

# AN EXPONENTIAL LOWER BOUND FOR THE LATEST DETERMINISTIC STRATEGY ITERATION ALGORITHMS

OLIVER FRIEDMANN  
UNIVERSITY OF MUNICH

*e-mail address:* OLIVER.FRIEDMANN@GMAIL.COM

---

**ABSTRACT.** This paper presents a new exponential lower bound for the two most popular deterministic variants of the strategy improvement algorithms for solving parity, mean payoff, discounted payoff and simple stochastic games. The first variant improves every node in each step maximizing the current valuation locally, whereas the second variant computes the globally optimal improvement in each step. We outline families of games on which both variants require exponentially many strategy iterations.

## 1. INTRODUCTION

In this paper, we study lower bounds for strategy improvement algorithms for solving parity games, mean payoff games, discounted payoff games as well as simple stochastic games. These are two-player games of perfect information played on directed graphs, and are related by a chain of polynomial-time reductions.

Parity games can be reduced to mean payoff games [Pur95], mean payoff games to discounted payoff games, and the latter ones to simple stochastic games [ZP96]. Solving games of any of these classes is one of the few combinatorial problems that belongs to the complexity class  $NP \cap coNP$  and that is not (yet) known to belong to  $P$  [EJS93, Con92]. It has also been shown that solving parity games as well as mean and discounted payoff games belongs to  $UP \cap coUP$  [Jur98].

We mainly consider parity games in this paper. They are played on a directed graph that is partitioned into two node sets associated with the two players; the nodes are labeled with natural numbers, called priorities. A play in a parity game is an infinite sequence of nodes whose winner is determined by the parity of the highest priority that occurs infinitely often, giving parity games their name.

The reason why parity games seem to be the most appropriate class of games, when trying to construct a worst-case family for one of the four classes, is that the effect of each node in a parity game is very clear: a higher priority dominates all lower priorities (in a play), no matter how many there are. By showing that the strategy iteration on our family of parity games directly corresponds to the strategy iteration that solves the other classes of games, we get the lower bounds for these by applying the standard reductions to our games.

Parity games occur in several fields of theoretical computer science, e.g. as solution to the problem of emptiness of tree automata [GTW02, EJ91] or as algorithmic backend to the model checking problem of the modal  $\mu$ -calculus [EJS93, Sti95].

There are many algorithms that solve parity games, such as the recursive decomposing algorithm due to Zielonka [Zie98] and its recent improvement by Jurdziński, Paterson and Zwick [JPZ06], the small progress measures algorithm due to Jurdziński [Jur00] with its recent improvement by Schewe [Sch07], the model-checking algorithm due to Stevens and Stirling [SS98] and finally the two strategy improvement algorithms by Vöge and Jurdziński [VJ00] and Schewe [Sch08].

All mentioned algorithms except for the two deterministic subexponential algorithms [JPZ06, Sch07] and except for the two strategy improvement algorithms have been shown to have a super-polynomial or exponential worst-case runtime complexity at best [Jur00, Fri10b, Fri10a]. The currently best known upper bound on the deterministic solution of parity games is  $\mathcal{O}(|E| \cdot |V|^{\frac{1}{3}|\text{ran}\Omega|})$  due to Schewe’s big-step algorithm [Sch07].

The *strategy improvement*, *strategy iteration* or *policy iteration* technique is the most general approach that can be applied as a solving procedure for all of these game classes. It was introduced by Howard [How60] for solving problems on Markov decision processes and has been adapted by several other authors for solving nonterminating stochastic games [HK66], simple stochastic games [Con92], discounted and mean payoff games [Pur95, ZP96] as well as parity games [VJ00].

Strategy iteration is an algorithmic scheme that is parameterized by an *improvement policy* which basically defines how to select a successor strategy in the iteration process. There are two major kinds of improvement policies: deterministic and randomized approaches; we will investigate deterministic approaches in this paper.

For discounted payoff games, there is the deterministic algorithm due to Puri [Pur95] that can also be used to solve mean payoff games as well as parity games by reduction [ZP96, VJ00]. Vöge and Jurdziński’s improvement algorithm is a refined version of Puri’s on parity games that omits the use of high-precision rational numbers; there are at least two reasonable improvement policies for the Vöge-Jurdziński procedure appearing in the literature such as the standard *locally optimizing policy* and Schewe’s *globally optimizing policy*.

An example has been known for some time for which a sufficiently poor choice of a single-switch policy causes an exponential number of iterations of the strategy improvement algorithm [BV07], but there have been no games known so far on which the policies due to Vöge/Jurdziński or Schewe require more than linearly many iterations.

In this paper, we particularly investigate the locally optimizing policy – which is, by far, the most natural choice for a multi-switching improvement policy – for solving parity games as it is applied by default in the original paper of Vöge and Jurdziński. We present a family of games comprising a linear number of nodes and a quadratic number of edges such that the strategy improvement algorithm using this policy requires an exponential number of iterations on them. We explain how these games can be refined in such a way that they only comprise a linear number of edges resulting in an undeniable exponential lower bound. Additionally, we describe what parts of the games have to be altered in order to get a family that results in exponentially many iterations when solved by Schewe’s strategy improvement algorithm.

Finally, we show that the parity game strategy iteration on our games directly corresponds to the strategy iteration that solves the associated mean payoff, discounted payoff as well as simple stochastic games, resulting in an exponential lower bound for the standard strategy improvement algorithms for all of these game classes.

Section 2 defines the basic notions of parity games and some notations that are employed throughout the paper. Section 3 recaps the strategy improvement algorithm by Vöge and Jurdziński; we define the two considered improvement policies in Section 4. In Section 5, we define a subclass of parity games called sink games that allows us to relate the lower bounds for parity games to the

other games classes. Section 6 presents a family of games on which the locally improving algorithm requires an exponential number of iterations. We discuss some improvements of the family in Section 7. In Section 8, we consider the modifications that have to be applied to our construction to obtain a lower bound for the globally optimizing policy. In Section 9, we show how to transfer the lower bounds to mean payoff, discounted payoff and simple stochastic games.

## 2. PARITY GAMES

A *parity game* is a tuple  $G = (V, V_0, V_1, E, \Omega)$  where  $(V, E)$  forms a directed graph whose node set is partitioned into  $V = V_0 \cup V_1$  with  $V_0 \cap V_1 = \emptyset$ , and  $\Omega : V \rightarrow \mathbb{N}$  is the *priority function* that assigns to each node a natural number called the *priority* of the node. We assume the graph to be total, i.e. for every  $v \in V$  there is a  $w \in V$  s.t.  $(v, w) \in E$ .

In the following we will restrict ourselves to finite parity games. W.l.o.g. we assume  $\Omega$  to be injective, i.e. there are no two different nodes with the same priority.

We also use infix notation  $vEw$  instead of  $(v, w) \in E$  and define the set of all *successors* of  $v$  as  $vE := \{w \mid vEw\}$ . The size  $|G|$  of a parity game  $G = (V, V_0, V_1, E, \Omega)$  is defined to be the cardinality of  $E$ , i.e.  $|G| := |E|$ ; since we assume parity games to be total w.r.t.  $E$ , this is a reasonable way to measure the size.

The game is played between two players called 0 and 1: starting in a node  $v_0 \in V$ , they construct an infinite path through the graph as follows. If the construction so far has yielded a finite sequence  $v_0 \dots v_n$  and  $v_n \in V_i$  then player  $i$  selects a  $w \in v_n E$  and the play continues with  $v_0 \dots v_n w$ .

Every play has a unique winner given by the *parity* of the greatest priority that occurs infinitely often. The winner of the play  $v_0 v_1 v_2 \dots$  is player  $i$  iff  $\max\{p \mid \forall j \in \mathbb{N} \exists k \geq j : \Omega(v_k) = p\} \equiv_2 i$  (where  $i \equiv_k j$  holds iff  $|i - j| \bmod k = 0$ ). That is, player 0 tries to make an even priority occur infinitely often without any greater odd priorities occurring infinitely often, player 1 attempts the converse.

We depict parity games as directed graphs where nodes owned by player 0 are drawn as circles and nodes owned by player 1 are drawn as rectangles; all nodes are labeled with their respective priority, and – if needed – with their name.

A *strategy* for player  $i$  is a – possibly partial – function  $\sigma : V^*V_i \rightarrow V$ , s.t. for all sequences  $v_0 \dots v_n$  with  $v_{j+1} \in v_j E$  for all  $j = 0, \dots, n-1$ , and all  $v_n \in V_i$  we have:  $\sigma(v_0 \dots v_n) \in v_n E$ . A play  $v_0 v_1 \dots$  *conforms* to a strategy  $\sigma$  for player  $i$  if for all  $j \in \mathbb{N}$  we have: if  $v_j \in V_i$  then  $v_{j+1} = \sigma(v_0 \dots v_j)$ . Intuitively, conforming to a strategy means to always make those choices that are prescribed by the strategy. A strategy  $\sigma$  for player  $i$  is a *winning strategy* in node  $v$  if player  $i$  wins every play that begins in  $v$  and conforms to  $\sigma$ .

A strategy  $\sigma$  for player  $i$  is called *positional* if for all  $v_0 \dots v_n \in V^*V_i$  and all  $w_0 \dots w_m \in V^*V_i$  we have: if  $v_n = w_m$  then  $\sigma(v_0 \dots v_n) = \sigma(w_0 \dots w_m)$ . That is, the choice of the strategy on a finite path only depends on the last node on that path.

With  $G$  we associate two sets  $W_0, W_1 \subseteq V$ ;  $W_i$  is the set of all nodes  $v$  s.t. player  $i$  wins the game  $G$  starting in  $v$ . Here we restrict ourselves to positional strategies because it is well-known that a player has a (general) winning strategy iff she has a positional winning strategy for a given game. In fact, parity games enjoy positional determinacy meaning that for every node  $v$  in the game either  $v \in W_0$  or  $v \in W_1$  [EJ91]. Furthermore, it is not difficult to show that, whenever player  $i$  has winning strategies  $\sigma_v$  for all  $v \in U$  for some  $U \subseteq V$ , then there is also a single strategy  $\sigma$  that is winning for player  $i$  from every node in  $U$ .

The problem of solving a parity game is to compute  $W_0$  and  $W_1$  as well as corresponding winning strategies  $\sigma_0$  and  $\sigma_1$  for the players on their respective winning regions.

A strategy  $\sigma$  for player  $i$  induces a *strategy subgame*  $G|_\sigma := (V, V_0, V_1, E|_\sigma, \Omega)$  where  $E|_\sigma := \{(u, v) \in E \mid u \in \text{dom}(\sigma) \Rightarrow \sigma(u) = v\}$ . Such a subgame  $G|_\sigma$  is basically the same game as  $G$  with the restriction that whenever  $\sigma$  provides a strategy decision for a node  $u \in V_i$ , all transitions from  $u$  but  $\sigma(u)$  are no longer accessible. The set of strategies for player  $i$  is denoted by  $\mathcal{S}_i(G)$ .

### 3. STRATEGY IMPROVEMENT

We briefly recap the basic definitions of the strategy improvement algorithm. For a given parity game  $G = (V, V_0, V_1, E, \Omega)$ , the *reward* of node  $v$  is defined as follows:  $\text{rew}_G(v) := \Omega(v)$  if  $\Omega(v) \equiv_2 0$  and  $\text{rew}_G(v) := -\Omega(v)$  otherwise. The set of *even resp. odd priority nodes* is defined to be  $V_\oplus := \{v \in V \mid \Omega(v) \equiv_2 0\}$  resp.  $V_\ominus := \{v \in V \mid \Omega(v) \equiv_2 1\}$ .

The *relevance ordering*  $<$  on  $V$  is induced by  $\Omega$ :  $v < u : \iff \Omega(v) < \Omega(u)$ ; additionally one defines the *reward ordering*  $\prec$  on  $V$  by  $v \prec u : \iff \text{rew}_G(v) < \text{rew}_G(u)$ . Note that both orderings are total due to injectivity of the priority function.

Let  $\pi$  be a path  $\sigma$  be a strategy for player  $i$ . We say that  $\pi$  *conforms* to  $\sigma$  iff for every  $j$  with  $\pi(j) \in V_i$  we have  $\sigma(\pi(j)) = \pi(j+1)$ .

Let  $v$  be a node,  $\sigma$  be a positional player 0 strategy and  $\tau$  be a positional player 1 strategy. Starting in  $v$ , there is exactly one path  $\pi_{\sigma, \tau, v}$  that conforms to  $\sigma$  and  $\tau$ . Since  $\sigma$  and  $\tau$  are positional strategies, this path can be uniquely written as follows.

$$\pi_{\sigma, \tau, v} = v_1 \dots v_k (w_1 \dots w_l)^\omega$$

with  $v_1 = v$ ,  $v_i \neq w_1$  for all  $1 \leq i \leq k$  and  $\Omega(w_1) > \Omega(w_j)$  for all  $1 < j \leq l$ . Note that the uniqueness follows from the fact that all nodes on the cycle have different priorities and we choose  $w_1$  to be the node with highest priority.

Discrete strategy improvement relies on a more abstract description of such a play  $\pi_{\sigma, \tau, v}$ . In fact, we only consider the *dominating cycle node*  $w_1$ , the set of *more relevant nodes* – i.e. all  $v_i > w_1$  – on the path to the cycle node, and the *length*  $k$  of the path leading to the cycle node. More formally, the *node valuation* of  $v$  w.r.t.  $\sigma$  and  $\tau$  is defined as follows.

$$\vartheta_{\sigma, \tau, v} := (w_1, \{v_i > w_1 \mid 1 \leq i \leq k\}, k)$$

Given a node valuation  $\vartheta$ , we refer to  $w_1$  as the *cycle component*, to  $\{v_i > w_1 \mid 1 \leq i \leq k\}$  as the *path component*, and to  $k$  as the *length component* of  $\vartheta$ .

In order to compare node valuations with each other, we introduce a total ordering on the set of node valuations. For that reason, we need to define a total ordering  $\prec$  on the second component of node valuations – i.e. on subsets of  $V$  – first. To compare two different sets  $M$  and  $N$  of nodes, we order all nodes lexicographically w.r.t. to their relevance and consider the first position in which the two lexicographically ordered sets differ, i.e. there is a node  $v \in M$  and a node  $w \in N$  with  $v \neq w$  s.t.  $u \in M$  iff  $u \in N$  for all  $u > v$  and all  $u > w$ . Now  $N$  is better than  $M$  iff  $v \prec w$ , i.e. the set which gives the higher reward in the first differing position is superior to the other set.

In other words, to determine which set of nodes is better w.r.t.  $\prec$ , one considers the node with the highest priority that occurs in only one of the two sets. The set owning that node is greater than the other if and only if that node has an even priority. More formally:

$$M \prec N : \iff M \Delta N \neq \emptyset \text{ and } \max_{\prec} (M \Delta N) \in ((N \cap V_\oplus) \cup (M \cap V_\ominus))$$

where  $M \Delta N$  denotes the symmetric difference of both sets.

Now we are able to extend the total ordering on sets of nodes to node valuations. The motivation behind this ordering is a lexicographic measurement of the profitability of a positional play w.r.t. player 0: the most prominent part of a positional play is the cycle in which the plays eventually stays, and here it is the reward ordering on the dominating cycle node that defines the profitability for player 0. The second important part is the loopless path that leads to the dominating cycle node. Here, we measure the profitability of a loopless path by a *lexicographic* ordering on the *relevancy* of the nodes on path, applying the *reward* ordering on each component in the lexicographic ordering. Finally, we consider the length, and the intuition behind the definition is that, assuming we have an even-priority dominating cycle node, it is better to reach the cycle fast whereas it is better to stay as long as possible out of the cycle otherwise. More formally:

$$(u, M, e) \prec (v, N, f) : \iff \begin{cases} (u \prec v) \text{ or } (u = v \text{ and } M \prec N) \text{ or} \\ (u = v \text{ and } M = N \text{ and } e < f \text{ and } u \in V_{\ominus}) \text{ or} \\ (u = v \text{ and } M = N \text{ and } e > f \text{ and } u \in V_{\oplus}) \end{cases}$$

Given a player 0 strategy  $\sigma$ , it is our goal to find a best response counterstrategy  $\tau$  that minimizes the associated node valuations. A strategy  $\tau$  is an *optimal counterstrategy* w.r.t.  $\sigma$  iff for every opponent strategy  $\tau'$  and for every node  $v$  we have:  $\vartheta_{\sigma, \tau, v} \preceq \vartheta_{\sigma, \tau', v}$ .

It is well-known that an optimal counterstrategy always exists and that it is efficiently computable.

**Lemma 3.1** ([VJ00]). Let  $G$  be a parity game and  $\sigma$  be a player 0 strategy. An optimal counterstrategy for player 1 w.r.t.  $\sigma$  exists and can be computed in polynomial time.

A fixed but arbitrary optimal counterstrategy will be denoted by  $\tau_{\sigma}$  from now on. The associated *game valuation*  $\Xi_{\sigma}$  is a map that assigns to each node the node valuation w.r.t.  $\sigma$  and  $\tau_{\sigma}$ :

$$\Xi_{\sigma} : v \mapsto \vartheta_{\sigma, \tau_{\sigma}, v}$$

Game valuations are used to measure the performance of a strategy of player 0: for a fixed strategy  $\sigma$  of player 0 and a node  $v$ , the associated valuation essentially states which is the worst cycle that can be reached from  $v$  conforming to  $\sigma$  as well as the worst loopless path leading to that cycle (also conforming to  $\sigma$ ).

We also write  $v \prec_{\sigma} u$  to compare the  $\Xi_{\sigma}$ -valuations of two nodes, i.e. to abbreviate  $\Xi_{\sigma}(v) \prec \Xi_{\sigma}(u)$ .

A run of the strategy improvement algorithm can be expressed by a sequence of *improving* game valuations; a partial ordering on game valuations is quite naturally defined as follows:

$$\Xi \triangleleft \Xi' : \iff (\Xi(v) \preceq \Xi'(v) \text{ for all } v \in V) \text{ and } (\Xi \neq \Xi')$$

A valuation  $\Xi_{\sigma}$  can be used to create a new strategy of player 0. The strategy improvement algorithm is only allowed to select new strategy decisions for player 0 occurring in the *improvement arena*  $\mathcal{A}_{G, \sigma} := (V, V_0, V_1, E', \Omega)$  where

$$vE'u : \iff \\ vEu \text{ and } (v \in V_1 \text{ or } (v \in V_0 \text{ and } \sigma(v) \preceq_{\sigma} u))$$

Thus all edges performing worse than the current strategy are removed from the game. A strategy  $\sigma$  is *improvable* iff there is a node  $v \in V_0$ , a node  $u \in V$  with  $vE'u$  s.t.  $\sigma(v) \prec_{\sigma} u$ .

An *improvement policy* now selects a strategy for player 0 in a given improvement arena. More formally: an improvement policy is a map  $\mathcal{I}_G : \mathcal{S}_0(G) \rightarrow \mathcal{S}_0(G)$  fulfilling the following two conditions for every strategy  $\sigma$ .

- (1) For every node  $v \in V_0$  it holds that  $(v, \mathcal{I}_G(\sigma)(v))$  is an edge in  $\mathcal{A}_{G,\sigma}$ .
- (2) If  $\sigma$  is improvable then there is a node  $v \in V_0$  s.t.  $\sigma(v) \prec_\sigma \mathcal{I}_G(\sigma)(v)$ .

We say that an edge  $(v, u)$  is an *improving edge* w.r.t.  $\sigma$  iff  $v \in V_0$ ,  $u \in vE$ ,  $\sigma(v) \neq u$  and  $\sigma(v) \prec_\sigma u$ .

Jurdziński and Vöge proved in their work that every strategy that is improved by an improvement policy can only result in strategies with valuations strictly better than the valuation of the original strategy.

**Theorem 3.2** ([VJ00]). Let  $G$  be a parity game,  $\sigma$  be an improvable strategy and  $\mathcal{I}_G$  be an improvement policy. We have  $\Xi_\sigma \triangleleft \Xi_{\mathcal{I}_G(\sigma)}$ .

If a strategy is not improvable, the strategy iteration procedure comes to an end and the winning sets for both players as well as associated winning strategies can be easily derived from the given valuation.

