V}_\sigma(s_0) \preceq \wp(s_0) + \mathcal{V}_\sigma(s_1) \preceq \wp(s_0 s_1) + \mathcal{V}_\sigma(s_2) \preceq \dots \preceq \wp(s_0 \dots s_l) + \mathcal{V}_\sigma(s_{l+1}).$$

From this easy fact, several important properties of direct improvements follow:

Corollary 1. *If σ is reasonable, then any 0-strategy $\sigma' \subseteq I_\sigma$ is reasonable, too.*

Corollary 2. *Let σ be a reasonable strategy. For a direct improvement σ' of σ we have that $\sigma \preceq \sigma'$. If σ' contains at least one strict improvement of σ, then this inequality is strict, i.e. $\sigma \prec \sigma'$.*

The preceding corollaries show that starting with an initial reasonable strategy σ_0, e.g. $\sigma_\bot$, we can generate a sequence $\sigma_0, \sigma_1, \sigma_2, \dots$ of reasonable strategies such that $\mathcal{V}_{\sigma_i}(s) \preceq \mathcal{V}_{\sigma_{i+1}}(s)$ for all $s \in V$, if we choose the strategy σ_{i+1} to be some direct improvement of σ_i. Further, we know, if σ_{i+1} uses at least one strict improvement (s,t) of σ_i, i.e. $(s,t) \in \sigma_{i+1} \cap S_{\sigma_i} \neq \emptyset$, then we have $\mathcal{V}_{\sigma_i}(s) \prec \mathcal{V}_{\sigma_{i+1}}(s)$, i.e. every possible reasonable strategy occurs at most once along the strategy improvement sequence. As already shown, we have always $\mathcal{V}_{\sigma_i}(s) \preceq \overline{\wp} \prec \infty$ for all nodes $s \in W_1$. The obvious question is now, if we can reach an optimal winning strategy by this procedure, i.e. is a reasonable strategy σ with $S_\sigma = \emptyset$ optimal? This is answered in the following lemma.

Lemma 3. *As long as there is a node $s \in W_0$ with $\mathcal{V}_\sigma(s) \prec \infty$, σ has at least one strict improvement.*

Due to this lemma, we know that, if a reasonable strategy σ has no strict improvements, i.e. $S_\sigma = \emptyset$, then we have $\mathcal{V}_\sigma(s) = \infty$ for at least all the nodes $s \in W_0$. On the other hand, for all nodes $s \in W_1$ we always have $\mathcal{V}_\sigma(s) \preceq \overline{\wp}$. Hence, by the determinacy of parity games, i.e. $W_1 = V \setminus W_0$, σ has to be an optimal winning strategy for player 0, if $S_\sigma = \emptyset$. By our construction such an optimal strategy σ with $S_\sigma = \emptyset$ might be non-deterministic. The following lemma shows how one can deduce an optimal deterministic strategy from such a σ.

Lemma 4. *Let σ be a reasonable strategy of player 0 in $\mathcal{A}_\perp$, and I_σ the strategy consisting of all improvements of σ. Then every deterministic strategy $\sigma' \subseteq I_\sigma$ with $\mathcal{V}_{I_\sigma}(s) = \wp(s) + \mathcal{V}_{I_\sigma}(t)$ for all $(s,t) \in \sigma'$ satisfies $\mathcal{V}_{I_\sigma} = \mathcal{V}_{\sigma'}$.*

Starting from $\sigma_\perp = \{(s,\perp) | s \in V_0\}$, if we improve the current strategy using at least one strict improvement in every step, we will end up with an optimal winning strategy for player 0. As in every step the valuation increases in at least those nodes at which a strict improvement exists, and as there are at most $(\frac{|V|}{d}+1)^d$ possible values a valuation can assign a given node, the number of improvement steps is bound by $|V| \cdot (\frac{|V|}{d}+1)^d$. The cost of every improvement step is given by the cost of the calculation of $\mathcal{V}_\sigma$, we thus get:

Theorem 3. *Let σ_0 be some reasonable 0-strategy. By iteratively taking σ_{i+1} to be some direct improvement of σ_i which uses at least one strict improvement, one obtains an optimal winning strategy after at most $|V| \cdot (\frac{|V|}{d}+1)^d$ iterations. The total running time is thus $O(|V|^2 \cdot |E| \cdot (\frac{|V|}{d}+1)^d)$.*

4.1 All Profitable Switches

In the previous subsection we have not said anything about which direct improvement should be taken in every improvement step. As no algorithms are known which determine for a given strategy such a direct improvement that the total number of improvement steps is minimal (we call such a direct improvement “globally optimal”), one usual resorts to heuristics for choosing a direct improvement, (see e.g. [1]).Most often the heuristic “all profitable switches” mentioned in the introduction is used. In the case of non-deterministic strategies this simply becomes taking I_σ as successor strategy. The interesting fact here is that I_σ is a “locally optimal” direct improvement for a given reasonable strategy σ, i.e. for all strategies $\sigma' \subseteq I_\sigma$ we have $\sigma' \preceq I_\sigma$. We remark that this has already been shown implicitly by Schewe in [13]:

Theorem 4. *Let σ be a reasonable strategy with I_σ its set of improvements. For any direct improvement of σ we have $\sigma' \preceq I_\sigma$.*

We like to give an easy proof for this theorem. We first note the following two properties of the operator F_σ:

Fact 2. **(i)** *For $\mathcal{V}, \mathcal{V}' : V \cup \{\perp\} \to \mathcal{P}$ with $\mathcal{V} \preceq \mathcal{V}'$ we have $F_\sigma[\mathcal{V}] \preceq F_\sigma[\mathcal{V}']$.*
(ii) *For two 0-strategies $\sigma_a \subseteq \sigma_b$ we have $F_{\sigma_a}[\mathcal{V}](s) \preceq F_{\sigma_b}[\mathcal{V}](s)$ for all $s \in V$.*

Using (i) and (ii) we get by induction

$$F^{i+1}_{\sigma_a}[\mathcal{V}_\perp] = F_{\sigma_a}[F^i_{\sigma_a}[\mathcal{V}_\perp]] \preceq F_{\sigma_a}[F^i_{\sigma_b}[\mathcal{V}_\perp]] \preceq F_{\sigma_b}[F^i_{\sigma_b}[\mathcal{V}_\perp]] = F^{i+1}_{\sigma_b}[\mathcal{V}_\perp],$$

and therefore the following lemma:

Lemma 5. *If σ_a and σ_b are reasonable and $\sigma_a \subseteq \sigma_b$, it holds that $\mathcal{V}_{\sigma_a} \preceq \mathcal{V}_{\sigma_b}$.*

Now, as the set of improvements I_σ of a given reasonable strategy σ is itself a (non-deterministic) strategy, and every direct improvement σ' of σ satisfies $\sigma' \subseteq I_\sigma$ by definition, the theorem from above follows. The algorithm of Schewe in [13] can therefore be described as an optimized implementation of non-deterministic strategy iteration using the "all profitable switches" heuristic.

