Mean-payoff games with partial observation

Reference:
Hunter Paul, Pauly Arno, Pérez Guillermo Alberto, Raskin Jean-François.- Mean-payoff games w ith partial observation
Theoretical computer science - ISSN 0304-3975 - 735:SI(2018), p. 82-110
Full text (Publisher's DOI): https://doi.org/10.1016/J.TCS.2017.03.038

Mean-payoff Games with Partial Observation*

Paul Hunter, Arno Pauly, Guillermo A. Pérez† and Jean-François Raskin
Département d'Informatique, Université libre de Bruxelles
{paul.hunter, apauly, guillermo.perez, jraskin} @ulb.ac.be

August 7, 2018

Abstract

Mean-payoff games are important quantitative models for open reactive systems. They have been widely studied as games of full observation. In this paper we investigate the algorithmic properties of several sub-classes of mean-payoff games where the players have asymmetric information about the state of the game. These games are in general undecidable and not determined according to the classical definition. We show that such games are determined under a more general notion of winning strategy. We also consider mean-payoff games where the winner can be determined by the winner of a finite cycle-forming game. This yields several decidable classes of mean-payoff games of asymmetric information that require only finite-memory strategies, including a generalization of full-observation games where positional strategies are sufficient. We give an exponential time algorithm for determining the winner of the latter.

1 Introduction

Mean-payoff games (MPGs) are two-player, infinite duration, turn-based games played on finite edge-weighted graphs. The two players alternately move a token around the graph; and one of the players (Eve) tries to maximize the (limit) average weight of the edges traversed, whilst the other player (Adam) attempts to minimize the average weight. Such games are particularly useful in the field of verification of models of reactive systems, and the full-observation versions of these games have been extensively studied [10, 4, 7, 8]. One of the major open questions in the field of verification is whether the following decision problem, known to be in the intersection of the classes NP and coNP [10][1], can be solved in polynomial time: Given a threshold ν, does Eve have a strategy to ensure a mean-payoff value of at least ν?

In game theory the concepts of partial and limited observation indicate situations where players are uncertain about the state of the game. In the context of verification games this partial knowledge is reflected in one or both players being unable to determine the precise location of the token amongst several equivalent states, and such games have also been extensively studied [21, 14, 3, 2, 9]. Adding partial observation to verification games results in an enormous increase in complexity, both algorithmically and in terms of strategy synthesis. For example, it was shown in [9] that for MPGs with partial observation, when the mean payoff value is defined using $\limsup$, the analogue of the above decision problem (*i.e.* the threshold problem) is undecidable; and whilst positional strategies suffice for MPGs with full observation, infinite memory may be required. The first result of this paper is to show that this is also the case when the mean payoff value is defined using the $\liminf$ operator, closing two open questions posed in [9].

These unfavourable results motivate the main investigation of this paper: identifying classes of MPGs with partial observation where determining the winner is decidable and where strategies with finite memory, possibly positional, are sufficient.

*Work partially supported by ERC Starting grant inVEST (FP7-279499).
†F.R.S.-FNRS Aspirant and FWA post-doc fellow
[1]From results in [23] and [12] it follows that the problem is also in UP ∩ coUP.

To simplify our definitions and algorithmic results we initially consider a restriction on the set of observations which we term *limited observation*. In games of limited observation the current observation contains only those states consistent with the observable history, that is the observations are the *belief set of Eve* (see, *e.g.* [6]). This is not too restrictive as any MPG with partial observation can be realized as a game of limited observation via a subset construction. In Section 9 we consider the extension of our definitions to MPGs with partial observation via this construction.

Our focus for the paper will be on games at the observation level, in particular we are interested in *observation-based strategies* for both players. Whilst observation-based strategies for Eve are usual in the literature, observation-based strategies for Adam have not, to the best of our knowledge, been considered. Such strategies are more advantageous for Adam as they encompass several simultaneous concrete strategies. Further, in games of limited observation there is guaranteed to be at least one concrete strategy consistent with an observation-based strategy. Our second result is to show that, although MPGs with partial observation are not determined under the usual definition of (concrete) strategy, they are determined when Adam can use an observation-based strategy.

In full-observation games, one aspect of MPGs that leads to simple (but not quite efficient) decision procedures is their equivalence to finite cycle-forming games. Such games are played as their infinite counterparts, however when the token revisits a state the game is stopped. The winner is determined by a finite analogue of the mean-payoff condition on the cycle now formed; that is, Eve wins if the average weight of the edges traversed in the cycle exceeds a given threshold. Ehrenfeucht and Mycielski [10] and Björklund et al. [4][2] used this equivalence to show that positional strategies are sufficient to win MPGs with full observation and this leads to an NP ∩ coNP procedure for determining the winner. Critically, a winning strategy in the finite game translates directly to a winning strategy in the MPG, so such games are especially useful for strategy synthesis.

We extend this idea to games of partial observation by introducing a finite, full-observation, cycle-forming game played at the observation level. That is, the game finishes when an observation is revisited (though not necessarily the first time). In this reachability game winning strategies can be translated to finite-memory winning strategies in the MPG. This leads to a large, natural subclass of MPGs with partial observation, *forcibly terminating* games, where determining the winner is decidable and finite-memory observation-based strategies suffice.

Unfortunately, recognizing if an MPG is a member of this class is undecidable, and although determining the winner is decidable, we show that this problem is complete (under polynomial-time reductions) for the class of all decidable problems. Motivated by these negative algorithmic results, we investigate two natural refinements of this class for which winner determination and class membership are decidable. The first, *forcibly first abstract cycle* games (forcibly FAC games, for short), is the natural class of games obtained when our cycle-forming game is restricted to simple cycles. Unlike the full-observation case, we show that winning strategies in this finite simple cycle-forming game may still require memory, though this memory is at most exponential in the size of the game. The second refinement, *first abstract cycle* (FAC) games, is a further structural refinement that guarantees a winner in the simple cycle-forming game. We show that in this class of games positional observation-based strategies suffice.

The sub-classes of MPGs with limited observation we study then give rise to sub-classes of MPGs with partial observation. For the class membership problem we show there is, as expected, an exponential blow-up in the complexity, however for the problem of determining the winner the algorithmic cost is significantly better.

Table 1 summarizes the results of this paper. An extended abstract of this work appeared in [11].

2 Preliminaries

Mean-payoff games A *mean-payoff game (MPG) with partial observation* is a tuple $G = (Q, q_I, \Sigma, \Delta, w, \mathsf{Obs})$, where Q is a finite set of states, $q_I \in Q$ is the initial state, Σ is a finite set of action symbols, $\Delta \subseteq Q \times \Sigma \times Q$ is

[2]A recent result of Aminof and Rubin [1] corrects some errors in [4].

		Sufficient memory	Class membership	Winner determination
Forcibly terminating		Finite (Thm. 5)	Undecidable (Thm. 6)	R-c (Thm. 7)
Forc. FAC	limited obs.	Exponential (Thm. 10)	PSPACE-c (Thm. 8)	PSPACE-c (Thm. 9)
	partial obs.	Doubly exponential (Thms. 10, 15)	NEXPTIME-h, in EXPSPACE (Thm. 16)	EXPTIME-c (Thm. 17)
FAC	limited obs.	Positional (Thm. 11)	CONP-c (Thm. 13)	NP $\cap$ CONP (Thm. 12)
	partial obs.	Exponential (Thms. 11, 15)	CONEXPTIME-c (Thm. 16)	EXPTIME-c (Thm. 17)

Table 1: Summary of results for the classes of games studied.

the transition relation, $w : \Delta \to \mathbb{Z}$ is the weight function, and $\mathsf{Obs} \subset 2^Q$ is a partition of Q into observations. We assume Δ is total, that is, for every $(q, \sigma) \in Q \times \Sigma$ there exists $q' \in Q$ such that $(q, \sigma, q') \in \Delta$. We say that G is a *mean-payoff game with limited observation* if additionally, Obs satisfies the following:

(1) $\{q_I\} \in \mathsf{Obs}$, and

(2) For each $(o, \sigma) \in \mathsf{Obs} \times \Sigma$ the set $\{q' \in Q \mid \exists q \in o \text{ and } (q, \sigma, q') \in \Delta\}$ is a union of elements of Obs.

