Synthesising Strategy Improvement and Recursive Algorithms for Solving 2.5 Player Parity Games

Ernst Moritz Hahn*, Sven Schewe†, Andrea Turrini*, Lijun Zhang*

*State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China
†University of Liverpool, United Kingdom

Abstract—2.5 player parity games combine the challenges posed by 2.5 player reachability games and the qualitative analysis of parity games. These two types of problems are best approached with different types of algorithms: strategy improvement algorithms for 2.5 player reachability games and recursive algorithms for the qualitative analysis of parity games. We present a method that—in contrast to existing techniques—tackles both aspects with the best suited approach and works exclusively on the 2.5 player game itself. The resulting technique is powerful enough to handle games with several million states.

I. INTRODUCTION

Parity games are non-terminating zero sum games between two players, Player 0 and Player 1. The players move a token along the edges of a finite graph without sinks. The vertices are coloured, i.e. labelled with a priority taken from the set of natural numbers. The infinite sequence of vertices visited by the token is called the run of a graph, and each run is coloured according to the minimum priority that appears infinitely often on the run. A run is winning for a player if the parity of its colour agrees with the parity of the player.

Parity games come in two flavours: games with random moves, also called 2.5 player games, and games without random moves, called 2 player games. For 2 player games, the adversarial objectives of the two players are to ensure that the lowest priority that occurs infinitely often is even (for Player 0) and odd (for Player 1), respectively. For 2.5 player games, the adversarial objectives of the two players are to maximise the likelihood that the lowest priority that occurs infinitely often is even resp. odd.

Solving parity games is the central and most expensive step in many model checking [1–5], satisfiability checking [1, 3, 6, 7], and synthesis [8] methods. As a result, efficient algorithms for 2 player parity games have been studied intensively [1, 10–26].

Parity games with 2.5 players have recently attracted attention [27–35]. This attention, however, does not mean that results are similarly rich or similarly diverse as for 2 player games. Results on the existence of pure strategies and on approximation algorithms [29, 31] are decades younger than similar results for 2 player games, while algorithmic solutions [27, 28] focus on strategy improvement techniques only.

The qualitative counterpart of 2.5 player games, where one of the players has the goal to win almost surely while the other one wants to win with a non-zero chance, can be reduced to 2 player parity games, cf. [36] or attacked directly on the 2.5 player game with recursive algorithm [37]. The more interesting quantitative analysis can be approached through a reduction to 2.5 player reachability games [38], which can then be attacked with strategy improvement algorithms [16, 17, 25, 26, 39]. Alternatively, entangled strategy improvement algorithms can also run concurrently the 2.5 player parity game directly (for the quantitative aspects) and on a reduction to 2 player parity games (for the qualitative aspects) [27, 28]. (Or, likewise, run on the larger game with an ordered quality measure that gives preference to the likelihood to win and uses the progress measure from [19] or [18] as a tie-breaker.)

This raises the question if strategy improvement techniques can be directly applied on 2.5 player parity games, especially as such games are memoryless determined and therefore satisfy a main prerequisite for the use of strategy improvement algorithms. The short answer is that strategy algorithms for 2.5 player parity games simply do not work. Classical strategy improvement algorithms follow a joint pattern. They start with an arbitrary strategy $f$ for one of the players (say Player 0). This strategy $f$ maps each vertex of Player 0 to a successor, and thus resolves all moves of Player 0. This strategy is then improved by changing the strategy $f$ at positions, where it is profitable to do so. The following steps are applied repeatedly until there is no improvement in Step 2.

1) Evaluate the simpler game resulting from fixing $f$.
2) Identify all changes to $f$ that, when applied once, lead to an improvement.
3) Obtain a new strategy $f'$ from $f$ by selecting some subset of these changes.

So where does this approach go wrong? The first step works fine. After fixing a strategy for Player 0, we obtain a 1.5 player parity game, which can be solved efficiently with standard techniques [40]. It is also not problematic to identify the profitable switches in the second step. The winning probability for the respective successor vertex provides a natural measure for the profitability of a switch. We will show in Section IV that, as usual for strategy improvement, any combination of such profitable switches will lead to an improvement.

The problem arises with the optimality guarantees. Strategy improvement algorithms guarantee that a strategy that cannot be improved is optimal. In the next paragraph, we will see an example, where this is not the case. Moreover, we will see that it can be necessary to change several decisions in a strategy $f$ in order to obtain an improvement, something which is against...
the principles of strategy improvement.

A. An illustrating example

Consider the example 2.5 player parity game $P_e$ depicted in Figure 1. Square vertices are controlled by Player 0, while triangular ones are controlled by Player 1. In circular vertices, a random successor vertex is chosen with the given probability. In $v_w$, Player 0 wins with certainty (and therefore in particular almost surely), while she loses with certainty in $v_1$. In $v_{0.55}$ (or $v_{0.95}$), Player 0 wins with probability 0.55 (or 0.95). For the nodes $v_0$ and $v_1$, we can see that the mutually optimal strategy for Player 0 and Player 1 are to play $e_{0.2}$ and $e_{1.1}$, respectively. Player 0 therefore wins with probability 0.95 when the game starts in $v_0$ and both players play optimally.

B. Naive strategy iteration

Strategy iteration algorithms start with an arbitrary strategy, and use an update rule to get profitable switches. These are edges, where the new target vertex has a higher probability of reaching the winning region (when applied once) compared to the current vertex. As usual with strategy improvement, any combination of profitable switches leads to a strictly better strategy for Player 0. We illustrate that, if done naively, it may lead to values that are only locally maximal. Assume that initially Player 0 chooses the edge $e_{0.1}$ from $v_0$, then the best counter strategy of Player 1 is to choose $e_{1.2}$ from $v_1$. The winning probability for Player 0 under these strategies is 0.55.

In strategy iteration, an update rule allows a player to switch actions only if the switching offers some improvement. Since by switching to the edge $e_{0.2}$ Player 0 would obtain the same winning probability, no strategy iteration can be applied, and the algorithm terminates with a sub-optimal solution.

Let us try to get some insights from this problem. Observe that Player 1 can entrap the play in the left vertices $v_0$ and $v_1$ when Player 0 chooses the edge $e_{0.2}$, such that the almost sure winning region of Player 0 cannot be reached. However, this comes to the cost of losing almost surely for Player 1, as the dominating colour on the resulting run is 0. Broadly speaking, Player 0 must find a strategy that maximises her chance of reaching her almost sure winning regions, but only under the constraint that the counter strategy of Player 1 does not introduce new almost sure winning regions for Player 0.

C. Solutions from the literature

In the literature, two different solutions to this problem have been discussed. Neither of these solutions works fully on the game graph of the 2.5 player parity game. Instead, one of them uses a reduction to reachability games through a simple gadget construction [38], while the other uses strategy improvement on two levels, for the qualitative update described above, and for an update within subgames of states that have the same value [27, 28]; this requires to keep a pair of entangled strategies.

Gadget construction for a reduction to reachability games: In [38], it is shown that 2.5 player parity games can be solved by reducing them to 2.5 player reachability games and solving them, e.g. by using a strategy improvement approach. For this reduction, one can use the simple gadgets shown in Figure 2. There, when a vertex is passed by, the token goes to an accepting sink with probability $w_{prob}$ and to a losing sink with probability $l_{prob}$, both depending on the priority of the node (and continues otherwise as in the parity game). For accordingly chosen $w_{prob}, l_{prob}$, any optimal strategy for this game is an optimal strategy for the parity game. To get this guarantee, however, the termination probabilities have to be very small indeed. In [38], they are constructed from the expression $(\nu!2^{2n+3}M^{2n^2})^{-1}$ where $n$ is the number of vertices and $M$ is an integer depending on the probabilities occurring in the model. Unfortunately, these small probabilities render this approach very inefficient and introduces numerical instability.

Classic strategy improvement for 2.5 player parity games: In [27, 28], the concept of strategy improvement algorithms has been extended to 2.5 player parity games. To overcome the problem that the natural quality measure—the likelihood of winning—is not fine enough, this approach constructs classical 2 player games on translations of the value classes (the set of vertices with the same likelihood of winning). These subgames are translated using a gadget construction similar to the one used for qualitative solutions for 2.5 player to a solution to 2 player games from [36]. This results in the 2 player game shown in Figure 3.

The strategy improvement algorithm keeps track of ‘wit-
vertices will translate to an update on the strategy of

we have seen in the example, this rule would not necessarily

player game. The first rule is a standard update rule for increas-

mnes \( \omega = (\pi, \pi_Q)^1 \), which consists of a strategy \( \pi \) on the

2.5 player parity game, and a strategy \( \pi_Q \) defined on the 2

player game \( Q \) obtained from this 2.5 player game using the

gadget construction from \([36]\). The strategies are entangled in

that \( \pi \) is the translation\(^1\) of \( \pi_Q \). That is, the strategies have
to concur on the nodes of Player 0 from the 2.5 player game,
and each update on \( \pi \) (resp. \( \pi_Q \)) on the decisions from these
vertices will translate to an update on the strategy of \( \pi_Q \) (resp.
\( \pi \)) on the same vertices.

The valuation of one of these vertices is an ordered pair,
consisting of the chance of obtaining the parity objective as
the primary measure, and the value obtained in the quantitative

game restricted to the individual value classes (vertices with
the same chance of obtaining the parity objective) as a
secondary measure \([27], [28]\).