**Theorem 3.3** ([VJ00]). Let  $G$  be a parity game and  $\sigma$  be a non-improvable strategy. Then the following holds:

- (1)  $W_0 = \{v \mid \Xi_\sigma(v) = (w, -, -) \text{ and } w \in V_\oplus\}$
- (2)  $W_1 = \{v \mid \Xi_\sigma(v) = (w, -, -) \text{ and } w \in V_\ominus\}$
- (3)  $\sigma$  is a winning strategy for player 0 on  $W_0$
- (4)  $\tau_\sigma$  is a winning strategy for player 1 on  $W_1$
- (5)  $\sigma$  is  $\preceq$ -optimal

The strategy iteration starts with an initial strategy  $\iota_G$  and runs for a given improvement policy  $\mathcal{I}_G$  as outlined in the pseudo-code of Algorithm 1.

---

**Algorithm 1** Strategy Iteration

---

- 1:  $\sigma \leftarrow \iota_G$
  - 2: **while**  $\sigma$  is improvable **do**
  - 3:      $\sigma \leftarrow \mathcal{I}_G(\sigma)$
  - 4: **end while**
  - 5: **return**  $W_0, W_1, \sigma, \tau$  as in Theorem 3.3
- 

#### 4. IMPROVEMENT POLICIES

There are two major deterministic improvement policies that we consider here, namely the *locally optimizing policy* due to Jurdziński and Vöge [VJ00] and the *globally optimizing policy* by Schewe [Sch08].

The *locally optimizing policy*  $\mathcal{I}_G^{1\text{oc}}$  selects a most profitable strategy decision in every point with respect to the current valuation. More formally, it holds for every strategy  $\sigma$ , every player 0 node  $v$  and every  $w \in vE$  that  $w \preceq_\sigma \mathcal{I}_G^{1\text{oc}}(\sigma)(v)$ .

**Lemma 4.1** ([VJ00]). The locally optimizing policy can be computed in polynomial time.

This policy is generally considered to be the most natural choice, particularly because it directly corresponds to the canonical versions of strategy iteration in related parts of game theory like discounted payoff games or simple stochastic games. We will present a family of games on which the algorithm parameterized with this policy requires exponentially many iterations.

The globally optimizing policy  $\mathcal{I}_G^{\text{glo}}$  on the other hand computes a globally optimal successor strategy in the sense that the associated valuation is the best under all allowed successor strategies. More formally, given a parity game  $G$ , an improvable strategy  $\sigma$  and the improved strategy  $\sigma^* = \mathcal{I}_G^{\text{glo}}(\sigma)$ , we have for an arbitrary strategy  $\sigma'$  in the arena  $\mathcal{A}_{G,\sigma}$  that  $\Xi_{\sigma'} \leq \Xi_{\sigma^*}$ .

The policy can be interpreted as providing strategy improvement with a *one-step lookahead*; it computes the optimal strategy under all possible strategies that can be reached by a single improvement step.

The interested reader is pointed to Schewe's paper [Sch08] for all the details on how to effectively compute the optimal strategy update.

**Theorem 4.2** ([Sch08]). The globally optimizing policy can be computed in polynomial time.

We will also explain how to adapt the presented family of games in order to enforce exponentially many strategy iterations on them when parameterized with the globally optimal policy.

Vöge mentions without proof in his thesis that there is an improvement policy that requires at most  $|V|$  many iterations to find its fixed point. We find this fact to be quite remarkable and give a short proof of it in the following.

**Lemma 4.3** ([Vög00]). Let  $G$  be a parity game. There is an improvement policy  $\mathcal{I}_G^{\text{lin}}$  s.t. the strategy improvement algorithm requires at most  $|V|$  many iterations.

*Proof.* Let  $G = (V, V_0, V_1, E, \Omega)$  be a parity game and let  $\sigma^*$  be a  $\leq$ -optimal strategy. We define the improvement policy  $\mathcal{I}_G^{\text{lin}}$  as follows.

$$\mathcal{I}_G^{\text{lin}}(\sigma)(v) := \begin{cases} \sigma^*(v) & \text{if } (v, \sigma^*(v)) \in \mathcal{A}_{G,\sigma} \\ \sigma(v) & \text{otherwise} \end{cases}$$

We will show that  $\mathcal{I}_G^{\text{lin}}$  is indeed an improvement policy and that the strategy iteration parameterized with  $\mathcal{I}_G^{\text{lin}}$  requires at most  $|V_0|$  iterations on  $G$  in one go by verifying that

$$m(\sigma) \subsetneq V_0 \quad \implies \quad m(\sigma) \subsetneq m(\mathcal{I}_G^{\text{lin}}(\sigma))$$

for all  $\sigma$  where  $m(\sigma) = \{v \in V_0 \mid \sigma(v) = \sigma^*(v)\}$ .

Let  $\sigma$  be a strategy s.t.  $m(\sigma) \subsetneq V_0$ . Since  $m(\sigma) \subseteq m(\mathcal{I}_G^{\text{lin}}(\sigma))$  holds by definition, we simply need to show that there is at least one node  $v \in V_0$  with  $\sigma(v) \neq \sigma^*(v)$  and  $(v, \sigma^*(v)) \in \mathcal{A}_{G,\sigma}$ . Consider the game  $G' = (V, V_0, V_1, F, \Omega)$  where

$$F = \{(v, w) \in E \mid v \in V_1 \text{ or } (v, w) \in \sigma \text{ or } (v, w) \in \sigma^*\}$$

It is easy to see that  $\mathcal{A}_{G',\sigma} \subseteq \mathcal{A}_{G,\sigma}$  and also that  $\sigma^*$  is a  $\leq$ -optimal strategy w.r.t.  $G'$ . As  $\sigma$  is not optimal, there must be at least one proper improvement edge  $(v, w) \in \mathcal{A}_{G',\sigma}$ . By definition of  $G'$ , it follows that  $\sigma(v) \neq w$  and  $\sigma^*(v) = w$ .  $\square$

One may be misled to combine the existence of an improvement policy  $\mathcal{I}_G^{\text{lin}}$  that enforces at most linearly many iterations with the existence of the improvement policy  $\mathcal{I}_G^{\text{glo}}$  that selects the optimal successor strategy in each iteration, in order to propose that  $\mathcal{I}_G^{\text{glo}}$  should also enforce linearly many iterations in the worst case.

The reason why this proposition is incorrect lies in the intransitivity of optimality of strategy updates. Although it is true that  $\mathcal{I}_G^{\text{lin}}(\sigma) \leq \mathcal{I}_G^{\text{glo}}(\sigma)$  for every strategy  $\sigma$ , this is not necessarily the case for iterated applications, i.e.  $\mathcal{I}_G^{\text{lin}}(\mathcal{I}_G^{\text{lin}}(\sigma)) \leq \mathcal{I}_G^{\text{glo}}(\mathcal{I}_G^{\text{glo}}(\sigma))$  does not necessarily hold for all strategies  $\sigma$ .

## 5. SINK GAMES

Every approach trying to construct a game family of polynomial size that requires super-polynomially many iterations to be solved by strategy iteration (no matter which policy the algorithm is parameterized with), needs to focus on the second component of game valuations: there are only linearly many different values for the first and third component while there are exponentially many for the second.

Particularly, as there are at most linearly many different cycle nodes that can occur in valuations during a run, there is no real benefit in actually using different cycle nodes. Hence our basic layout of a game exploiting exponential behavior consists of a complex structure leading to one single loop – the only cycle node that will occur in valuations (such structures can easily be identified by preprocessing, but obviously it is not very difficult to obfuscate the whole construction without really altering its effect on the strategy iteration). In this setting, the strategy iteration algorithm is just improving the paths leading to the cycle node.

More formally: we call a parity game  $G$  (in combination with an initial strategy  $\iota_G$ ) a *1-sink game* iff the following two properties hold:

- (1) *Sink Existence*: there is a node  $v^*$  (called the *1-sink* of  $G$ ) with  $v^*Ev^*$  and  $\Omega(v^*) = 1$  reachable from all nodes; also, there is no other node  $w$  with  $\Omega(w) \leq \Omega(v^*)$ .
- (2) *Sink Seeking*: for each player 0 strategy  $\sigma$  with  $\Xi_{\iota_G} \trianglelefteq \Xi_\sigma$  and each node  $w$  it holds that the cycle component of  $\Xi_\sigma(w)$  equals  $v^*$ .

Obviously, a 1-sink game is won by player 1. Note that comparing node valuations in a 1-sink game can be reduced to comparing the path components of the respective node valuations, for two reasons. First, the cycle component remains constant. Second, the path-length component equals the cardinality of the path component, because all nodes except the sink node are more relevant than the cycle node itself. In the case of a 1-sink game, we will therefore identify node valuations with their path component.

It is fairly easy to prove that a game is a 1-sink game indeed. One simply has to check that the sink existence property holds by looking at the graph, that the game is completely won by player 1, and that the 1-sink is the cycle component of all nodes of the initial strategy.

**Lemma 5.1.** Let  $G$  be a parity game fulfilling the sink existence property w.r.t.  $v^*$ .  $G$  is a 1-sink game iff  $G$  is completely won by player 1 (i.e.  $W_1 = V$ ) and for each node  $w$  it holds that the cycle component of  $\Xi_{\iota_G}(w)$  equals  $v^*$ .

*Proof.* The “only-if”-part is trivial. For the “if”-part, we need to show that the *sink seeking*-property holds. Let  $\sigma$  be a player 0 strategy with  $\Xi_{\iota_G} \trianglelefteq \Xi_\sigma$ ,  $w$  be an arbitrary node and  $u$  be the cycle component of  $\Xi_\sigma(w)$ . Due to the fact that  $G$  is completely won by player 1,  $u$  has to be of odd priority. Also, since  $\Xi_{\iota_G} \trianglelefteq \Xi_\sigma$ , it holds that  $\Omega(u) \leq \Omega(v^*)$  implying  $u = v^*$  by the *sink existence*-property. □

There is another reason why 1-sink games are an interesting subclass of parity games: we will see later that the strategy iteration on a discounted payoff game that has been induced by the canonic reduction from a 1-sink parity game, directly corresponds to the strategy iteration on the original 1-sink parity game. This connection between discounted payoff games and 1-sink parity games allows us to directly transfer the lower bound to discounted payoff games.



## 6. LOWER BOUND FOR THE LOCALLY OPTIMIZING POLICY

The lower bound construction for the locally optimizing policy is a family of 1-sink parity games that implement a binary counter. In order to reduce the overall complexity of the games, our construction relies on unbounded edge outdegree, yielding a quadratic number of edges in total. We will discuss in the next section how the number of edges can be reduced to a linear number and even how to get binary outdegree.

The implementation of the binary counter is based on a structure called *simple cycles* that allows us to encode a single bit state in a given strategy  $\sigma$ . By having  $n$  such simple cycles, we can represent every state of an  $n$ -bit binary counter. In order to allow strategy improvement the transitions of the binary counter, we need to embed the simple cycles in a more complicated structure called *cycle gadget*, connect the cycle gadgets of the different bits with each other, and with an additional structure called *deceleration lane*.

This section is organized as follows. First, we consider the three gadgets that will be used in our lower bound construction, namely *simple cycles*, the *deceleration lane* and *cycle gates*. Then, we present the full construction of our lower bound family and give a high-level description of strategy iteration on these games. Finally, we prove that strategy iteration on the games indeed follows the high-level description.

For the presentation of the gadgets, we assume the context of a 1-sink parity game. The labelings and priorities of the gadgets will match the final priorities of the lower bound family.

Gadgets consist of three kinds of nodes: *input nodes*, *output nodes* and *internal nodes*. Input nodes are nodes that will have incoming edges from outside of the gadget, output nodes will have outgoing edges to the outside of the gadget and internal nodes will not be directly connected to the outside of the gadget.

In the context of 1-sink game  $G$  and a strategy  $\sigma$ , we will sometimes say that a node  $v$  *reaches* a node  $w$  to denote the fact that  $w$  lies on the path  $\pi_{\sigma, \tau_{\sigma}, v}$ .

**6.1. Simple Cycles.** The binary counter will contain a representation of  $n$  bits that are realized by  $n$  instances of a gadget called a *cycle gate*. The most important part of a cycle gate is the *simple cycle* that we will introduce first. We fix some index  $i$  for the simple cycle gadget for the sake of this subsection in order to have consistent node labelings.

A simple cycle consists of one player 0 controlled internal node  $d_i$  that is connected to a set of external nodes  $D_i$  in the rest of the graph, and one player 1 controlled input node  $e_i$ . The node  $e_i$  itself is connected to  $d_i$  (therefore the name *simple cycle*) and to one output node  $h_i \notin D_i$ . We note that all  $e_i$  nodes are the *only* player 1 controlled nodes with real choices in the complete lower bound construction.

All priorities of the simple cycle are based on an odd priority  $p_i$ . Intuitively, the  $p_i$  is considered to be a very small priority compared to the priorities of the other nodes in the external graph that the simple cycle is connected to. We will implicitly assume this in the following.

See Figure 1 for a simple cycle of index 1 with  $p_1 = 3$ . The players, priorities and edges are described in Table 1.

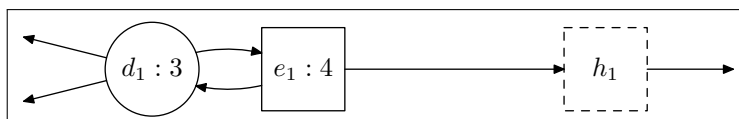


Figure 1: Simple Cycle

Node	Player	Priority	Successors
$d_i$	0	$p_i$	$\{e_i\} \cup D_i$
$e_i$	1	$p_i + 1$	$\{d_i, h_i\}$
$h_i$	?	$> p_i + 1$	?
$w \in D_i$	?	$> p_i + 1$	?

Table 1: Description of the Simple Cycle

Given a strategy  $\sigma$ , we say that the cycle is *closed* iff  $\sigma(d_i) = e_i$  and *open* otherwise. A closed cycle corresponds to a bit which is set while an open cycle corresponds to an unset bit.

The main idea now is to assign priorities to the simple cycle in such a way that the simple cycle is won by player 0, i.e. the most relevant node on the cycle needs to have an even priority. This has important consequences for the behaviour of the player 1 controlled node.

First, assume that  $\sigma(d_i) = e_i$ . The optimal counter-strategy here is  $\tau_\sigma(e_i) = h_i$ , since otherwise player 0 would win the cycle which is impossible with  $G$  being a 1-sink game. Player 0 is therefore able to *force* player 1 to move out of the cycle; in other words, *setting* a bit corresponds to *forcing* player 1 out of the cycle. In a set bit, the valuation of  $d_i$  is essentially the valuation of  $h_i$ , i.e.  $\Xi_\sigma(d_i) = \Xi_\sigma(h_i) \cup \{d_i, e_i\}$ .

Second, assume that  $\sigma(d_i) = w$  for some  $w \in D_i$ , and that  $w \prec_\sigma h_i$ . It follows that  $d_i \prec_\sigma h_i$ , hence  $\tau_\sigma(e_i) = d_i$ . The interesting part is now that  $\Xi_\sigma(e_i) = \Xi_\sigma(w) \cup \{d_i, e_i\}$ , i.e.  $e_i$  is an improving node for  $d_i$  (since  $\Xi_\sigma(w) \triangle \Xi_\sigma(e_i) = \{d_i, e_i\}$ ), but updating to  $e_i$  would yield a much greater reward than just  $\Xi_\sigma(e_i)$  (namely  $\Xi_\sigma(h_i) \cup \{e_i\}$  by forcing player 1 to leave the cycle).

Assume now that  $w' \in D_i$  with  $w \prec_\sigma w'$  but  $w' \prec_\sigma h_i$ . Obviously,  $w'$  and  $e_i$  are improving nodes for  $d_i$ , but  $e_i \prec_\sigma w'$ , hence by the locally improving policy, player 0 switches to  $w'$  although  $e_i$  might give a much better valuation. In other words, by moving to  $d_i$ , the player 1 node hides the fact that there is a highly profitable node on the other side.

We formalize the behaviour of the simple cycle in two lemmas. The first describes the valuation of  $e_i$  depending on the state of the simple cycle and the second explains the switching behaviour of the player 0 controlled node. The claimed result can easily be obtained by tracing the paths that the strategies take through the gadget, and then comparing valuations.

**Lemma 6.1.** Let  $\sigma$  be a strategy. The following holds:

- (1) If cycle  $i$  is closed, we have  $\tau_\sigma(e_i) = h_i$ .
- (2) If cycle  $i$  is open and  $h_i \prec_\sigma \sigma(d_i)$ , we have  $\tau_\sigma(e_i) = h_i$ .
- (3) If cycle  $i$  is open and  $\sigma(d_i) \prec_\sigma h_i$ , we have  $\tau_\sigma(e_i) = d_i$ .

**Lemma 6.2.** Let  $\sigma$  be a strategy and  $w = \max_{\prec_\sigma} D_i$ . Let  $\sigma' = \mathcal{I}^{\text{loc}}(\sigma)$ . The following holds:

- (1) If cycle  $i$  is closed and  $w \prec_\sigma h_i$ , we have cycle  $i$   $\sigma'$ -closed (“closed cycle remains closed”).
- (2) If cycle  $i$  is open,  $\sigma(d_i) \neq w$  or  $h_i \prec_\sigma w$ , we have  $\sigma'(d_i) = w$  (“open cycle remains open”).
- (3) If cycle  $i$  is open,  $\sigma(d_i) = w$  and  $w \prec_\sigma h_i$ , then cycle  $i$  is  $\sigma'$ -closed (“open cycle closes”).
- (4) If cycle  $i$  is closed and  $h_i \prec_\sigma w$ , we have  $\sigma'(d_i) = w$  (“closed cycle opens”).

Open simple cycles have the important property that we can postpone closing them by supplying them with new nodes  $w \in D_i$  in each iteration s.t.  $\sigma(d_i) \prec_\sigma w$ . We will use this property in the construction of our binary counter. Since we do not want to set all bits at the same time, rather one by one, we need to make sure that unset bits which are not supposed to be set remain unset for some time (more precisely, until the respective bit represents the least unset bit), and this will be realized

by this property. The device that supplies us with new best-valued external nodes in each iteration is called *deceleration lane* and will be described next.

**6.2. Deceleration Lane.** A *deceleration lane* has several, say  $m$ , input nodes and some output nodes, called *roots*. The lower bound construction will only require a deceleration lane with two roots  $s$  and  $r$ , however, it would be easy to generalize the construction of deceleration lanes to an arbitrary number of roots.

More formally, a deceleration lane consists of  $m$  internal nodes  $t_1, \dots, t_m$ , one additional internal node  $c$ ,  $m$  input nodes  $a_1, \dots, a_m$  and two output nodes  $s$  and  $r$ , called *roots* of the deceleration lane.

All priorities of the deceleration lane are based on some odd priority  $p$ . We assume that all root nodes have a priority greater than  $p + 2m + 1$ . See Figure 2 for a deceleration lane with  $m = 6$  and  $p = 15$ . The players, priorities and edges are described in Table 2.

Node	Player	Priority	Successors
$t_1$	0	$p$	$\{s, r, c\}$
$t_{i>1}$	0	$p + 2i - 2$	$\{s, r, t_{i-1}\}$
$c$	0	$p + 2m + 1$	$\{s, r\}$
$a_i$	1	$p + 2i - 1$	$\{t_i\}$
$s$	?	$> p + 2m + 1$	?
$r$	?	$> p + 2m + 1$	?