We close this section with a remark on the calculation of $\mathcal{V}_{I_\sigma}$. Schewe proposes an algorithm for calculating $\mathcal{V}_{I_\sigma}$ which uses $\mathcal{V}_\sigma$ to speed up the calculation leading to $O(|E|\log|V|)$ operations on color-profiles instead of $O(|E|\cdot|V|)$. For this, formulated in the notation of our algorithm, he introduces edge weights $w(u,v) := (\wp(u) + \mathcal{V}_\sigma(v)) - \mathcal{V}_\sigma(u)$, and calculates w.r.t. these edges an update $\delta = \mathcal{V}_{I_\sigma} - \mathcal{V}_\sigma$. We argue that one can use Dijkstra's algorithm for this, as we have $\mathcal{V}_\sigma(u) \preceq \wp(u) + \mathcal{V}_\sigma(v)$ along all edges $(u,v) \in I_\sigma$, and thus $w(u,v) \succeq ø$, i.e. all edge weights are non-negative.

Proposition 1. *$\mathcal{V}_{I_\sigma}$ can be calculated using Dijkstra's algorithm which needs $O(|V|^2)$ operations on color-profiles on dense graphs; for graphs whose out-degree is bound by some b this can be improved to $O(b\cdot|V|\cdot\log|V|)$ by using a heap* [1].

This gives us a running time of $O(|V|^3\cdot(\frac{|V|}{d}+1)^d)$, resp. $O(|V|^2\cdot b\cdot\log|V|\cdot(\frac{|V|}{d}+1)^d)$.

5 Comparison with the Algorithm by Jurdzinski and Vöge

This section compares the algorithm presented in this article with the one by Jurdzinski and Vöge [15]. We first give a short (slightly imprecise) description of the algorithm in [15]: This algorithm starts in each step with some *deterministic* 0-strategy σ. Using σ a valuation Ω_σ is calculated (see below for details about Ω_σ). Then, by means of this valuation possible strategy improvements are determined, and finally some non-empty subset of these improvements is chosen, but only one improvement per node at most, such that implementing these improvements yields a *deterministic* strategy again. This process is repeated until there are no improvements anymore w.r.t. the current strategy.

The valuation Ω_σ: We present a slightly "optimized" version of the valuation used in [15]. The valuation $\Omega_\sigma(s)$ of a deterministic 0-strategy σ consists of the the *cycle value* $z_\sigma(s)$, the *path profile* $\wp_\sigma(s)$, and the *path length* $l_\sigma(s)$ which are defined as follows:

- As σ is deterministic, all plays in $\mathcal{A}|_\sigma$ are determined by player 1. For every node z having odd color, we can decide whether there is at least one cycle in $\mathcal{A}|_\sigma$ such that this cycle is dominated by z. Let Z be the set of all odd colored nodes dominating a cycle in $\mathcal{A}|_\sigma$.
  Given a node s we define $z_\sigma(s)$ to be a node of maximal color in Z, which is reachable from s in $\mathcal{A}|_\sigma$; if no node in Z is reachable from s in $\mathcal{A}|_\sigma$, then s has to be won by player 0, and we set $z_\sigma(s) = \infty$.

[1] In [3] the authors propose another optimization to speed up the calculation of $\mathcal{V}_\sigma$ by restricting the re-calculation of $\mathcal{V}_\sigma$ to only those nodes where $\mathcal{V}_\sigma$ changes. Those nodes can be easily identified by calculating an attractor again in time $O(|E|)$. Unfortunately, combining this optimization with the one by Schewe ([13]) does not lead to a better asymptotic upper bound.

- If $z_\sigma(s)$ is some odd colored node, the second component $\wp_\sigma(s)$ becomes the color profile of a $\prec$-minimal play from s to $z_\sigma(s)$ in $\mathcal{A}|_\sigma$ – with the restriction that only nodes of color $\geq c(z_\sigma(s))$ are counted.
- Finally, if $\wp_\sigma(s)$ is defined, $l_\sigma(s)$ is the length of shortest play from s to $z_\sigma(s)$ w.r.t. $\wp_\sigma(s)$, if $z_\sigma(s)$ has odd color.

Remark 4. We assume here that z_σ is either ∞, if s is already won using σ, or the "worst" odd-dominated cycle into which player 1 can force a play starting from s. In [15], the authors even try to optimize $z_\sigma(s)$ when s is already won using σ. These improvements are obviously unnecessary, as we can always remove the attractor to these nodes from the arena in an intermediate step in order to obtain a smaller arena.

Further, it is assumed in [15] that every node is uniquely colored. Therefore, in [15] $\wp_\sigma$ is defined to be the *set* of nodes having higher color than $z_\sigma(s)$ on a "worst" path from s to $z_\sigma(s)$. Jurdzinski and Vöge already mention at the end of [15] that their algorithm also works when not assuming that every vertex is uniquely colored, but do not present the adapted data structures needed in this case. This was done in [2]: If the same color is used for several vertices, it is sufficient to only count the number of nodes having a color $k \geq z_\sigma(s)$ along such a "worst" path from s to $z_\sigma(s)$ where "worst" path simply means a $\prec$-minimal path then. Therefore, the color profiles used in this article are a direct generalization of the path profiles used in [15].

In [15] an edge $(s,t) \in E_0$ is now called a strict improvement over $(s, \sigma(s))$, if $\Omega_\sigma(t)$ is strictly better than $\Omega_\sigma(\sigma(s))$, i.e. either the "worst" cycle improves, or the worst play to it improves, or the length of a worst play becomes longer ("the longer player 0 can stay away from z_σ the better for him"). A deterministic strategy σ' is then a direct improvement of a given deterministic strategy σ w.r.t. [15], if it differs from σ only in strict improvements.