Note that condition (2) is equivalent to saying that if $q \in o$, $q' \in o'$ and $(q, \sigma, q') \in \Delta$ then for every $r' \in o'$ there exists $r \in o$ such that $(r, \sigma, r') \in \Delta$. If every element of Obs is a singleton, then we say G is a *mean-payoff game with full observation*. For simplicity, we denote by $\mathsf{post}_\sigma(s) = \{q' \in Q \mid \exists q \in s : (q, \sigma, q') \in \Delta\}$ the set of σ-successors of a set of states $s \subseteq Q$.

Figure 1 gives an example of an MPG with limited observation, with $\Sigma = \{a, b\}$ and $\mathsf{Obs} = \{\{q_0\}, \{q_1, q_2\}, \{q_3\}\}$. In this work, unless explicitly stated otherwise, we depict states from an MPG with partial observation as circles and transitions as arrows labelled by an action-weight pair: σ, w. Observations are represented by dashed boxes.

Abstract & concrete paths A *concrete path* in an MPG with partial observation is a sequence $q_0\sigma_0q_1\sigma_1\ldots$ where for all $i \geq 0$ we have $q_i \in Q$, $\sigma_i \in \Sigma$ and $(q_i, \sigma_i, q_{i+1}) \in \Delta$. An *abstract path* is a sequence $o_0\sigma_0o_1\sigma_1\ldots$ where $o_i \in \mathsf{Obs}$, $\sigma_i \in \Sigma$ and for all $i \geq 0$ there exists $q_i \in o_i$ and $q_{i+1} \in o_{i+1}$ with $(q_i, \sigma_i, q_{i+1}) \in \Delta$. Given an abstract path ψ, let $\gamma(\psi)$ be the (possibly empty) set of concrete paths that agree with the observation and action sequence. In other words $\gamma(\psi) = \{q_0\sigma_0q_1\sigma_1\cdots \mid \forall i \geq 0 : q_i \in o_i \text{ and } (q_i, \sigma, q_{i+1}) \in \Delta\}$. Note that in games of limited observation this set is never empty. Also, given an abstract (respectively concrete) path ψ, let $\psi[..n]$ represent the prefix of ψ up to the $(n+1)$-th observation (state), which we express as $\psi[n]$; similarly, we denote by $\psi[\ell..]$ the suffix of ψ starting from the $(\ell+1)$-th observation (state) and by $\psi[\ell..n]$ the finite sub-sequence starting and ending in the aforementioned locations.

Cycles An *abstract (respectively concrete) cycle* is an abstract (concrete) path $\chi = o_0\sigma_0\ldots o_n$ where $o_0 = o_n$. We say χ is *simple* if $o_j \neq o_i$ for $0 \leq i < j < n$. Given $k \in \mathbb{N}$ define χ^k to be the abstract (concrete) cycle obtained by traversing k times χ. That is, $\chi^k = o'_0\sigma'_0\ldots o'_{nk}$ where for all $0 \leq j \leq nk$ we have that $o'_j = o_{j \pmod n}$ and $\sigma'_j = \sigma_{j \pmod n}$. A *cyclic permutation* of χ is an abstract (concrete) cycle $o'_0\sigma'_0\ldots o'_n$ such that $o'_j = o_{j+k \pmod n}$ and $\sigma'_j = \sigma_{j+k \pmod n}$ for some $k \in \mathbb{N}$. If $\chi' = o'_0\sigma'_0\ldots o'_m$ is a cycle with $o'_0 = o_i$ for some $0 \leq i < n$, the *interleaving* of χ and χ' is the cycle $o_0\sigma_0\ldots o_i\sigma'_0\ldots o'_m\sigma_i\ldots o_n$.

The mean payoff Given an infinite concrete path $\pi = q_0\sigma_0q_1\sigma_1\ldots$, the *payoff* up to the $(n+1)$-th element is given by

$$w(\pi[..n]) = \sum_{i=0}^{n-1} w(q_i, \sigma_i, q_{i+1}).$$

If π is infinite, we define two *mean payoff* values $\underline{\mathsf{MP}}$ and $\overline{\mathsf{MP}}$ as:

$$\underline{\mathsf{MP}}(\pi) = \liminf_{n\to\infty} \frac{1}{n} w(\pi[..n]) \qquad \overline{\mathsf{MP}}(\pi) = \limsup_{n\to\infty} \frac{1}{n} w(\pi[..n])$$

Plays & strategies A play in an MPG with partial observation G is an infinite abstract path starting at $o_I \in \mathsf{Obs}$ where $q_I \in o_I$. Denote by $\mathsf{Plays}(G)$ the set of all plays and by $\mathsf{Prefs}(G)$ the set of all finite prefixes of such plays ending in an observation. Let $\gamma(\mathsf{Plays}(G))$ be the set of concrete paths of all plays in the game, and $\gamma(\mathsf{Prefs}(G))$ be the set of all finite prefixes of all concrete paths.

An *observation-based strategy for Eve* is a function from finite prefixes of plays to actions, *i.e.* $\lambda_\exists : \mathsf{Prefs}(G) \to \Sigma$. A play $\psi = o_0\sigma_0o_1\sigma_1\ldots$ is *consistent* with $\lambda_\exists$ if $\sigma_i = \lambda_\exists(\psi[..i])$ for all $i \geq 0$. An *observation-based strategy for Adam* is a function $\lambda_\forall : \mathsf{Prefs}(G) \times \Sigma \to \mathsf{Obs}$ such that for any prefix $\varrho = o_0\sigma_0 \ldots o_n \in \mathsf{Prefs}(G)$ and action σ, $\lambda_\forall(\varrho, \sigma) \cap \mathsf{post}_\sigma(\varrho[n]) \neq \varnothing$. A play $\psi = o_0\sigma_0o_1\sigma_1\ldots$ is consistent with $\lambda_\forall$ if $o_{i+1} = \lambda_\forall(\psi[..i], \sigma_i)$ for all $i \geq 0$. A *concrete strategy for Adam* is a function $\mu_\forall : \gamma(\mathsf{Prefs}(G)) \times \Sigma \to Q$ such that for any concrete prefix $\pi = q_0\sigma_0 \ldots q_n \in \gamma(\mathsf{Prefs}(G))$ and action σ, $\mu_\forall(\pi, \sigma) \in \mathsf{post}_\sigma(\{\pi[n]\})$. A play $\psi = o_0\sigma_0o_1\sigma_1\ldots$ is consistent with $\mu_\forall$ if there exists a concrete path $\pi \in \gamma(\psi)$ such that $\mu_\forall(\pi[..i], \sigma_i) = \pi[i+1]$ for all $i \geq 0$.

An observation-based strategy for Eve $\lambda_\exists$ can be encoded into a *finite Mealy machine* if there is a finite set M, an element $m_0 \in M$, and functions $\alpha_u : M \times \mathsf{Obs} \to M$ and $\alpha_o : M \times \mathsf{Obs} \to \Sigma$ such that for any play prefix $\psi = o_0\sigma_0 \ldots o_n$ we have $\sigma_i = \lambda_\exists(\psi) = \alpha_o(m_n, o_n)$, where m_n is defined inductively by $m_{i+1} = \alpha_u(m_i, o_i)$ for $i \geq 0$. Similarly, an observation-based strategy for Adam $\lambda_\forall$ can be encoded into a finite Mealy machine if there is a finite set M, an element $m_0 \in M$, and functions $\alpha_u : M \times \mathsf{Obs} \times \Sigma \to M$ and $\alpha_o : M \times \mathsf{Obs} \times \Sigma \to \mathsf{Obs}$ such that for any play prefix ending in an action $\psi = o_0\sigma_0 \ldots o_n\sigma_n$, we have $o_{i+1} = \lambda_\forall(\psi) = \alpha_o(m_n, o_n, \sigma_n)$, where m_n is defined inductively by $m_{i+1} = \alpha_u(m_i, o_i, \sigma_i)$ for $i \geq 0$. In both cases we say the observation-based strategy has *memory* $|M|$. An observation-based strategy (for either player) with memory 1 is *positional*.

Remark 1. Note that for any concrete strategy μ for Adam there is a unique observation-based strategy λ_μ for him such that all plays consistent with μ are consistent with λ_μ. Conversely there may be several, but possibly no, concrete strategies that correspond to a single observation-based strategy. In games of limited observation there is guaranteed to be at least one concrete strategy for every observation-based strategy.