Table 2: Description of the Deceleration Lane

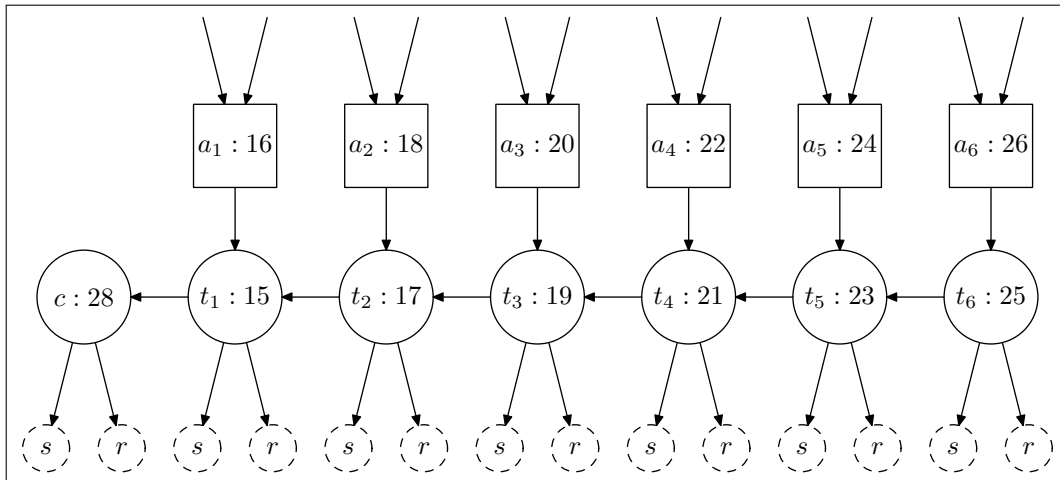


Figure 2: A Deceleration Lane (with  $m = 6$  and  $p = 15$ )

A deceleration lane serves the following purpose. Assume that one of the output nodes, say  $r$ , has the better valuation compared to the other root node, and assume further that this setting sustains for some iterations.

The input nodes, say  $a_1, \dots, a_m$ , now serve as an entry point, and all reach the best valued root  $-r-$  by some internal nodes. The valuation ordering of all input nodes depends on the iteration: at first,  $a_1$  has a better valuation than all other input nodes. Then,  $a_2$  has a better valuation than all other input nodes and so on.

This process continues until the other output node, say  $s$ , has a better valuation than  $r$ . Within the next iteration, the internal nodes perform a *resetting* step s.t. all input nodes eventually reach the new root node. One iteration after that,  $a_1$  has the best valuation compared to all other input nodes again.

In other words, by giving one of the roots, say  $s$ , a better valuation than another root, say  $r$ , it is possible to reset and therefore reuse the lane again. In fact, the lower bound construction will use a deceleration lane with two roots  $s$  and  $r$ , and will employ  $s$  only for resetting, i.e. after some iterations with  $r \succ_\sigma s$ , there will be one iteration with  $s \succ_\sigma r$  and right after that again  $r \succ_\sigma s$ .

From an abstract point of view, we describe the state of a deceleration lane by which of the two roots is chosen and by how many  $t_i$  nodes are already moving down to  $c$ . Formally, we say that  $\sigma$  is in *deceleration state*  $(x, j)$  (where  $x \in \{s, r\}$  and  $0 < j \leq m + 1$  a natural number) iff

- (1)  $\sigma(c) = x$ ,
- (2)  $\sigma(t_1) = c$  if  $j > 1$ ,
- (3)  $\sigma(t_i) = t_{i-1}$  for all  $1 < i < j$ , and
- (4)  $\sigma(t_i) = x$  for all  $j \leq i$ .

We say that the deceleration lane is *rooted in*  $x$  if  $\sigma$  is in state  $(x, *)$ , and that the *index is*  $i$  if  $\sigma$  is in state  $(*, i)$ . Whenever a strategy  $\sigma$  is in state  $(x, i)$ , we define  $root(\sigma) = x$  and  $ind(\sigma) = i$ . In this case, we say that the strategy is *well-behaved*.

We formalize the behaviour of the deceleration lane in two lemmas. The first describes the ordering of the valuations of the input nodes depending on the state  $(x, i)$  of the deceleration lane: (1) if the ordering of the root nodes changes, all input nodes have a worse valuation than the better root, and (2) otherwise the best valued input node is  $a_{i-1}$ . The second explains the switching behaviour of the player 0 controlled nodes: (1) if the ordering of the root node changes, than the whole lane resets, and (2) otherwise the lane assembles further, providing a new best-valued input node.

**Lemma 6.3.** Let  $\sigma$  be a strategy in deceleration state  $(x, i)$ . Let  $\bar{x}$  denote the other root. Then

- (1)  $x \prec_\sigma \bar{x}$  implies  $a_j \prec_\sigma \bar{x}$  for all  $j$  (“resetting results in unprofitable lane”).
- (2)  $\bar{x} \prec_\sigma x$  implies  $x \prec_\sigma a_i \prec_\sigma \dots \prec_\sigma a_m \prec_\sigma c \prec_\sigma a_1 \prec_\sigma \dots \prec_\sigma a_{i-1}$  (“new best-valued node in each iteration”).

**Lemma 6.4.** Let  $\sigma$  be a strategy that is in deceleration state  $(x, i)$ . Let  $\bar{x}$  denote the other root. Let  $\sigma' = \mathcal{I}^{1 \circ c}(\sigma)$ . Then

- (1)  $x \prec_\sigma \bar{x}$  implies that  $\sigma'$  is in state  $(\bar{x}, 1)$  (“lane resets”).
- (2)  $\bar{x} \prec_\sigma x$  implies that  $\sigma'$  is in state  $(x, \min(i, m) + 1)$  (“lane assembles one step at a time”).
- (3)  $\sigma'$  is well-behaved (“always ending up with well-behaved strategies”).

The main purpose of a deceleration lane is to absorb the update activity of other nodes in such a way that wise (i.e. edges that will result in much better valuations *after* switching and reevaluating) strategy updates are postponed. Consider a node for instance that has more than one proper improving switch; the locally optimizing policy will select the edge with the best valuation to be switched. In order to prevent that one particular improving switch is applied for some iterations, one can connect the node to the input nodes of the deceleration lane.

The particular scenario in which we will use the deceleration lane are *simple cycles* as described in the previous subsection. We will connect the simple cycles encoding the bits of our counter to the deceleration lane in such a way, that lower cycles have less edges entering the deceleration lane. This construction ensures that lower open cycles (representing unset bits) will close (i.e. set the corresponding bit) before higher open cycles (representing higher unset bits) have their turn to close.

**6.3. Cycle Gate.** The simple cycles will appear in a more complicated gadget, called *cycle gate*. We will have  $n$  different cycle gates in the game number  $n$  of the lower bound family, hence we fix some index  $i$  for the cycle gate gadget for the sake of this subsection in order to have consistent node labelings.

Formally, a cycle gate consists of two internal nodes  $e_i$  and  $h_i$ , two input nodes  $f_i$  and  $g_i$ , and two output nodes  $d_i$  and  $k_i$ . The output node  $d_i$  will be connected to a set of other nodes  $D_i$  in the game graph, and  $k_i$  to some other set  $K_i$  as well. The two nodes  $d_i$  and  $e_i$  form a simple cycle as described earlier.

All priorities of the cycle gate are based on two odd priorities  $p_i$  and  $p'_i$ . See Figure 3 for a cycle gate of index 1 with  $p'_1 = 3$  and  $p_1 = 33$ . The players, priorities and edges are described in Table 3.

Node	Player	Priority	Successors
$d_i$	0	$p'_i$	$\{e_i\} \cup D_i$
$e_i$	1	$p'_i + 1$	$\{d_i, h_i\}$
$g_i$	0	$p'_i + 3$	$\{f_i, k_i\}$
$k_i$	0	$p_i$	$K_i$
$f_i$	1	$p_i + 2$	$\{e_i\}$
$h_i$	1	$p_i + 3$	$\{k_i\}$

Table 3: Description of the Cycle Gate

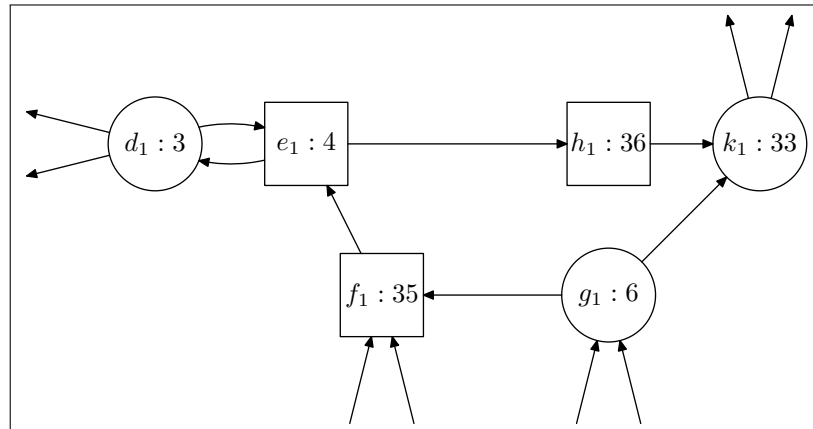


Figure 3: A Cycle Gate (index 1 with  $p'_1 = 3$  and  $p_1 = 33$ )

The main idea behind a cycle gate is to have a pass-through structure *controlled by the simple cycle* that is either very profitable or quite unprofitable. The pass-through structure of the cycle gate has one major input node, named  $g_i$ , and one major output node, named  $k_i$ . The input node is controlled by player 0 and connected via two paths with the output node; there is a direct edge and a longer path leading through the interior of the cycle gate.

However, the longer path only leads to the output node if the simple cycle, consisting of one player 0 node  $d_i$  and one player 1 node  $e_i$ , is *closed*. In this case, it is possible and profitable to reach the output node via the internal path; otherwise, this path is not accessible, and hence, the input node has to select the unprofitable direct way to reach the output node.

We will have one additional input node, named  $f_i$ , that can only access the path leading through the interior of the cycle gate, for the following purpose. Assume that the simple cycle has just been closed and now the path leading through the interior becomes highly profitable. Hence, the next switching event to happen will be the node  $g_i$  switching from the direct path to the path through the interior. However, it will be useful to be able to reach the highly profitable path from some parts of the outside graph one iteration before it is accessible via  $g_i$ . For this reason, we include an additional input node  $f_i$  that immediately accesses the interior path.

We say that a cycle gate is *closed* resp. *open* iff the interior simple cycle is closed resp. open. Similarly, we say that a cycle gate is *accessed* resp. *skipped* iff the access control node  $g_i$  moves through the interior ( $\sigma(g_i) = f_i$ ) resp. directly to  $k_i$ .

From an abstract point of view, we describe the state of a cycle gate by a pair  $(\beta_i(\sigma), \alpha_i(\sigma)) \in \{0, 1\}^2$ . The first component describes the state of the simple cycle, and the second component gives the state of the access control node. Formally, we have the following.

- (1)  $\beta_i(\sigma) = 1$  iff the  $i$ -th cycle gate is closed, and
- (2)  $\alpha_i(\sigma) = 1$  iff the  $i$ -th cycle gate is accessed.

We formalize the behaviour of the cycle gate in two lemmas. The first describes the valuation of all important nodes of the cycle gate, using our knowledge of simple cycles of Lemma 6.1. The second explains the switching behaviour of the access control node. The behaviour of the simple cycle contained in the cycle gate is described by Lemma 6.2.

**Lemma 6.5.** Let  $\sigma$  be a strategy. Then

- (1) If gate  $i$  is open, we have  $f_i \prec_\sigma \sigma(d_i)$ .
- (2) If gate  $i$  is closed, we have  $\sigma(k_i) \prec_\sigma f_i$ .
- (3) If gate  $i$  is closed and skipped, we have  $g_i \prec_\sigma f_i$ .
- (4) If gate  $i$  is accessed, we have  $f_i \prec_\sigma g_i$ .
- (5) If gate  $i$  is skipped, we have  $\sigma(k_i) \prec_\sigma g_i$ .

**Lemma 6.6.** Let  $\sigma$  be a strategy and  $\sigma' = \mathcal{I}^{1oc}(\sigma)$ .

- (1) If gate  $i$  is  $\sigma$ -closed, then gate  $i$  is  $\sigma'$ -accessed (“closed gates will be accessed”).
- (2) If gate  $i$  is  $\sigma$ -open and  $\sigma(d_i) \prec_\sigma h_i$ , then gate  $i$  is  $\sigma'$ -skipped (“open gates with unprofitable exit nodes will be skipped”).
- (3) If gate  $i$  is  $\sigma$ -open and  $h_i \prec_\sigma \sigma(d_i)$ , then gate  $i$  is  $\sigma'$ -accessed (“open gates with profitable exit nodes will be accessed”).

The last two items of Lemma 6.6 are based on the uniqueness of priorities in the game, implying that there are no priorities between  $f_i$  and  $h_i$ .

We will use cycle gates to represent the bit states of a binary counter: unset bits will correspond to cycle gates with the state  $(0, 0)$ , set bits to the state  $(1, 1)$ . Setting and resetting bits therefore traverses more than one phase, more precisely, from  $(0, 0)$  over  $(1, 0)$  to  $(1, 1)$ , and from the latter

again over  $(0, 1)$  to  $(0, 0)$ . Particularly, it can be observed that the second component of the cycle gate states switches one iteration after the first component in both cases.

**6.4. Lower Bound Construction.** In this subsection, we provide the complete construction of the lower bound family. It essentially consists of a 1-sink  $x$ , a deceleration lane of length  $2n$  that is connected to the two roots  $s$  and  $r$ , and  $n$  cycle gates. The simple cycles of the cycle gates are connected to the roots and to the deceleration lane with the important detail, that lower cycle gates have less edges to the deceleration lane. This construction ensures that lower open cycle gates will close before higher open cycle gates.

The output node of a cycle gate is connected to the 1-sink and to the  $g_*$ -input nodes of all higher cycle gates. The  $s$  root node is connected to all  $f_*$ -input nodes, the  $r$  root node is connected to all  $g_*$ -input nodes.

We now give the formal construction. The games are denoted by  $G_n = (V_n, V_{n,0}, V_{n,1}, E_n, \Omega_n)$ . The sets of nodes are

$$V_n := \{x, s, c, r\} \cup \{t_i, a_i \mid 1 \leq i \leq 2n\} \cup \{d_i, e_i, g_i, k_i, f_i, h_i \mid 1 \leq i \leq n\}$$

The players, priorities and edges are described in Table 4. The game  $G_3$  is depicted in Figure 4.

Node	Player	Priority	Successors
$t_1$	0	$4n + 3$	$\{s, r, c\}$
$t_{i>1}$	0	$4n + 2i + 1$	$\{s, r, t_{i-1}\}$
$a_i$	1	$4n + 2i + 2$	$\{t_i\}$
$c$	0	$8n + 4$	$\{s, r\}$
$d_i$	0	$4i + 1$	$\{s, e_i, r\} \cup \{a_j \mid j < 2i + 1\}$
$e_i$	1	$4i + 2$	$\{d_i, h_i\}$
$g_i$	0	$4i + 4$	$\{f_i, k_i\}$
$k_i$	0	$8n + 4i + 7$	$\{x\} \cup \{g_j \mid i < j \leq n\}$
$f_i$	1	$8n + 4i + 9$	$\{e_i\}$
$h_i$	1	$8n + 4i + 10$	$\{k_i\}$
$s$	0	$8n + 6$	$\{f_j \mid j \leq n\} \cup \{x\}$
$r$	0	$8n + 8$	$\{g_j \mid j \leq n\} \cup \{x\}$
$x$	1	1	$\{x\}$

Table 4: Lower Bound Construction for the Locally Optimizing Policy

**Fact 1.** The game  $G_n$  has  $10 \cdot n + 4$  nodes,  $1.5 \cdot n^2 + 20.5 \cdot n + 5$  edges and  $12 \cdot n + 8$  as highest priority. In particular,  $|G_n| = \mathcal{O}(n^2)$ .

As an initial strategy we select the following  $\iota_{G_n}$ . It will correspond to the global counter state in which no bit has been set.

$$\iota_{G_n}(t_1) = c \quad \iota_{G_n}(g_i) = k_i \quad \iota_{G_n}(* \in \{t_{i>1}, c, d_i\}) = r \quad \iota_{G_n}(* \in \{k_i, s, r\}) = x$$

Note that  $\iota_{G_n}$  particularly is well-behaved. Hence, by Lemma 6.4(3) we know that all strategies that will occur in a run of the strategy improvement algorithm will be well-behaved.

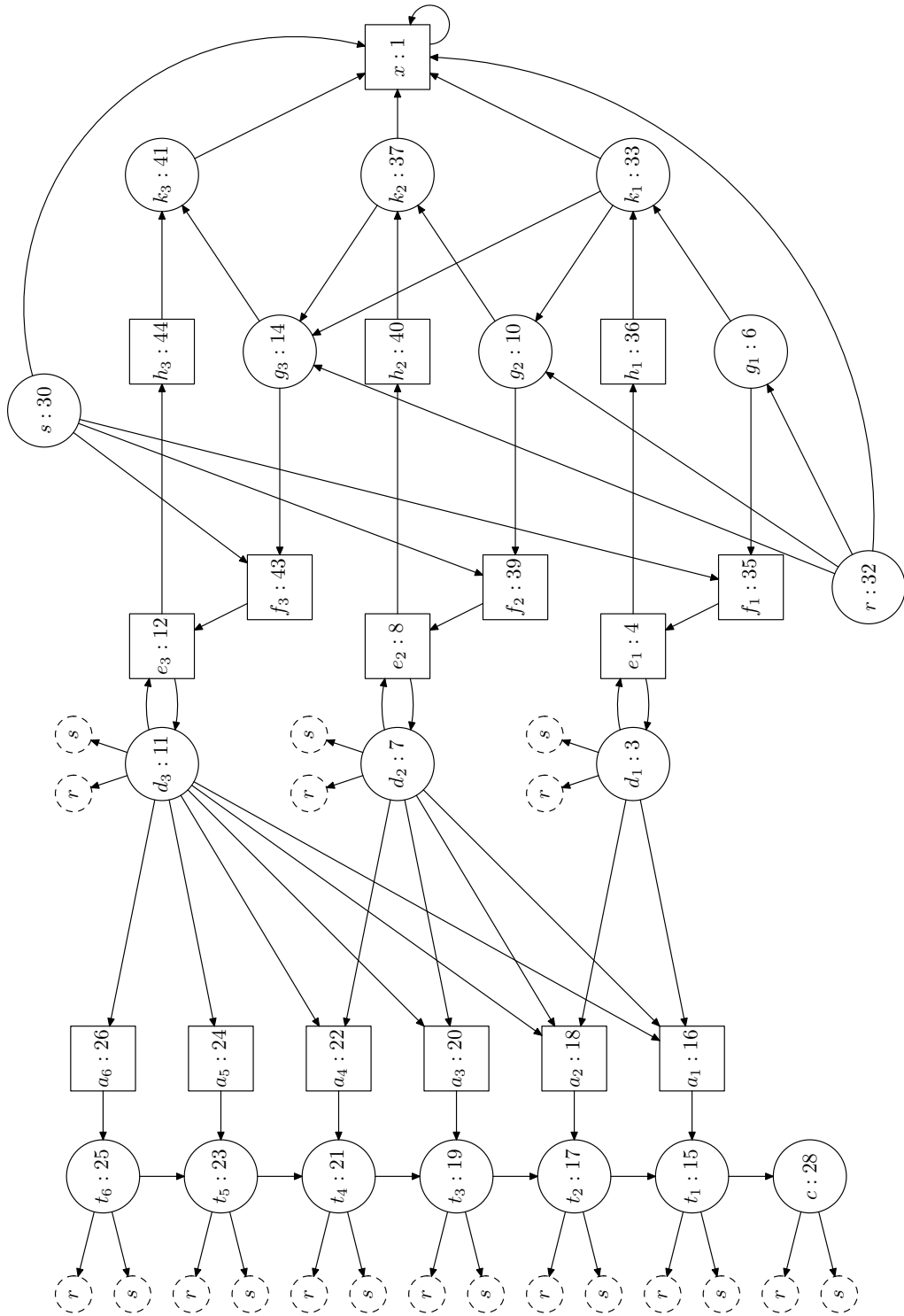


Figure 4: Locally Optimizing Lower Bound Game  $G_3$



We will see in the next section, how the family  $G_n$  can be refined in such a way that it only comprises a linear number of edges. The reason why we present the games with a quadratic number of edges first is that the refined family looks even more confusing and obfuscates the general principle.

**Lemma 6.7.** Let  $n > 0$ .

- (1) The game  $G_n$  is completely won by player 1.
- (2)  $x$  is the 1-sink of  $G_n$  and the cycle component of  $\Xi_{\iota_{G_n}}(w)$  equals  $x$  for all  $w$ .

*Proof.* Let  $n > 0$ .