Definition 5. *For a given parity game arena $\mathcal{A} = (V, E, c, o)$, set*

$$\mathcal{A}^\perp := (V \cup \{\perp\}, E \cup V_0 \times \{\perp\} \cup \{(\perp,\perp)\}, c \cup \{(\perp,-1)\}, o \cup \{(\perp,0)\}).$$

$\mathcal{A}^\perp$ results from $\mathcal{A}_\perp$ by simply adding a loop to $\perp$, and giving $\perp$ the color -1 so that $\mathcal{A}^\perp$ is a parity game arena where $\perp$ is the cycle dominated by the least odd color. A straight-forward adaption of the proof of Theorem 2 shows that player 0 wins a node s in $\mathcal{A}$ iff he wins it in $\mathcal{A}^\perp$.

Now, as the strategy improvement algorithm in [15] tries to play to the "best" possible cycle, an optimal strategy (obtained by the algorithm) will always choose to play to $\perp$ from a node s, if s cannot be won by player 0, as every other 1-dominated cycle has at least 1 as maximal color. A strategy σ of player 0 is therefore "reasonable" w.r.t. to the algorithm by Jurdzinski and Vöge, if $(\perp,\perp)$ is the only 1-dominated cycle in $\mathcal{A}^\perp|_\sigma$.

Obviously, we now have an one-to-one correspondence between reasonable strategies σ in $\mathcal{A}_\perp$, and reasonable strategies σ in $\mathcal{A}^\perp$ of player 0: we simply have to remove or add the edge $(\perp,\perp)$ to move from $\mathcal{A}_\perp$ to $\mathcal{A}^\perp$ and vice versa. We therefore may identify these strategies in the following as one strategy.

This allows us to compare the improvement step of the algorithm presented in this article with that of [15]. Indeed, as the color of $\perp$ is -1 (recall that all other nodes have colors ≥ 0), we have $\wp_\sigma(s) = \mathcal{V}_\sigma(s)$ for all nodes with $z_\sigma(s) = \bot$, and $\mathcal{V}_\sigma(s) = \infty$, if $z_\sigma(s) \neq \bot$. This proves the following proposition:

Proposition 2. *Any (deterministic) direct improvement σ' of σ identified by [15] is a subset of I_σ. Therefore $\sigma' \preceq I_\sigma$.*

In other words, the algorithm presented here always chooses *locally* a direct improvement of σ which is at least as good as any deterministic direct improvement obtainable by [15]. In the appendix, a small example can be found illustrating this.

5.1 Bound on the number of Improvement Steps

We finish this section by giving an upper bound on the total number of improvement steps when using the "all profitable switches"-heuristic. In the case of an arena with out-degree two, one can show that the number of improvement steps done by the algorithm in [15] is bounded by $O(\frac{2^{|V_0|}}{|V_0|})$ (cf. [1]).

When considering non-deterministic strategies the heuristic "all profitable switches" naturally generalizes to simply taking I_σ as successor strategy in every iteration. Here we can show the following upper bound:

Theorem 5. *Let $\mathcal{A}_\bot$ be a escape-parity-game arena where every node of player 0 has at most two successor. Then the number of improvement steps needed to reach an optimal winning strategy is bound by $3 \cdot 1.724^{|V_0|}$ when using non-deterministic strategy iteration and the "all profitable switches"-heuristic.*

Remark 5. To the best of our knowledge this is the best upper bound known for any deterministic strategy-improvement algorithm. In [1] a similar bound is only obtained by using randomization.

6 Conclusions

In the first part of the article, we presented an extended version of the algorithm by [3] which (i) allows the use of non-deterministic strategies, and (ii) works directly on the given parity game arena without requiring a reduction to a mean payoff game as an intermediate step. For (ii), we used the path profiles introduced in [15], resp. a generalized version of it called color profiles (see also [2]).

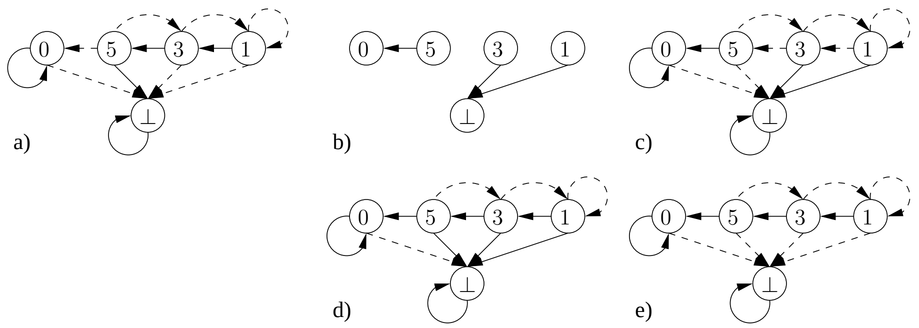
We then showed that the heuristic "all profitable switches" in the setting of non-deterministic strategies leads to the locally best direct improvement, and therefore to the algorithm presented in [13].We further identified the fast calculation of the valuation proposed by Schewe as the Dijkstra algorithm.

Finally, we turned to the comparison of the algorithm presented here to the one by Jurdzinski and Vöge [15]. As our algorithm works directly on parity games in contrast to [3, 13], we could show that the valuations used in both coincide for parity game arenas with escape for player 0. We finished the article by adapting results from [10] which allowed us to show that using the "all profitable switches"-heuristic in the setting of non-deterministic strategies allows to obtain an upper bound of $O(1.724^{|V_0|})$ on the total number of improvement steps. This bound also carries over to the algorithm in [13]. This bound was previously only attainable using randomization [1].