D. Novel strategy iteration algorithm

We show that we can apply strategy improvement tech-
niques with two different update rules directly on the 2.5
player game. The first rule is a standard update rule for increas-
ing the chance of reaching the almost sure winning region.
As we have seen in the example, this rule would not necessarily
find the optimum: it would not find the improvement from
edge \( e_{0,1} \) to \( e_{0,2} \). To overcome this problem, we introduce a
second rule that handles the problem that Player 1 can reduce
the chances of reaching the almost sure winning region of

1In the notation of \([27], [28]\), \( \pi = \text{Tr}_{\text{almost}}(\pi_Q) \).

Player 0 by playing a strategy that leads to a larger almost
sure winning region for Player 0. This step uses a reduction
to the qualitative evaluation of these games. Player 0 changes
her strategy in a way that she would win on the subgame
that consists only of the edges of Player 0 and Player 1 that
are neutral. For both players, these are the edges that lead to
successor states with the same chance of winning under the
current strategy. If this provides a larger almost sure winning
region for Player 0 than \( f \), then update \( f \) in this new winning
region accordingly leads to a strictly better strategy \( f' \).

While the first rule alone is not powerful enough, the two
rules together provide the guarantee that a strategy that cannot
be improved by either of them is optimal.

Note that the second rule is a non-standard rule for strategy
improvement. Not only does it not rely on an improvement that
is obtained when a change is applied once, it also requires to
apply a fixed set of changes (in the new region) in one step
for correctness. This is quite unusual for strategy improvement
algorithms, where the combination of updates selected is
irrelevant for correctness.

A further significant difference to the method from \([27],
[28]\) is that we do not have to revert to solving trans-
formed games. Instead, we use the new generalisation of
McNaughton’s algorithm to the qualitative solution of 2.5
player parity games \([37]\). This method seems to maintain
the good practical performance known for classic recursive

techniques, which have proven to be much faster than strategy
improvement for the qualitative analysis of parity games \([41]\).

A consequence of this choice is that we solve the qualitative
games completely when there is no progress through the naïve
update step, which reduces the number of times that qualitative
updates have to be considered.

This way, we use strategy improvement for the quantitative
part of the analysis, where it has its strengths, while leaning
on a variation \([27]\) of McNaughton’s algorithm \([10], [12],
[14]\) for the qualitative part of the analysis, where prior
research suggests that recursive algorithms outperform strategy
improvement \([41]\).

Note that our quality measure strategy improvement is
the same as the primary measure used in classical strategy
improvement for 2.5 player parity games \([27], [28]\). Different
from that approach, we do not need to resort to gadget
constructions for progressing within value classes, but can
overcome the lack of progress w.r.t. the primary measure
through invoking a performant algorithm for solving 2.5 player
games quantitatively \([37]\).

E. Organisation of the Paper

We first introduce the standard terms and concepts in
Section II. We then recall the strategy improvement algorithms
in Section III, describe our algorithm in Section IV, show its
correctness in Section V, and offer an experimental evaluation
in Section VI.

II. TERMS AND CONCEPTS

A probability distribution over a finite set \( A \) is a function
\( \mu: A \to [0, 1] \cap \mathbb{Q} \) with \( \sum_{a \in A} \mu(a) = 1 \). By \( \text{Distr}(A) \) we
denote the set of probability distributions over $A$.

**Definition 1:** An arena is a tuple $\mathfrak{A} = (V_0, V_1, V_r, E, \text{prob})$, where

- $V_0$, $V_1$, and $V_r$ are three finite disjoint sets of vertices owned by the three players: Player 0, Player 1, and Player random, respectively. Let $V \overset{\text{def}}{=} V_0 \cup V_1 \cup V_r$;
- $E \subseteq V \times V$ is a set of edges such that $(V, E)$ is a sinkless directed graph, i.e., for each $v \in V$ there exists $v' \in V$ such that $(v, v') \in E$; for $\sigma \in \{0, 1, r\}$ we let $E_\sigma \overset{\text{def}}{=} E \cap (V_\sigma \times V)$.
- prob: $V_r \rightarrow \text{Distr}(V)$ is the successor distribution function. We require that for each $v \in V_r$ and each $v' \in V$, prob$(v)(v') > 0$ if and only if $(v, v') \in E$.

If $V_0 = \emptyset$ or $V_1 = \emptyset$, we call $\mathfrak{A}$ a Markov decision process (MDP) or 1.5 player game. If both $V_0 = V_1 = \emptyset$, we call $\mathfrak{A}$ a Markov chain (MC). Given an arena $\mathfrak{A} = (V_0, V_1, V_r, E, \text{prob})$, we define the following concepts:

- A **play** is an infinite sequence $\pi = v_0v_1v_2v_3\ldots$ such that $(v_i, v_{i+1}) \in E$ for all $i \in \mathbb{N}$. We define $\pi(i) \overset{\text{def}}{=} v_i$. We denote by $\text{Play}(\mathfrak{A})$ the set of all plays of $\mathfrak{A}$.
- For $\sigma \in \{0, 1\}$, a (pure memoryless) **strategy** $f_\sigma$ of Player $\sigma$ is a mapping $f_\sigma: V_\sigma \rightarrow V_r$ from the vertices of $V_r$ to Player $\sigma$’s successor states, i.e., for each $v \in V_\sigma$, $(f_\sigma(v), v) \in E$. We denote the set of Player 0 and 1 strategies by $\text{Strats}_0$ and $\text{Strats}_1$, respectively.
- Given a strategy $f_0$ for Player 0, we define the **induced MDP** as $\mathfrak{A}_{f_0} = (\emptyset, V_1, V_r \cup V_0, E_{f_0}, \text{prob}_{f_0})$ with $E_{f_0} \overset{\text{def}}{=} (E \setminus V_0 \times V) \cup \{(v, f_0(v)) \mid v \in V_0\}$ and
  \[
  \text{prob}_{f_0}(v)(v') = \begin{cases} 
  \text{prob}(v)(v') & \text{if } v \in V_r, \\
  1 & \text{if } v \in V_0 \text{ and } v' = f_0(v), \\
  0 & \text{otherwise},
  \end{cases}
  \]
and similarly for Player 1.
- Given strategies $f_0, f_1$ for Player 0 and Player 1, respectively, we denote by $\mathfrak{A}_{f_0, f_1} \overset{\text{def}}{=} (\mathfrak{A}_{f_0})_{f_1}$ the **induced MC** of the strategies.
- If $\mathfrak{A}$ is an MC, we denote by $P^\mathfrak{A}(v): \Sigma^\mathfrak{A} \rightarrow [0, 1]$ the uniquely induced probability measure on $\Sigma^\mathfrak{A}$, the $\sigma$-algebra on the cylinder sets of the plays of $\mathfrak{A}$, under the condition that the initial node is $v$. For general $\mathfrak{A}$, we let $P^\mathfrak{A}_{f_0, f_1}(v) \overset{\text{def}}{=} P^\mathfrak{A}_{f_0, f_1}(v)$.

**Definition 2:** A **2.5 player game**, also referred to as Markov game (MG), is a tuple $\mathcal{P} = (V_0, V_1, V_r, E, \text{prob}, \text{win})$, where $\mathfrak{A} = (V_0, V_1, V_r, E, \text{prob})$ is an arena and $\text{win} \subseteq \text{Play}(\mathfrak{A})$ is the winning condition for Player 0, the set of plays for which Player 0 wins.

The notions of plays, strategies, induced 1.5 player games, etc. extend to 2.5 player games by considering their underlying arena.

We consider two types of winning conditions, reachability and parity objectives.

**Definition 3:** A 2.5 player reachability game is a 2.5 player game $\mathcal{P}$ in which the winning condition win is defined by a target set $R \subseteq V$. Then, we have $\text{win} = \{ \pi \in \text{Play}(\mathcal{P}) \mid \exists i \geq 0 : \pi(i) \in R \}$. For 2.5 player reachability games, we also use the notation $\mathcal{P} = (V_0, V_1, V_r, E, \text{prob}, \text{R})$.

**Definition 4:** A 2.5 player parity game (MPG) is a 2.5 player game $\mathcal{P}$ in which the winning condition win is defined by the **priority function** $\text{pri}: V \rightarrow \mathbb{N}$ mapping each vertex to a natural number. We call the image of $\pi$ the set of priorities (or: colours), denoted by $C$. Note that, since $V$ is finite, $C$ is finite as well. We extend $\text{pri}$ to plays, using $\text{pri}(\pi) = \lim_{i \rightarrow \infty} \text{pri}(\pi(i))$. Then, we have $\text{win} = \{ \pi \in \text{Play}(\mathcal{P}) \mid \text{pri}(\pi) \text{ is even} \}$. For 2.5 player parity games, we also use the notation $\mathcal{P} = (V_0, V_1, V_r, E, \text{prob}, \text{pri})$. We denote with $|\mathcal{P}|$ the size of a 2.5 player parity game, referring to the space its overall representation takes.

Note that in the above discussion we have defined strategies as mappings from vertices of the respective player to successor vertices. More general definitions of strategies exist that e.g. use randomised choices (imposing a probability distributions over the edges chosen) or take the complete history of the game so far into account. However, it is known that, for finite 2.5 player parity and reachability games, the simple pure memoryless strategies we have introduced above suffice to obtain mutually optimal infima and suprema [356].