Winning an MPG Given a threshold $\nu \in \mathbb{R}$, we say a play ψ is *winning for Eve* if $\underline{\mathsf{MP}}(\pi) \geq \nu$ for all concrete paths $\pi \in \gamma(\psi)$, otherwise it is *winning for Adam*. Given ν, one can construct an equivalent game in which Eve wins if and only if $\underline{\mathsf{MP}}(\pi) \geq 0$ if and only if she wins the original game, so without loss of generality we will assume $\nu = 0$. A strategy λ is *winning* for a player if all plays consistent with λ are winning for that player. We say that a player *wins* G if (s)he has a winning strategy.

Remark 2. It was shown in [9] that in MPGs with partial observation where finite-memory strategies suffice Eve wins the $\underline{\mathsf{MP}}$ version of the game if and only if she wins the $\overline{\mathsf{MP}}$ version. As the majority of games considered in this paper only require finite memory, we can take either definition. For simplicity and consistency with Section 3 we will use $\underline{\mathsf{MP}}$.

Non-zero-sum reachability games A *reachability game* $G = (Q, q_I, \Sigma, \Delta, \mathcal{T}_\exists, \mathcal{T}_\forall)$ is a tuple where Q is a (not necessarily finite) set of states; Σ is a finite set of actions; $\Delta \subseteq Q \times \Sigma \times Q$ is a finitary transition function (that is, for any $q \in Q$ and $\sigma \in \Sigma$ there are finitely many $q' \in Q$ such that $(q, \sigma, q') \in \Delta$); $q_I \in Q$ is the initial state; and $\mathcal{T}_\exists, \mathcal{T}_\forall \subseteq Q$ are the terminating states. The game is played as follows. We place a token on $q_I \in Q$ and start the game. Eve chooses an action $\sigma \in \Sigma$ and Adam chooses a σ-successor of the current

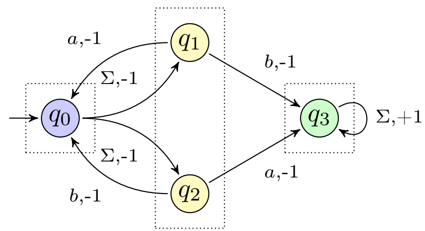

Figure 1: A non-determined MPG with limited observation ($\Sigma = \{a, b\}$)

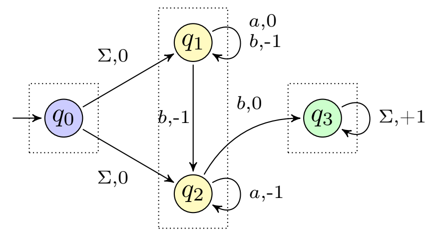

Figure 2: A limited-observation MPG in which Eve requires infinite memory to win

state as determined by Δ. The process is repeated until the game reaches a state in $\mathcal{T}_\exists$ or $\mathcal{T}_\forall$. In the first case we declare Eve as the winner whereas the latter corresponds to Adam winning the game. Notice that the game, in general, might not terminate, in which case neither player wins. Notions of plays and strategies in the reachability game follow the definitions for mean-payoff games, however we extend plays to include finite paths that end in $\mathcal{T}_\exists \cup \mathcal{T}_\forall$.

3 Undecidability of Liminf Games

Mean-payoff games with partial observation were extensively studied in [9]. In that paper the authors show that, with the mean payoff condition defined using $\underline{\mathsf{MP}}$ and $>$, determining whether Eve has a winning observation-based strategy is undecidable and when defined using $\overline{\mathsf{MP}}$ and $\geq$, strategies with infinite memory may be necessary. The analogous, and more general, questions using $\underline{\mathsf{MP}}$ and $\geq$ were left open. In this section we answer these questions, showing that both results still hold.

Proposition 1. *There exist MPGs with partial observation for which Eve requires infinite-memory observation-based strategies to ensure* $\underline{\mathsf{MP}} \geq 0$.

Proof. Consider the game G in Figure 2. We will show that Eve has an infinite-memory observation-based strategy to win this game, but no finite-memory observation-based strategy.

Consider the observation-based strategy that plays (regardless of the witnessed observations) $aba^2ba^3ba^4b\dots$ As b is played infinitely often by this strategy, the only concrete paths consistent with it are $\pi = q_0q_1^\omega$ and $\pi = q_0 \cdot q_1^k \cdot q_2^l \cdot q_3^\omega$ for non-negative integers k, l. In the first case we see that $\frac{1}{n}w(\pi[..n]) \to 0$ as $n \to \infty$, and for all paths matching the second case we have $\frac{1}{n}w(\pi[..n]) \to 1$ as $n \to \infty$. Thus $\underline{\mathsf{MP}} \geq 0$ and so the strategy is winning.

Now suppose Eve has a finite-memory observation-based winning strategy for G. We will define a concrete strategy for Adam such that a concrete path with negative mean payoff and consistent with both strategies exists. The strategy for Adam is such that the game remains in $\{q_1, q_2\}$. The resulting play can now be seen as choosing a word $w \in \{a, b\}^\omega$, but as Eve's strategy has finite memory, this word must be ultimately periodic, that is $w = w_0 \cdot v^\omega$ for words $w_0, v \in \{a, b\}^*$. We now describe the concrete strategy for Adam. If w contains finitely many b's then Adam moves to q_2 on the final b and $\frac{1}{n}w(\pi[..n]) \to -1$ as $n \to \infty$. Otherwise Adam remains in q_1 and $\frac{1}{n}w(\pi[..n]) \to -\frac{m}{|v|}$ as $n \to \infty$ where m is the number of b's in v. ☐

Theorem 1. *Let G be an MPG with partial observation. Determining whether Eve has an observation-based strategy to ensure* $\underline{\mathsf{MP}} \geq 0$ *is undecidable.*

The proof of this result is based on a similar construction to the one used in the proof of Proposition 20, so we defer it to Section 6.1.

4 Observable Determinacy

One of the key features of MPGs with full observation is that they are determined, that is, it is always the case that one player has a winning strategy. This is not true in games of partial or limited observation as can be seen in Figure 1. Any concrete strategy of Adam reveals to Eve the successor of q_0 and she can use this information to play to q_3. Conversely Adam can defeat any strategy of Eve by playing to whichever of q_1 or q_2 means the play returns to q_0 on Eve's next choice (recall Eve cannot distinguish q_1 and q_2 and must therefore choose an action to apply to the observation $\{q_1, q_2\}$). This strategy of Adam can be encoded as an observation-based strategy: "from $\{q_1, q_2\}$ with action a or b, play to $\{q_0\}$". It transpires that any such counter-play by Adam is always encodable as an observable strategy. We formalize these claims in the sequel.

Let us recall the definition of the Borel hierarchy of sets. For a detailed description of both the hierarchy and its properties we refer the reader to [13].

Definition 1 (Borel hierarchy & (co-)Suslin sets)**.** For a (possibly infinite) alphabet A, let A^ω and A^* denote the set of infinite and finite words on A, respectively. The *Borel hierarchy* is inductively defined as follows.

- $\Sigma_1^0 = \{W \cdot A^\omega \mid W \subseteq A^*\}$ is the set of open sets.
- For all $n \geq 1$, $\Pi_n^0 = \{A^\omega \setminus L \mid L \in \Sigma_n^0\}$ consists of the complement of sets in Σ_n^0.
- For all $n \geq 1$, $\Sigma_{n+1}^0 = \{\bigcup_{i \in \mathbb{N}} L_i \mid \forall i \in \mathbb{N} : L_i \in \Pi_n^0\}$ is the set obtained by countable unions of sets in Π_n^0.
- Finally, we write $\Delta_n^0 = \Sigma_n^0 \cap \Pi_n^0$, for all $n \geq 0$.

The first level of the *Projective hierarchy* consists of Σ_1^1 (Suslin) sets, which are those whose preimage is a Borel set, *i.e.* all sets that can be defined as a projection of a Borel set, and Π_1^1 (co-Suslin) sets: those sets whose complement is the image of a Borel set.

4.1 (Full-observation) Determinacy

Let us first consider MPGs with full observation and recall the well-known determinacy result that applies to them. Note that in games with full observation a play in fact corresponds to a unique infinite concrete path. Furthermore, the distinction between observation-based and concrete strategies is unnecessary. For clarity, in the remaining of this section we speak of *concrete plays* in full-observation games and *abstract plays* in partial-observation games. Given a strategy $\lambda_\exists$ for Eve and a strategy $\lambda_\forall$ for Adam in an MPG, we denote by $\mathsf{Out}(\lambda_\exists, \lambda_\forall)$ the unique play consistent with both strategies.