References

1. H. Björklund, S. Sandberg, and S. Vorobyov. Optimization on completely unimodal hypercubes. Technical Report 2002–18, Department of Information Technology, Uppsala University, 2002.
2. H. Björklund, S. Sandberg, and S. Vorobyov. A discrete subexponential algorithm for parity games. In *STACS'03*, LNCS 2607, pages 663–674. Springer, 2003.
3. H. Björklund, S. Sandberg, and S. Vorobyov. A combinatorial strongly subexponential strategy improvement algorithm for mean payoff games. In *MFCS'04*, LNCS 3153, pages 673–685. Springer, 2004.
4. E.A. Emerson and C.S. Jutla. Tree automata, mu-calculus and determinacy (extended abstract). In *FOCS'91*. IEEE Computer Society Press, 1991.
5. A. Hoffman and R. Karp. On nonterminating stochastic games. *Management Science*, 12, 1966.
6. Ronald A. Howard. Dynamic programming and markov processes. *The M.I.T. Press*, 1960.
7. M. Jurdziński, M. Paterson, and U. Zwick. A deterministic subexponential algorithm for solving parity games. In *SODA'06*. ACM/SIAM, 2006.
8. Marcin Jurdziński. Deciding the winner in parity games is in UP $\cap$ co-UP. *Information Processing Letters*, 68(3):119–124, November 1998.
9. Marcin Jurdziński. Small progress measures for solving parity games. In *STACS 2000*, volume 1770 of *LNCS*, 2000.
10. Y. Mansour and S. Singh. On the complexity of policy iteration. In *UAI 1999*, 1999.
11. A.W. Mostowski. Games with forbidden positions. Technical Report 78, University of Gdańsk, 1991.
12. Anuj Puri. *Theory of Hybrid Systems and Discrete Event Systems*. PhD thesis, Electronic Research Laboratory, College of Engineering, University of California, Berkley, 1995.
13. Sven Schewe. An optimal strategy improvement algorithm for solving parity games. Technical Report 28, Universität Saarbrücken, 2007.
14. H. Seidl and T. Gawlitza. Precise relational invariants through strategy iteration. In *CSL'07*, LNCS, 2007.
15. Jens Vöge and Marcin Jurdziński. A discrete strategy improvement algorithm for solving parity games (Extended abstract). In *CAV'00*, volume 1855 of *LNCS*, 2000.

A Example: Comparison with the Algorithm by Jurdzinski and Vöge



a) depicts an arena $\mathcal{A}^{\perp}$ where bold arrows represent the edges of a 0-strategy σ, and dashed arrows represent edges not included in σ. Further, all nodes belong to player 0, where the numbers inside the nodes represent the colors. **b)** shows the set S_σ of strict improvements w.r.t. σ. **c)** The heuristic applied usually for choosing a deterministic direct improvement of σ is to take a maximal subset of S_σ so that for every node, for which a strict improvement exists, there is exactly one strict improvement chosen. In this example this leads to the strategy depicted in c). **d)** The algorithm presented in this article, on the other hand, chooses the non-deterministic strategy $I_\sigma = \sigma \cup S_\sigma$, as shown in d). **e)** Calculating the valuation of both I_σ, and the strategy shown in e) shows that both strategy are equivalent w.r.t. their valuation (see also lemma 4). This means the strategy I_σ is already optimal in difference to c).

B Missing Proofs

B.1 Preliminaries

Definition 6. *Given an arena $\mathcal{A} = (V, E, o)$ and a target set $T \subseteq V$ of nodes, we define the i-attractor Attr$_i[\mathcal{A}](T)$ to T in $\mathcal{A}$ by*

$$\begin{array}{ll} A_0 & := T \\ A_{i+1} & := A_i \cup \{s \in V_i | sE \cap A_i \neq \emptyset\} \cup \{s \in V_{1-i} | sE \subseteq A_i\} \\ \mathit{Attr}_0[\mathcal{A}](T) & := \bigcup_{i \geq 0} A_i. \end{array}$$

The rank $r(s) \in \mathbb{N} \cup \{\infty\}$ of a node s w.r.t. to Attr$_0[\mathcal{A}](T)$ is given by

$$\min\{i \in \mathbb{N} | s \in A_i\}$$

where we assume that $\min \emptyset = \infty$.

A strategy $\sigma \subseteq E_i$ is then an i-attractor strategy to T, if for every $(s, t) \in \sigma$ the rank decreases along (s, t) as long as s has finite, non-zero rank.

Remark 6. Obviously, player i can use any i-attractor strategy to force any play starting from a node with finite rank into T on an acyclic path as the rank is strictly decreasing until T is hit.

B.2 Parity Game Arenas with Escape for Player 0

Lemma 1. *Assume that $\chi = s_0 s_1 \ldots s_n$ is a non-empty cycle in the parity game arena $\mathcal{A}$, i.e. $s_0 \in s_n E$ and $n \geq 0$. χ is 0-dominated, i.e. the highest color in χ is even if and only if $\wp(\chi) \succ$ ø. χ is 1-dominated if and only if $\wp(\chi) \prec$ ø.*

Proof. Wlog. we may assume that s_0 has the dominating color in χ. As all remaining nodes in χ have at most color $c(s_0)$, the color profile $\wp(\chi)$ is 0 for all colors $> c(s_0)$. Hence, the highest color in which $\wp(\chi)$ and ø differ is $c(s)$. If $c(s)$ is even, then $\wp(\chi) \succ$ ø by definition, otherwise $\wp(\chi) \prec$ ø, as $\wp(\chi)_{c(s_0)} > 0$. The other direction is shown similarly. □

Theorem 2. *Player i wins the node s in $\mathcal{A}$ iff he wins it in $\mathcal{A}_\perp$.*

Proof. Let σ_i^* be the optimal, memoryless winning strategy in the parity game $\mathcal{A}$, and W_i the winning set of of player i w.r.t. σ_i^*.

First consider the case $s \in W_0$. As only player 0 can choose to move to $\perp$, any play π in $\mathcal{A}_\perp$ w.r.t. σ_0^* is a play in $\mathcal{A}$, too. Hence, π is infinite, and won by player 0 w.r.t. the parity game winning condition. Thus, π has the value ∞.

Assume now that $s \in W_1$. Player 1 can use his optimal strategy to force player 0 starting from s into a play such that every cycle visited is 1-dominated. If player 0 does not move to $\perp$, the infinite play also exists in the original parity game arena, is therefore won by player 1, and, hence, has the value $-\infty$ in the escape game. On the other hand, in the escape parity game $\mathcal{A}_\perp$ player 0 has now the option to escape any such infinite play by opting to terminate the game by moving to $\perp$. Consider therefore a finite play $\pi = s_0 s_1 \ldots s_n \perp$. Assume that this path is not acyclic. Thus, as we are only counting how often a given color appears along the path, we may split π into a simple path π' from s_0 to $\top$ and several cycles $\chi_1, \ldots, \chi_l$. By using his winning strategy σ_1^* player 1 can make sure that every such cycle has an odd color as maximal color. It is now easy to see that $\wp(\chi_j) \prec$ ø by definition of $\prec$. Thus, we have

$$\wp(\pi) = \wp(\pi') + \wp(\chi_1) + \ldots + \wp(\chi_l) \prec \wp(\pi') \preceq \overline{\wp}.$$