We also use the common intersection and subtraction operations on directed graphs for arenas and games: given an MG $\mathcal{P}$ with arena $\mathfrak{A} = (V_0, V_1, V_r, E, \text{prob})$,

- $\mathfrak{A} \cap \mathfrak{A}'$ denotes the MG $\mathcal{P}'$ we obtain when we restrict the arena $\mathfrak{A}$ to $\mathfrak{A} \cap \mathfrak{A}' \overset{\text{def}}{=} (V_0 \cap V', V_1 \cap V', V_r \cap V', E \cap (V' \times V'), \text{prob}_{\mathfrak{A} \cap \mathfrak{A}'})$.
- for $E' \subseteq E_r$, we denote by $\mathfrak{A} \cap E'$ the MG $\mathcal{P}'$ we obtain when restricting arena $\mathfrak{A}$ to $\mathfrak{A} \cap E' \overset{\text{def}}{=} (V_0, V_1, V_r, E \cap E', \text{prob})$.

Note that the result of such an intersection may or may not be substoichiastic or contain sinks. While we use these operations freely in intermediate constructions, we make sure that, whenever they are treated as games, they have no sinks and are not substoichiastic.

**Definition 5:** Let $\mathcal{P} = (V_0, V_1, V_r, E, \text{prob}, \text{win})$ be a 2.5 player game, and let $f_0$ and $f_1$ two strategies for player 0 and 1 respectively. The value $\text{val}_{f_0, f_1}: V \rightarrow [0, 1]$ is defined as

$$\text{val}_{f_0, f_1}(v) \overset{\text{def}}{=} \text{P}^\mathcal{P}_{f_0}((\{ \pi \in \text{Play}(\mathcal{P}) \mid \pi \in \text{win} \}).$$

We also define

$$\text{val}_{f_0}(v) \overset{\text{def}}{=} \inf_{f_1 \in \text{Strats}_1} \text{val}_{f_0, f_1}(v),$$
$$\text{val}_{f_1}(v) \overset{\text{def}}{=} \sup_{f_0 \in \text{Strats}_0} \text{val}_{f_0, f_1}(v),$$
$$\text{val}^\mathcal{P}(v) \overset{\text{def}}{=} \sup_{f_0 \in \text{Strats}_0} \inf_{f_1 \in \text{Strats}_1} \text{val}_{f_0, f_1}(v).$$

We write $\text{val}^\mathcal{P} \geq \text{val}^\mathcal{P}_{f_r}$ if, for all $v \in V$, $\text{val}^\mathcal{P}_{f_r}(v) \geq \text{val}^\mathcal{P}(v)$ holds, and $\text{val}^\mathcal{P} > \text{val}^\mathcal{P}_{f_r}$ if $\text{val}^\mathcal{P}_{f_r} \geq \text{val}^\mathcal{P}$ and $\text{val}^\mathcal{P}_{f_r} \neq \text{val}^\mathcal{P}$ hold.

**Definition 6:** Given a vertex $v \in V$, a strategy $f_\sigma$ for Player $\sigma$ is called $\nu$-winning if, starting from $v$, Player $\sigma$ wins almost surely in the MDP defined by $f_\sigma$ (that is,
changes that are neutral for one player, in our case of Player 0, and either infers that the strategy is optimal, or offers a family \( I_f \) of strategies, such that, for all strategies \( f' \in I_f \), \( \text{val}^f_{P}(v) > \text{val}^{f'}_{P}(v) \).

The family \( I_f \) is usually given through profitable switches. In such a case, \( I_f \) is defined as follows.

**Definition 7:** Given a 2.5 player game \( P = (V_0, V_1, V_r, E, \text{prob}, \text{win}) \) and a strategy \( f \) for Player 0, the profitable switches, denoted \( \text{profit}(P, f) \), for Player 0 are the edges that offer a strictly higher chance of succeeding (under the given strategy). That is, \( \text{profit}(P, f) = \{ (v, v') \in E_0 \mid \text{val}^f_{P}(v') > \text{val}^f_{P}(v) \} \).

We also define the unprofitable switches according to profit as loss \( (P, f) = \{ (v, v') \in E_0 \mid \text{val}^f_{P}(v') < \text{val}^f_{P}(v) \} \).

\( I_f \) is the set of strategies that can be obtained from \( f \) by applying one or more profitable switches to \( f \): \( I_f = \{ f' \in \text{Strats} \mid f' \neq f \text{ and } \forall v \in V_0 : f'(v') = f(v) \text{ or } (v, f'(v')) \in \text{profit}(P, f) \} \).

Strategy improvement methods can usually start with an arbitrary strategy \( f_0 \), which is then updated iteratively by selecting some \( f_{i+1} \in I_{f_i} \) until \( I_{f_i} \) is eventually empty. This \( f_i \) is then guaranteed to be optimal. The update policy with which the profitable switches or switches are selected is not relevant for the correctness of the method, although it does impact on the performance and complexity of the algorithms. In our implementation, we use a greedy approach to update policy, that is we perform an update operation and terminate with the strategy that achieves the locally optimal switch.

For 2.5 player reachability games, strategy improvement algorithms provide optimal strategies.

**Theorem 1 (cf. [39]):** For a 2.5 player reachability game \( P \), a strategy improvement algorithm with the profitable switches \( I \) terminates with an optimal strategy for Player 0.

In the strategy improvement step, for all \( v \in V \) and all \( f' \in I_{f_f} \), it holds that \( \text{val}^P_{f_f}(v) = \text{val}^P_{f'}(f'(v)) \geq \text{val}^P_{f_f}(f(v)) = \text{val}^P_{f_0}(v) \). Moreover, strict inequality is obtained at some vertex in \( V \). As we have seen in the introduction, this is not the case for 2.5 player parity games: in the example from Figure 1 for a strategy \( f \) with \( f(v_0) = v_{0,55} \), the switch from edge \( e_{0,1} \) to \( e_{0,2} \) is not profitable. Note, however, that it is not unprofitable either.

**IV. ALGORITHM**

We observe that situations where the naive strategy improvement algorithm described in the previous section gets stuck are not always neutral; an improvement would be available, but among changes that are neutral in that applying them once would neither lead to an increased nor to a decreased likelihood of winning. As usual with strategy improvement algorithms, neutral switches cannot generally be added to the profitable switches: not only would one lose the guarantee to improve, one can also reduce the likelihood of winning when applying such changes.

Overcoming this problem is the main reason why strategy improvement techniques for MPG would currently have to use a reduction to 2.5 player player reachability games (or other reductions), with the disadvantages discussed in the introduction. We treat these tableaux directly and avoid reductions. We first make formal what neutral edges are.

**Definition 8:** Given a 2.5 player game \( P = (V_0, V_1, V_r, E, \text{prob}, \text{win}) \) and a strategy \( f \) for Player 0, we define the set of neutral edges \( \text{neutral}(P, f) \) as follows:

\[
\text{neutral}(P, f) = E_0 \cup \{ (v, v') \in E_0 \cup E_1 \mid \text{val}^f_{P}(v') = \text{val}^f_{P}(v) \}.
\]

Based on these neutral edges, we define an update policy on the subgame played on the neutral edges.

**Definition 9:** Given a 2.5 player game \( P = (V_0, V_1, V_r, E, \text{prob}, \text{win}) \) and a strategy \( f \) for Player 0, we define the neutral subgame of \( P \) for \( f \) as \( P' = P \cap \text{neutral}(P, f) \). Based on \( P' \) we define the set \( I_f' \) of additional strategy improvements as follows.

Let \( W_0 \) and \( W_0' \) be the winning regions of Player 0 on \( P \) and \( P' \), respectively. If \( W_0 = W_0' \), then \( I_f' = \emptyset \). Otherwise, let \( W \) be the set of strategies that are \( v \)-winning for Player 0 on \( P' \) for all vertices \( v \in W_0' \). Then we set

\[
I_f' = \left\{ f_0 \in \text{Strats} \mid \exists f_w \in W : \forall v \in W_0' : f_0(v) = f_w(v) \right\},
\]

\[
I_f' = \left\{ f' \in I_f' \mid \forall v \in W' : f'(v) = f(v) \right\}.
\]

We remark that \( W_0 \subseteq W_0' \) always holds. Intuitively, we apply a qualitative analysis on the neutral subgame, and if the winning region of Player 0 on the neutral subgame is larger than her winning region on the full game, we then perform the new winning strategy on the new part of the winning region. Intuitively, this forces Player 1 to leave this area eventually (or to lose almost surely). As he cannot do this through neutral edges, the new strategy for Player 0 is superior over the old one.

**Example 1:** Consider again the example MPG \( P_e \) from Figure 1 and the strategy such that \( f_0(v_0) = v_{0,55} \). Under this strategy, \( \text{neutral}(P_e, f_0) = E_r \cup \{(v_0, v_{0,55}), (v_0, v_1), (v_1, v_0)\} \); the resulting neutral subgame \( P_e' \) is the same as \( P_e \) except for the edge \( e_{1,1} \). In \( P_e' \), the winning region \( W_0' \) is \( W_0 = \{v_0, v_1, v_2\} \), while the original region was \( W_0 = \{v_0\} \). The two sets \( I_{f_0} \) and \( I_{f_0}' \) contain only the strategy \( f_0 \) such that \( f_0(v_0) = v_1 \). In order to avoid to lose almost surely in \( W_0' \), Player 1 has to change his strategy from \( f_1(v_1) = v_0 \) to \( f_1'(v_1) = v_{0,95} \) in \( P_e' \). Consequently, strategy \( f_0 \) is superior to \( f_0 \): the resulting winning probability is not 0.55 but 0.95 for \( v_0 \) and \( v_1 \).