- (1) Note that the only nodes owned by player 1 with an outdegree greater than 1 are  $e_1, \dots, e_n$ . Consider the player 1 strategy  $\tau$  which selects to move to  $h_i$  from  $e_i$  for all  $i$ . Now it is the case that  $G_n|_\tau$  contains exactly one cycle that is eventually reached no matter what player 0 does, namely the self-cycle at  $x$  which is won by player 1.
- (2) The self-cycle at  $x$  obviously is the 1-sink since it can be reached from all other nodes and has the smallest priority 1. Since  $xEx$  is the only cycle won by player 1 in  $G_n|_{\iota_{G_n}}$ ,  $x$  must be the cycle component of each node valuation w.r.t.  $\iota_{G_n}$ . □

By Lemma 5.1 it follows that  $G_n$  is a 1-sink game, hence it is safe to identify the valuation of a node with its path component from now on.

**6.5. Lower Bound Description and Phases.** Here, we describe how the binary counter performs the task of counting by strategy improvement. Our games implement a full binary counter in which every bit is represented by a simple cycle encapsulated in a cycle gate. An unset bit  $i$  corresponds to an open simple cycle in cycle gate  $i$ , a set bit  $i$  corresponds to a closed simple cycle in cycle gate  $i$ .

Formally, we represent the bit state of the counter by elements from  $B_n = \{0, 1\}^n$ . For  $b = (b_n, \dots, b_1) \in B_n$ , let  $b_i$  denote the  $i$ -th component in  $b$  for every  $i \leq n$ , where  $b_n$  denotes the most and  $b_1$  denotes the least significant bit. By  $b \oplus 1$ , we denote the increment of the number represented by  $b$  by 1. The least resp. greatest bit states are denoted by  $\mathbf{0}_n$  resp.  $\mathbf{1}_n$ . We refer to the least unset bit by  $\mu b := \min(\{n+1\} \cup \{i \leq n \mid b_i = 0\})$ , and similarly to the least set bit by  $\nu b := \min(\{n+1\} \cup \{i \leq n \mid b_i = 1\})$ .

From the most abstract point of view, our lower bound construction performs counting on  $B_n$ . However, the increment of a global bit state requires more than one strategy iteration, more precisely four different phases that will be described next (with one phase of dynamic length).

Every phase is defined w.r.t. a given global counter state  $b \in B_n$ . Let  $b \in B_n$  be a global bit state different from  $\mathbf{1}_n$ .

An abstract counter performs the increment from  $b$  to  $b \oplus 1$  by computing  $b[\mu b \mapsto 1][j < \mu b \mapsto 0]$ , i.e. by setting bit  $\mu b$  and by resetting all lower bits  $j < \mu b$ . In the context of the games, we start in phase 1 corresponding to  $b$ , and then proceed to phase 2 and phase 3 corresponding to  $b[\mu b \mapsto 1]$ , from phase 3 to phase 4 corresponding to  $b[\mu b \mapsto 1][j < \mu b \mapsto 0]$ , and finally from phase 4 to phase 1 again. The transition from phase 2 to phase 3 and from phase 4 to phase 1 handles the correction of the internal structure connecting the cycles with each other.

For the sake of this subsection, let  $\sigma$  be a strategy and  $b \in B_n$  be a global counter state. All phases will be defined w.r.t.  $\sigma$  and  $b \in B_n$ . Let  $\sigma' = \mathcal{I}^{\text{loc}}(\sigma)$ .

To keep everything as simple as possible and to be able to prove all the lemmas without considering special cases, we will assume that  $b$  is different from  $\mathbf{0}_n$  and that the two highest bits in  $b$

are zero and remain zero, i.e. we will only use the first  $n - 2$  bits for counting. Note however, that every bit works as intended in the counter.

Given a strategy  $\sigma$ , we denote the *associated simple cycle state*  $(\beta_n(\sigma), \dots, \beta_1(\sigma))$  by  $b_\sigma$ , and the *associated access state*  $(\alpha_n(\sigma), \dots, \alpha_1(\sigma))$  by  $a_\sigma$ .

Recall that every strategy  $\sigma$  occurring will be well-behaved. In addition to the deceleration lane and the cycle gates, we have two more structures that are controlled by a strategy  $\sigma$ , namely the two roots  $r$  and  $s$ , and the cycle gate output nodes  $k_i$ . We write  $\sigma(r) = i$  to denote that  $\sigma(r) = g_i$ , and  $\sigma(r) = n$  if  $\sigma(r) = x$ ; we write  $\sigma(s) = i$  to denote that  $\sigma(s) = f_i$ , and  $\sigma(s) = n$  if  $\sigma(s) = x$ ; we write  $\sigma(k_i) = j$  to denote that  $\sigma(k_i) = g_j$ , and  $\sigma(k_i) = n$  if  $\sigma(k_i) = x$ . We also use a more compact notation for the strategy decision of  $d_i$ -nodes of open cycles. We write  $\sigma(d_i) = j$  if  $\sigma(d_i) = a_j$ .

Recall that we say that a strategy  $\sigma$  is *rooted* in  $s$  or  $r$ , if every path in the deceleration lane conforming to  $\sigma$  eventually exits to  $s$  resp.  $r$ . Likewise, we say that  $\sigma$  has index  $i$  if all nodes of the deceleration lane with smaller index  $j < i$  are moving down the lane by  $\sigma$ , and that  $i$  is the first index which is directly exiting through the root.

The first phase, called the *waiting* phase, corresponds to a stable strategy  $\sigma$  in which open cycles are busy waiting to be closed while the deceleration lane is assembling. Cycle gates that correspond to set bits are closed and accessed, while cycle gates of unset bits are open and skipped, i.e.  $b = b_\sigma = a_\sigma$ . The *selector nodes*  $k_i$  move to the next higher cycle gate corresponding to a set bit, and both roots are connected to the least set bit  $\nu b$ .

More formally, we say that  $\sigma$  is a  $b$ -phase 1 strategy iff all the following conditions hold:

- (1)  $b = b_\sigma = a_\sigma$ , i.e. set bits correspond to closed and accessed cycle gates, while unset bits correspond to open and skipped cycle gates,
- (2)  $\text{root}(\sigma) = r$ , i.e. the strategy is rooted in  $r$ ,
- (3)  $\sigma(s) = \sigma(r) = \nu b$ , i.e. both roots are connected to the least set bit,
- (4)  $\sigma(k_i) = \min(\{j > i \mid b_j = 1\} \cup \{n + 1\})$ , i.e. the selector nodes move to the next set bit,
- (5)  $\text{ind}(\sigma) \leq 2\mu b + 2$ , i.e. the deceleration lane has not passed the least unset bit, and
- (6)  $\sigma(d_j) \neq \text{ind}(\sigma) - 1$  for all  $j$  with  $b_j = 0$ , i.e. every open cycle node is not connected to the best-valued node of the lane.

The only improving switches in the first phase are edges of open simple cycles and edges of the deceleration lane.

**Lemma 6.8.** Let  $\sigma$  be a  $b$ -phase 1 strategy with  $\text{ind}(\sigma) < 2\mu b + 2$ . Then  $\sigma'$  is a  $b$ -phase 1 strategy with  $\text{ind}(\sigma') = \text{ind}(\sigma) + 1$ , and if  $\text{ind}(\sigma) > 1$ , then  $\sigma'(d_{\mu b}) = \text{ind}(\sigma) - 1$ .

*Proof.* Let  $\sigma$  be a  $b$ -phase 1 strategy,  $\Xi := \Xi_\sigma$  and  $\text{ind}(\sigma) < 2\mu b + 2$ .

We first compute the valuations for all those nodes directly that do not involve any complicated strategy decision of player 1. Obviously,  $\Xi(x) = \emptyset$ . By Lemma 6.1(1) we know that for all set bits  $i$  (i.e.  $b_i = 1$ ) we have the following.

$$\Xi(e_i) = \{e_i\} \cup \Xi(h_i) \quad \Xi(d_i) = \{e_i, d_i\} \cup \Xi(h_i) \quad \Xi(f_i) = \{e_i, d_i, f_i\} \cup \Xi(h_i)$$

Using these equations, we are able to compute many other valuations that do not involve any complicated strategy decision of player 1. Let  $U_j = \{g_j, f_j, e_j, h_j, k_j\}$ . The following holds (by

$\text{CF}_p(A)$  we denote the function that returns  $A$  if  $p$  holds and  $\emptyset$  otherwise):

$$\begin{aligned}
 \Xi(k_i) &= \{k_i\} \cup \bigcup \{U_j \mid j > i, b_j = 1\} & \Xi(h_i) &= \{h_i, k_i\} \cup \bigcup \{U_j \mid j > i, b_j = 1\} \\
 \Xi(g_i) &= \{g_i, k_i\} \cup \bigcup \{U_j \mid j \geq i, b_j = 1\} & \Xi(r) &= \{r\} \cup \bigcup \{U_j \mid b_j = 1\} \\
 \Xi(s) &= \{s\} \cup \text{CF}_{b \neq \mathbf{0}_n}(\bigcup \{U_j \mid b_j = 1\} \setminus \{g_{\nu b}\}) & \Xi(c) &= \{c, r\} \cup \bigcup \{U_j \mid b_j = 1\} \\
 \Xi(t_i) &= \{t_i\} \cup \Xi(r) \cup \text{CF}_{i < \text{ind}(\sigma)}(\{t_j \mid j < i\} \cup \{c\}) & \Xi(a_i) &= \{a_i\} \cup \Xi(t_i)
 \end{aligned}$$

It is easy to see that we have the following orderings on the nodes specified above.

$$s \prec_\sigma r \prec_\sigma a_* \prec_\sigma h_* \tag{a}$$

By Lemma 6.1(2), it follows from (a) that  $\tau_\sigma(e_i) = d_i$  for all unset bits  $i$  (i.e.  $b_i = 0$ ), hence we are able to compute the valuations of the remaining nodes.

$$\Xi(e_i) = \{e_i\} \cup \Xi(d_i) \quad \Xi(f_i) = \{e_i, f_i\} \cup \Xi(d_i)$$

This completes the valuation of  $\Xi$  for all nodes.

It is easy to see that for every  $i$  with  $b_i = 0$  and every  $j$  with  $b_j = 1$  s.t. there is no  $i < i' < j$  with  $b_{i'} = 1$ , the following holds:

$$f_i \prec_\sigma f_j \quad g_i \prec_\sigma g_j \tag{b}$$

Also, for  $i > j$  with  $b_i = 1$  and  $b_j = 1$  we have

$$f_i \prec_\sigma f_j \quad g_i \prec_\sigma g_j \tag{c}$$

By (a) and Lemma 6.3(2) we obtain that the following holds:

$$a_{\text{ind}(\sigma)} \prec_\sigma \dots \prec_\sigma a_{2n} \prec_\sigma a_1 \prec_\sigma \dots \prec_\sigma a_{\text{ind}(\sigma)-1} \tag{d}$$

We are now ready to prove that  $\sigma'$  is of the desired form.

- (1) By Lemma 6.2(1) and (a) we derive that closed cycles remain closed. By Lemma 6.6(1) we derive that closed cycles remain accessed. By (a) and Lemma 6.6(2) we derive that open cycles remain skipped.

By phase 1 condition (5), phase 1 condition (6), (d), it follows that for every  $j$  with  $b_j = 0$ , there is an improving node  $a_*$  for  $d_j$ . By Lemma 6.2(2), we conclude that open cycles remain open.

- (2) By (a) and Lemma 6.4(2).
- (3) By (b) and (c).
- (4) By (b) and (c).
- (5) By Lemma 6.4(2).
- (6) By Lemma 6.4(2) and Lemma 6.3(2).

By Lemma 6.4(2) it follows that  $\text{ind}(\sigma') = \text{ind}(\sigma) + 1$ .

If  $\text{ind}(\sigma) > 1$ , then we have by (a) and (d) that  $\sigma'(d_{\mu b}) = \text{ind}(\sigma) - 1$ . □

The first phase ends, when a simple cycle corresponding to an unset bit has no more edges leading to the deceleration lane that keeps it busy waiting, and closes. Since lower bits have less edges going to the lane, it is clear that this will be the least unset bit  $\mu b$ .

The second phase, called the *set* phase, corresponds to a strategy  $\sigma$  in which the least unset bit has just been set, i.e. to the global state  $b[\mu b \mapsto 1] = b_\sigma$ . The selector nodes and roots are as in phase 1 and also the access states, i.e.  $b = a_\sigma$ .

More formally, we say that  $\sigma$  is a  $b$ -phase 2 strategy iff all the following conditions hold:

- (1)  $b[\mu b \mapsto 1] = b_\sigma$  and  $b = a_\sigma$ , i.e. set bits correspond to closed and accessed (for all set bits except for  $\mu b$ ) cycle gates, while unset bits correspond to open and skipped cycle gates,
- (2)  $root(\sigma) = r$ , i.e. the strategy is rooted in  $r$ ,
- (3)  $\sigma(s) = \sigma(r) = \nu b$ , i.e. both roots are connected to the former least set bit,
- (4)  $\sigma(k_i) = \min(\{j > i \mid b_j = 1\} \cup \{n + 1\})$ , i.e. the selector nodes move to the next set bit,
- (5)  $ind(\sigma) \leq 2\mu b + 3$ , i.e. the deceleration lane has not passed the next bit, and
- (6)  $\sigma(d_j) \neq ind(\sigma) - 1$  for all  $j > \mu b$  with  $b_j = 0$ , i.e. every higher open cycle node is not connected to the best-valued node of the lane.

**Lemma 6.9.** Let  $\sigma$  be a  $b$ -phase 1 strategy with  $ind(\sigma) = 2\mu b + 2$  and  $\sigma(d_{\mu b}) = ind(\sigma)$ . Then  $\sigma'$  is a  $b$ -phase 2 strategy.

*Proof.* This can be shown essentially the same way as Lemma 6.8. The only difference now is that  $d_{\mu b}$  has no more improving switches to the deceleration lane and hence, by Lemma 6.2(3), we learn that the  $\mu b$ -cycle has to close.  $\square$

In phase 2, the deceleration lane is still assembling, and the improving switches again include edges of open simple cycles and edges of the deceleration lane. Additionally, it is improving for the cycle gate  $\mu b$  to be accessed and for the root  $s$  to update to cycle gate  $\mu b$ . By performing all these switches, we enter phase three.

The third phase, called the *access* phase, is defined by a renewed correspondence of the cycle gate structure again, i.e.  $b[\mu b \mapsto 1] = b_\sigma = a_\sigma$ . The  $s$  root is connected to  $\mu b$  while  $r$  is still connected to  $\nu b$ . This implies that  $s$  now has a much better valuation than  $r$ .

More formally, we say that  $\sigma$  is a  $b$ -phase 3 strategy iff all the following conditions hold:

- (1)  $b[\mu b \mapsto 1] = b_\sigma = a_\sigma$ , i.e. set bits correspond to closed and accessed cycle gates, while unset bits correspond to open and skipped cycle gates,
- (2)  $root(\sigma) = r$ , i.e. the strategy is rooted in  $r$ ,
- (3)  $\sigma(s) = \mu b$  and  $\sigma(r) = \nu b$ , i.e. one root is connected to the new set bit and the other one is still connected to the former least set bit,
- (4)  $\sigma(k_i) = \min(\{j > i \mid b_j = 1\} \cup \{n + 1\})$ , i.e. the selectors move to the former next set bit,
- (5)  $\sigma(d_j) \neq s$  for all  $j > \mu b$  with  $b_j = 0$ , i.e. every higher open cycle node is not connected to the best-valued root node.

**Lemma 6.10.** Let  $\sigma$  be a  $b$ -phase 2 strategy. Then  $\sigma'$  is a  $b$ -phase 3 strategy.

*Proof.* Again, this can be shown essentially as the previous Lemmas 6.8 and 6.9. The main difference is that now  $f_i \prec_\sigma f_{\mu b}$  for all  $i \neq \mu b$  which is why  $\sigma'(s) = \mu b$ , and that by Lemma 6.6(1) we have that the  $\mu b$ -th gate is  $\sigma'$ -accessed.  $\square$

The cycle gate with the best valuation is now  $\mu b$ , hence, there are many improving switches, that eventually lead to cycle gate  $\mu b$ . First, there are all nodes of the deceleration lane that have improving switches to  $s$ . Second,  $r$  has an improving switch to  $\mu b$ . Third, lower closed cycles (all lower cycles should be closed!) have an improving switch to  $\mu b$  (opening them again). Fourth, all

lower selector nodes have an improving switch to  $\mu b$ . By performing all these switches, we enter phase four.

The fourth and last phase, called the *reset* phase, corresponds to a strategy  $\sigma$  that performed the full increment, i.e.  $b_\sigma = b \oplus 1$ . However, the access states are not reset, i.e.  $a_\sigma = b[\mu b \mapsto 1]$  and the deceleration lane is moving to root  $s$ .

More formally, we say that  $\sigma$  is a  $b$ -phase 4 strategy iff all the following conditions hold:

- (1)  $b \oplus 1 = b_\sigma$  and  $b[\mu b \mapsto 1] = a_\sigma$ , i.e. set bits correspond to closed and accessed cycle gates, while unset bits correspond to open and skipped ( $> \mu b$ ) resp. accessed ( $< \mu b$ ) cycles gates,
- (2)  $root(\sigma) = s$ , i.e. the strategy is rooted in  $s$ ,
- (3)  $\sigma(s) = \sigma(r) = \mu b$ , i.e. both roots are connected to the new set bit,
- (4)  $\sigma(k_i) = \min(\{j > i \mid (b \oplus 1)_j = 1\} \cup \{n + 1\})$ , i.e. the selectors move to the new next set bit,
- (5)  $ind(\sigma) = 0$ , i.e. the deceleration lane has reset, and
- (6)  $\sigma(d_j) = s$  for all  $j$  with  $(b \oplus 1)_j = 0$ , i.e. every open cycle node is connected to the  $s$  root.

**Lemma 6.11.** Let  $\sigma$  be a  $b$ -phase 3 strategy. Then  $\sigma'$  is a  $b$ -phase 4 strategy.

*Proof.* Let  $\sigma$  be a  $b$ -phase 3 strategy,  $\Xi := \Xi_\sigma$  and  $b' := b[\mu b \mapsto 1]$ .

We first compute the valuations for all those nodes directly that do not involve any complicated strategy decision of player 1. Obviously,  $\Xi(x) = \emptyset$ . By Lemma 6.1(1) we know that for all set bits  $i$  (i.e.  $b'_i = 1$ ) we have the following.

$$\Xi(e_i) = \{e_i\} \cup \Xi(h_i) \quad \Xi(d_i) = \{e_i, d_i\} \cup \Xi(h_i) \quad \Xi(f_i) = \{e_i, d_i, f_i\} \cup \Xi(h_i)$$

Using these equations, we are able to compute many other valuations that do not involve any complicated strategy decision of player 1. Let  $U_j = \{g_j, f_j, e_j, h_j, k_j\}$ .

$$\begin{aligned} \Xi(k_i) &= \{k_i\} \cup \bigcup \{U_j \mid j > i, b_j = 1\} & \Xi(h_i) &= \{h_i, k_i\} \cup \bigcup \{U_j \mid j > i, b_j = 1\} \\ \Xi(g_i) &= \{g_i, k_i\} \cup \bigcup \{U_j \mid j \geq i, b_j = 1\} \cup \text{CF}_{i=\mu b} U_i & \Xi(r) &= \{r\} \cup \bigcup \{U_j \mid b_j = 1\} \\ \Xi(s) &= \{s\} \cup \bigcup \{U_j \mid j \geq \mu b, b'_j = 1\} \setminus \{g_{\mu b}\} & \Xi(c) &= \{c, r\} \cup \bigcup \{U_j \mid b_j = 1\} \\ \Xi(t_i) &= \{t_i\} \cup \Xi(r) \cup \text{CF}_{i < ind(\sigma)} (\{t_j \mid j < i\} \cup \{c\}) & \Xi(a_i) &= \{a_i\} \cup \Xi(t_i) \end{aligned}$$

We have the following orderings on the nodes specified above.

$$r \prec_\sigma a_* \prec_\sigma h_{* < \mu b} \prec_\sigma s \prec_\sigma h_{* \geq \mu b} \quad (\text{a})$$

Note that the last inequality  $s \prec_\sigma h_{i \geq \mu b}$  holds for the following reason: If  $i$  corresponds to a set bit, then the path from  $s$  eventually reaches the node  $h_i$ , but the highest priority on the way to  $h_i$  is  $f_i$ , which is odd. If  $i$  on the other hand corresponds to an unset bit, then path from  $s$  to the sink shares the common postfix with  $h_i$ , which starts with the node  $\sigma(k_i)$ . Comparing the two differing prefix show that the most significant difference is  $h_i$  itself, which is even.