□

B.3 Strategy Improvement

Lemma 2. *Let $\sigma \subseteq E_0$ be a reasonable strategy of player 0. We define $\mathcal{V}_\perp : V \cup \{\perp\} \to \mathcal{P}$ by $\mathcal{V}_\perp(\perp) :=$ ø, and $\mathcal{V}_\perp(s) = \infty$ for all $s \in V$, and the operator $F_\sigma : (V \cup \{\perp\} \to \mathcal{P}) \to (V \cup \{\perp\} \to \mathcal{P})$ by*

$$\begin{aligned} F_\sigma[\mathcal{V}](\perp) &:= ø \\ F_\sigma[\mathcal{V}](s) &:= \wp(s) + \min{}^{\preceq}\{\mathcal{V}(t) \mid (s,t) \in E_1\} \quad \textit{if } s \in V_1, \\ F_\sigma[\mathcal{V}](s) &:= \wp(s) + \max{}^{\preceq}\{\mathcal{V}(t) \mid (s,t) \in \sigma\} \quad \textit{if } s \in V_0, \end{aligned}$$

for any $\mathcal{V} : V \cup \{\perp\} \to \mathcal{P}$.

Then, the valuation $\mathcal{V}_\sigma$ of σ is given as the limit of the sequence $F_\sigma^i[\mathcal{V}_\perp]$ for $i \to \infty$, and this limit is reached after at most $|V|$ iterations.

Proof. For all $\mathcal{V}, \mathcal{V}' : V \cup \{\bot\} \to \mathcal{P}$ with $\mathcal{V}(s) \preceq \mathcal{V}'(s)$ for $s \in V \cup \{\bot\}$ we have $F_\sigma[\mathcal{V}](s) \preceq F_\sigma[\mathcal{V}'](s)$, too, i.e. F_σ is monotone. Obviously, we have $F_\sigma[\mathcal{V}_\bot](s) \preceq \mathcal{V}_\bot(s)$ for all $s \in V \cup \{\bot\}$. Therefore, $F_\sigma^i[\mathcal{V}_\bot](s)$ is monotonically decreasing for $i \to \infty$.

As σ is reasonable, $\mathcal{V}_\sigma(s) \succ -\infty$, and it can only be finite, if s is in the 1-attractor to $\bot$ in $\mathcal{A}_\bot|_\sigma$. Further, for $\mathcal{V}_\sigma(s) \prec \infty$, $\mathcal{V}_\sigma(s)$ has to be the value of an acyclic play π in $\mathcal{A}_\bot|_\sigma$. One therefore checks easily that $\mathcal{V}_\sigma$ is a fixed point of F_σ; hence, by the monotonicity of F_σ, and $\mathcal{V}_\sigma \preceq \mathcal{V}_\bot$, we have $\mathcal{V}_\sigma \preceq F^i[\mathcal{V}_\bot]$ for all $i \in \mathbb{N}$.

Let C_i be the set of nodes $s \in V \cup \{\bot\}$ such that $F_\sigma^i[\mathcal{V}_\bot](s) = \mathcal{V}_\sigma(s)$. Obviously, we have $\bot \in C_i$ for all $i \in \mathbb{N}$. As $F_\sigma^i[\mathcal{V}_\bot]$ is monotonically decreasing, and bounded from below by $\mathcal{V}_\sigma$, we have $C_i \subseteq C_{i+1}$.

Define B_i to be the boundary of C_i, i.e. the set of nodes $s \in V \setminus C_i$ with $sE \cap C_i \neq \emptyset \wedge sE \cap V \setminus C_i \neq \emptyset$.

If $B_i \subseteq V_0$, then player 0 has a strategy to stay away from $\bot \in C_i$ for every node $s \in V \setminus C_i$. It is easy to see that $F^i[\mathcal{V}_\bot](s) = \infty$ for all $s \in V \setminus C_i$ in this case.

Thus, assume $B_i \cap V_1 \neq \emptyset$. As player 1 eventually needs to enter C_i in order to reach $\bot$, he has to use an edge from a node $s' \in V_1 \cap B_i$ to C_i. At least for this node s' we have to have $s' \in C_{i+1}$.

Hence, we have to have $C_i = V$ for some $i \leq |V|$, implying $F_\sigma^{i+1}[\mathcal{V}_\bot] = F_\sigma^i[\mathcal{V}_\bot]$.
□

Definition 7. *We write $\tau_\sigma \subseteq E_1$ for the 1-strategy consisting of the edges (s,t) with $\mathcal{V}_\sigma(s) = \wp(s) + \mathcal{V}_\sigma(t)$.*

Corollary 1. *If σ is reasonable, then any direct improvement σ' of σ is reasonable, too.*

Proof. For any cycle $s_0 s_1 \dots s_l$ with $s_0 \in s_l E_\sigma$, we have

$$\mathcal{V}_\sigma(s_0) \preceq \wp(s_0 \dots s_l) + \mathcal{V}_\sigma(s_0), \text{ i.e. } ø \prec \wp(s_0 \dots s_l).$$

□

Corollary 2. *Let σ be a reasonable strategy.*

(a) For a direct improvement σ' of σ we have that $\mathcal{V}_\sigma(s) \preceq \mathcal{V}_{\sigma'}(s)$ for all $s \in V$.
(b) If $(s,t) \in \sigma'$ is a strict improvement of σ, then $\mathcal{V}_\sigma(s) \prec \mathcal{V}_{\sigma'}(s)$.