Proposition 2. *In every MPG with full observation exactly one of the following assertions holds.*

1. *There exists a strategy $\lambda_\exists$ for Eve such that, for all strategies $\lambda_\forall$ for Adam, the concrete play $\mathsf{Out}(\lambda_\exists, \lambda_\forall)$ is winning for Eve.*
2. *There exists a strategy $\lambda_\forall$ for Adam such that, for all strategies $\lambda_\exists$ for Eve, the concrete play $\mathsf{Out}(\lambda_\exists, \lambda_\forall)$ is winning for Adam.*

The proof of the above determinacy result follows from the fact that the set of winning plays in any MPG is a Borel set. More precisely, the statement that the limit inferior of a given sequence $(a_n)_{n \in \mathbb{N}}$ is non-negative is a Π_3^0-statement (for every k there exists a t such that for all $n \geq t \;\; a_n \geq -2^{-k}$). Similarly, for the limit superior we get a Σ_2^0-statement. Hence, by Borel determinacy [15], all mean-payoff games with full observation are determined.

4.2 Determinacy & partial-observation games

In MPGs with partial observation, authors usually focus on observation-based strategies for Eve and concrete strategies for Adam. Using this asymmetric point of view, we will now state the well-known non-determinacy of games with partial observation. Given an observation-based strategy $\lambda_\exists$ for Eve and a concrete strategy $\mu_\forall$ for Adam in an MPG with partial observation, we denote by $\mathsf{Out}(\lambda_\exists, \mu_\forall)$ the unique abstract play consistent with both strategies. Remark that here we can no longer assume a play is one unique concrete path. Furthermore, recall that an abstract play is winning (for a player) if all its concretizations are winning (for the player)

Proposition 3. *There are MPGs with limited observation for which none of the following assertions hold.*

1. *There exists an observation-based strategy $\lambda_\exists$ for Eve such that, for all concrete strategies $\mu_\forall$ for Adam, the abstract play $\mathsf{Out}(\lambda_\exists, \mu_\forall)$ is winning for Eve.*
2. *There exists a concrete strategy $\mu_\forall$ for Adam such that, for all observation-based strategies $\lambda_\exists$ for Eve, the abstract play $\mathsf{Out}(\lambda_\exists, \mu_\forall)$ is winning for Adam.*

One such game is shown in Figure 1. Intuitively, the second statement is too strong for it to be implied by the negation of the first. Indeed, the game remains asymmetric—in favour of Adam—just as long as Eve does not know the concrete path corresponding to the current play prefix. This is not the case in the second statement because of the order of quantifications over the strategies and the fact Adam is using a concrete strategy. That is, since she knows his concrete strategy and the current play prefix, she knows the current concrete state of the game as well.

We presently show that if one considers the "more symmetrical" statements in which both players use observation-based strategies, then we recover determinacy.

Theorem 2 (Observable determinacy)**.** *In every MPG with limited observation (defined with $\underline{\mathsf{MP}}$ or $\overline{\mathsf{MP}}$) exactly one of the following assertions holds.*

1. *There exists an observation-based strategy $\lambda_\exists$ for Eve such that, for all observation-based strategies $\lambda_\forall$ for Adam, the abstract play $\mathsf{Out}(\lambda_\exists, \lambda_\forall)$ is winning for Eve.*
2. *There exists an observation-based strategy $\lambda_\forall$ for Adam such that, for all observation-based strategies $\lambda_\exists$ for Eve, the abstract play $\mathsf{Out}(\lambda_\exists, \lambda_\forall)$ is winning for Adam.*

In what follows we will first show how to construct a non-deterministic mean-payoff automaton that recognizes as its language the set of all concrete plays that are winning for Adam in a given MPG with limited observation. We then show that the language of the automaton is a Borel set. The result will thus follow from Lemma 4, Corollaries 6 and 10, the fact that Borel sets are closed under complement, and Borel determinacy [15].

Determinacy usually enables to simplify proofs. Without any sort of determinacy, game reductions can become tedious and confusing. However, with determinacy, we can simply transfer winning strategies for both players between games and that directly implies both players win in one game if and only if they win in the game we reduce to (from). We remark that although our results in Section 5 already imply that the games we consider from then onwards are observably determined, the above result is more general. That is to say, all partial-observation MPGs which do not fit into the classes we consider later in this work, are observably determined. It is also worth noting that observation-based strategies for Adam only really make sense for games with limited observation, since in a general partial-observation game he may be able to choose an observation which yields an abstract path with an empty set of concretizations. (The latter is not possible in a limited-observation game.) Since, for a given partial-observation game, the equivalent limited-observation game may be of size exponential w.r.t. to the original game, these kind of determinacy results may be useful in instances where one is interested in decidability of game-related problems but not necessarily when interested in establishing complexity bounds.

4.3 Borelness of losing plays

In [5] the authors consider parity objectives and show that, given a game with partial observation, one can construct a non-deterministic automaton (with the negation of the game's objective as acceptance condition) that recognizes the set of plays that are winning for Adam. We adapt their construction for MPGs. Let $G = (Q, q_I, \Sigma, \Delta, w, \mathsf{Obs})$ be a limited-observation liminf (respectively, limsup) MPG. We construct a *mean-payoff automaton* $\mathcal{A} = (Q, q_I, A, T, c)$ where:

- $A = \Sigma \times \mathsf{Obs}$ is the alphabet of the automaton,
- $T = \{(p, (\sigma, o), q) \mid (p, \sigma, q) \in \Delta \wedge q \in o\}$ is the transition relation, and

- c is a weight function such that, $(p, (\sigma, o), q) \mapsto -w(p, \sigma, q)$.

A *run* of $\mathcal{A}$ over an infinite word $\alpha = a_0 a_1 \cdots \in A^\omega$ is an infinite sequence $\varrho = q_0 a_0 q_1 a_1 \ldots$ such that $q_0 = q_I$ and $(q_i, a_i, q_{i+1}) \in T$ for all $i \geq 0$. We say ϱ is *accepting* if the limit superior (resp. limit inferior) of the sequence $(c(q_i, a_i, q_{i+1}))_{i \in \mathbb{N}}$ is strictly positive. Depending on its acceptance condition, we say the constructed machine is a limsup (resp. liminf) mean-payoff automaton. Finally, the *language* of a mean-payoff automaton *recognizes* is the set $\{\sigma_0 o_0 \sigma_1 o_1 \cdots \mid$ there is an accepting run of $\mathcal{A}$ over $(\sigma_0, o_0)(\sigma_1, o_1) \ldots\}$.

Clearly, if we write $\mathcal{L}_\mathcal{A}$ for the language of an automaton $\mathcal{A}$ constructed for an MPG with limited observation $G = (Q, q_I, \Sigma, \Delta, w, \mathsf{Obs})$, then the set $\{q_I\} \cdot \mathcal{L}_A \subseteq \mathsf{Plays}(G)$ is the set of all plays in G which are **not winning for Eve**. Intuitively, $\mathcal{A}$ receives the choice of action σ for Eve and observation o for Adam and then "guesses" the actual state chosen by a concrete strategy for Adam, thus constructing all concretizations of a play in parallel and accepting if one of them is losing for Eve (*i.e.*, winning for Adam).

Lemma 4. *The set of winning plays in a limited-observation MPG is recognizable by a non-deterministic mean-payoff automaton.*

We will now show that the language of any non-deterministic mean-payoff automaton is a Borel set.

Liminf mean-payoff automata Recall that the statement that the limit inferior of a given sequence is a Π^0_3-statement. Thus, any set recognized by a deterministic liminf mean-payoff automaton is Σ^0_4. The jump from Π^0_3 is due to the strictness of the inequality (if there exists $b > 0$ such that $\underline{\mathsf{MP}} \geq b$, then $\underline{\mathsf{MP}} > 0$). Moving to non-deterministic liminf mean-payoff automata, however, adds an existential quantification over all runs—and hence looks like it could go as high as Σ^1_1. (Whether or not all games with Σ^1_1 winning-play sets are determined is independent of ZFC. A positive answer follows, *e.g.*, from the existence of a measurable cardinal [16].) However, in the following we shall see that non-deterministic liminf mean-payoff automata still only recognize Σ^0_4-sets.