By Lemma 6.1(3), it follows from (a) that  $\tau_\sigma(e_i) = d_i$  for all unset bits  $i$  (i.e.  $b'_i = 0$ ), hence we are able to compute the valuations of the remaining nodes.

$$\Xi(e_i) = \{e_i\} \cup \Xi(d_i) \quad \Xi(f_i) = \{e_i, f_i\} \cup \Xi(d_i)$$

It is easy to see that for every  $i$  with  $(b \oplus 1)_i = 0$  and every  $j$  with  $(b \oplus 1)_j = 1$  s.t. there is no  $i < i' < j$  with  $(b \oplus 1)_{i'} = 1$ , the following holds:

$$f_i \prec_\sigma f_j \quad g_i \prec_\sigma g_j \quad (\text{b})$$

Similarly, for  $i > j$  with  $(b \oplus 1)_i = 1$  and  $(b \oplus 1)_j = 1$  we have

$$f_i \prec_\sigma f_j \quad g_i \prec_\sigma g_j \quad (\text{c})$$

We are now ready to prove that  $\sigma'$  is of the desired form.

- (1) By Lemma 6.2(1) and (a) we derive that closed cycles with index  $i \geq \mu b$  remain closed. By Lemma 6.2(4) and (a) we derive that closed cycles with index  $i < \mu b$  open. By Lemma 6.6(1) we derive that closed cycles remain accessed. By (a) and Lemma 6.6(2) we derive that open cycles remain skipped.

By phase 3 condition (5) and (a), it follows that for every  $j$  with  $b_j = 0$ , there is the improving node  $s$  for  $d_j$ . By Lemma 6.2(2), we conclude that open cycles remain open.

- (2) By (a) and Lemma 6.4(1).
- (3) By (b) and (c).
- (4) By (b) and (c).
- (5) By Lemma 6.4(1).
- (6) By (a) and Lemma 6.2(2).

□

By switching the lane back to the initial configuration and the access states to match the simple cycles states, we end up in phase 1 again that corresponds to the incremented global counter state.

**Lemma 6.12.** Let  $\sigma$  be a  $b$ -phase 4 strategy and  $b \oplus 1 \neq \mathbf{1}_n$ . Then  $\sigma'$  is a  $b \oplus 1$ -phase 1 strategy with  $\text{ind}(\sigma') = 1$ .

*Proof.* Let  $\sigma$  be a  $b$ -phase 4 strategy,  $\Xi := \Xi_\sigma$  and  $b' = b \oplus 1$ .

We first compute the valuations for all those nodes directly that do not involve any complicated strategy decision of player 1. Obviously,  $\Xi(x) = \emptyset$ . By Lemma 6.1(1) we know that for all set bits  $i$  (i.e.  $b'_i = 1$ ) we have the following.

$$\Xi(e_i) = \{e_i\} \cup \Xi(h_i) \quad \Xi(d_i) = \{e_i, d_i\} \cup \Xi(h_i) \quad \Xi(f_i) = \{e_i, d_i, f_i\} \cup \Xi(h_i)$$

Using these equations, we are able to compute many other valuations that do not involve any complicated strategy decision of player 1. Let  $U_j = \{g_j, f_j, e_j, h_j, k_j\}$ . The following holds:

$$\begin{aligned} \Xi(k_i) &= \{k_i\} \cup \bigcup \{U_j \mid j > i, b'_j = 1\} & \Xi(h_i) &= \{h_i, k_i\} \cup \bigcup \{U_j \mid j > i, b'_j = 1\} \\ \Xi(r) &= \{r\} \cup \bigcup \{U_j \mid b'_j = 1\} & \Xi(s) &= \{s\} \cup (\bigcup \{U_j \mid b'_j = 1\} \setminus \{g_{\mu b}\}) \\ \Xi(c) &= \{c, s\} \cup \bigcup \{U_j \mid b'_j = 1\} & \Xi(t_i) &= \{t_i\} \cup \Xi(s) \\ \Xi(a_i) &= \{a_i, t_i\} \cup \Xi(s) \end{aligned}$$

Additionally for all  $i \geq \mu b$ , we have:

$$\Xi(g_i) = \{g_i, k_i\} \cup \bigcup \{U_j \mid j \geq i, b'_j = 1\}$$

It is easy to see that we have the following orderings on the nodes specified above.

$$s \prec_\sigma a_* \prec_\sigma r \prec_\sigma h_* \quad (\text{a})$$

By Lemma 6.1(2), it follows from (a) that  $\tau_\sigma(e_i) = d_i$  for all unset bits  $i$  (i.e.  $b'_i = 0$ ), hence we are able to compute the valuations of the remaining nodes.

$$\Xi(d_i) = \{d_i\} \cup \Xi(s) \quad \Xi(e_i) = \{e_i, d_i\} \cup \Xi(s) \quad \Xi(f_i) = \{f_i, e_i, d_i\} \cup \Xi(s)$$

Additionally for all  $i < \mu b$ , we have:

$$\Xi(g_i) = \{g_i, f_i, e_i, d_i\} \cup \Xi(s)$$

This completes the valuation of  $\Xi$  for all nodes.

It is easy to see that for every  $i$  with  $b'_i = 0$  and every  $j$  with  $b'_j = 1$  s.t. there is no  $i < i' < j$  with  $b'_{i'} = 1$ , the following holds:

$$f_i \prec_\sigma f_j \quad g_i \prec_\sigma g_j \quad (b)$$

Similarly, for  $i > j$  with  $b'_i = 1$  and  $b'_j = 1$  we have

$$f_i \prec_\sigma f_j \quad g_i \prec_\sigma g_j \quad (c)$$

We are now ready to prove that  $\sigma'$  is of the desired form.

- (1) By Lemma 6.2(1) and (a) we derive that closed cycles remain closed. By Lemma 6.6(1) we derive that closed cycles remain accessed. By (a) and Lemma 6.6(2) we derive that open cycles remain or will be skipped.

By Lemma 6.2(2) and (a), we conclude that open cycles remain open.

- (2) By (a) and Lemma 6.4(2).
- (3) By (b) and (c).
- (4) By (b) and (c).
- (5) By Lemma 6.4(1) it follows that  $\text{ind}(\sigma') = 1$ .
- (6) By (a) it follows that  $\sigma'(d_i) = r$  for every  $i$  with  $b'_i = 0$ .

□

**6.6. Lower Bound Proof.** Finally, we are ready to prove that our family of games really implements a binary counter. From Lemmas 6.8, 6.9, 6.10, 6.11 and 6.12, we immediately derive the following.

**Lemma 6.13.** Let  $\sigma$  be a phase 1 strategy and  $b_\sigma \neq \mathbf{1}_n$ . There is some  $k \geq 4$  s.t.  $\sigma' = (\mathcal{I}^{\text{loc}})^k(\sigma)$  is a phase 1 strategy and  $b_{\sigma'} = b_\sigma \oplus \mathbf{1}$ .

Particularly, we conclude that strategy improvement with the locally optimizing policy requires exponentially many iterations on  $G_n$ .

**Theorem 6.14.** Let  $n > 0$ . The Strategy Improvement Algorithm with the  $\mathcal{I}^{\text{loc}}$ -policy requires at least  $2^n$  improvement steps on  $G_n$  starting with  $\iota_{G_n}$ .

**6.7. Remarks.** One could conjecture that 1-sink games form a “degenerate” class of parity games as they are always won by player 1. Remember that the problem of solving parity games is to determine the complete winning sets for both players. Given a strategy  $\sigma$  of player 0, we know by Theorem 3.3 that both winning sets can be directly inferred if  $\sigma$  is the optimal strategy. But it is also possible to derive some information about player 0’s winning set given a non-optimal strategy. More precisely,  $W_0 \supseteq \{v \mid \Xi_\sigma(v) = (w, -, -) \text{ and } w \in V_\oplus\}$ .

In other words: Is there a family of games on which the strategy improvement algorithm requires exponentially many iterations to find a player 0 strategy that wins at least one node in the game?

The answer to this question is positive. Simply take our lower bound games  $G_n$  and remove the edge from  $e_n$  to  $h_n$ . Remember that the first time player 1 wants to use this edge by best response is when the binary counter is about to flip bit  $n$ , i.e. after it processed  $2^{n-1}$  many counting steps.

Eventually, the player 0 strategy is updated s.t.  $\sigma(d_n) = e_n$ , forcing player 1 by best response to move to  $h_n$ . Removing this edge leaves player 1 no choice but to stay in the cycle which is dominated by player 0.

## 7. IMPROVING THE LOWER BOUND CONSTRUCTION

We briefly address two improvements of our construction. First, we explain how to reduce the number of edges s.t. the overall size of the games is linear in  $n$ . Second, we describe how to obtain a lower bound construction with binary edge outdegree.

**7.1. Linear Number of Edges.** Consider the lower bound construction again. It consists of a deceleration lane, cycle gates, two roots and connectives between these structures. All three kinds of structures only have linearly many edges when considered on their own. The quadratic number of edges is solely due to the  $d_*$ -nodes of the simple cycles of the cycle gates that are connected to the deceleration lane and due to the  $k_*$ -nodes of the cycle gates that are connected to all higher cycle gates.

We focus on the edges connecting the  $d_*$ -nodes with the deceleration lane first. Their purpose is twofold: lower cycle gates have less edges to the deceleration lane (so they close first), and as long as an open cycle gate should be prevented from closing, there must be a directly accessible lane input node in every iteration with a better valuation than the currently chosen lane input node.

Instead of connecting  $d_i$  to all  $a_j$  with  $j < 2i + 1$  nodes, it would suffice to connect  $d_i$  to two intermediate nodes, say  $y_i$  and  $z_i$ , that are controlled by player 0 with negligible priorities. We connect  $z_i$  to all  $a_j$  with even  $j < 2i + 1$  and  $y_i$  to all  $a_j$  with odd  $j < 2i + 1$ . By this construction, we shift the “busy updating”-part alternately to  $y_i$  and  $z_i$ , and  $d_i$  remains updating as well by switching from  $y_i$  to  $z_i$  and vice versa in every iteration.

Next, we observe that the edges connecting  $y_i$  (resp.  $z_i$ ) to the lane are a proper subset of the edges connecting  $y_{i+1}$  (resp.  $z_{i+1}$ ) to the lane, and hence we adapt our construction in the following way. Instead of connecting  $y_{i+1}$  (and similarly  $z_{i+1}$ ) to all  $a_j$  with even  $j < 2i + 3$ , we simply connect  $y_{i+1}$  to  $a_{2i+1}$  and to  $y_i$ . In order to ensure proper resetting of the two intermediate lanes constituted by  $y_*$  and  $z_*$  in concordance with the original deceleration, we need to connect every additional node to  $c$ . See Figure 5 for the construction (note that by introducing new nodes with “negligible priorities”, we simply shift all other priorities in the game).

Second, we consider the edges connecting lower cycle gates with higher cycle gates. As the set of edges connecting  $k_{i+1}$  with higher  $g_j$  is a proper subset of  $k_i$ , we can apply a similar construction by attaching an additional lane to cycle gate connections that subsumes shared edges.

**7.2. Binary Outdegree.** Every parity game can be linear-time reduced to an equivalent (in the sense that winning sets and strategies can be easily related to winning sets and strategies in the original game) parity game with an edge outdegree bounded by two. See Figure 6 for an example of such a transformation.

However, not *every* such transformation that can be applied to our construction (for clarity of presentation, we start with our original construction again) yields games on which strategy iteration still requires an exponential number of iterations. We discuss the necessary transformations for every player 0 controlled node in the following, although we omit the exact priorities of additional helper nodes. It suffices to assign arbitrary even priorities to the additional nodes that lie below the priorities of all other nodes of the original game (except for the 1-sink).



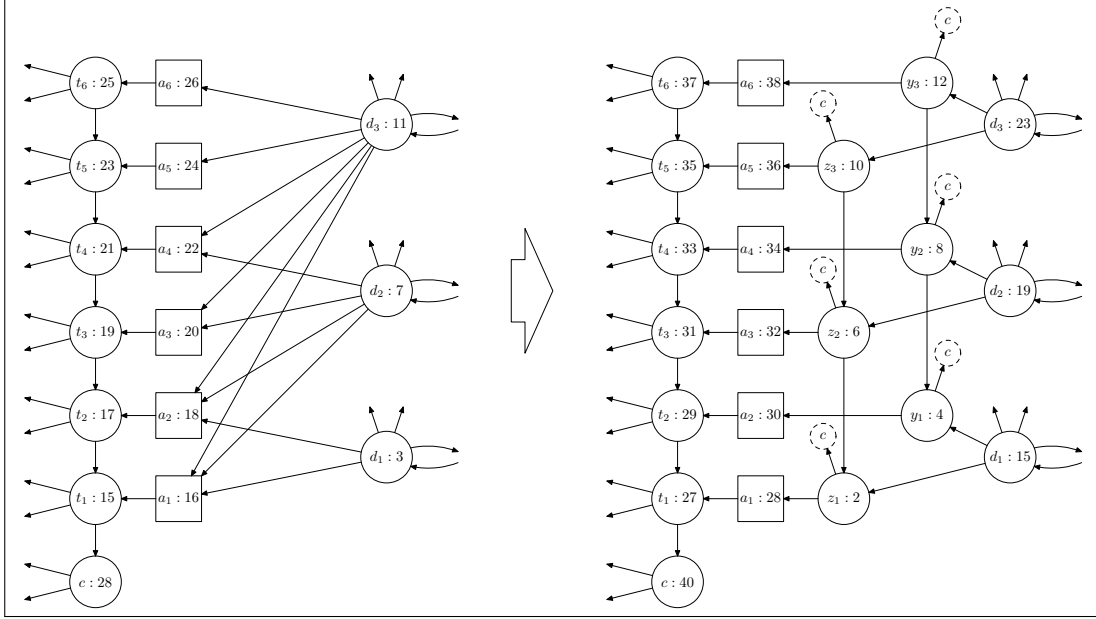


Figure 5: Intermediate Layer

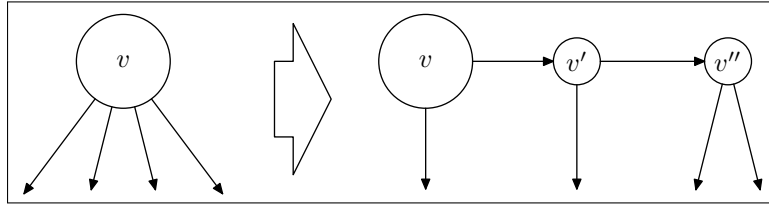


Figure 6: Binary Outdegree Transformation

First, we consider the two root nodes  $s$  and  $r$ , that are connected to the 1-sink  $x$  and to  $f_1, \dots, f_n$  resp.  $g_1, \dots, g_n$ . As  $r$  copies the decision (see the transition from the *access* to the *reset* phase) of  $s$ , it suffices to describe how the outdegree-two transformation is to be applied to  $s$ . We introduce  $n$  additional helper nodes  $s'_1, \dots, s'_n$ , replace the outgoing edges of  $s$  by  $x$  and  $s'_n$ , connect  $s'_{i+1}$  with  $f_{i+1}$  and  $s'_i$ , and finally  $s'_1$  simply with  $f_1$ .

It is still possible to show that  $s$  reaches the best valued  $f_i$  after one iteration. Assume that  $s$  currently reaches some cycle gate  $i$  via the ladder that is given by the helper nodes. Let  $j$  be the next best-valued cycle gate that just has been set. If  $j > i$ , it follows that  $s$  currently reaches  $s'_j$  that moves to  $s'_{j-1}$ , but updates within one iteration to  $f_j$ . If  $j < i$ , it must be the case that  $j = 1$  ( $i$  is the least bit which was set;  $j$  is the least bit which was unset). Moreover,  $s$  currently reaches  $s'_i$  that moves to  $f_i$ . All lower  $s'_{k+1}$  with  $k + 1 < i$  move to  $s'_k$  since lower unset cycle gates are more profitable than higher unset cycle gates (unset cycle gates eventually reach one of the roots via the unprofitable  $f_*$  nodes). Hence,  $s'_i$  updates within one iteration to  $s'_{i-1}$ .

Second, there are the output nodes of cycle gates  $k_1, \dots, k_n$ . We apply a very similar ladder-style construction here. For every  $k_i$ , we introduce  $n - i$  additional helper nodes  $k'_{i,j}$  with  $i < j \leq n$ , replace the outgoing edges of  $k_i$  by  $x$  and  $k'_{i,i+1}$ , connect  $k'_{i,j}$  with  $g_j$  and  $k'_{i,j+1}$  (if  $j < n$ ). The argument why this construction suffices runs similarly as for the root nodes.

Third, there are the nodes  $t_1, \dots, t_{2n}$  of the deceleration lane that are connected to three nodes. Again, we introduce an additional helper node  $t'_i$  for every  $t_i$ , and replace the two edges to  $r$  and  $t_{i-1}$  resp.  $c$  by an edge to  $t'_i$  that is connected to  $r$  and  $t_{i-1}$  resp.  $c$  instead. It is not hard to see that this slightly modified deceleration lane still provides the same functionality.

Finally, there are the player 0 controlled nodes  $d_1, \dots, d_n$  of the simple cycles of the cycle gates. Essentially, two transformations are possible here. Both replace  $d_i$  by as many helper nodes  $d'_{i,x}$  as there are edges from  $d_i$  to any other node  $x$  but  $e_i$ . Then, every  $d'_{i,x}$  is connected to the target node  $x$ .

The first possible transformation connects every  $d'_{i,x}$  with  $e_i$  and vice versa, yielding a multicycle with  $e_i$  as the center of each cycle. The second possible transformation connects  $e_i$  with the first  $d'_{i,x_1}$ ,  $d'_{i,x_1}$  with  $d'_{i,x_2}$  etc. and the last  $d'_{i,x_l}$  again with  $e_i$ , yielding one large cycle. Both replacements behave exactly as the original simple cycle.

The transformation described here results in a quadratic number of nodes since we started with a game with a quadratic number of edges. We note, however, that a similar transformation can be applied to the version of the game with linearly many edges, resulting in a game with binary outdegree of linear size.

## 8. LOWER BOUND FOR THE GLOBALLY OPTIMIZING POLICY

The lower bound construction for the globally optimizing policy again is a family of 1-sink parity games that implement a binary counter by a combination of a (modified) deceleration lane and a chain of (modified) cycle gates

This section is organized as follows. First, we discuss the modifications of the deceleration lane and the cycle gates and why they are required to obtain a lower bound for the globally optimizing policy. Then, we present the full construction along with some remarks to the correctness.

The main difference between the locally optimizing policy and the globally optimizing policy is that the latter takes cross-effects of improving switches into account. It is aware of the impact of any combination of profitable edges, in contrast to the locally optimizing policy that only sees the local valuations, but not the effects.

One primary example that separates both policies are the simple cycles of the previous sections: the locally optimizing policy sees that closing a cycle is an improvement, but not that the actual profitability of closing a cycle is much higher than updating to another node of the deceleration lane.

The globally optimizing policy, on the other hand, is well aware of the profitability of closing the cycle in one step. In some sense, the policy has the ability of a *one-step lookahead*. However, our lower bound for the globally optimizing policy is not so different from the original construction – the trick is to hide very profitable choices by structures that cannot be solved by a single strategy iteration. In other words, we simply need to replace the gadgets that can be solved with a one-step lookahead by slightly more complicated variations that cannot be solved within one iteration *and* that maintain this property for as long as it is necessary.

**8.1. Modified Deceleration Lane.** The modified deceleration lane looks almost the same as the original deceleration lane. It has again several, say  $m$ , input nodes  $a_1, \dots, a_m$  along with some special input node  $c$ . We have two output *roots*,  $r$  and  $s$ , this time with a slightly different connotation. We call  $r$  the *default root* and  $s$  the *reset root*.

More formally, a modified deceleration lane consists of  $m$  (in our case,  $m$  will be  $6 \cdot n - 2$ ) internal nodes  $t_1, \dots, t_m$ ,  $m$  input nodes  $a_1, \dots, a_m$ , one additional input node  $c$ , the default root output node  $r$  and the reset root output node  $s$ .