Note that using \( I_f' \) in the strategy iteration has the same effect. Once a run has reached \( W_0' \) in the neutral subgame, it cannot leave it. Thus, changing the strategy \( f_0 \)
from $\mathcal{I}_f''$ to a strategy $f'$ with $f'(v) = f(v)$ for $v \in W_0$ and $f'(v) = f_0(v)$ for $v \notin W_0$ will not change the chance of winning: $\val^P_{f'} = \val^P_{f}$ and $\val^P_{f_0} = \val^P_{f}$. This also implies $\mathcal{I}_f'' \neq \emptyset \Rightarrow \mathcal{I}_f' \neq \emptyset$, since $\mathcal{I}_f'$ contains all strategies that belong to $\mathcal{I}_f''$ and that agree with $f$ only on the original winning region $W_0$. Using $\mathcal{I}_f'$ simplifies the proof of Lemma 1, but it also emphasizes that one does not need to re-calculate the strategy on a region that is already winning.

Our extended strategy improvement algorithm applies updates from either of these constructions until no further improvement is possible. That is, we can start with an arbitrary Player 0 strategy $f_0$ and then apply $f_{i+1} \in \mathcal{I}_{f_i} \cup \mathcal{I}_{f_0}$ until $\mathcal{I}_{f_i} = \emptyset$. We will show that therefore $f_i$ is an optimal Player 0 strategy.

For the algorithm, we need to calculate $\mathcal{I}_{f_0}$ and $\mathcal{I}_f'$. Calculating $\mathcal{I}_{f_0}$ requires only to solve 1.5 player parity games [40], and we use IsCASMC [43, 44] to do so. Calculating $\mathcal{I}_f'$ requires only qualitative solutions of neutral subgame $\mathcal{P}'$. For this, we apply the algorithm from [37].

A more algorithmic representation of our algorithm with a number of minor design decisions is provided in Appendix A. The main design decision is to favour improvements from $\mathcal{I}_f'$ over those from $\mathcal{I}_{f_0}$. This allows for calculating $\mathcal{I}_f'$ only if $\mathcal{I}_{f_0}$ is empty. Starting with calculating $\mathcal{I}_{f_0}$ first is a design decision, which is slightly arbitrary. We have made it because solving 1.5 player games quantitatively is cheaper than solving 2.5 player games qualitatively and we believe that the guidance for the search is, in practice, better in case of quantitative results. Likewise, we have implemented a ‘greedy switch all’ improvement strategy, simply because this is believed to behave well in practice. We have, however, not collected evidence for either decision and acknowledge that finding a good update policy is an interesting line of future research.

V. CORRECTNESS

A. Correctness proof in a nutshell

The correctness proof combines two arguments: the correctness of all basic strategy improvement algorithms for reachability games and a reduction from 2.5 player parity games to 2.5 player reachability games with arbitrarily close winning probabilities for similar strategy pairs. In a nutshell, if we approximate close enough, then three properties hold for a game $\mathcal{P}$ and a strategy $f$ of Player 0:

1) all ‘normal’ strategy improvements of the parity game correspond to strategy improvements in the reachability game (Corollary 2);
2) if Player 0 has a larger winning region $W'_0$ in the neutral subgame (cf. Definition 9 for $\mathcal{P} \cap \text{neutral}(\mathcal{P}, f)$) than for $\mathcal{P}_f$, then replacing $f$ by a winning strategy in $\mathcal{I}_{f_0}$ leads to an improved strategy in the reachability game (Lemma 1); and
3) if neither of these two types of strategy improvements are left, then a strategy improvement step on the related 2.5 player reachability game will not lead to a change in the winning probability on the 2.5 player parity game (Lemma 2).

B. Two game transformations

In this subsection we discuss two game transformations that change the likelihood of winning only marginally and preserve the probability of winning, respectively. The first transformation turns 2.5 player parity games into 2.5 player reachability games such that a strategy that is optimal strategy for the reachability game is also optimal for the parity game (cf. [38]).

Definition 10: Let $\mathcal{P} = (V_0, V_1, V_*, E, \text{prob}, \text{pri})$, and let $\varepsilon \in (0, 1)$ and $n \in \mathbb{N}$. We define the 2.5 player reachability game $\mathcal{P}_{\varepsilon,n} = (V_0, V_1, V_{\varepsilon,n}, E', \text{prob}', \text{win})$ with

- $V_{\varepsilon,n} = V_* \cup V' \cup \\{\text{won}, \text{lost}\}$, where $(i) V'$ contains primed copies of the vertices for ease of notation, the copy of a vertex $v$ is referred to as $v'$ in this construction; (ii) won and lost are fresh vertices; they are winning and a losing sink, respectively;
- $E' = \{(v, w') | (v, w) \in E\} \cup \{(\text{won}, \text{won}), (\text{lost}, \text{lost})\}$;
- $E'' = E' \cup \{(v', v) | v \in V\} \cup \{(v', \text{won}) | v \in V\} \cup \{(v', \text{lost}) | v \in V\}$;
- $\text{prob}'(v')(w') = \text{prob}(v)(w)$ for all $v \in V$ and $(v, w) \in E$;
- $\text{prob}'(v')(\text{won}) = \text{wprob}(\varepsilon, n, \text{pri}(v))$;
- $\text{prob}'(v')(\text{lost}) = \text{lprob}(\varepsilon, n, \text{pri}(v))$
- $\text{prob}'(v')(v') = 1 - \text{wprob}(\varepsilon, n, \text{pri}(v)) - \text{lprob}(\varepsilon, n, \text{pri}(v))$ for all $v \in V$; and
- $\text{prob}'(\text{won})(\text{won}) = \text{prob}'(\text{lost})(\text{lost}) = 1$.

where $\text{lprob}$, $\text{wprob} : (0, 1) \times \mathbb{N} \times \mathbb{N} \rightarrow [0, 1]$ are two functions with $\text{lprob}(\varepsilon, n, c) + \text{wprob}(\varepsilon, n, c) \leq 1$ for all $\varepsilon \in (0, 1)$ and $n, c \in \mathbb{N}$.

Intuitively, this translation replaces all the vertices by the gadgets from Figure 2.

Note that $\mathcal{P}_{\varepsilon,n}$ and $\mathcal{P}$ have similar memoryless strategies. By a slight abuse of the term, we say that a strategy $f_\sigma$ of Player $\sigma$ on $\mathcal{P}_{\varepsilon,n}$ is similar to her strategy $f'_\sigma$ on $\mathcal{P}$ if $f'_\sigma : v \mapsto f_\sigma(v')$ holds, i.e. when $v$ is mapped to $w$ by $f_\sigma$, then $v$ is mapped to $w'$ by $f'_\sigma$.

Theorem 2 (cf. [38]): Let $\mathcal{P} = (V_0, V_1, V_*, E, \text{prob}, \text{pri})$ be a 2.5 player parity game. Then, there exists $\varepsilon \in (0, 1)$, $n \geq |\mathcal{P}|$ such that we can construct $\mathcal{P}_{\varepsilon,n}$ and the following holds: For all strategies $f_0 \in \text{Strats}_0, f_1 \in \text{Strats}_1$, and all vertices $v \in V$, $|\val^P_{f_0,f_1}(v) - \val^P_{f_0,f_1}(v')| < \varepsilon$, $|\val^P_{f_0,f_1}(v) - \val^P_{f_0,f_1}(v')| < \varepsilon$, $|\val^P_{f_0}(v) - \val^P_{f_0}(v')| < \varepsilon$, $|\val^P_{f_0}(v) - \val^P_{f_0}(v')| < \varepsilon$, $|\val^P_{f_1}(v) - \val^P_{f_1}(v')| < \varepsilon$, and $|\val^P_{f_1}(v) - \val^P_{f_1}(v')| < \varepsilon$ holds, where $f_0$ resp. $f_1$ are similar to $f_0$ resp. $f_1$.

The results of [38] are stronger in that they show that the probabilities grow sufficiently slow for the reduction to be polynomial, but we use this construction only for correctness proofs and do not apply it in our algorithms. For this reason, existence is enough for our purpose. As [38] does not contain a theorem that directly makes the statement above, we have included a simple construction (without tractability claim) with a correctness proof in Appendix B.