Proof. (a) Let s be any node. For any play $\pi = s_0 s_1 \dots s_n \bot$ starting from s in $\mathcal{A}_\bot|_\sigma$ we have already shown:

$$\mathcal{V}_\sigma(s) \preceq \wp(\pi) + \mathcal{V}_\sigma(\bot) = \wp(\pi) \preceq \mathcal{V}_{\sigma'}(s).$$

(b) As (s,t) is a strict improvement of σ, we have (i) $\mathcal{V}_\sigma(s) \prec \wp(s) + \mathcal{V}_\sigma(t)$, (ii) $s \in V_0$, and, hence, (iii) $\mathcal{V}_{\sigma'}(s) = \max^{\prec}\{\wp(s) + \mathcal{V}_{\sigma'}(t') \mid (s,t') \in \sigma'\}$. With the result from (a) it follows that

$$\mathcal{V}_\sigma(s) \prec \wp(s) + \mathcal{V}_\sigma(t) \preceq \wp(s) + \mathcal{V}_{\sigma'}(t) \preceq \mathcal{V}_{\sigma'}(s).$$

□

Lemma 3. *As long as there is a node $s \in W_0$ with $\mathcal{V}_\sigma(s) \prec \infty$, σ has at least one strict improvement.*

Proof. Let A be the set of nodes t with $\mathcal{V}_\sigma(t) \prec \infty$, i.e. A is the 1-attractor to $\perp$ in $\mathcal{A}_\perp|_\sigma$. By assumption we have $W_0 \cap A \neq \emptyset$. Assume $s \in A \cap W_0$. Let π be any play determined by τ_σ and σ_0^*. As σ_0^* is optimal and $s \in W_0$, π stays in W_0 forever, i.e. the play is infinite.

First, assume π does not leave A. Every time π uses an edge (u, v) which does not exist in $\mathcal{A}_\perp|_\sigma$ it has to hold that $u \in V_0$. Hence, as σ is not strict improvable, we have to have $\mathcal{V}_\sigma(u) \succeq \wp(u) + \mathcal{V}_\sigma(v)$ for all edges $(u, v) \in \sigma_0^*$. On the other hand, we have $\mathcal{V}_\sigma(u) = \wp(u) + \mathcal{V}_\sigma(v)$ along edges $(u, v) \in \tau_\sigma$. Thus, the value of any cycle visited by π is $\prec$ ø – a contradiction.

Therefore, consider the case that π leaves A. This also has to happen along an edge (u, v) with $u \in V_0$. As $u \in A$ and $v \in V \setminus A$ we have $\mathcal{V}_\sigma(u) \prec \infty = \mathcal{V}_\sigma(v)$. Hence, (u, v) is a strict improvement. □

Lemma 4. *Let σ be a reasonable strategy of player 0 in $\mathcal{A}_\perp$, and I_σ the strategy consisting of all improvements of σ.*

Then every deterministic strategy $\sigma' \subseteq I_\sigma$ with $\mathcal{V}_{I_\sigma}(s) = \wp(s) + \mathcal{V}_{I_\sigma}(t)$ for all $(s, t) \in \sigma'$ satisfies $\mathcal{V}_{I_\sigma} = \mathcal{V}_{\sigma'}$.

Proof. By definition, σ' is a direct improvement of I_σ, hence, we have $\mathcal{V}_{I_\sigma}(s) \preceq \mathcal{V}_{\sigma'}(s)$ for all nodes s.

On the other hand, σ' is also a direct improvement of σ, as $\sigma' \subseteq I_\sigma$. Thus, we have $\mathcal{V}_{\sigma'}(s) \preceq \mathcal{V}_{I_\sigma}(s)$ for all $s \in V$. □

Lemma 6. *(a) For σ_a and σ_b two reasonable strategies of player 0, we define the strategy σ_{ab} by*

$$(s, t) \in \sigma_{ab} :\Leftrightarrow \max\{\mathcal{V}_{\sigma_a}(s), \mathcal{V}_{\sigma_b}(s)\} \preceq \wp(s) + \overset{\prec}{\max}\{\mathcal{V}_{\sigma_a}(t), \mathcal{V}_{\sigma_b}(t)\}.$$

Then $\max^{\prec}\{\mathcal{V}_{\sigma_a}(s), \mathcal{V}_{\sigma_b}(s)\} \preceq \mathcal{V}_{\sigma_{ab}}(s)$ for all $s \in V$, i.e. there is a strategy $\hat{\sigma}$ such that for all other strategies σ we have $\mathcal{V}_\sigma(s) \preceq \mathcal{V}_{\hat{\sigma}}(s)$ for all $s \in V$.

(b) If $\mathcal{V}_\sigma(s) \prec \mathcal{V}_{\hat{\sigma}}(s)$ for at least one $s \in V$, then σ has a strict improvement.

Proof. (a) We first show that σ_{ab} is indeed a strategy. Consider any $s \in V_0$. Then there is at least one t_a s.t. $(s, t_a) \in \sigma_a$ and $\mathcal{V}_{\sigma_a}(s) = \wp(s) + \mathcal{V}_{\sigma_a}(t_a)$, and similarly a t_b with the same properties w.r.t. σ_b. Assume $\mathcal{V}_{\sigma_a}(s) \preceq \mathcal{V}_{\sigma_b}(s)$ – the other case being similar. By definition of $\mathcal{V}_\sigma$ we then have

$$\mathcal{V}_{\sigma_a}(t_b) \preceq \mathcal{V}_{\sigma_a}(t_a) = \wp(s) + \mathcal{V}_{\sigma_a}(s) \preceq \wp(s) + \mathcal{V}_{\sigma_b}(s) = \mathcal{V}_{\sigma_b}(t_b),$$

i.e. $(s, t_b) \in \sigma_{ab}$.

By definition, we have

$$\max\{\mathcal{V}_{\sigma_a}(s), \mathcal{V}_{\sigma_b}(s)\} \preceq \wp(s) + \overset{\prec}{\max}\{\mathcal{V}_{\sigma_a}(s), \mathcal{V}_{\sigma_b}(s)\} \quad (*)$$

along every edge $(s, t) \in \sigma_{ab}$. For any edge $(s, t) \in E_1$, we have

$$\mathcal{V}_{\sigma_a}(s) \preceq \wp(s) + \mathcal{V}_{\sigma_a}(t) \text{ and } \mathcal{V}_{\sigma_b}(s) \preceq \wp(s) + \mathcal{V}_{\sigma_b}(t).$$

Hence, $(*)$ holds along every edge of $\mathcal{A}_\perp|_{\sigma_{ab}}$. Therefore, any cycle in $\mathcal{A}_\perp|_{\sigma_{ab}}$ has to be 0-dominated, again, i.e. σ_{ab} is reasonable, too.

If $\mathcal{V}_{\sigma_{ab}}(s) = \infty$, there is nothing to show. Assume $\mathcal{V}_{\sigma_{ab}}(s) \prec \infty$, first, and let $\pi = s_0 s_1 \dots s_n \perp$ be any acyclic play with $\wp(\pi) = \mathcal{V}_{\sigma_{ab}}(s)$. Because of $(*)$ we then have $\max\{\mathcal{V}_{\sigma_a}(s), \mathcal{V}_{\sigma_b}(s)\} \preceq \wp(\pi) = \mathcal{V}_{\sigma_{ab}}(s)$, again.