All priorities of the modified deceleration lane are based on some odd priority  $p$ . We assume that all root nodes have a priority greater than  $p + 2m + 1$ . The structural difference between the modified deceleration lane and the original one is that the lane base  $c$  only has one outgoing edge leading to the default root  $r$ . See Figure 7 for a deceleration lane with  $m = 5$  and  $p = 27$ . The players, priorities and edges are described in Table 5.

Node	Player	Priority	Successors
$t_1$	0	$p$	$\{s, r, c\}$
$t_{i>1}$	0	$p + 2i - 2$	$\{s, r, t_{i-1}\}$
$c$	1	$p + 2m + 1$	$\{r\}$
$a_i$	1	$p + 2i - 1$	$\{t_i\}$
$s$	?	$> p + 2m + 1$	?
$r$	?	$> p + 2m + 1$	?

Table 5: Description of the Modified Deceleration Lane

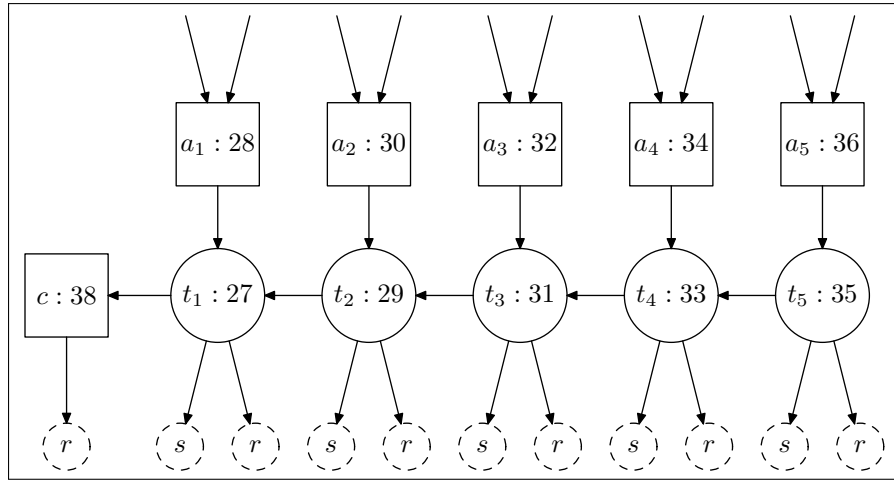


Figure 7: A Modified Deceleration Lane (with  $m = 5$  and  $p = 27$ )

The intuition behind the two roots is the same as before. The default root  $r$  serves as an entry point to the cycle gate structure and the reset root  $s$  is only used for a short time to reset the whole deceleration lane structure.

We describe the state of a modified deceleration lane again by a tuple specifying which root has been chosen and by how many  $t_i$  nodes are already moving down to  $c$ . Formally, we say that  $\sigma$  is in deceleration state  $(x, j)$  (where  $x \in \{s, r\}$  and  $0 < j \leq m + 1$  a natural number) iff

- (1)  $\sigma(t_1) = c$  if  $j > 1$ ,
- (2)  $\sigma(t_i) = t_{i-1}$  for all  $1 < i < j$ , and
- (3)  $\sigma(t_i) = x$  for all  $j \leq i$ .

The modified deceleration lane treats the two roots differently. If the currently best-valued root is the reset root, it is the optimal choice for all  $t_*$ -nodes to directly move to the reset root. In other words, no matter what state the deceleration lane is currently in, if the reset root provides the best valuation, it requires exactly one improvement step to reach the optimal setting.

If the currently best-valued root is the default root, however, it is profitable to reach the root via the lane base  $c$ . The globally optimizing policy behaves in this case just like the locally optimizing policy, because the deceleration lane has exactly one improving switch at a time which is also globally profitable.

The following lemma formalizes the intuitive description of the deceleration lane's behaviour: a change in the ordering of the root valuations leads to a reset of the deceleration lane, otherwise the lane continues to align its edges to eventually reach the best-valued root node via  $c$ .

It is notable that resetting the lane by an external event (i.e. by giving  $s$  a better valuation than  $r$ ) is a bit more difficult than in the case of the locally optimizing policy. Let  $\sigma$  be a strategy and  $\sigma' = \mathcal{I}^{\mathbb{G}^{1\circ}}(\sigma)$ . Assume that the current state of the deceleration lane is  $(r, i)$  and now we have that  $s$  has a better valuation than  $r$ , i.e.  $s \succ_{\sigma} r$ . Assume further – which for instance applies to our original lower bound construction – that the next strategy  $\sigma'$  assigns a better valuation to  $r$  again, i.e.  $r \succ_{\sigma'} s$ . Therefore, it would *not* be the globally optimal choice to reset the deceleration lane to  $s$ , but instead just to keep the original root  $r$ .

In other words, the globally optimizing policy refrains from resetting the lane if the resetting event persists for only one iteration. The solution to fool the policy, however, is not too difficult: we just alter our construction in such a way that the resetting root will have a better valuation than the default root for *two* iterations.

**Lemma 8.1.** Let  $\sigma$  be a strategy that is in deceleration state  $(x, i)$ . Let  $\bar{x}$  denote the other root. Let  $\sigma' = \mathcal{I}^{\mathbb{G}^{1\circ}}(\sigma)$ . Then

- (1)  $r \succ_{\sigma} s$ ,  $x = r$  implies that  $\sigma'$  is in state  $(r, \min(m, i) + 1)$ .
- (2)  $\bar{x} \succ_{\sigma} x$  and  $\bar{x} \succ_{\sigma'} x$  implies that  $\sigma'$  is in state  $(\bar{x}, 1)$ .

The purpose of the modified deceleration lane is exactly the same as before: we absorb the update activity of cyclic structures that represent the counting bits of the lower bound construction.

**8.2. Stubborn Cycles.** With the locally optimizing policy, we employed simple cycles and hid the fact that the improving edge leading into the simple cycle results in a much better valuation than updating to the next best-valued node of the deceleration lane.

However, simple cycles do not suffice to fool the globally optimizing policy. If it is possible to close the cycle within one iteration, the policy sees that closing the cycle is much more profitable than updating to the deceleration lane.

The solution to this problem is to replace the simple cycle structure by a cycle consisting of more than one player 0 node s.t. it is impossible to close the cycle within one iteration. More precisely, we use a structure consisting again of one player 1 node  $e$  and three player 0 nodes  $d^1$ ,  $d^2$  and  $d^3$ , called *stubborn cycle*. We connect all four nodes with each other in such a way that they form a cycle, and connect all player 0 nodes with the deceleration lane. See Figure 8 for an example of such a situation.

More precisely, we connect the player 0 nodes in a *round robin* manner to the deceleration lane, for instance  $d^1$  to  $a_3, a_6, \dots$ ,  $d^2$  to  $a_2, a_5, \dots$ , and  $d^3$  to  $a_1, a_4, \dots$ . We assume that it is more profitable for player 1 to move into the cyclic structure as long as it is not closed.

Now let  $\sigma$  be a strategy s.t.  $\sigma$  is in state  $(r, 6)$  and  $\sigma(d^1) = a_3$ ,  $\sigma(d^2) = d^3$  and  $\sigma(d^3) = a_4$ . There are exactly two improving switches here:  $d^2$  to  $a_5$  (which is the best-valued deceleration

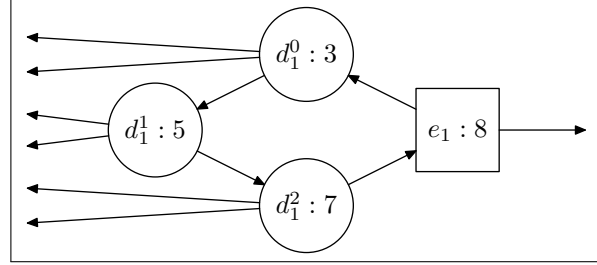


Figure 8: A Stubborn Cycle

node) and  $d^1$  to  $d^2$  (because  $d^2$  currently reaches  $a_4$  via  $d^3$  which has a better valuation than  $a_3$ ). In fact, the combination of both switches is the optimal choice.

A close observation reveals that the improved strategy has essentially the same structure as the original strategy  $\sigma$ : two nodes leave the stubborn cycle to the deceleration lane and one node moves into the stubborn cycle. By this construction, we can ensure that cycles are not closed within one iteration. In other words, the global policy makes no progress towards closing the cycle (it switches one edge towards the cycle, and one edge away from the cycle, leaving it in the exact same position).

**8.3. Modified Cycle Gate.** We again use a slightly modified version of the cycle gates as a pass-through structure that is either very profitable or quite unprofitable. Essentially, we apply two modifications. First, we replace the simple cycle by a stubborn cycle, for the reasons outlined in the previous subsection. Second, we put an additional player 0 controlled internal node  $y_i$  between the input node  $g_i$  and the internal node  $f_i$ . It will delay the update of  $g_i$  to move to the stubborn cycle after closing the cycle by one iteration. By this, we ensure that the modified deceleration lane will have enough time to reset itself.

Formally, a modified cycle gate consists of three internal nodes  $e_i$ ,  $h_i$  and  $y_i$ , two input nodes  $f_i$  and  $g_i$ , and four output nodes  $d_i^1$ ,  $d_i^2$ ,  $d_i^3$  and  $k_i$ . The output node  $d_i^1$  (resp.  $d_i^2$  and  $d_i^3$ ) will be connected to a set of other nodes  $D_i^1$  (resp.  $D_i^2$  and  $D_i^3$ ) in the game graph, and  $k_i$  to some set  $K_i$ .

All priorities of the cycle gate are based on two odd priorities  $p_i$  and  $p'_i$ . See Figure 9 for a cycle gate of index 1 with  $p'_1 = 3$  and  $p_1 = 65$ . The players, priorities and edges are described in Table 6.

Node	Player	Priority	Successors
$d_i^1$	0	$p'_i$	$\{d_i^2\} \cup D_i^1$
$d_i^2$	0	$p'_i + 2$	$\{d_i^3\} \cup D_i^2$
$d_i^3$	0	$p'_i + 4$	$\{e_i\} \cup D_i^3$
$e_i$	1	$p'_i + 5$	$\{d_i^1, h_i\}$
$y_i$	0	$p'_i + 6$	$\{f_i, k_i\}$
$g_i$	0	$p'_i + 7$	$\{y_i, k_i\}$
$k_i$	0	$p_i$	$K_i$
$f_i$	1	$p_i + 2$	$\{e_i\}$
$h_i$	1	$p_i + 3$	$\{k_i\}$

Table 6: Description of the Modified Cycle Gate

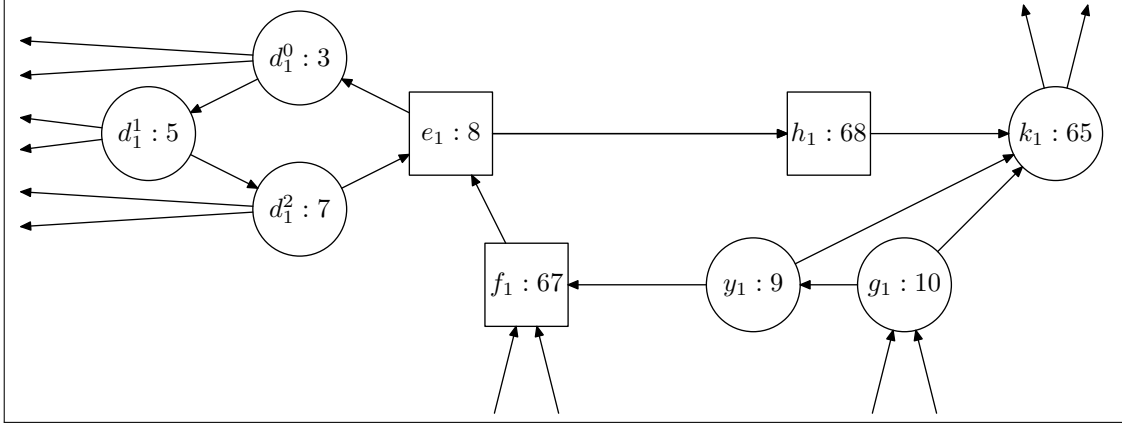


Figure 9: A Modified Cycle Gate (index 1 with  $p'_1 = 3$  and  $p_1 = 65$ )

From an abstract point of view, we describe the state of a modified cycle gate again by a pair  $(\beta_i(\sigma), \alpha_i(\sigma)) \in \{0, 1, 2, 3\} \times \{0, 1, 2\}$ . The first component describes the state of the stubborn cycle, counting the number of edges pointing into the cycle, and the second component gives the state of the two access control nodes. Formally, we have the following.

$$\beta_i(\sigma) = |\{d_i^j \mid \sigma(d_i^j) \notin D_i^j\}| \quad \alpha_i(\sigma) = \begin{cases} 2 & \text{if } \sigma(g_i) = y_i \\ 0 & \text{if } \sigma(g_i) = \sigma(y_i) = k_i \\ 1 & \text{otherwise} \end{cases}$$

The behaviour is formalized in terms of modified cycle gate states as follows. Intuitively, it functions as the original cycle gates: if the cycle is  $\sigma$ -closed and remains closed, it is profitable to go through the cycle gate. If the cycle opens by some external event and remains open, it is more profitable to directly move to the output node instead.

**Lemma 8.2.** Let  $\sigma$  be a strategy and  $\sigma' = \mathcal{I}^{\text{gl}}(\sigma)$ .

- (1) If  $\beta_i(\sigma) = \beta_i(\sigma') = 3$ , we have  $\alpha_i(\sigma') = \min(\alpha_i(\sigma) + 1, 2)$  (“closed gates will be successively accessed”).
- (2) If  $\beta_i(\sigma) < 3$ ,  $\beta_i(\sigma') < 3$  and  $\sigma(k_i) \succ_{\sigma'} f_i$ , we have  $\alpha_i(\sigma') = 0$  (“open gates with unprofitable exit nodes will be skipped”).

We use modified cycle gates again to represent the bit states of a binary counter: unset bits will correspond to modified cycle gates with the state  $(1, 0)$ , set bits to the state  $(3, 2)$ . Setting and resetting bits therefore traverses more than one phase, more precisely, from  $(1, 0)$  over  $(2, 0)$ ,  $(3, 0)$  and  $(3, 1)$  to  $(3, 2)$ , and from the latter again over  $(1, 2)$  to  $(1, 0)$ .

**8.4. Modified Construction.** In this subsection, we provide the complete construction of the lower bound family for the globally optimizing policy. It again consists of a 1-sink  $x$ , a modified deceleration lane of length  $6n - 3$  that is connected to the two roots  $s$  and  $r$ , and  $n$  modified cycle gates. The stubborn cycles of the cycle gates are connected to the  $r$  root, the lane base  $c$  and to the deceleration lane. The modified cycle gates are connected to each other in the same manner as in the original lower bound structure for the locally optimizing policy.

The way the stubborn cycles are connected to the deceleration lane is more involved as in the previous lower bound construction. Remember that for all open stubborn cycles, we need to

maintain the setting in which two edges point to the deceleration lane while the other points into the cycle. We achieve this task by assigning the three nodes of the respective stubborn cycle to the input nodes of the deceleration lane in a round-robin fashion.

We now give the formal construction. The games are denoted by  $H_n = (V_n, V_{n,0}, V_{n,1}, E_n, \Omega_n)$ . The sets of nodes are

$$V_n := \{x, s, c, r\} \cup \{a_i, t_i \mid 0 < i \leq 6n - 2\} \cup \{d_i^1, d_i^2, d_i^3, e_i, f_i, h_i, g_i, y_i, k_i \mid 0 < i \leq n\}$$

The players, priorities and edges are described in Table 7. The game  $H_3$  is depicted in Figure 10. However, the edges connecting the cycle gates with the deceleration lane are not included in the figure.

Node	Player	Priority	Successors
$t_1$	0	$8n + 3$	$\{s, r, c\}$
$t_{i>1}$	0	$8n + 2i + 1$	$\{s, r, t_{i-1}\}$
$a_i$	1	$8n + 2i + 2$	$\{t_i\}$
$c$	1	$20n$	$\{r\}$
$d_i^1$	0	$8i + 1$	$\{s, c, d_i^2\} \cup \{a_{3j+3} \mid j \leq 2i - 2\}$
$d_i^2$	0	$8i + 3$	$\{d_i^3\} \cup \{a_{3j+2} \mid j \leq 2i - 2\}$
$d_i^3$	0	$8i + 5$	$\{e_i\} \cup \{a_{3j+1} \mid j \leq 2i - 1\}$
$e_i$	1	$8i + 6$	$\{d_i^1, h_i\}$
$y_i$	0	$8i + 7$	$\{f_i, k_i\}$
$g_i$	0	$8i + 8$	$\{y_i, k_i\}$
$k_i$	0	$20n + 4i + 3$	$\{x\} \cup \{g_j \mid i < j \leq n\}$
$f_i$	1	$20n + 4i + 5$	$\{e_i\}$
$h_i$	1	$20n + 4i + 6$	$\{k_i\}$
$s$	0	$20n + 2$	$\{f_j \mid j \leq n\} \cup \{x\}$
$r$	0	$20n + 4$	$\{g_j \mid j \leq n\} \cup \{x\}$
$x$	1	1	$\{x\}$

Table 7: Lower Bound Construction for the Globally Optimizing Policy

**Fact 2.** The game  $H_n$  has  $21 \cdot n$  nodes,  $3.5 \cdot n^2 + 40.5 \cdot n - 4$  edges and  $24 \cdot n + 6$  as highest priority. In particular,  $|H_n| = \mathcal{O}(n^2)$ .

As an initial strategy we select the following strategy  $\iota_{H_n}$ . Again, it corresponds to a global counter setting in which no bit has been set.

$$\begin{array}{lll} \iota_{H_n}(t_1) = c & \iota_{H_n}(t_{1 < i \leq 3}) = t_{i-1} & \iota_{H_n}(t_{i > 3}) = r \\ \iota_{H_n}(c) = r & \iota_{H_n}(d_i^1) = d_i^2 & \iota_{H_n}(d_i^2) = a_2 \\ \iota_{H_n}(d_i^3) = a_1 & \iota_{H_n}(g_i) = k_i & \iota_{H_n}(y_i) = k_i \\ \iota_{H_n}(k_i) = x & \iota_{H_n}(s) = x & \iota_{H_n}(r) = x \end{array}$$

It is easy to see that the  $H_n$  family again is a family of 1-sink games.