We will now introduce a second transformation that allows us to consider changes in the strategies in many vertices at the same time.
Definition 11: Let $\mathcal{P} = (V_0, V_1, V_r, E, \text{prob, win})$ and a region $R \subseteq V$. Let $\mathcal{F}_R = \{ f : R \cap V_0 \to V | \forall v \in R : (v, f(v)) \in E \}$ denote the set of memoryless strategies for Player 0 restricted to $R$. The transformation results in a parity game $\mathcal{P}^R = (V'_0, V'_1, V'_r, E', \text{prob}', \text{pri}')$ such that

- $V'_0 = V_0 \cap R$, $V'_0 = (V_0 \cap R) \times \mathcal{F}_R$, and $V'_0 = V'_0 \cup V''_0$;
- $V'_1 = V_1 \cap R$, $V'_1 = (V_1 \cap R) \times \mathcal{F}_R$, and $V'_1 = V'_1 \cup V''_1$;
- $V'_r = V_r \cap R$, $V'_r = (V_r \cap R) \times \mathcal{F}_R$, and $V'_r = V'_r \cup V''_r$;
- $E' = \{(v, w) \in E | v \in V \cap R \}$ or $\{(v, w, f) | v, w \in R, (v, w) \in E \}$ and either $v \notin V_0$ or $f(v) = w$ or $\{(v, f) | v \in R, (v, w) \in E \}$ and either $v \notin V_0$ or $f(v) = w$;
- $\text{prob}'(v)(w) = \text{prob}(v)(w)$, $\text{prob}'((v, f))(w) = \text{prob}(v)(w)$, and $\text{prob}'((v, f), (v, f))(w) = \text{prob}(v)(w)$; and
- $\text{pri}'(v) = \text{pri}(v)$ for all $v \in V$ and $\text{pri}'((v, f)) = \text{pri}(v)$ otherwise.

Intuitively, the transformation changes the game so that, every time $R$ is entered, Player 0 has to fix her memoryless strategy in the game. The fact that in the resulting game the strategy $f$ for Player 0 is fixed entering $R$ is due to the jump from the original vertex $v$ to $(v, f)$ whenever $v \in R$. Once in $R$, either the part of $v$ is under the control of Player 1 or Player random, i.e. $v \notin V_0$, so it behaves as in $\mathcal{P}$, or the next state $w$ if $(w, f)$ if $w \in R$ is the outcome of $f$, i.e. $w = f(v)$.

It is quite obvious that this transformation does not impact on the likelihood of winning. In fact, Player 0 can simulate every memoryless strategy $f : V_0 \to V$ by playing a strategy $f_R : V'_0 \to V'$ that copies $f$ outside of $R$ (i.e. for each $v \in V_0 \cap R$, $f_R(v) = f(v)$) and moves to the $f | R$ for each $v \in V_0 \cap R$, $f_R(v) = (v, f | R)$ there is a one-to-one correspondence between playing in $\mathcal{P}$ with strategy $f$ and playing in $\mathcal{P}^R$ with strategy $f_R$ when starting in $V_0$.

Theorem 3: For all $v \in V$, all $R \subseteq V$, and all memoryless Player 0 strategies $f$, $\text{val}^P(v) = \text{val}^P((v, f | R))$,

\[ \text{val}^P(v) = \sup_{f \in \text{Strats}_0(\mathcal{P})} \text{val}^P((v, f | R)), \quad \text{val}^P(v) = \text{val}^P((v, f | R)) \]

C. Correctness proof

For a given game $\mathcal{P}$, we call an $\varepsilon \in (0, 1)$ small if it is at most $\frac{1}{2}$ of the smallest difference between all probabilities of winning that can occur on any strategy pair for any state in any game $\mathcal{P}^R$ for any $R \subseteq V$. For every small $\varepsilon$, we get the following corollary from Theorem 3.

Corollary 1 (preservation of profitable and unprofitable switches): Let $n \geq |\mathcal{P}|$, let $f$ be a Player 0 strategy for $\mathcal{P}$, $f'$ the corresponding strategy for $\mathcal{P}_{\varepsilon,n}$, $\varepsilon$ small, $v \in V$, $w = f(v)$, and $(v, u) \in E$. Then $\text{val}^P(v) > \text{val}^F(w)$ implies $\text{val}^P(v, n) > \text{val}^F(w, n)$, and $\text{val}^P(v) < \text{val}^F(w)$ implies $\text{val}^P(v, n) < \text{val}^F(w, n)$.

It immediately follows that all combinations of profitable switches can be applied, and will lead to an improved strategy: for small $\varepsilon$, a profitable switch for $f_1$ from $f_1(v) = w$ to $f_1(v) = u$ implies $\text{val}^P(v) > \text{val}^P(w) + 5\varepsilon$ since by definition, we have that $\text{val}^P(v) > \text{val}^P(w)$ (as the switch is profitable); in particular, $\text{val}^P(v) = \text{val}^P(w) + \delta$ with $\delta \in \mathbb{R}^+0$, since $\varepsilon \leq \frac{1}{2}$, we have that $\text{val}^P(v) \geq \text{val}^P(w) + 5\varepsilon$.

Consequently, since under $f_{i+1}$ we have that $\text{val}^P_{i+1}(v') = \text{val}^P_{i+1}(v')$, it follows that $\text{val}^P_{i+1}(v) \geq \text{val}^P_{i+1}(v) + 3\varepsilon$, and, using triangulation again, we get $\text{val}^P_{i+1}(v) \geq \text{val}^P_{i+1}(v) + \varepsilon$.

Thus, we have the following corollary.

Corollary 2: Let $\mathcal{P}$ be a given 2.5 player parity game, and $f_i$ be a strategy with profitable switches ($\text{profit}(f_i) \neq \emptyset$). Then, $\mathcal{I}_{f_i} \neq \emptyset$, and for all $f_{i+1} \in \mathcal{I}_{f_i}$, $\text{val}^P_{i+1} > \text{val}^P_{i}$.

We now turn to the case where there are no profitable switches for $f$ in the game $\mathcal{P}$. Corollary 1 shows that, for the corresponding strategy $f'$ in $\mathcal{P}_{\varepsilon,n}$, all profitable switches lie within the neutral edges for $f$ in $\mathcal{P}$, provided $f$ has no profitable switches.

We expand the game by fixing the strategy of Player 0 for the vertices in $R \cap V_0$ for a region $R \subseteq V$. The region we are interested in is the winning region of Player 0 in the neutral subgame $\mathcal{P} \cap \text{neutral}(\mathcal{P}, f)$. The game is played as follows.

For every strategy $f_R : R \cap V_0 \to V$ such that $(r, f_R(r)) \in E$ holds for all $r \in R$, the game has a copy of the original game interacted with $R$, where the choices of Player 0 on the vertices in $R$ are fixed to the single choice defined by the respective strategy $f_R$. We define $\|\mathcal{P}\| = \max\{ |R| | R \subseteq V \}$.

We consider the case where the almost sure winning region of Player 0 in the neutral subgame $\mathcal{P}' = \mathcal{P} \cap \text{neutral}(\mathcal{P}, f)$ is strictly larger than her winning region in $\mathcal{P}_f$.

Lemma 1: Let $\mathcal{P}$ be a given 2.5 player parity game, and $f_i$ be a strategy such that the winning region $W_0^i$ for Player 0 in the neutral subgame $\mathcal{P}' = \mathcal{P} \cap \text{neutral}(\mathcal{P}, f_i)$ is strictly larger than her winning region $W_0$ in $\mathcal{P}_f$. Then $\mathcal{I}_{f_i} \neq \emptyset$, and $\forall f_{i+1} \in \mathcal{I}_{f_i}$, $\text{val}^P_{i+1} > \text{val}^P_i$.

Proof. The argument is an extension of the common argument for strategy improvement made for the modified reachability game. We first recall that the strategies in $\mathcal{I}_{f_i}$ differ from $f_i$ only on the winning region $W_0^i$ of Player 0 in the neutral subgame $\mathcal{P}'$. Assume that we apply the change once: the first time $W_0^i$ is entered, we play the new strategy, and after it is left, we play the old strategy. If the reaction of Player 1 is to stay in $W_0^i$, Player 0 will win almost surely in $\mathcal{P}$. If he leaves it, the value is improved due to the fact that Player 1 has to take a disadvantageous edge to leave it.

Consider the game $\mathcal{P}_{W_0}^i$ and fix $f_{i+1} \in \mathcal{I}_{f_i}$. Using Theorem 3 this implies that, when first in a state $v \in W_0^i$, Player 0 moves to $(v, f_{i+1})$ for some $f_{i+1} \in \mathcal{I}_{f_i}$, then the likelihood of winning is either improved or 1. For any counter strategy of Player 1. For all $v \in W_0^i \setminus W_0$, this implies a strict improvement. For an $n \geq |\mathcal{P}|$ and a small $\varepsilon$, we can now follow the same arguments as for the Corollaries 1 and 2 on $\mathcal{P}_{W_0}^i$ to establish that $\text{val}^P_{i+1}(v) > \text{val}^P_i(v)$.
\[ \text{val}^{P_{W_0}^n}_{f_1}\text{w}_0 \text{ holds, where the inequality is obtained through the same steps: } \text{val}^{P_{W_0}^n}_{f_1}\text{w}_0((v,f_1+1|w_0)) > \text{val}^{P_{W_0}^n}_{f_1}\text{w}_0(v) \text{ implies } \text{val}^{P_{W_0}^n}_{f_1}\text{w}_0((v,f_1+1|w_0)) \geq \text{val}^{P_{W_0}^n}_{f_1}\text{w}_0(v) + 5\varepsilon; \text{ this implies } \text{val}^{P_{W_0}^n}_{f_1}\text{w}_0((v,f_1+1|w_0)) \geq \text{val}^{P_{W_0}^n}_{f_1}\text{w}_0(v) + 3\varepsilon; \text{ and this implies } \text{val}^{P_{W_0}^n}_{f_1}\text{w}_0(v) + 3\varepsilon \text{ and we finally get } \text{val}^{P_{W_0}^n}_{f_1}\text{w}_0((v,f_1+1|w_0)) \geq \text{val}^{P_{W_0}^n}_{f_1}\text{w}_0(v). \]

With Theorem 3 we obtain that \( \text{val}_{f_{i+1}}^P > \text{val}_{f_i}^P \) holds. □

Let us finally consider the case where there are no profitable switches for Player 0 in \( P_{f_i} \) and her winning region on the neutral subgame \( \mathcal{P} \cap \text{neutral}(\mathcal{P}, f_i) \) coincides with her winning region in \( P_{f_i} \).