(b) If there is some node $s \in V$ with $\mathcal{V}_\sigma(s) \prec \mathcal{V}_{\hat{\sigma}}(s) = \infty$, we already know that σ has a strict improvement as it is not optimal (s is won by $\hat{\sigma}$ but not by σ).

Therefore assume that $\mathcal{V}_{\hat{\sigma}}(s') = \infty$ implies $\mathcal{V}_\sigma(s') = \infty$ for all nodes s', and let s be a node with $\mathcal{V}_{\hat{\sigma}}(s) \prec \infty$. Let π again be an acyclic play in $\mathcal{A}_\perp|_{\hat{\sigma},\tau_\sigma}$ with $\wp(\pi) = \mathcal{V}_{\hat{\sigma}}(s)$, i.e. player 0 uses $\hat{\sigma}$ and player 1 his response-strategy τ_σ for σ.

As σ has no strict improvements, we have $\mathcal{V}_\sigma(s) \succeq \wp(s) + \mathcal{V}_\sigma(t)$ for all edges $(s,t) \in E_0$; on the other hand, along the edges $(s,t) \in \tau_\sigma$ we have $\mathcal{V}_\sigma(s) = \wp(s) + \mathcal{V}_\sigma(t)$ by definition of τ_σ.

Hence, we get $\mathcal{V}_{\hat{\sigma}}(s) \succeq \mathcal{V}_\sigma(s) \succeq \wp(\pi) = \mathcal{V}_{\hat{\sigma}}(s)$, if σ has no strict improvements.

□

Proposition 1. *$\mathcal{V}_{I_\sigma}$ can be calculated using Dijkstra's algorithm which needs $O(|V|^2)$ operations on color-profiles on dense graphs; for graphs whose out-degree is bound by some b this can be improved to $O(b \cdot |V| \cdot \log |V|)$ by using a heap.*

Proof. Let σ be a reasonable σ strategy of player 0, and A the 1-attractor to $\perp$ in $\mathcal{A}_\perp|_{I_\sigma}$. For all nodes $s \in V \setminus A$, we have $\mathcal{V}_{I_\sigma}(s) = \infty$. We therefore have only to consider the graph $(A, E_{I_\sigma} \cap A \times A)$ in order to calculate $\mathcal{V}_{I_\sigma}$ for the nodes in A.

Recall that we have for every edge (u,v) in $\mathcal{A}_\perp|_{I_\sigma}$ that $\mathcal{V}_\sigma(u) \preceq \wp(u) + \mathcal{V}_\sigma(v)$. Define now for $(u,v) \in E_{I_\sigma} \cap A \times A$ the function w by $w(u,v) := (\wp(u) + \mathcal{V}_\sigma(v)) - \mathcal{V}_\sigma(u) \succeq \text{ø}$. Hence, for any path $\pi' = t_0 t_1 \dots t_n \perp$ in $(A, E_{I_\sigma} \cap A \times A)$ we have

$$\begin{aligned}
&\wp(\pi') - \mathcal{V}_\sigma(t_0)\\
=\ &\wp(t_0) + \dots + \wp(t_n) + (\mathcal{V}_\sigma(t_1) - \mathcal{V}_\sigma(t_1)) + \dots + (\mathcal{V}_\sigma(t_n) - \mathcal{V}_\sigma(t_n)) - \mathcal{V}_\sigma(t_0)\\
=\ &(\wp(t_0) + \mathcal{V}_\sigma(t_1) - \mathcal{V}_\sigma(t_0)) + \dots + (\wp(t_n) + \mathcal{V}_\sigma(\perp) - \mathcal{V}_\sigma(t_n))\\
=\ &w(t_0, t_1) + w(t_1, t_2) + \dots + w(t_n, \perp).
\end{aligned}$$

Therefore, for any $s \in A$ we have that $\mathcal{V}_{I_\sigma}(s)$, i.e. the $\prec$-minimal value player 1 can guarantee to achieve in a play starting from s, has to be $\mathcal{V}_\sigma(s)$ plus the $\prec$-minimal value $\delta_\sigma(s)$ player 1 can guarantee starting from s in the edge-weighted graph $(A, E_{I_\sigma} \cap A \times A, w)$.

As $w(u,v) \succeq \text{ø}$, we can use Dijkstra's algorithm to find $\delta_\sigma(s)$ with the restriction that we only may add a node controlled by player 0 to the boundary in every step of Dijkstra's algorithm, if all successors of this node have already been evaluated. We then have $\delta_\sigma(s) = \mathcal{V}_{I_\sigma}(s) - \mathcal{V}_\sigma(s)$. □

B.4 Comparison with the Algorithm by Jurdzinski and Vöge

Theorem 5. *Let $\mathcal{A}_\perp$ be a escape-parity-game arena where every node of player 0 has at most two successor. Then the number of improvement steps needed to reach an optimal winning strategy is bound by $3 \cdot 1.724^{|V_0|}$.*

Proof. **Assumption 2.** *We assume that player* 0 *can only choose between at most two different successors in every state controlled by him, i.e.* $\forall v \in V_0 : |vE| \in \{1, 2\}$.

Let $(\sigma_\perp = \sigma_0) \prec \sigma_1 \prec \ldots \prec (\sigma_l = \hat{\sigma})$ be the sequence of strategies produced by the strategy-improvement algorithm presented in this article. As already shown, we may assume that σ_i is deterministic.

For σ_i let k_i be the number of nodes $s \in V_0$ such that there is at least one strict improvement of σ at s, i.e.

$$k_i := |\text{src}(S_{\sigma_i})| \text{ with } \text{src}(S_{\sigma_i}) := \{s \in V_0 \mid \exists(s,t) \in S_{\sigma_i}\}.$$

(Recall that S_σ is defined to be the set of strict improvements of a given strategy σ.)

Then there are at least $2^{k_i} - 1$ deterministic direct improvements σ' of σ_i with $\sigma_i \prec \sigma'$ and $\sigma' \setminus \sigma_i \subseteq S_{\sigma_i}$.[2]

We then have $\sigma_i \prec \sigma' \preceq \sigma_{i+1}$ for every such σ'. Now, as $\sigma_i \prec \sigma_{i+1}$, we know that every such σ' has not been considered in a previous step ($< i$) nor will it be considered in any following step ($> i$). Therefore, at least $2^{k_i} - 1$ new deterministic strategies can be ruled out as candidates for optimal winning strategies.