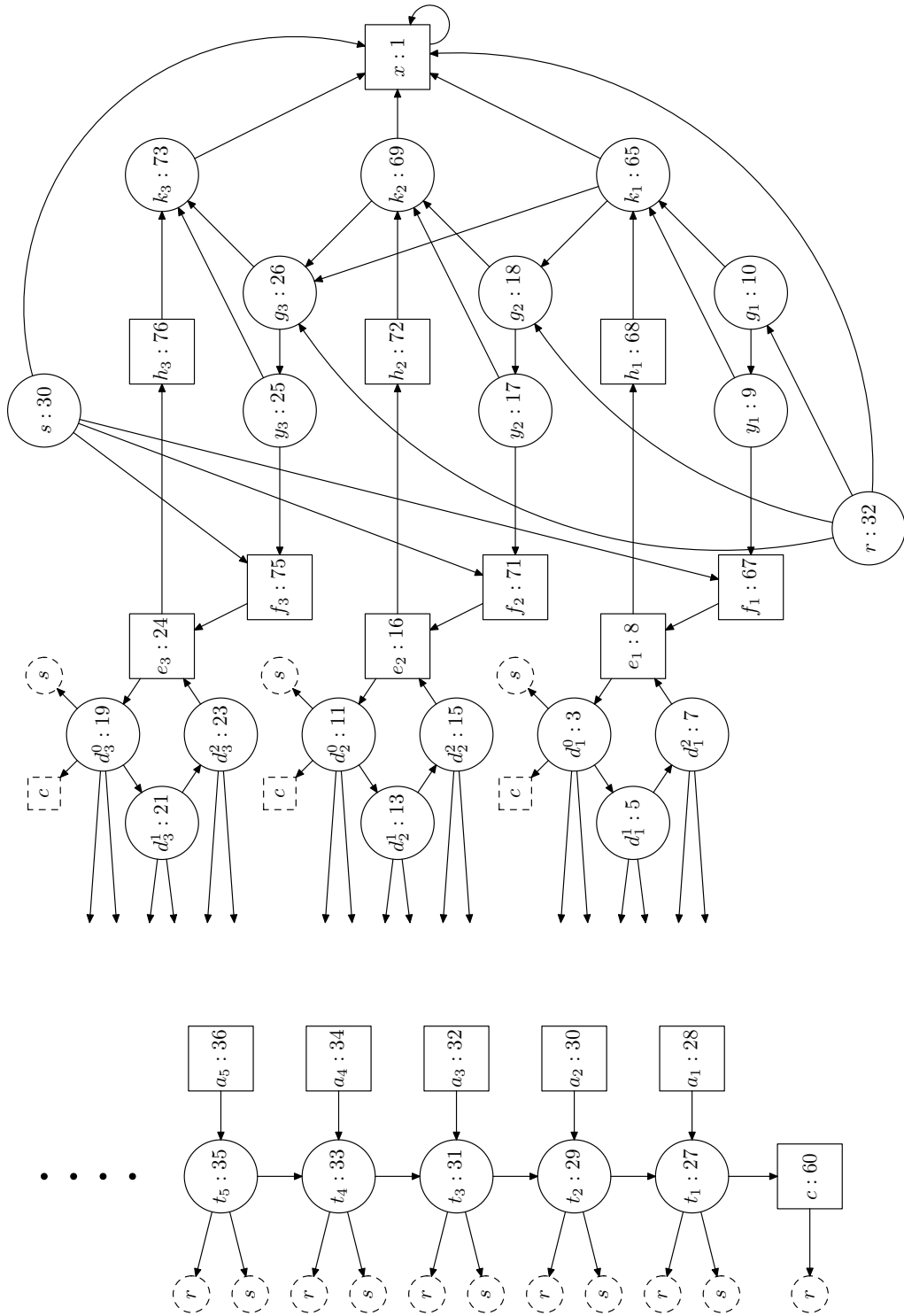


Figure 10: Globally Optimizing Lower Bound Game  $H_3$



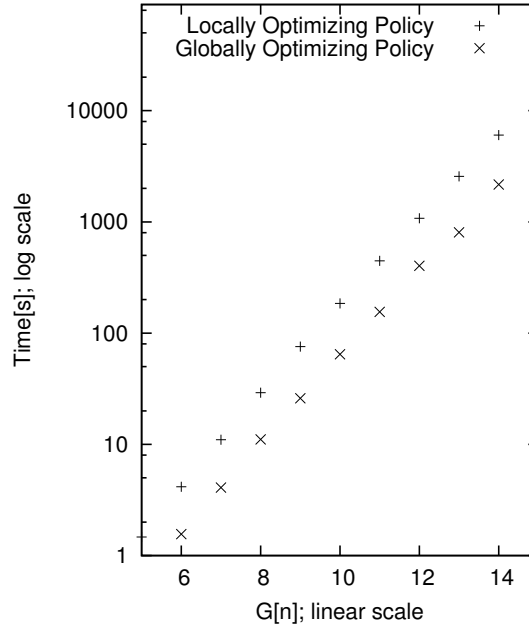


Figure 11: Empirical Evaluation

**Lemma 8.3.** Let  $n > 0$ .

- (1) The game  $H_n$  is completely won by player 1.
- (2)  $x$  is the 1-sink of  $H_n$  and the cycle component of  $\Xi_{\nu_{H_n}}(w)$  equals  $x$  for all  $w$ .

Again, we note that it is possible to refine the family  $H_n$  in such a way that it only comprises a linear number of edges and only outdegree two.

**8.5. Remarks.** The way to prove the construction corrects runs almost exactly the same as for the locally optimizing policy. Every global counting step is separated into some counting iterations of the deceleration lane with busy updating of the open stubborn cycles of the cycle gates until the least significant open cycle closes. Then, resetting of the lane, reopening of lower cycles and alignment of connecting edges is carried out.

**Theorem 8.4.** Let  $n > 0$ . The Strategy Improvement Algorithm with the  $\mathcal{I}^{\text{glo}}$ -policy requires at least  $2^n$  improvement steps on  $H_n$  starting with  $\nu_{H_n}$ .

Our publicly available PGSOLVER Collection [FL09a] of parity game solvers contains implementations of strategy iteration, and particularly parameterizations with the locally and globally optimizing policy. Additionally, the platform features a number of game generators, including all the games and extensions that are presented here. Benchmarking both strategy iteration variants with our lower bound constructions results in exponential run-time behavior as can be seen in Figure 11.

## 9. MEAN PAYOFF, DISCOUNTED PAYOFF AND SIMPLE STOCHASTIC GAMES

We now show that the standard reductions [Pur95, ZP96] from parity games to mean payoff, discounted payoff as well as simple stochastic games can be used to derive worst-case families for all the other game classes.

A *mean payoff game* is a tuple  $G = (V, V_0, V_1, E, r)$  where  $V$ ,  $V_0$ ,  $V_1$  and  $E$  are as in the definition of parity games and  $r : V \rightarrow \mathbb{R}$  is the so-called *reward function*. A *discounted payoff game* is a tuple  $G = (V, V_0, V_1, E, r, \beta)$  where  $(V, V_0, V_1, E, r)$  is a mean payoff game and  $0 < \beta < 1$  is the so-called *discount factor*. Whenever we do not want to distinguish between a discounted and a mean payoff game, we simply write payoff game.

Strategies and plays are defined exactly the same as in the definition of parity games. Given a play  $\pi$ , the *payoff of the play*  $R_G(\pi)$  is defined as follows. For a mean payoff game  $G = (V, V_0, V_1, E, r)$ , it is

$$R_G(\pi) := \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n r(\pi_k)$$

and in the case of a discounted payoff game  $G = (V, V_0, V_1, E, r, \beta)$ , it is

$$R_G(\pi) := \sum_{k=0}^{\infty} \beta^k \cdot r(\pi_k)$$

Let  $G$  be a payoff game. For a given node  $v$ , a player 0 strategy  $\sigma$  and a player 1 strategy  $\varrho$ , let  $\pi_{v,\sigma,\varrho}$  denote the unique play that starts in  $v$  and conforms to  $\sigma$  and  $\varrho$ . We say that a node  $v$  *has a value* iff  $\sup_{\sigma} \inf_{\varrho} R_G(\pi_{v,\sigma,\varrho})$  and  $\inf_{\varrho} \sup_{\sigma} R_G(\pi_{v,\sigma,\varrho})$  exist, and

$$\sup_{\sigma} \inf_{\varrho} R_G(\pi_{v,\sigma,\varrho}) = \inf_{\varrho} \sup_{\sigma} R_G(\pi_{v,\sigma,\varrho})$$

Whenever a node  $v$  has a value, we write  $\vartheta_G(v) := \sup_{\sigma} \inf_{\varrho} R_G(\pi_{v,\sigma,\varrho})$  to refer to it. If every node has a value, we say that a player 0 strategy  $\sigma$  is *optimal* iff  $\inf_{\varrho} R_G(\pi_{v,\sigma,\varrho}) \geq \inf_{\varrho} R_G(\pi_{v,\sigma',\varrho})$  for every node  $v$  and every player 0 strategy  $\sigma'$  and similarly for player 1.

**Theorem 9.1** ([EM79]). Let  $G$  be a mean payoff game. Every node  $v$  has a value and there are *optimal* positional strategies  $\sigma$  and  $\varrho$  s.t.  $\vartheta_G(v) = R_G(\pi_{v,\sigma,\varrho})$  for every  $v$ .

Note that given two optimal positional strategies, it is fairly easy to compute the associated values.

Parity Games can be easily polynomial-time reduced to mean payoff games s.t. optimal strategies correspond to winning strategies and the values of the nodes directly induce corresponding winning sets in the original parity game. Given a parity game  $G = (V, V_0, V_1, E, \Omega)$ , the  *$G$ -induced mean payoff game*  $IndMPG(G) = (V, V_0, V_1, E, r_{\Omega})$  operates on the same graph and defines the reward function  $r_{\Omega}$  as follows.

$$r_{\Omega} : v \mapsto (-|V|)^{\Omega(v)}$$

**Theorem 9.2** ([Pur95]). Let  $G$  be a parity game and let  $\sigma$  and  $\varrho$  be optimal positional strategies w.r.t.  $IndMPG(G)$ . Then the following holds.

- (1)  $W_0 = \{v \in V \mid \vartheta_{IndMPG(G)}(v) \geq 0\}$  is the  $G$ -winning set of player 0
- (2)  $W_1 = \{v \in V \mid \vartheta_{IndMPG(G)}(v) < 0\}$  is the  $G$ -winning set of player 1
- (3)  $\sigma$  is a  $G$ -winning strategy for player 0 on  $W_0$
- (4)  $\varrho$  is a  $G$ -winning strategy for player 1 on  $W_1$

Mean payoff games can be reduced to discounted payoff games by specifying a discount factor that is sufficiently close to 1. Given a mean payoff game  $G = (V, V_0, V_1, E, r)$ , the  $G$ -induced discounted payoff game  $IndDPG(G) = (V, V_0, V_1, E, r, \beta_G)$  operates on the same graph with the same reward function and defines the discount factor  $\beta_G$  as follows. We assume that  $r : V \rightarrow \mathbb{Z}$ .

$$\beta_G := 1 - \frac{1}{4 \cdot |V|^3 \cdot \max\{|r(v)| \mid v \in V\}}$$

Every parity game  $G$  obviously also induces a discounted payoff game via an intermediate mean payoff game.

Let  $v$  be a node in a mean payoff game  $G$ , let  $\vartheta(v)$  be the value of  $v$  in  $G$ , and let  $\vartheta_\beta(v)$  be the value of  $v$  in  $IndDPG(G)$ . Zwick and Paterson [ZP96] show that the value  $\vartheta(v)$  can be essentially bounded by  $\vartheta_\beta(v)$ , i.e.  $|\vartheta_\beta(v) - \vartheta(v)| \leq \frac{1-\beta}{2|V|^2(1-\beta_G)}$ . By choosing  $\beta \geq \beta_G$ , it follows that  $\vartheta(v)$  can be obtained from  $\vartheta_\beta(v)$  by rounding to the nearest rational with a denominator less than  $|V|$ . It follows that optimal strategies in an induced discounted payoff game coincide with optimal strategies in the original mean payoff game.

**Theorem 9.3** ([ZP96]). Let  $G$  be a mean payoff game and let  $\sigma$  and  $\varrho$  be optimal positional strategies w.r.t.  $IndDPG(G)$ . Then  $\sigma$  and  $\varrho$  are also optimal positional strategies w.r.t.  $G$ .

Let  $G$  be a discounted payoff game. Every player 0 strategy  $\sigma$  induces an optimal (not necessarily unique) counterstrategy  $\varrho_\sigma$  s.t.  $R_G(\pi_{v,\sigma,\varrho}) \leq R_G(\pi_{v,\sigma,\varrho'})$  for all other player 1 strategies  $\varrho'$  and all nodes  $v$ . Note that  $\varrho_\sigma$  can be computed by solving an LP-problem as described in Algorithm 2.

---

**Algorithm 2** Computation of Optimal Counter Strategy in a DPG

---

Maximize  $\sum_{v \in V} \varphi(v)$  w.r.t.

$$\begin{aligned} \varphi(v) &= r(v) + \beta \cdot \varphi(\sigma(v)) && \text{for all } v \in V_0 \\ \varphi(v) &\leq r(v) + \beta \cdot \varphi(u) && \text{for all } v \in V_1 \text{ and } u \in vE \end{aligned}$$


---

The value assignment  $\varphi$  can be computed in strongly polynomial time by applying the algorithm of Madani, Thorup and Zwick [MTZ10] for instance. Given  $\varphi$ , an optimal counterstrategy  $\varrho_\sigma$  can be easily induced.

We say that a strategy  $\sigma$  is *improvable* iff there is a node  $v \in V_0$  and a node  $u \in vE$  s.t.  $R_G(\pi_{\sigma(v),\sigma,\varrho_\sigma}) < R_G(\pi_{u,\sigma,\varrho_\sigma})$ . Again, an *improvement policy* is a function  $\mathcal{I}_G : \mathcal{S}_0(G) \rightarrow \mathcal{S}_0(G)$  that satisfies the following two conditions for every strategy  $\sigma$ .

- (1) For every node  $v \in V_0$  it holds that  $R_G(\pi_{\sigma(v),\sigma,\varrho_\sigma}) \leq R_G(\pi_{\mathcal{I}_G(\sigma)(v),\sigma,\varrho_\sigma})$ .
- (2) If  $\sigma$  is improvable then there is a node  $v \in V_0$  s.t.  $R_G(\pi_{\sigma(v),\sigma,\varrho_\sigma}) < R_G(\pi_{\mathcal{I}_G(\sigma)(v),\sigma,\varrho_\sigma})$ .

As with parity game strategy improvement, it is the case that improving a strategy following improvement edges results indeed in an improved strategy.

**Theorem 9.4** ([Pur95]). Let  $G$  be a discounted payoff game,  $\sigma$  be a player 0 strategy and  $\mathcal{I}_G$  be an improvement policy. Then  $R_G(\pi_{v,\sigma,\varrho_\sigma}) \leq R_G(\pi_{v,\mathcal{I}_G(\sigma),\varrho_{\mathcal{I}_G(\sigma)}})$  for every node  $v$ . If  $\sigma$  is not optimal, then  $\sigma$  is improvable.

Puri's algorithm for solving discounted payoff games – as well as mean payoff and parity games via the standard reductions – starts with an initial strategy  $\iota_G$  and runs for a given improvement policy  $\mathcal{I}_G$  as outlined in Algorithm 3. Note that the algorithmic scheme is exactly the same as the discrete version for solving parity games.

---

**Algorithm 3** Puri's Algorithm for Solving Discounted Payoff Games
 

---

```

1:  $\sigma \leftarrow \iota_G$ 
2: while  $\sigma$  is improvable do
3:    $\sigma \leftarrow \mathcal{I}_G(\sigma)$ 
4: end while
5: return  $\sigma, \varrho_\sigma$ 

```

---

Next, we will show that the strategy iteration for discounted payoff games behaves exactly the same as the strategy iteration for 1-sink-parity games.

Vöge proves in his thesis [Vög00] the following theorem that relates parity game strategy iteration to Puri's Algorithm for solving the induced discounted payoff game.

**Theorem 9.5** ([Vög00]). Let  $G$  be a parity game,  $H = \text{IndDPG}(\text{IndMPG}(G))$  be the induced discounted payoff game and  $\sigma$  be a player 0 strategy. For every two nodes  $v$  and  $u$  the following holds.

$$\Xi_\sigma(v) \prec \Xi_\sigma(u) \quad \Rightarrow \quad R_H(\pi_{v,\sigma,\varrho_\sigma}) < R_H(\pi_{u,\sigma,\varrho_\sigma})$$

In other words, every improving switch in the original parity game is also an improving switch in the induced discounted payoff game. The reason why this holds true is that by the reduction from parity games to mean payoff games, the priorities are mapped to such extremely large rewards that the largest reward that occurs on a path dominates all lower ones, the largest reward on a cycle dominates all other ones and that the cycle itself dominates all finite paths leading into it.

Theorem 9.5 is almost what we need to show that strategy iteration for discounted payoff games behaves exactly the same on  $\text{IndDPG}(\text{IndMPG}(G_n))$  as the discrete strategy iteration algorithm on  $G_n$ . Essentially, we need to show the conversion which is equivalent to showing

$$\Xi_\sigma(v) = \Xi_\sigma(u) \quad \Rightarrow \quad R_H(\pi_{v,\sigma,\varrho_\sigma}) = R_H(\pi_{u,\sigma,\varrho_\sigma})$$

However, this statement is not true for every parity game. The reason why a run of the strategy improvement algorithm on general parity games may differ from a run on the induced discounted payoff game is that the parity game strategy iteration does not care about the priority of all nodes on its path to the dominating cycle node that are less relevant. In case of 1-sink-parity games, the only occurring dominating cycle node has the least priority in the game, and therefore all priorities occurring in paths influence the valuations. Also, the strategy iteration on arbitrary parity games does not consider the priorities of all the nodes on a cycle appearing in a node valuation.

First, we show that optimal player 1 counter strategies in the induced discounted payoff game also eventually reach the 1-sink.

**Lemma 9.6.** Let  $G$  be a 1-sink-parity game with  $v^*$  being the 1-sink,  $H = \text{IndDPG}(\text{IndMPG}(G))$  be the induced discounted payoff game, and  $\sigma$  be a player 0 strategy s.t.  $\Xi_{\iota_G} \leq \Xi_\sigma$ . Let  $v_0 \neq v^*$  be an arbitrary node. Then,  $\pi_{v_0,\sigma,\varrho_\sigma}$  is of the following form:

$$\pi_{v_0,\sigma,\varrho_\sigma} = v_0 v_1 \dots v_{l-1} (v^*)^\omega$$

*Proof.* Consider the games  $G' := G|_\sigma$  and  $H' := H|_\sigma$  and note that  $\tau_\sigma^G = \tau_\sigma^{G'}$  as well as  $\varrho_\sigma^H = \varrho_\sigma^{H'}$ . Note that  $G'$  is won by player 1 following  $\tau_\sigma^{G'}$  since  $G$  is 1-sink parity game.

By Theorems 9.2 and 9.3 it follows that  $\varrho_\sigma^{H'}$  must be also a player 1 winning strategy for the whole game  $G'$ . Therefore, it follows that every play  $\pi_{v_0,\sigma,\varrho_\sigma}$  eventually ends in cycle with a dominating cycle node  $w^*$  of odd priority, hence  $\Omega(w^*) \geq \Omega(v^*)$ .

If  $\Omega(w^*) > \Omega(v^*)$ , it follows that there is a  $w^*$ -dominated cycle reachable in  $G'$  starting from  $v_0$ . But since  $\Xi_\sigma(v_0) = (v^*, -, -)$ , this cannot be the case. Hence  $\Omega(w^*) = \Omega(v^*)$ , implying that  $w^* = v^*$ .  $\square$

Second, we show that the value ordering between two different paths leading to the 1-sink again depends solely on the most relevant node in the symmetric difference of the paths.

**Lemma 9.7.** Let  $G$  be a 1-sink-parity game with  $v^*$  being the 1-sink,  $H = \text{IndDPG}(\text{IndMPG}(G))$  be the induced discounted payoff game. Let  $\pi$  and  $\xi$  be two paths of the form  $\pi = u_0 u_1 \dots u_{l-1} (v^*)^\omega$  and  $\xi = w_0 w_1 \dots w_{k-1} (v^*)^\omega$  and let  $U = \{u_0, \dots, u_{l-1}\}$  and  $W = \{w_0, \dots, w_{k-1}\}$ . Then  $U \prec W$  implies  $R_H(\pi) \prec R_H(\xi)$ .