**Lemma 2:** Let \( \mathcal{P} \) be an MPG and \( f_i \) be a strategy such that the set of profitable switches is empty and the neutral subgame \( \mathcal{P} \cap \text{neutral}(\mathcal{P}, f_i) \) has the same winning region for Player 0 as her winning region in \( P_{f_i} \). Then, every individual profitable switch in the reachability game \( \mathcal{P}_{\varepsilon,n} \) from \( f_i \) to \( f_{i+1} \) implies \( \text{val}^P_{f_{i+1}} = \text{val}^P_{f_i} \) and \( \text{neutral}(\mathcal{P}, f_{i+1}) = \text{neutral}(\mathcal{P}, f_i) \).

**Proof.** When there are no profitable switches in the parity game \( P \) for \( f_i \), then all profitable switches in the reachability game \( \mathcal{P}_{\varepsilon,n} \) for \( f_i \) (if any) must be within the set of neutral edges \( \text{neutral}(\mathcal{P}, f_i) \) in the parity game \( P \). We apply one of these profitable switches at a time. By our definitions, this profitable switch is neutral in the 2.5 player parity game.

Taking this profitable (in the reachability game \( \mathcal{P}_{\varepsilon,n} \) for a small \( \varepsilon \) and some \( n \geq \|P\| \)) switch will improve the likelihood of winning for Player 0 in the reachability game. By our definition of \( \varepsilon \), this implies that the likelihood of winning cannot be decreased on any position in the parity game.

To see that the quality of the resulting strategy cannot be higher for Player 0 in the 2.5 player parity game, recall that Player 1 can simply follow his optimal strategy on the neutral subgame. The likelihood of winning for Player 0 is the likelihood of reaching her winning region, and this winning region has not changed. Moreover, consider the evaluation of the likelihood of reaching this winning region: since by fixing the strategy for Player 1 the resulting game is an MDP, such an evaluation can be obtained by solving a linear programming problem (cf. the arXiv version for more details). The old minimal non-negative solution to the resulting linear programming problem is a solution to the new linear programming problem, as it satisfies all constraints.

Putting these arguments together, likelihood of winning in the parity game is not altered in any vertex by this change. Hence, the set of neutral edges is not altered. □

This lemma implies that none of the subsequently applied improvement steps applied on the 2.5 player reachability game has any effect on the quality of the resulting strategy on the 2.5 player parity game. Together, the above lemmas and corollaries therefore provide the correctness argument.

**Theorem 4:** The algorithm is correct.

**Proof.** Lemma 2 shows that, when \( I_{f_i} \) and \( I_{f'_i} \) are empty (i.e. when the algorithm terminates), then the updates in the related 2.5 player reachability game will henceforth (and thus until termination) not change the valuation for the 2.5 player parity game. With Theorems 1 and 2 and our selection of small \( \varepsilon \), it follows that \( f_i \) is an optimal strategy. The earlier lemmas and corollaries in this subsection show that every strategy \( f_{i+1} \in I_{f_i} \cup I_{f'_i} \) satisfies \( \text{val}_{f_{i+1}} > \text{val}_{f_i} \). Thus, the algorithm produces strategies with strictly increasing quality in each step until it terminates. As the game is finite, then also the set of strategies is finite, thus the algorithm will terminate after finitely many improvement steps with an optimal strategy. □

As usual with strategy improvement algorithms, we cannot provide good bounds on the number of iterations. As reachability games are a special case of 2.5 player games, all selection rules considered by Friedmann [45], [46] will have exponential lower bounds.

**VI. IMPLEMENTATION AND EXPERIMENTAL RESULTS**

We have written a prototypical implementation for the approach of this paper. Our tool supports the input language of the probabilistic model checker PRISM-GAMES [47], an extension of PRISM [48] to stochastic Markov games. As case study, we consider a battlefield consisting of \( n \times n \) square tiles, surrounded by a solid wall. On the battlefield there are two robots, \( R_0 \) and \( R_1 \), and four marked zones \( z_{01}, \ldots, z_{04} \) at the corners, each of size \( 3 \times 3 \). Each tile can be occupied by at most one robot at a time. The robots act in strict alternation.

When it is the turn of a robot, this robot can move as follows: decide a direction and move one field forward; decide a direction and attempt to move two fields forward. In the latter case, the robot moves two fields forward with a probability of 50%, but only one field forward with a probability of 50%. If the robot would run into a wall or into the other robot, it stops at the field before the obstacle. Robot \( R_1 \) can also shoot \( R_0 \) instead of moving, which is destroyed with probability \( p_{\text{dstr}} \) where \( p_{\text{dstr}} \) is the probability of destroying the robot and \( d \) is the Euclidean distance between the two robots. Once destroyed, \( R_0 \) cannot move any more. We assume that we are in control of \( R_0 \) but cannot control the behaviour of \( R_1 \). Our goal is to maximise, under any possible behaviour of \( R_1 \), the probability of fulfilling a certain objective depending on the zones, such as repeatedly visiting all zones infinitely often, visiting the zones in a specific order, performing such visits without entering other zones in the meanwhile, and so on. As an example, we can specify that the robot eventually reaches each zone by means of the probabilistic LTL (PLTL) formula \( \langle\langle R_0 \rangle\rangle P_{\max=?}\left[\bigwedge_{i=1,...,4} F z_{0i}\right] \) requiring to maximise the probability of satisfying \( \bigwedge_{i=1,...,4} F z_{0i} \) by controlling \( R_0 \) only.

The machine we used for the experiments is a 3.6 GHz Intel Core i7-4790 with 16GB 1600 MHz DDR3 RAM of which
We have combined a recursive algorithm for the quantitative solution of 2.5 player parity games with a strategy improvement algorithm, which lifts these results to the qualitative solution of 2.5 player parity games. This shift is motivated by the significant acceleration in the quantitative solution of 2.5 player parity games: while [49] scaled to a few thousand vertices, [37] scales to tens of millions of states. This changes the playing field and makes qualitative synthesis a realistic target. It also raises the question if this technique can be incorporated smoothly into a quantitative solver.

Previous approaches [27], [28] have focused on developing techniques that give preference to the likelihood of winning, and overcome stalling by performing strategy improvement on the larger qualitative game from [36] on the value classes. This approach was reasonable at the time, where the updates benefited from memorising the recently successful strategies. Moreover, focussing on value classes keeps the part of the qualitative game under consideration small, which is a reasonable approach when the cost of qualitative strategy improvement is considered significant. Building on a fast solver for the qualitative analysis, we can afford to progress in larger steps.

The main advancement, however, is as simple as it is effective. We use strategy improvement where it has a simple direct meaning (the likelihood to win), and we do not use it where the progress measure is indirect (progress measure within a value class). This has allowed us to transfer the recent performance gains from qualitative solutions of 2.5 player parity games [37] to their quantitative solution.

The difference in performance also explains the difference in the approach regarding complexity. Just as the deterministic subexponential complexity of solving 2.5 player games qualitatively is not very relevant in [37] (as this approach would be very slow in practice), the expected subexponential complexity in [27] is bought by exploiting a random facet

12GB assigned to the tool; the timeout has been set to 30 minutes. We have applied our tool on a number of properties that require the robot $R_0$ to visit the different zones in a certain order. In Table I we report the performance measurements for these properties. Column “property” shows the PLTL formula we consider, column “$n$” the width of the battlefield instance, and column “$b$” the number of bullets $R_1$ can shoot. For the “MPG” part, we present the number of “vertices” of the resulting MPG and the number of “colours”. In the remaining columns, for each value of “$p_{detr}$”, we report the achieved maximum probability “$p_{max}$” and the time “$t_{sol}$” in seconds needed to solve the game. Note that we cannot compare to PRISM-GAMES because it does not support general PLTL formulas, and we are not aware of other tools to compare with.

As we can see, the algorithm performs quite well on MPG with few million states. It is worth mentioning that a large share of the time spent is due to the evaluation of the 1.5 player parity games in the construction of the profitable switches. For instance, such an evaluation required 137 seconds out of 172 for the case $n = 9, b = 5,$ and $p_{detr} = 0.1$. Since a large part of these 1.5 player games are similar, we are investigating how to avoid the repeated evaluation of similar parts to reduce the running time. Generally, all improvements in the quantitative solution of 1.5 player parity games and the qualitative solution of 2.5 player parity games will reduce the running time of our algorithm.