Proposition 5. *The following are equivalent for a non-deterministic liminf mean-payoff automaton:*

1. *There exists a run over α with non-negative liminf mean-payoff.*
2. *For any k there exists a run p_k over α and a position $t_k \in \mathbb{N}$ such that the mean payoff along p_k never falls below -2^{-k} after position $t_k \in \mathbb{N}$.*

Essentially, the difference between (1.) and (2.) is that the existential quantifier over the runs is moved inwards. In particular, it is obvious that (1.) implies (2.), but the converse direction is non-trivial. The basic idea of the proof is that we construct a new run p from the runs p_k by always following some run for some time, and then switching to a run for higher k, and so on. We are faced with two problems: We can only switch from a run to another if they are at the same state of the automaton at the same time, so we might get *stuck* in a run which never meets another run for higher k. Moreover, a run p_k could at some position t have much higher current mean payoff than a run $p_{k'}$ with $k > k'$, and proceed to lose a lot of payoff—which $p_{k'}$ could not afford.

Thus, in order to construct our run p, we need to make sure that we always have the option available to switch to a suitable run for higher k at some position where the two runs have very similar current mean payoff. The existence of suitable collections will be proven by iterative applications of Ramsey's theorem:

Theorem 3 (Infinite Ramsey's theorem)**.** *Let $\mathcal{P}(\mathbb{N})_r$ denote the set of r-element subsets of $\mathbb{N}$. Then for any* colouring function $c : \mathcal{P}(\mathbb{N})_r \to \{0, 1, \ldots, \kappa\}$, *where $\kappa \in \mathbb{N}$, there exists an infinite subset $H \subseteq \mathbb{N}$ such that for any two $A, B \in \mathcal{P}(\mathbb{N})_r$ with $A \subseteq H$, $B \subseteq H$ we find that $c(A) = c(B)$. Such an H is called c-homogenous.*

Proof of Proposition 5. Assume that for any k there exists a run p_k and a position t_k such that the mean payoff along p_k never falls below -2^{-k} after position t_k. Let states in the automaton be labelled 0 to n. W.l.o.g., assume that the payoff values are from $[-1, 1]$. For $x \in [-1, 1]$ and $\ell \in \mathbb{N}$, let $b_{x,\ell} := \lceil 2^\ell (x + 1) \rceil$. Note that $b_{x,\ell} \in \{0, \ldots, 2^{\ell+1}\}$. Let $\langle\ ,\ \rangle : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ denote a standard pairing function, *i.e.* an encoding of a pair of natural numbers into one single natural number.

To use Ramsey's theorem, one must colour subsets of $\mathbb{N}$, not tuples. Since we will use it on sets $\{k, t\}$ with k the index of a run and t a position, we must somehow decide which number in a given set of size 2 is the run index and which one is the position. Since for any position $t \geq t_k$ the mean payoff of p_k never falls below -2^{-k} after position t, the larger number in a pair will always be understood as the position while the smaller will be the run index.

We iteratively define colourings c_i of 2-element subsets of $\mathbb{N}$, to which we apply Ramsey's theorem in order to obtain c_i-homogenous sets H_i and derived infinite sets S_i. For all $k < t$, let $c_0(\{k,t\}) := \langle v, b_{x,1}\rangle$ where v is the state the run p_k is in at position t, and x is the current mean payoff of p_k at position t. Let H_0 be an infinite c_0-homogenous set. Let $S_0 := H_0$. Once we have obtained S_i, let $(m^i_j)_{j\in\mathbb{N}}$ be a monotone (increasing) sequence enumerating S_i. Then let $c_{i+1}(\{k,t\})$, again for $k < t$, be $\langle v, b_{x,i+2}\rangle$ where v is the state the run $p_{m^i_k}$ is in at position t and x is the current mean payoff of $p_{m^i_k}$ at position t. Let H_{i+1} be an infinite c_{i+1}-homogenous set. Let $S_{i+1} = \{m^i_k \mid k \in H_{i+1}\}$.

This construction ensures that for all $i \in \mathbb{N}$, for all $k_1, k_2 \in S_i$, and for all sufficiently large $t \in H_i$, the runs p_{k_1} and p_{k_2} will be at the same state at position t and their current mean payoff at position t will differ by at most 2^{-i-1}. If $i > 0$, by sufficiently large we mean any t larger than the indices j, ℓ we assign to k_1 and k_2 in the construction of H_i, *i.e.* $t > m^i_j, m^i_\ell$ where $k_1 = m^i_j$ and $k_2 = m^i_\ell$. Since the sequence $(m^i_j)_{j\in\mathbb{N}}$ is monotone, it suffices to take $t > \max\{k_1, k_2\}$. (The latter also allows us to claim the property holds for $i = 0$.) Finally, also note that $S_{i+1} \subseteq S_i$.

The run p we need to construct will first follow some p_{k_0} with $k_0 \in S_0$ until sufficient time $s_0 \in H_0$ has passed, then switch to some p_{k_1} with $k_1 > k_0$ and $k_1 \in S_1$, again until sufficient (total) time $s_1 \in H_1$ has passed, then switch to p_{k_2} with $k_2 > k_1$ and $k_2 \in S_2$, and so forth.

It remains to specify what *sufficient time* means for the position s_i, and to show that this condition ensures that the mean payoff of p is non-negative. For the latter, we will ensure that after position s_i the current mean payoff of p never again drops below -2^{-i+1}. The sufficient condition for s_i will include the sufficiency condition for any runs with indices from S_i having the same current vertex and current mean-payoff difference at most 2^{-i-1}. Moreover, we need that $s_i \geq t_{k_{i+1}}$.

Let us now consider the current mean payoff of p at some position t with $s_i \leq t < s_{i+1}$. By summing up the loss of mean payoff through the changes, we see that the current mean payoff of p differs by at most $2^{-2}\frac{s_0}{t} + \ldots + 2^{-i-2}\frac{s_i}{t}$ from that of p_{k_i}, which in turn is at least -2^{-k_i}. Note that, since we have chosen our indices so that $k_i \leq k_{i+1}$ for all $i \in \mathbb{N}$, we have that $k_i \geq i$ and therefore $-2^{-k_i} \geq -2^{-i}$. Thus, once $s_0, \ldots, s_{i-1}$ have been chosen, we just need make sure that s_i is large enough so that:

- $s_i \geq t_{k_{i+1}}$,
- $s_i \geq \max\{k_i, k_{i+1}\}$, and finally
- $\sum_{j=0}^{i} 2^{-j-1}\frac{s_j}{s_i} \leq 2^{-i-1}$ so that the mean payoff of p never again drops below -2^{-i+1}.

The first two items are trivial. For the third one, note that, as the left hand side goes to 0 for $s_i \to \infty$, this can always be ensured by staying increasingly longer with each run we switch to. □

Corollary 6. *Any set recognized by a liminf mean-payoff automaton is* Σ^0_4.

Proof. Using Proposition 5, it suffices to argue that the second equivalent condition is Π^0_3. This in turn follows from the observation that for fixed $k, t \in \mathbb{N}$ the condition *there exists a run whose mean payoff never falls below* -2^k *after position* t is by Weak König's Lemma a Π^0_1 condition. □

Limsup mean-payoff automata We now show an analogue of Corollary 6 holds for limsup mean-payoff automata. Once more, to simplify the argument, we focus on the non-strict acceptance condition. We show languages recognized by such automata are Π^0_3 and, thus, those recognized by limsup mean-payoff automata are Σ^4_0.

Proposition 7. *The following are equivalent for a non-deterministic limsup mean-payoff automaton:*

1. *There exists a run over α with non-negative limsup mean payoff.*
2. *For any k there exists a run p_k over α such that for all positions t there exists a position $t' > t$ such that the mean payoff of p_k at position t' is at least -2^{-k}.*

Proof. That (1.) implies (2.) follows immediately from the definition of limsup mean payoff and the fact that the witnessing run from (1.) is also a witness for all k in (2.). That (2.) implies (1.) will be shown using Ramsey's theorem, similar to the argument in the proof of Proposition 5. We do not need iterative applications here, though.