*Proof.* Let  $V = \{v_0, \dots, v_{n-1}\}$  s.t.  $p_{n-1} > p_{n-2} > \dots > p_0$  with  $p_i = \Omega(v_i)$  and  $v_0 = v^*$ , and let  $\beta$  be the discount factor of  $H$ . W.l.o.g. assume that  $n > 2$  since otherwise both paths are necessarily the same. Let  $a : \{1, \dots, n-1\} \rightarrow \{0, \dots, n-2, \perp\}$  be a map s.t.

$$a(i) = \begin{cases} j & \text{if } u_j = v_i \\ \perp & \text{if there is no } j \text{ s.t. } u_j = v_i \end{cases}$$

and let  $b : \{1, \dots, n-1\} \rightarrow \{0, \dots, n-2, \perp\}$  be defined accordingly for  $w_j$ . Set  $\beta^\perp := 0$ . Note that the following holds.

$$R_H(\pi) = \sum_{i=1}^{n-1} \beta^{a(i)} \cdot (-n)^{p_i} + \frac{n \cdot \beta^l}{\beta - 1} \quad R_H(\xi) = \sum_{i=1}^{n-1} \beta^{b(i)} \cdot (-n)^{p_i} + \frac{n \cdot \beta^k}{\beta - 1}$$

Let  $m = \max\{i \mid (a(i) = \perp \text{ and } b(j) \neq \perp) \text{ or } (a(i) \neq \perp \text{ and } b(j) = \perp)\}$  and note that  $m$  indeed is well-defined. Set

$$\Delta := R_H(\xi) - R_H(\pi) = \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$$

where

$$\begin{aligned} \Delta_1 &:= \sum_{i=m+1}^{n-1} (\beta^{b(i)} - \beta^{a(i)}) \cdot (-n)^{p_i} \\ \Delta_2 &:= (\beta^{b(m)} - \beta^{a(m)}) \cdot (-n)^{p_m} \\ \Delta_3 &:= \sum_{i=1}^{m-1} (\beta^{b(i)} - \beta^{a(i)}) \cdot (-n)^{p_i} \\ \Delta_4 &:= \frac{n \cdot (\beta^k - \beta^l)}{\beta - 1} \end{aligned}$$

Regarding  $\Delta_1$ , let  $m < i < n$  and consider that  $|\beta^{b(i)} - \beta^{a(i)}| \leq |1 - \beta^{n-2}|$ . The following holds.

$$\begin{aligned} |\beta^{b(i)} - \beta^{a(i)}| \cdot n^{p_i} &\leq |1 - \beta^{n-2}| \cdot n^{p_i} = \left( \sum_{j=0}^{n-3} \beta^j \right) \cdot (1 - \beta) \cdot n^{p_i} \\ &\leq n \cdot (1 - \beta) \cdot n^{p_i} = \frac{n^{p_i+1}}{4 \cdot n^3 \cdot n^{p_{n-1}}} \leq n^{-2} \end{aligned}$$

We conclude that  $|\Delta_1| \leq \frac{n-1-m}{n^2} \leq 1$ .

Regarding  $\Delta_2$ , note that  $b(m) \neq \perp$  implies that  $p_m$  is even and  $b(m) = \perp$  implies that  $p_m$  is odd. Let  $c = b(m)$  iff  $b(m) \neq \perp$  and  $c = a(m)$  otherwise. Hence the following holds.

$$\begin{aligned}\Delta_2 &= \beta^c \cdot n^{p_m} \geq \beta^{n-1} \cdot n^{p_m} = (\beta^{n-1} - 1) \cdot n^{p_m} + n^{p_m} \\ &= \left( \sum_{j=0}^{n-2} \beta^j \right) \cdot (\beta - 1) \cdot n^{p_m} + n^{p_m} \geq (1 - n) \cdot (1 - \beta) \cdot n^{p_m} + n^{p_m} \\ &= \frac{1 - n}{4 \cdot n^{p_{n-1}+3}} \cdot n^{p_m} + n^{p_m} \geq \frac{3}{4} \cdot n^{p_m}\end{aligned}$$

Regarding  $\Delta_3$ , let  $0 < i < m$  and consider that  $|\beta^{b(i)} - \beta^{a(i)}| \leq 1$ . The following holds.

$$|\Delta_3| \leq \sum_{i=1}^{m-1} |\beta^{b(i)} - \beta^{a(i)}| \cdot n^{p_i} \leq \sum_{i=1}^{m-1} n^{p_i}$$

Now we need to distinguish on whether  $p_m = 2$ . If so, note that  $m = 1$ ,  $b(m) \neq \perp$  and  $k = l + 1$ . Hence, regarding  $\Delta_4$ , the following holds.

$$|\Delta_4| = \frac{n \cdot |\beta^{l+1} - \beta^l|}{|\beta - 1|} = n \cdot \beta \leq n$$

Therefore we conclude (remember that  $n > 2$ )

$$\Delta \geq \Delta_2 - |\Delta_1| - |\Delta_3| - |\Delta_4| \geq \frac{3}{4} \cdot n^2 - 1 - 0 - n > 0$$

Otherwise, if  $p_m > 2$ , it holds that  $|\beta^k - \beta^l| \leq |1 - \beta^{n-1}| \leq (n-1) \cdot (1 - \beta)$  and hence  $|\Delta_4| \leq n^2 - n$ . Additionally, consider  $\Delta_3$  again.

$$|\Delta_3| \leq \sum_{i=1}^{m-1} n^{p_i} \leq \sum_{i=2}^{p_m-1} n^i = \sum_{i=0}^{p_m-1} n^i - 1 - n = \frac{n^{p_m} - 1}{n - 1} - 1 - n$$

We conclude the following (remember again that  $n > 2$ ).

$$\begin{aligned}\Delta &\geq \Delta_2 - |\Delta_1| - |\Delta_3| - |\Delta_4| \\ &\geq \frac{3}{4} \cdot n^{p_m} - 1 - \frac{n^{p_m} - 1}{n - 1} + 1 + n - n^2 + n \\ &= \frac{3}{4} \cdot n^{p_m} - \frac{n^{p_m} - 1}{n - 1} - (n - 1)^2 + 1 \\ &\geq \frac{3}{4} \cdot n^{p_m} - \frac{n^{p_m}}{2} - (n - 1)^2 + 1 \\ &= \frac{1}{4} \cdot n^{p_m} - (n - 1)^2 + 1 > 0\end{aligned}$$

□

Third, we derive that the strategy iteration for discounted payoff games behaves exactly the same as the strategy iteration for 1-sink-parity games.

**Theorem 9.8.** Let  $G$  be a 1-sink-parity game,  $v$  be a node,  $H = \text{IndDPG}(\text{IndMPG}(G))$  be the induced discounted payoff game and  $\sigma$  be a player 0 strategy s.t.  $\Xi_{v_G} \preceq \Xi_\sigma$ . Then  $\varrho_\sigma = \tau_\sigma$ .

*Proof.* Assume by contradiction that  $\varrho_\sigma \neq \tau_\sigma$ . Hence, there is a node  $v$  s.t.  $\pi_{v,\sigma,\tau_\sigma} \neq \pi_{v,\sigma,\varrho_\sigma}$ . Since  $G$  is a 1-sink parity game and  $\Xi_{v_G} \preceq \Xi_\sigma$ , it follows by Lemma 9.6 that  $\pi_{v,\sigma,\varrho_\sigma}$  eventually reaches the 1-sink. It follows that  $R_H(\pi_{v,\sigma,\varrho_\sigma}) < R_H(\pi_{v,\sigma,\tau_\sigma})$  which is impossible due to Lemma 9.7. □

**Corollary 9.9.** Let  $G$  be a 1-sink-parity game,  $H = \text{IndDPG}(\text{IndMPG}(G))$  be the induced discounted payoff game and  $\sigma$  be a player 0 strategy s.t.  $\Xi_{\iota_G} \leq \Xi_\sigma$ . For every two nodes  $v$  and  $u$  the following holds.

$$\Xi_\sigma(v) < \Xi_\sigma(u) \iff R_H(\pi_{v,\sigma,\varrho_\sigma}) < R_H(\pi_{u,\sigma,\varrho_\sigma})$$

**Corollary 9.10.** Puri’s algorithm for solving payoff games requires exponentially many iterations in the worst case when parameterized with the locally or the globally optimal policy.

We note that it is possible to define strategy iteration for mean payoff games directly, i.e. without applying the reduction to discounted payoff games first. Unfortunately, with mean payoff games, it is not the case that if  $\sigma$  is not optimal then there necessarily exists at least one switch that *strictly* improves the reward. There are several way to remedy this situation; most of them are based on a lexicographic ordering again with the first component being the reward and the second component being a description of the nodes leading to the cycle, usually called *potential*. We note without proof that our results translate to this variant of strategy iteration as well.

Finally, we relate our results to simple stochastic games. Particularly, we consider simple stochastic games with arbitrary outdegree and arbitrary probabilities that halt almost surely. Zwick, Paterson and Condon show that there is direct correspondence between this version of simple stochastic games and the original one [ZP96, Con92].

A *simple stochastic game* is a tuple  $G = (V, V_{min}, V_{max}, V_{avg}, 0, 1, E, p)$  s.t.  $V_{min}, V_{max}, V_{avg}, \{0\}$  and  $\{1\}$  are a partition of  $V$ ,  $(V, E)$  is a directed graph with exactly two sinks 0 and 1, and  $p : E \cap (V_{avg} \times V) \rightarrow [0; 1]$  is the probability mapping s.t.  $\sum_{u \in vE} p(v, u) = 1$  for all  $v \in V_{avg}$ .

We say that a simple stochastic game *halts with probability 1* iff every node  $v$  in  $G|_{\sigma,\tau}$  has a path with non-negligible probabilities to a sink for every pair of strategies  $\sigma$  and  $\tau$ . Every simple stochastic game can be reduced to an equivalent simple stochastic game that halts with probability 1 in polynomial time [Con92]. We assume from now on that every given simple stochastic game halts with priority 1.

Given a simple stochastic game and a play  $\pi$ , we say that player *Max* wins  $\pi$  iff it ends in the 1-sink and similarly that player *Min* wins  $\pi$  if it ends in the 0-sink. Let  $R_G(v, \sigma, \varrho)$  denote the probability that player *Max* wins starting from  $v$  conforming to the *Max*-strategy  $\sigma$  and the *Min*-strategy  $\varrho$ .

We say that a node  $v$  *has a value* iff  $\sup_\sigma \inf_\varrho R_G(\pi_{v,\sigma,\varrho})$  and  $\inf_\varrho \sup_\sigma R_G(\pi_{v,\sigma,\varrho})$  exist, and

$$\sup_\sigma \inf_\varrho R_G(\pi_{v,\sigma,\varrho}) = \inf_\varrho \sup_\sigma R_G(\pi_{v,\sigma,\varrho})$$

Whenever a node  $v$  has a value, we write  $\vartheta_G(v) := \sup_\sigma \inf_\varrho R_G(\pi_{v,\sigma,\varrho})$  to refer to it. If every node has a value, we say that a player 0 strategy  $\sigma$  is *optimal* iff  $\inf_\varrho R_G(\pi_{v,\sigma,\varrho}) \geq \inf_\varrho R_G(\pi_{v,\sigma',\varrho})$  for every node  $v$  and every player 0 strategy  $\sigma'$  and similarly for player 1.

**Theorem 9.11** ([Con92]). Let  $G$  be a simple stochastic game. Every node  $v$  has a value and there are *optimal* positional strategies  $\sigma$  and  $\varrho$  s.t.  $\vartheta_G(v) = R_G(\pi_{v,\sigma,\varrho})$  for every  $v$ .

Again, simple stochastic games can be solved by strategy iteration. Given a player *Max* strategy  $\sigma$ , an (not necessarily unique) optimal counterstrategy  $\varrho_\sigma$  – i.e.  $R_G(v, \sigma, \varrho) \leq R_G(v, \sigma, \varrho')$  for all other player *Min* strategies  $\varrho'$  and all nodes  $v$  – can be computed by solving an LP-problem as described in Algorithm 4.

The value assignment  $\varphi$  can be computed in polynomial time by applying Khachiyan’s algorithm [Kha79] for instance. Given  $\varphi$ , an optimal counterstrategy  $\varrho_\sigma$  can be efficiently deduced. The strategy iteration that solves the simple stochastic games runs exactly the same as for discounted payoff games.

**Algorithm 4** Computation of Optimal Counter Strategy in a SSGMaximize  $\sum_{v \in V} \varphi(v)$  w.r.t.

$$\varphi(v) = \varphi(\sigma(v)) \quad \text{for all } v \in V_{max}$$

$$\varphi(v) \leq \varphi(u) \quad \text{for all } v \in V_{min} \text{ and } u \in vE$$

$$\varphi(v) = \sum_{u \in vE} p(v, u) \cdot \varphi(u) \quad \text{for all } v \in V_{avg}$$

$$\varphi(1) = 1$$

$$\varphi(0) = 0$$

Zwick and Paterson [ZP96] describe a simple reduction from discounted payoff games to simple stochastic games that halt with probability 1. Let  $G = (V, V_0, V_1, E, r, \beta)$  be a discounted payoff game and let  $l = \min\{r(v) \mid v \in V\}$ ,  $u = \max\{r(v) \mid v \in V\}$  and  $d = \max(1, u - l)$ .

The  $G$ -induced simple stochastic game is the game  $IndSSG(G) = (V', V_{min}, V_{max}, V_{avg}, 0, 1, E', p)$  where  $V_{min} = V_1$ ,  $V_{max} = V_0$ ,  $V_{avg} = E$ ,  $V' = V_{min} \cup V_{max} \cup V_{avg} \cup \{0, 1\}$  and

$$E' = \{(v, (v, u)), ((v, u), u), ((v, u), 0), ((v, u), 1) \mid v \in V \text{ and } u \in E\}$$

$$p : \begin{cases} ((v, u), u) \mapsto \beta \\ ((v, u), 1) \mapsto (1 - \beta) \cdot \frac{r(v) - l}{d} \\ ((v, u), 0) \mapsto (1 - \beta) \cdot (1 - \frac{r(v) - l}{d}) \end{cases}$$

Clearly, the induced simple stochastic game halts with probability 1. As Zwick and Paterson pointed out, the values of the induced simple stochastic game directly correspond to the values of the original discounted payoff game.

**Lemma 9.12** ([ZP96]). Let  $G$  be a discounted payoff game and  $G' = IndSSG(G)$ . Let  $\sigma$  be a player 0 strategy and  $\varrho$  be a player 1 strategy. Then  $(1 - \beta) \cdot R_G(\pi_{v, \sigma, \varrho}) = d \cdot R_{G'}(v, \sigma, \varrho) + l$  for every node  $v$  where  $l = \min\{r(v) \mid v \in V\}$ ,  $u = \max\{r(v) \mid v \in V\}$  and  $d = \max(1, u - l)$ .

This particularly implies that  $R_G(\pi_{v, \sigma, \varrho}) = \frac{d}{1 - \beta} \cdot R_{G'}(v, \sigma, \varrho) + l$  with  $\frac{d}{1 - \beta} > 0$ , i.e. the values of the original discounted payoff game correspond to the values of the induced simple stochastic game by an affine transformation that preserves the ordering.

**Corollary 9.13.** The standard strategy iteration for simple stochastic games requires exponentially many iterations in the worst case.

## 10. CONCLUSION

We have presented a family of games on which the deterministic strategy improvement algorithm for parity games requires exponentially many iterations. Additionally, we have shown how to adapt this family to prove an exponential lower bound on Schewe's policy.

Finally, we have shown that the presented family can be used to transfer the exponential lower bound to mean payoff, discounted payoff and simple stochastic games by applying the standard reductions.

Although there are many preprocessing techniques that could be used to simplify the family of games presented here – e.g. decomposition into strongly connected components, compression



of priorities, direct-solving of simple cycles, see [FL09b] for instance – they are no solution to the general weakness of strategy iteration on these games, simply due to the fact that all known preprocessing techniques can be fooled quite easily without really touching the inner structure of the games.

Parity games are widely believed to be solvable in polynomial time, yet there is no algorithm known that is performing better than superpolynomially. Jurdziński and Vöge presented the strategy iteration technique for parity games over ten years ago, and this class of solving procedures is generally supposed to be the best candidate to give rise to an algorithm that solves parity games in polynomial time since then. Unfortunately, the locally and the globally optimizing technique are not capable of achieving this goal.

We think that the strategy iteration still is a promising candidate for a polynomial time algorithm, however it may be necessary to alter more of it than just the improvement policy.

**Acknowledgements.** I am very thankful to Martin Lange and Martin Hofmann for their guidance and numerous inspiring discussions on the subject. Also, I would like to thank the anonymous referees for their thorough reports containing many comments that helped to improve the presentation of this paper.

## REFERENCES

- [BV07] Henrik Björklund and Sergei Vorobyov. A combinatorial strongly subexponential strategy improvement algorithm for mean payoff games. *Discrete Appl. Math.*, 155(2):210–229, 2007.
- [Con92] Anne Condon. The complexity of stochastic games. *Information and Computation*, 96:203–224, 1992.
- [EJ91] E. Emerson and C. Jutla. Tree automata,  $\mu$ -calculus and determinacy. In *Proc. 32nd Symp. on Foundations of Computer Science*, pages 368–377, San Juan, 1991. IEEE.
- [EJS93] E. Emerson, C. Jutla, and A. Sistla. On model-checking for fragments of  $\mu$ -calculus. In *Proc. 5th Conf. on CAV, CAV’93*, volume 697 of *LNCS*, pages 385–396. Springer, 1993.
- [EM79] A. Ehrenfeucht and J. Mycielski. Positional strategies for mean payoff games. *Internat. J. Game Theory*, 8:109–113, 1979.
- [FL09a] Oliver Friedmann and Martin Lange. The PGSolver collection of parity game solvers, 2009. University of Munich. Available from <http://www.tcs.ifi.lmu.de/pgsolver>.
- [FL09b] Oliver Friedmann and Martin Lange. Solving parity games in practice. In *ATVA*, pages 182–196, 2009.
- [Fri10a] O. Friedmann. Recursive solving of parity games requires exponential time. A preprint available from <http://www.tcs.ifi.lmu.de/~friedman>, 2010.
- [Fri10b] Oliver Friedmann. The stevens-stirling-algorithm for solving parity games locally requires exponential time. *Int. J. Found. Comput. Sci.*, 21(3):277–287, 2010.
- [GTW02] E. Grädel, W. Thomas, and Th. Wilke, editors. *Automata, Logics, and Infinite Games*, LNCS. Springer, 2002.
- [HK66] A. J. Hofmann and R. M. Karp. On nonterminating stochastic games. *Management Science*, 12(5):359–370, 1966.
- [How60] Ronald Howard. *Dynamic Programming and Markov Processes*. The M.I.T. Press, 1960.
- [JPZ06] M. Jurdziński, M. Paterson, and U. Zwick. A deterministic subexponential algorithm for solving parity games. In *Proc. 17th Ann. ACM-SIAM Symp. on Discrete Algorithm, SODA’06*, pages 117–123. ACM, 2006.
- [Jur98] M. Jurdziński. Deciding the winner in parity games is in  $\text{UP} \cap \text{coUP}$ . *Inf. Process. Lett.*, 68(3):119–124, 1998.
- [Jur00] M. Jurdziński. Small progress measures for solving parity games. In H. Reichel and S. Tison, editors, *Proc. 17th Ann. Symp. on Theo. Aspects of Computer Science, STACS’00*, volume 1770 of *LNCS*, pages 290–301. Springer, 2000.
- [Kha79] Leonid Khachiyan. A polynomial algorithm in linear programming. *Soviet Math. Dokl.*, 20:191–194, 1979.
- [MTZ10] Omid Madani, Mikkel Thorup, and Uri Zwick. Discounted deterministic markov decision processes and discounted all-pairs shortest paths. *ACM Transactions on Algorithms*, 6(2), 2010.
- [Pur95] Anuj Puri. *Theory of Hybrid Systems and Discrete Event Systems*. PhD thesis, University of California, Berkeley, 1995.
- [Sch07] Sven Schewe. Solving parity games in big steps. In *Proc. FST TCS*. Springer-Verlag, 2007.

- [Sch08] S. Schewe. An optimal strategy improvement algorithm for solving parity and payoff games. In *17th Annual Conference on Computer Science Logic (CSL 2008)*, 2008.
- [SS98] P. Stevens and C. Stirling. Practical model-checking using games. In B. Steffen, editor, *Proc. 4th Int. Conf. on Tools and Alg. for the Constr. and Analysis of Systems, TACAS'98*, volume 1384 of *LNCS*, pages 85–101. Springer, 1998.
- [Sti95] C. Stirling. Local model checking games. In *Proc. 6th Conf. on Concurrency Theory, CONCUR'95*, volume 962 of *LNCS*, pages 1–11. Springer, 1995.
- [VJ00] J. Vöge and M. Jurdzinski. A discrete strategy improvement algorithm for solving parity games. In *Proc. 12th Int. Conf. on Computer Aided Verification, CAV'00*, volume 1855 of *LNCS*, pages 202–215. Springer, 2000.
- [Vög00] Jens Vöge. *Strategiesynthese für Paritätsspiele auf endlichen Graphen*. PhD thesis, University of Aachen, 2000.
- [Zie98] W. Zielonka. Infinite games on finitely coloured graphs with applications to automata on infinite trees. *TCS*, 200(1–2):135–183, 1998.
- [ZP96] Uri Zwick and Mike Paterson. The complexity of mean payoff games on graphs. *Theoretical Computer Science*, 158(1-2):343–359, 1996.