VII. DISCUSSION

We have combined a recursive algorithm for the quantitative solution of 2.5 player parity games with a strategy improvement algorithm, which lifts these results to the qualitative solution of 2.5 player parity games. This shift is motivated by the significant acceleration in the quantitative solution of 2.5 player parity games: while [49] scaled to a few thousand vertices, [37] scales to tens of millions of states. This changes the playing field and makes qualitative synthesis a realistic target. It also raises the question if this technique can be incorporated smoothly into a quantitative solver.

Previous approaches [27], [28] have focused on developing a progress measure that allows for joining the two objective. This has been achieved in studying strategy improvement techniques that give preference to the likelihood of winning, and overcome stalling by performing strategy improvement on the larger qualitative game from [36] on the value classes. This approach was reasonable at the time, where the updates benefited from memorising the recently successful strategies on the qualitative game. Moreover, focussing on value classes keeps the part of the qualitative game under consideration small, which is a reasonable approach when the cost of qualitative strategy improvement is considered significant. Building on a fast solver for the qualitative analysis, we can afford to progress in larger steps.

The main advancement, however, is as simple as it is effective. We use strategy improvement where it has a simple direct meaning (the likelihood to win), and we do not use it where the progress measure is indirect (progress measure within a value class). This has allowed us to transfer the recent performance gains from qualitative solutions of 2.5 player parity games [37] to their quantitative solution.

The difference in performance also explains the difference in the approach regarding complexity. Just as the deterministic subexponential complexity of solving 2.5 player games qualitatively is not very relevant in [37] (as this approach would be very slow in practice), the expected subexponential complexity in [27] is bought by exploiting a random facet
APPENDIX

In this section we provide the details of the algorithm presented in the main part of the paper. It is an implementation of the algorithm we have described in Section IV and contains our design decisions. They are not relevant for correctness. We consider the extended game in Figure 4 for our running example.

We start with an initialisation, where we solve 2.5 player parity games qualitatively, and may require the qualitative solution of subgames. For this initialisation, we first define an extended qualitative solution of 2.5 player game as a memoryless strategy for a player, which guarantees that s/he wins almost surely on his or her almost sure winning region and only loses almost surely on the almost sure winning region of his or her opponent (cf. Section A).

We initialise our strategy improvement algorithm with an extended qualitative solution. To obtain such a solution, we can essentially use the algorithms known from ordinary qualitative solutions, cf. Definition 12.

The normal improvement step is an instance of standard strategy improvement algorithms. We evaluate the likelihood of winning for Player 0 for her current strategy, by computing the value optimal for Player 1 against this strategy. If we can obtain an improvement by changing a decision in a Player 0 state, we do so.

The correctness proof in the main part of the paper uses the related 2.5 player reachability game as a comparison point in the correctness argument. It shows that, for sufficiently small \( \varepsilon \), each improvement selected by the algorithm is also an improvement in the related reachability game, while the stopping condition guarantees that further improvements in the reachability games do not translate to further improvements in the Markov parity game.

Consequently, our technique not only avoids using the tiny \( \varepsilon \), it also avoids the slow progression through updates that are stale (lead to no improvement) on the parity game while leading to an improvement on the reachability game resulting from the translation.

Our main algorithm is given as MAIN(\( \cdot \)) in Algorithm 1.

It makes use of the auxiliary algorithms from Definition 12.

Definition 12: For an MPG \( \mathcal{P} = (V_0, V_1, V_r, E, \text{pri}) \) we denote by QUALISOLVE(\( \mathcal{P} \)) = (\( W, f \)) a method which computes the winning regions \( W \) of Player 0 and a Player 0 strategy \( f \) winning in \( W \) and arbitrary defined elsewhere.

Further, for \( A \subseteq V \), we let \( \text{REACH}(\mathcal{P}, A) = \text{val} \) denote the result of computing mutually optimal reachability probabilities, that is \( \text{val} = \text{val}' \mathcal{P}' \) where \( \mathcal{P}' \) is the reachability game \( \mathcal{P}' = (V_0, V_1, V_r, E, A) \).

For an MPG \( \mathcal{P} = (V_0, \emptyset, V_r, E, \text{pri}) \), whose arena is an MDP, we let EVALUA TedMDP(\( \mathcal{P} \)) = \( \text{val} \) denote the value of the MDP, that is \( \text{val} = \text{val}' \mathcal{P} \).

QUALISOLVE(\( \mathcal{P} \)) can be effectively implemented by 37 without having to construct intermediate 2 player games.

Note that efficient procedures for evaluating parity MDPs exist 40. Having to control only a single player, such algorithms merely need to determine the almost sure winning region of this single player (which is simple) and compute the maximal probability to reach this region (cf. Appendix B). (This does not hold for general 2.5 player parity games.)

In our setting, we obtain a parity MDP by fixing the strategy for Player 0. We therefore have to compute the minimal values for reachability. We can, however, transform this parity game by adding 1 to the parity labels, so to complement the winning condition, computing the maximal winning values for such a complement, and finally returning 1 minus the value computed for each node.

A. Initialisation

1) Determine the almost sure winning region \( W \) (and a winning strategy for it) for Player 0 (Line 5). If the winning region is empty, i.e. \( W = \emptyset \), return (Line 7-9).
2) Solve the remaining game as reachability game with the reachability objective to reach the winning region \( W \) (Line 10), obtaining the value \( \text{val} \) and the strategy \( g \) for the vertices with non-zero value.
3) Improve the strategy according to the reachability computation (Line 11-13). Call this algorithm recursively for the sub-game that contains only the states with reachability probability 0, improving the strategy using results from the recursive calls (Line 14-18).

The main task of the initialisation phase performed by INITIALIZE(\( \mathcal{P} \)) is to provide a strategy \( f \) for Player 0 under which the winning probability \( \text{val}' \mathcal{P}'(v) \) lies in \([0,1]\) for each vertex \( v \) having the optimal winning probability \( \text{val}' \mathcal{P}'(v) \) in \([0,1]\) while the strategy is winning in the region \( W \) where Player 0 wins almost surely. We call such strategies realising.

Note that the correctness of the algorithm does not rely on using realising strategies, it is merely a heuristic. It is chosen to avoid that the algorithm gets stuck by a too large initial winning region of player 1.

Example 2: We apply INITIALISE(\( \cdot \)) on the MPG \( \mathcal{P}_x \) from Figure 4. The most significant steps of the algorithm on \( \mathcal{P}_x \) are depicted in the pictures shown in Algorithm 1. The winning region from Line 6 is \( W = \{v_0\} \), depicted as a dashed box. Because \( v_0 \) and \( v_3 \) are outside the winning region, the
Algorithm 1 Quantitative parity game solving algorithm

1: function $\text{MAIN}(\mathcal{P})$
2:   $f \leftarrow \text{INITIALISE}(\mathcal{P})$
3:   return $\text{IMPROVE}(\mathcal{P}, f)$
4: end function

5: function $\text{INITIALISE}(\mathcal{P})$
   where $\mathcal{P} = (V_0, V_1, V_r, E, \text{prob, pri})$ and $V = V_0 \cup V_1 \cup V_r$
6:   $(W, f) \leftarrow \text{QUALISOLVE}(\mathcal{P})$
7:   if $W = \emptyset$ then
8:      return $f$
9:   end if
10:  $(\text{val, } g) \leftarrow \text{REACH}(\mathcal{P}, W)$
11:  for $v \in V_0 \setminus W$ do
12:     $f(v) \leftarrow g(v)$
13:  end for
14:  $V' \leftarrow \{ v \in V \mid \text{val}(v) = 0 \}$
15:  $g \leftarrow \text{INITIALISE}(\mathcal{P} \cap V')$
16:  for $v \in V_0 \cap V'$ do
17:     $f(v) \leftarrow g(v)$
18:  end for
19:  return $f$
20: end function

21: function $\text{IMPROVE}(\mathcal{P}, f)$
   where $\mathcal{P} = (V_0, V_1, V_r, E, \text{prob, pri})$
   $V = V_0 \cup V_1 \cup V_r$, $E_0 = E \cap V_0 \times V$, $f$ is a strategy
22: repeat
23:    $g \leftarrow f$
24:    $\mathcal{P}' \leftarrow \mathcal{P}_f$
25:    $\text{val} \leftarrow \text{EVALUATEMDP}(\mathcal{P}')$
26:    for $(v, v') \in E_0$ with $\text{val}(v') > \text{val}(v)$ do
27:       $f(v) \leftarrow v'$
28:    end for
29:    if $f = g$ then
30:       $\mathcal{P}' \leftarrow \mathcal{P} \cap \text{neutral}(\mathcal{P}, \text{val})$
31:       $(W, h) \leftarrow \text{QUALISOLVE}(\mathcal{P}')$
32:       for $v \in W$ do
33:          $f(v) \leftarrow h(v)$
34:       end for
35:    end if
36: until $f = g$
37: return $f$
38: end function
choice of their edges is arbitrary and we can assume that QualiSolve(·) returns a strategy in which the edges \( e_{0,2} \) and \( e_{3,1} \) are chosen, depicted as full edges. The dashed edges are those under the control of Player 1; the dash-dotted edges are those randomly taken. As the winning region is nonempty, we do not return in Lines 12. The reachability computation in Line 10 and the following updates now set the choice for \( v_3 \) to \( e_{0,1} \); regarding \( v_3 \), since initially the value for \( v_3 \) is 0, the initial choice for \( v_2 \) is \( e_{2,2} \), so for \( v_3 \) the choice between \( e_{3,1} \) and \( e_{3,2} \) is irrelevant. After the updates, the only nodes with winning probability 0 are \( v_1 \), \( v_2 \), and \( v_3 \); this means that, in Line 15 we call the function recursively with the game consisting of the nodes \( \{v_1, v_2, v_3\} \). The winning region is now \( W = \{v_2, v_3\} \), obtained by Player 0 by choosing \( e_{3,2} \), so the following reachability computation in Line 10 and updates maintain the choice for \( v_3 \) to \( e_{3,2} \). The only state still with value 0 is \( v_1 \), so the recursive call has as argument a game consisting only of the node \( v_1 \). However, this recursive step is already left at Line 8 because the winning region is empty. As there are no Player 0 nodes in such a one-node game, the recursive call does not lead to further updates of the strategy.