We define a colouring c by setting $c(\{t, k\}) = v$ where $k < t$ and the run p_k is in the state v at position t, and obtain an infinite c-homogenous set S. If $k, k' \in S$, then for any position $t \in S$ with $t > \max\{k, k'\}$ the runs p_k and $p_{k'}$ are at the same state at position t, and hence we are allowed to switch from one run to the other. Let $(k_m)_{m\in\mathbb{N}}$ be a monotone sequence enumerating S. We first follow k_0 for a while, then switch to k_1, and so on. This constructs the witnessing run p.

By assumption, the run p_{k_0} will eventually reach a current mean payoff of at least -2^{-k_0} at some position t_0. We pick some $t'_0 \in S$ with $t'_0 > \max\{t_0, k_0, k_1\}$, and follow p_{k_0} until position t'_0, and then switch to the run p_{k_1}. At the time of the switch, there is some c_0 such that the current mean payoff of p_{k_0} is not more than c_0 below the current mean payoff of p_{k_1}. (Our intention is to make the difference c_0, along with all future ones, disappear by staying longer and longer with runs we switch to.) This implies that at positions $t > t'_0$ (but prior to the next switch) the mean payoff of p is at least the mean payoff of p_{k_1} minus $\frac{t'_0}{t}c$, for some c. We pick $t_1 > t'_0$ such that $\frac{t'_0}{t_1}c \leq 2^{-k_1}$, and then some $t'_1 \geq t_1$ such that p_{k_1} at position t'_1 has a mean payoff of at least -2^{-k_1}. Then p has a current mean payoff of at least -2^{-k_1+1} at position t'_1. Then let $t''_1 \in S$ such that $t''_1 > \max\{t'_1, k_1, k_2\}$. The run p follows p_{k_1} until position t''_1, and then switches to p_{k_2}. Again, we will follow p_{k_2} long enough so that the accumulated difference of the mean payoff caused by the switches is small enough, and then until p_{k_2} realizes its bound next, and then switch to p_{k_3} at the next possible chance, and so on. As we keep reaching mean payoff values closer and closer to 0, the limsup mean payoff of p is non-negative, as intended. ◻

To obtain an analogue of Corollary 10 here, we must still argue that the statement φ: *there exists a run over α whose mean payoff is at least -2^k infinitely often*, is Borel. In the sequel we show how to encode all run prefixes of the mean-payoff automaton over a given word α into a DAG. The DAG is constructed to have special edges witnessing the existence of a run prefix with mean payoff of at least -2^k. We show the DAG has an infinite path that traverses such edges infinitely often if and only if φ holds. To conclude, we then argue the set of all such DAGs is ω-regular and thus Borel.

Definition 2. Given an automaton M with n states, an input α and some precision parameter $k \in \mathbb{N}$, we define a directed acyclic graph (DAG, for short) $D_k(\alpha)$ over $\{1, \dots, n\} \times \mathbb{N}$ with the structural constraint that there are only edges from (i, ℓ) to $(j, \ell + 1)$ or from (i, ℓ) to $(j, \ell + 2)$. We fix some sequence $(t_\ell)_{\ell\in\mathbb{N}}$ such that $\frac{t_\ell}{t_{\ell+1}} \leq 2^{-k-1}$. There is an edge from (i, ℓ) to $(j, \ell + 1)$ if there is an run of M starting from state i, reading in α from position t_ℓ to position $t_{\ell+1}$ and ending in j. There is an edge from (i, ℓ) to $(j, \ell + 2)$ if there is an run of M starting from state i, reading in α from position t_ℓ to position $t_{\ell+2}$ and ending in j, such that at some point t with $t_\ell < t < t_{\ell+2}$ the current mean payoff x satisfies that $\frac{t-t_\ell}{t}x - \frac{t_\ell}{t} \geq -2^{-k}$. We call the latter *long edges*.

The following remark will be useful in the sequel to establish that languages recognized by limsup mean-payoff automata are Borel. Intuitively, we will argue that the set of DAGs containing infinite paths with infinitely many long edges form a Borel set.

Remark 3. We can code the DAGs $D_k(\alpha)$ for a fixed M into an infinite sequence over a finite alphabet, by letting the l-th symbol code which of the finitely many potential edges from some (i, ℓ) to $(j, \ell+1)$ and from (i, ℓ) to $(i, \ell + 2)$ are available.

We will now sketch how to recognize such DAGs using a *Büchi automaton*. That is, an infinite-word non-deterministic automaton (S, s_I, A, T, B) with finite set of states S, initial state s_I, alphabet A, transition relation $T \subseteq S \times A \times S$ and *accepting transition set* $B \subseteq T$. The notion of run is as for mean-payoff automata. We say a run of a Büchi automaton is accepting if it contains infinitely many accepting transitions, and define its language as for mean-payoff automata.

Remark 4. There exists a finite non-deterministic Büchi automaton that reads in DAGs (represented as described in Remark 3), and accepts exactly those DAGs that admit a path containing infinitely many long edges starting from $(i_0, 0)$, where i_0 is the initial state of M. For instance, the Büchi automaton can be defined using the same state-space as the original mean-payoff automaton, with a transition from i to j on input edge $((i,\ell),(i,\ell'))$ if j is reachable from i in the original automaton. The transition is then marked as accepting if the edge being read is long.

Lemma 8. *If $D_k(\alpha)$ admits a path containing infinitely many long edges starting from $(i_0, 0)$, where i_0 is the initial state of M, then M has a run over α which has a mean payoff of at least -2^{-k} infinitely many times.*

Proof. Any edge in $D_k(\alpha)$ is witnessed by a partial run of M, in such a way that an infinite path through $D_k(\alpha)$ starting from $(i_0, 0)$ gives rise to a full run of M on input α. Given the condition for adding a long edge, *i.e.* of the form $(i,\ell) \Rightarrow (j, \ell+2)$, we note that at the witnessing position t, the current mean payoff of M is of the form $\frac{t_\ell}{t}y + \frac{t-t_\ell}{t}x$, where x is the mean payoff from position t to position t_ℓ, and y the mean payoff from the start to position t. As we assume all payoffs to be from $[-1, 1]$, this is bounded from below by $\frac{t-t_\ell}{t}x - \frac{t_\ell}{t} \geq -2^{-k}$, hence the claim follows. □

Lemma 9. *If M has a run over α reaching a mean payoff of at least -2^{-k-1} infinitely often, then $D_k(\alpha)$ admits a path containing infinitely many long edges starting from $(i_0, 0)$.*

Proof. Following a run p through M, we can construct an infinite path through $D_k(\alpha)$ as follows: Start from $(i_0, 0)$. If we are currently at (i,ℓ), and there is an edge available to $(j, \ell+2)$ where p is in state j at position $t_{\ell+2}$, take that edge. Else, take the edge to $(j', \ell+1)$, where j' is the state p is in a position $t_{\ell+1}$ (by construction of $D_k(\alpha)$, the latter always exists. We claim that if the mean payoff of p is at least -2^{-k-1} infinitely often, then the former case occurs infinitely many times.

Assume that p has mean payoff at least -2^{-k-1} at position t with $t_{\ell+1} \leq t < t_{\ell+2}$. Let p be in state i at position t_ℓ and at state j at position $t_{\ell+2}$. We claim that $D_k(\alpha)$ has an edge from (i,ℓ) to $(j, \ell+2)$. To see that, note that the mean payoff of p at position t is of the form $x\frac{t-t_\ell}{t} + y\frac{t_\ell}{t}$, where x is the mean payoff from position t_ℓ to position t, and y the mean payoff from the start to position t_ℓ. As $y \geq -1$, we can conclude that $x\frac{t-t_1}{t} \geq -2^{-k-1} - \frac{t_\ell}{t}$. As $t \geq t_{\ell+1}$ and the constraint on the choice of $(t_\ell)_{\ell\in\mathbb{N}}$ that $\frac{t_\ell}{t_{\ell+1}} \leq 2^{-k-1}$, we in turn find that $x\frac{t-t_1}{t} \geq -2^{-k}$ – hence the claimed edge exists.

The only reason why we might be unable to choose such an edge from (i,ℓ) to $(j, \ell+2)$ is if we choose some edge from $(i', \ell-1)$ to $(j', \ell+1)$ earlier. But that means that the availability of infinitely many such edges implies that our construction will choose them, hence showing the claim. □

Corollary 10. *Any set recognized by a limsup mean-payoff automaton is Σ^0_4.*

Proof. By Lemmas 8 and 9 we find that for some input α and limsup mean payoff automaton M the following are equivalent:

1. For all k there exists a run p_k over α such that for all positions t there exists a position $t' > t$ such that the mean payoff of p_k at position t' is at least -2^{-k}.
2. For all k the DAG $D_k(\alpha)$ has a path containing infinitely many long edges starting from $(i_0, 0)$.