Hence, if S_k is the number of deterministic strategies which have at most k nodes at which there exists at least one strict improvement, we get as an upper bound for the number of improvement steps

$$S_k + \frac{2^{|V_0|}}{2^{k+1} - 1} \leq S_k + 2^{|V_0| - k}.$$

The next lemma bounds the number S_{k_i} of strategies σ_i having the same value for k_i:

Lemma 7. *Let* $(\sigma_i)_{0 \leq i \leq l} = \sigma_\perp = \sigma_0 \prec \sigma_1 \prec \ldots \prec \sigma_l = \hat{\sigma}$ *be the sequence of reasonable deterministic strategies generated by the strategy improvement algorithm.*

For an arena $\mathcal{A}_\perp$ *with* $|sE| \leq 2$ *for all* $s \in V_0$ *it holds that there are most* $\binom{|V_0|}{k'}$ *strategies in* $(\sigma_i)_{0 \leq i \leq l}$ *with* $|src(S_{\sigma_i})| = k'$.

Proof. First note the following easy fact: As along any edge $(s,t) \in \sigma$ holds, we have $\mathcal{V}_\sigma(s) \succeq \wp(s) + \mathcal{V}_\sigma(t)$ by definition of F_σ. Thus, for any strategy $\sigma \subseteq E_0$ of player 0 it holds that $S_\sigma \cap \sigma = \emptyset$.

Next, let σ_a and σ_b be two reasonable strategies of player 0 in $\mathcal{A}_\perp$. We claim that it holds that

(a) If $S_{\sigma_b} \cap \sigma_a = \emptyset$, we have $\sigma_a \preceq \sigma_b$.
(b) Assume that $|sE| \leq 2$ for all $s \in V_0$. If $\text{src}(S_{\sigma_b}) \subseteq \text{src}(S_{\sigma_a})$, it holds that $\sigma_a \preceq \sigma_b$.

Before given the proofs to these two claims, note that (b) already implies that we can have at most $\binom{|V_0|}{k'}$-many strategies σ_i with $k_i = k'$, as this is the number of disjoint subsets of V_0 with k' distinct elements.

In order to show (b), we first need to show (a): (a) Let $\mathcal{A}'_\perp$ be the arena resulting from $\mathcal{A}_\perp$ by removing all strict improvements of σ_b from E, i.e. $E' = E \setminus S_{\sigma_b}$. Both σ_a and σ_b are reasonable strategies of player 0 in $\mathcal{A}'_\perp$, as we only remove edges and these

[2] Note that we do not claim that σ_{i+1} is one of these strategies σ'.

edges are neither used by σ nor by σ'. This also means that the operators F_σ and $F_{\sigma'}$ stay unchanged, implying that the valuations of σ (reps. σ') on $\mathcal{A}_\perp$ and $\mathcal{A}'_\perp$ coincide. But as σ_b has no strict improvements in $\mathcal{A}'_\perp$, it has to hold that σ_b is an optimal winning strategy in $\mathcal{A}'_\perp$, meaning that $\sigma_a \preceq \sigma_b$ (cf. lemma 6).

(b) Set $C = S_{\sigma_b} \cap \sigma_a$. For every $s \in \text{src}(C)$ we find a t_C such that $(s, t_s^C) \in C$, a $t_s^{\sigma_b}$ with $(s, t_s^{\sigma_b}) \in \sigma_b$ (as σ_b is a strategy), and a $t_s^{S_{\sigma_a}}$ with $(s, t_s^{S_{\sigma_a}}) \in S_{\sigma_a}$ (as $\text{src}(S_{\sigma_b}) \subseteq \text{src}(S_{\sigma_a})$).

Now, because of $S_\sigma \cap \sigma = \emptyset$ for any strategy σ, we may conclude that $t^C \neq t_s^{\sigma_b}$, and $t^C \neq t_s^{S_{\sigma_a}}$ for all $s \in \text{src}(C)$. Thus, as we assume that $|sE| \leq 2$, it has to hold that $t_s^{S_{\sigma_a}} = t_s^{\sigma_b}$ for all $s \in \text{src}(C)$. We define therefore $C' = \{(s, t_s^{\sigma_b}) \mid s \in \text{src}(C)\}$, and

$$\sigma' := C' \cup \sigma_a \setminus C.$$

As $C' \subseteq S_{\sigma_a}$, we have $\sigma_a \preceq \sigma'$. Further $\sigma' \preceq \sigma_b$, as $\sigma' \cap S_{\sigma_b} = \emptyset$. □

The last lemma can be found in [10] for Markov decision processes.

As long as $1 \leq k \leq \frac{|V_0|}{3}$, we have

$$S_k \leq \sum_{k'=0}^{k} \binom{|V_0|}{k'} \leq 2\binom{|V_0|}{k} \leq 2\left(\frac{|V_0|}{k} \cdot e\right)^k.$$

What remains is to find a $1 \leq k \leq \frac{|V_0|}{3}$ such that

$$2\left(\frac{|V_0|}{k} \cdot e\right)^k + 2^{|V_0|-k}$$

is minimal. For this set $b = \frac{|V_0|}{k}$ with $b \geq 3$, yielding

$$2 \cdot e^{|V_0| \cdot \frac{1+\ln b}{b}} + e^{\ln 2 \cdot |V_0| \cdot \frac{b-1}{b}}.$$

As $\frac{1+\ln b}{b}$ is strictly decreasing and $\frac{b-1}{b}$ is strictly increasing, we need to look for the largest $b \geq 3$ such that

$$\frac{1+\ln b}{b} \geq \ln 2 \cdot \frac{b-1}{b}.$$

Using e.g. Newton's method one can easily check that $b \in (4.6, 4.7)$ with $b \approx 4.66438$. We therefore get

$$3 \cdot e^{0.545 \cdot |V_0|} \leq 3 \cdot 1.724^{|V_0|} \leq 3 \cdot 1.313^{|V|}$$

as an alternative upper bound for the number of improvement steps for an arena with out-degree two [3].

□

[3] Using a more detailed analysis in the spirit of [1] one can even show an upper bound of $O(1.71^{|V_0|})$.