B. Strategy improvement step

Input is a strategy and a parity game. Output is a superior strategy and a parity game – or an optimality result for the given strategy.

1) take a strategy \( f \) for Player 0 (Line 21)
2) construct the parity MDP for it (Line 22)
3) evaluate the parity MDP (Line 29)
4) if there are profitable switches: choose & return update among them (Lines 26-28) else (Lines 29-35)
   a) build the sub-game that only uses neutral edges (Line 30)
   b) determine almost sure winning region & strategy (Line 31) on it for Player 0
   c) if the region is not empty, update the strategy accordingly. That is, in this winning region, update the strategy such that it is winning. (Lines 32-54)
   d) if the region is empty, terminate \( f \) is optimal (Line 55)

This step is repeated until \( f \) is found to be optimal by the algorithm.

Example 3: We reconsider the MPG \( P_\epsilon \) from Figure 4 and the strategy \( f \) from Example 2 with \( f(v_0) = e_{0,1} \). The evaluation of the induced MDP in Line 29 leads to a value of 0.55 in \( v_0 \) and \( v_1 \). There are no profitable switches, so Lines 26-28 do not lead to any changes of \( f \). The subgame \( P' \) computed in Line 30 does not contain \( e_{1,1} \), because choosing this edge would lead to a value of 0.95, which is worse than 0.55 for Player 1. Evaluating \( P' \) in Line 31 we see that \( v_0 \) and \( v_1 \) are now part of the winning region, because Player 1 cannot leave it using \( e_{1,1} \). The choice for \( v_0 \) is thus updated to \( e_{0,2} \). In the next iteration of the improvement loop, there are neither profitable switches, nor does the subgame of neutral edges lead to any improvement. Thus, the algorithm terminates at this point.

Consider a reachability MDP \( P = (V_0, 0, V_r, E, \text{prob}, R) \). Then we have that, for each \( v \in V \), we have \( \text{val}^P(v) = w_v \), where \( w \) is the solution vector obtained from the following linear programming problem:

\[
\begin{align*}
\text{minimise} & \sum \{w_v \ | \ v \in V\} \\
\text{subject to} & \notag \\
& w_v \geq 0 \quad \forall v \in V \\
& w_v \leq 1 \quad \forall v \in R \\
& w_v \geq w_{v'} \quad \forall v \in V_0, (v, v') \in E \\
& w_v \geq \sum \{\text{prob}(v)(v') \cdot w_{v'} \ | \ (v, v') \in E\} \quad \forall v \in V_r
\end{align*}
\]

The reduction from parity to reachability games from [38] focuses on the tractability of the reductions. This has left us in a tight spot between re-doing a simplified version of the construction – which is arguably not required – and referring to a complicated construction that consists of many steps and that does not provide a theorem, which directly makes the claim we make in Theorem 2. For this reason, we provide a reduction below. Note that we make no claim regarding tractability. This is not required, as the resulting game only occurs in proofs, but is not used in our algorithm.

We use the translation from \( P \) to \( P_{\epsilon, n} \) with

\[
l_{\text{prob}}(\epsilon, n, c) = \begin{cases} 
0 & \text{if } c \text{ is even,} \\
\delta^{c+1} & \text{if } c \text{ is odd}
\end{cases}
\]

and

\[
w_{\text{prob}}(\epsilon, n, c) = \begin{cases} 
\delta^{c+1} & \text{if } c \text{ is even,} \\
0 & \text{if } c \text{ is odd}
\end{cases}
\]

where a suitable \( \delta \in (0, 1) \) exists with the properties we require for given \( n \) and \( c \). (Details follows.)

That is, we obtain the gadget construction from Figure 5. We refer to the translation as \( P_{\delta}^\epsilon \). The main observation when looking at \( \delta > 0 \) is to follow what happens if we let \( \delta \) shrink towards 0.

Let \( f_0 \) and \( f_1 \) be strategies of Player 0 and 1 for \( P \) and \( f_0' \) and \( f_1' \) their similar strategies for \( P_{\delta}^\epsilon \) (defined in the same way as Theorem 2). Let \( L \subseteq V \) be a leaf component (a strongly connected component without outgoing edges) in \( P_{f_0, f_1} \). Recall that runs almost surely reach (and then get trapped) in some leaf component for Markov chains. For Markov chains with a parity acceptance condition, the runs that reach a leaf component \( L \) are almost surely accepting if the minimal priority \( c_L = \min \{\text{pri}(v) \ | \ v \in L\} \) of the states in \( L \) is even, and they are almost surely losing if \( c_L \) is odd.

We make the following simple observations for \( P_{f_0', f_1'}^\delta \).

1) If \( \delta \) goes to 0, the chance of reaching states in \( L' = L \cup \{v' \ | \ v \in L\} \) in \( P_{f_0', f_1'}^\delta \) from a state \( v_0 \in V \) goes to the chance of reaching \( L \) in \( P_{f_0, f_1} \) from \( v_0 \).

2) When starting in \( L' \), there is a \( \delta' \in (0, 0.5) \) such that, for all \( \delta \in (0, \delta') \), the chance of leaving to won or lost
before visiting a vertex \( v' \) (the primed copy of \( v \)) with \( \text{pri}(v) = c_L \) is smaller than \( \delta^{c+1.5} \).

(3) When starting in \( L' \) where \( c_L \) is even, there is a \( \delta' > 0 \) such that, for all \( \delta \in (0, \delta') \), the chance of reaching lost \( L \) is at most \( \sqrt[3]{\delta} \) times the chance of reaching won.

(4) When starting in \( L' \) where \( c_L \) is odd, there is a \( \delta' > 0 \) such that, for all \( \delta \in (0, \delta') \), the chance of reaching won \( L \) is at most \( \sqrt[3]{\delta} \) times the chance of reaching lost.

(5) When \( \delta \) goes to 0, the chance of reaching the won from a state \( v_0 \in V \) goes to the chance of winning in \( \mathcal{P}_{f_0,f_1} \) from \( v_0 \).

To see (1), if we want to approach the likelihood with precision \( 2\varepsilon \), we choose an \( n \) such that \( L' \) is reached after more than \( n \) states with chance \( < n \), and then choose \( \delta \in (0, 1) \) such that \( (1 - \delta)^n > 1 - \varepsilon \). The latter property provides that the difference in the chance of reaching \( L \) in \( n \) steps and reaching \( L' \) in \( 2n \) steps is less than \( \varepsilon \).

To see (2), consider that, as \( L \) is a leaf component, there is positive probability from every vertex \( w \in L \) to reach a vertex \( v \) with minimal \( \text{pri}(v) = c_L \) within \( n = |L| \) steps in \( \mathcal{P}_{f,g} \). Let \( p_{\text{min}} > 0 \) be the smallest such probability. Then, in \( L' \), a vertex \( v' \), which is the primed copy of a vertex \( v \) with minimal \( \text{pri}(v) = c_L \), can be reached within \( 2n \) steps in \( \mathcal{P}_{f',g'}^\delta \) with chance at least \( p_{\text{min}} (1 - \delta^{c+2})^n > \frac{p_{\text{min}}}{2^n} \). The chance to reach won or lost within \( n \) steps and without reaching such a vertex \( v' \) first is at most \( n \cdot \delta^{c+2} \).

Consequently, the chance of reaching won or lost prior to reaching a vertex \( v' \) which is the primed copy of a vertex \( v \) with minimal colour \( \text{pri}(v) = c_L \) is at most \( \frac{n \cdot \delta^{c+2}}{n \cdot \delta^{c+2} + \frac{p_{\text{min}}}{2^n}} \). If we choose \( \delta \) small enough that \( \delta^{c+2} < \frac{p_{\text{min}}}{n^2} \) and \( \sqrt[3]{\delta} < \frac{2p_{\text{min}}}{n^2} \), then we get

\[
\frac{n \cdot \delta^{c+2}}{n \cdot \delta^{c+2} + \frac{p_{\text{min}}}{2^n}} < \frac{n \cdot 2^n \cdot \delta^{c+2}}{2p_{\text{min}}} < \delta^{c+1.5},
\]

which provides the claim.

(3) and (4) follow immediately when \( \delta \) is small enough such that \( \frac{\delta^{c+1.5}}{1-\delta^{c+1.5}} = \frac{1}{\sqrt[3]{\delta}} < \sqrt[3]{\delta} \) holds.

(5) finally follows with the previous points and the observation that won and lost are the only leaf components in \( \mathcal{P}_{f_0,f_1}^{L'} \), and are therefore reached almost surely.

This establishes all properties we need for Theorem 2.