By Proposition 7, the former is equivalent to M accepting α, and by Remarks 3 and 4, and [20], the latter is a universal quantification over a Δ^0_3-set, hence a Π^0_3-set. Thus, with the strict acceptance condition, a Σ^4_0 set. □

5 Strategy Transfer

In this section we will construct a reachability game from an MPG with limited observation in which winning strategies for either player are sufficient (but not necessary) for observation-based winning strategies in the original MPG.

Let us fix a limited-observation MPG $G = (Q, q_I, \Sigma, \Delta, w, \mathsf{Obs})$. We will define a reachability game on the weighted unfolding of G.

Belief functions Let $\mathcal{B}$ be the set of (*belief*) functions $f : Q \to \mathbb{Z} \cup \{+\infty, \bot\}$. Our intention is to use functions in $\mathcal{B}$ to keep track of the minimum payoff of all concrete paths ending in the given state. A function value of $\bot$ indicates that the given state is not in the current observation, and a function value of $+\infty$ is used to indicate to Eve that the token is not located at such a state. Intuitively, $+\infty$ will allow our reachability winning condition to include games where Adam wins by ignoring paths going through the given state.

The *support* of $f \in \mathcal{B}$ is $\mathsf{supp}(f) := \{q \in Q \mid f(q) \neq \bot\}$. We define a family of partial orders $\preceq_k \subseteq \mathcal{B} \times \mathcal{B}$ for all $k \in \mathbb{N}$. Formally, $f \preceq_k f'$ if:

- $\mathsf{supp}(f) = \mathsf{supp}(f')$ and
- $f(q) + k \leq f'(q)$ for all $q \in \mathsf{supp}(f)$

where $+\infty + k = +\infty$.

(Proper) successor functions Given two functions $f, f' \in \mathcal{B}$, we say f' a σ-successor of f if:

- $\mathsf{supp}(f') \in \mathsf{Obs}$;
- $\mathsf{supp}(f') \subseteq \mathsf{post}_\sigma(\mathsf{supp}(f))$; and
- for all $q \in \mathsf{supp}(f')$ either
  - $f'(q) = \min\{f(q') + w(q', \sigma, q) \mid q' \in \mathsf{supp}(f) \text{ and } (q', \sigma, q) \in \Delta\}$, or
  - $f'(q) = +\infty$.

Moreover, if f' is a $\preceq_0$-minimal σ-successor of f, we say it is a *proper* σ-successor of f.

Function-action sequences Let us denote by $\mathcal{F}(G)$ the set of all sequences $f_0\sigma_0 f_1 \ldots f_n \in (\mathcal{B} \cdot \Sigma)^*\mathcal{B}$ such that for all $0 \leq i < n$, f_{i+1} is a σ_i-successor of f_i. Observe that for each $\varphi = f_0\sigma_0 \ldots f_n \in \mathcal{F}(G)$ there is a unique abstract path $\mathsf{supp}(\varphi) := o_0\sigma_0 \ldots o_n$ such that $o_i = \mathsf{supp}(f_i)$ for all i. Conversely for each finite abstract path $\psi = o_0\sigma_0 \ldots o_n$ there may be many corresponding function-action sequences $\mathsf{supp}^{-1}(\psi) := \{\varphi \in \mathcal{F}(G) \mid \mathsf{supp}(\varphi) = \psi\}$. Of particular interest are function-action sequences that are minimal with respect to $\preceq_0$. Given a finite abstract path $\psi = o_0\sigma_0 \ldots o_n$ and a function $f_0 \in \mathcal{B}$ such that $\mathsf{supp}(f_0) \subseteq o_0$, let $\mathsf{prop}(\psi, f_0)$ denote the unique (pointwise) $\preceq_0$-minimal function-action sequence $f_0\sigma_0 \ldots f_n \in \mathsf{supp}^{-1}(\psi)$. That is to say, if $\mathsf{prop}(\psi, f_0) = f_0\sigma_0 \ldots f_n$, then for all $g_0\sigma_0 \ldots g_n \in \mathsf{supp}^{-1}(\psi)$ it holds that $f_i \preceq_0 g_i$ for all $0 \leq i \leq n$. Observe the latter holds if and only if f_{i+1} is a proper σ_i-successor of f_i, for all $0 \leq i < n$.

We extend $\mathsf{supp}(\cdot)$, $\mathsf{supp}^{-1}(\cdot)$, and $\mathsf{prop}(\cdot, \cdot)$ to infinite sequences in the obvious way.

The weighted unfolding of an MPG The *reachability game associated with* G, *i.e.* $\Gamma = (\Pi, \Sigma, f_I, \delta, \mathcal{T}_\exists, \mathcal{T}_\forall)$, is formally defined as follows. The initial state $f_I \in \mathcal{B}$ is the function for which $q_I \mapsto 0$, and $q \mapsto \bot$ for all $q \neq q_I$. The state-set Π is the subset of $\mathcal{F}(G)$ where for all $f_0\sigma_0 f_1 \ldots f_n \in \Pi$ we have:

- $f_0 = f_I$; and
- for all $0 \leq i < j < n$,
  - $f_i \not\preceq_0 f_j$ and
  - $f_j \not\preceq_1 f_i$.

The transition function δ is such that if x and $x \cdot \sigma \cdot f$ are elements of Π then $(x, \sigma, x \cdot \sigma \cdot f) \in \delta$. For the terminating states we have

$$\begin{aligned}
\mathcal{T}_\exists &= \{f_0\sigma_0 \ldots f_n \in \Pi \mid \text{for some } 0 \leq i < n : f_i \preceq_0 f_n\}; \text{ and} \\
\mathcal{T}_\forall &= \{f_0\sigma_0 \ldots f_n \in \Pi \mid \text{for some } 0 \leq i < n : f_n \preceq_1 f_i, \text{ and} \\
&\qquad\qquad \text{for some } q \in \mathsf{supp}(f_i) : f_i(q) \neq +\infty\}.
\end{aligned}$$

Note that the directed graph defined by Π and δ is a tree, but not necessarily finite.

Good and bad cycles To gain an intuition about Γ, let us say an abstract cycle χ is *good* if:

- there exists $f_0\sigma_0 \ldots f_n \in \mathsf{supp}^{-1}(\chi)$ such that $f_i(q) \neq +\infty$ for all $q \in Q$ and all $0 \leq i \leq n$, and
- $f_0 \preceq_0 f_n$.

Let us say χ is *bad* if:

- there exists $f_0\sigma_0 \ldots f_n \in \mathsf{supp}^{-1}(\chi)$ such that $f_0(q) \neq +\infty$ for some $q \in \mathsf{supp}(f_0)$, and
- $f_n \preceq_1 f_0$.

Then it is not difficult to see that Γ is essentially an abstract cycle-forming game played on G which is winning for Eve if a good abstract cycle is formed and winning for Adam if a bad abstract cycle is formed.

Our main result for this section is the following:

Theorem 4. *Let G be an MPG with limited observation and let Γ be the associated reachability game. If Adam (Eve) has a winning strategy in Γ then (s)he has a finite-memory observation-based winning strategy in G.*

The idea behind the observation-based strategies for the MPG is straightforward. If Eve wins the reachability game then she can transform her strategy into one that plays indefinitely by returning, whenever the play reaches $\mathcal{T}_\exists$, to the natural previous position—namely the position that witnesses the membership of $\mathcal{T}_\exists$. By continually playing her winning strategy in this way Eve perpetually completes good abstract cycles and this ensures that all concrete paths consistent with the play have non-negative mean-payoff value. Likewise if Adam has a winning strategy in the reachability game, he can continually play his strategy by returning to the natural position whenever the play reaches $\mathcal{T}_\forall$. By doing this he perpetually completes bad abstract cycles and this ensures that there is a concrete path consistent with the play that has strictly negative mean-payoff value.

We will repeatedly use the next result which follows by induction immediately from the definition of a σ-successor.

Lemma 11. *Let $\varphi = f_0\sigma_0 \ldots f_n \in \mathcal{F}(G)$ be a sequence such that f_{i+1} is a proper σ_i-successor of f_i, for all i. Then for all $q \in \mathsf{supp}(f_n)$,*

$$f_n(q) = \min\{f_0(\pi[0]) + w(\pi) \mid \pi \in \gamma(\mathsf{supp}(\varphi)) \text{ and } \pi[n] = q\}.$$

The following simple facts about $\preceq_n$ will also be useful:

Lemma 12. *For any two functions $f_1, f_2 \in \mathcal{B}$ with $f_1 \preceq_k f_2$:*

(i) For all $k' \leq k$, $f_1 \preceq_{k'} f_2$,

(ii) For all $k' \geq 0$, if $f_2 \preceq_{k'} f_3$ for some $f_3 \in \mathcal{F}$ then $f_1 \preceq_{k+k'} f_3$, and

(iii) If f_1' is a proper σ-successor of f_1 and f_2' is a σ-successor of f_2 with $\mathsf{supp}(f_2') = \mathsf{supp}(f_1')$, then $f_1' \preceq_k f_2'$.

Proof. Items (i) and (ii) are trivial. For (iii), let $d_{i,j} = w(q_i, \sigma, q_j)$ for $q_i \in \mathsf{supp}(f_1)$ and $q_j \in \mathsf{supp}(f_1')$ where such a transition exists and $+\infty$ otherwise. We now observe that as f_1' is $\preceq_0$-minimal, $f_1'(q_j)$ can be defined as $\min\{f_1(q_i) + d_{i,j} \mid q_i \in \mathsf{supp}(f_1)\}$ for all $q_j \in \mathsf{supp}(f_1')$. As $f_1(q_i) \leq f_2(q_i) - k$ for any $q_i \in \mathsf{supp}(f_1)$, it follows that

$$f_1'(q_j) \leq \min\{f_2(q_i) + d_{i,j} \mid q_i \in \mathsf{supp}(f_1)\} - k \leq f_2'(q_j) - k,$$

where the second inequality follows from the definition of a σ-successor. Thus $f_1' \preceq_k f_2'$. $\square$

Although the following results are not used until Section 7, they already give an intuition towards the correctness of the strategies described above. In words, we will show that repeating good cycles is itself, in some sense, good, while repeating bad ones is bad.

Lemma 13. *Let χ be an abstract cycle.*

(i) If χ is good (bad) then an interleaving of χ with another good (bad) cycle is also good (bad).

(ii) If χ is good then for all k and all concrete cycles $\pi \in \gamma(\chi^k)$, $w(\pi) \geq 0$.

(iii) If χ is bad then $\exists k \geq 0, \pi \in \gamma(\chi^k)$ such that $w(\pi) < 0$.

Proof. Item (i) follows from Lemma 12. For (ii), let $f_0\sigma_0 \dots f_n \in \mathsf{supp}^{-1}(\chi)$ be such that $f_i(q) \neq +\infty$ for all i and q and $f_0 \preceq_0 f_n$. In particular this means that f_{i+1} is a proper σ_i-successor of f_i. Now fix k and let $\pi \in \gamma(\chi^k)$ be a concrete cycle. From Lemma 11 we have, for all $0 \leq i < k$,

$$w(\pi[n \cdot i..n(i+1)]) \geq f_n(\pi[n(i+1)]) - f_0(\pi[n \cdot i])$$

and

$$f_n(\pi[n(i+1)]) - f_0(\pi[n \cdot i]) \geq f_0(\pi[n(i+1)]) - f_0(\pi[n \cdot i]).$$

Hence

$$w(\pi) = \sum_{i=1}^{k} w(\pi[n \cdot i..n(i+1)]) \geq f_0(\pi[n \cdot k]) - f_0(\pi[0]) = 0.$$

We now prove item (iii) holds. Let $f_0\sigma_0 \dots f_n \in \mathsf{supp}^{-1}(\chi)$ and $q_0 \in \mathsf{supp}(f_0)$ be such that $f_0(q_0) \neq +\infty$ and $f_n \preceq_1 f_0$. It follows that $f_n(q_0) < +\infty$. From the definition of a σ-successor, it follows that there exists $r \in \mathsf{supp}(f_{n-1})$ such that $f_{n-1}(r) < +\infty$, and there is an edge from r to q_0 with weight $f_n(q_0) - f_{n-1}(r)$. Proceeding this way inductively we find there is a $q_1 \in \mathsf{supp}(f_0)$ with $f_0(q_1) < +\infty$ and a concrete path $\pi_0 \in \gamma(\chi)$ from q_1 to q_0 with $w(\pi_0) = f_n(q_0) - f_0(q_1)$. As $f_0(q_1) < +\infty$ and $f_n \preceq_1 f_0$ we have $f_n(q_1) \leq f_0(q_1) - 1 < +\infty$. Repeating the argument yields a sequence of states $q_0, q_1, \dots$ such that there is a concrete path $\pi_i \in \gamma(\chi)$ from q_{i+1} to q_i with

$$w(\pi_i) = f_n(q_i) - f_0(q_{i+1}) \leq f_0(q_i) - f_0(q_{i+1}) - 1.$$

As Q is finite it follows that there exists $i < j$ such that $q_i = q_j$. Then the concrete path $\pi = \pi_j \cdot \pi_{j-1} \dots \pi_{i+1} \in \gamma(\chi^{j-i})$ is a concrete cycle with

$$w(\pi) = \sum_{k=i+1}^{j} w(\pi_k) \leq f_0(q_i) - f_0(q_j) - (j-i) < 0.$$

$\square$

Corollary 14. *No cyclic permutation of a good abstract cycle is bad.*

Restricting Γ w.r.t. a strategy We note that, as a play prefix in Γ is completely described by the last state in the sequence, it suffices to consider positional strategies for both players. Thus, when speaking of winning strategies for either player in Γ, we will assume they are positional.

Let $\Pi\lfloor_\lambda$ denote the set of states from Π that can be reached via plays consistent with λ. At this point, we can already show that a winning strategy for either player in Γ will reach a terminating state in a bounded number of steps. This will later allow us to argue that the strategies we construct for Eve or Adam in G based on their strategies in Γ use finite memory.

Lemma 15. *If λ is a winning strategy for Adam or Eve in Γ, then there exists $N \in \mathbb{N}$ such that for all plays π consistent with λ we have $|\pi| \leq N$.*

Proof. Suppose there is no bound on the size of $\Pi\lfloor_\lambda$. As Γ, is acyclic, it follows that $\Pi\lfloor_\lambda$ contains infinitely many states. However, as Γ is finitely-branching, it follows from König's lemma that there exists an infinite path in Γ. As this play is not winning for either player and it is consistent with λ, this contradicts the fact that λ is a winning strategy. $\square$

5.1 Strategy transfer for Eve

Suppose Eve has a winning positional strategy λ in Γ. Let $M = \Pi\lfloor_\lambda$ be the corresponding restriction of Π. From Lemma 15, M is finite. We will define an observation-based strategy λ^* with memory $|M|$ for Eve in G. Given a memory state $\mu = f_0\sigma_0 \dots f_n \in M$ let

$$\mu' = \begin{cases} \text{the proper prefix } f_0\sigma_0 \dots f_\ell \text{ of } \mu \text{ such that } f_\ell \preceq_0 f_n & \text{if } \mu \in \mathcal{T}_\exists \\ \mu & \text{otherwise.} \end{cases}$$

The initial memory state is $\mu_0 := f_I$. Let us write $\mu' = f'_0\sigma'_0 \dots f'_m$. We define the output function $\alpha_o : M \times \mathsf{Obs} \to \Sigma$ as $\alpha_o(\mu, o) = \lambda(\mu')$. Finally we define the update function $\alpha_u : M \times \mathsf{Obs} \to M$ as $\alpha_u(\mu, o) = \mu' \cdot \lambda(\mu') \cdot g$ where g is the proper $\lambda(\mu')$-successor of f'_m with $\mathsf{supp}(g) = o$. Observe that we maintain the invariant that the current observation is $\mathsf{supp}(f'_m)$, consequently the Obs input to α_o is not used.

We will show shortly that λ^* is a winning observation-based strategy for Eve in G. First, we require a result about plays consistent with λ^*. We will argue that, by following λ^* in G, Eve ensures the belief functions from proper function-action sequences induced by play prefixes consistent with it are $\preceq_0$-smaller than her current memory state.

Lemma 16. *Let $\psi = o_0\sigma_0 \dots \in \mathsf{Plays}(G)$ and $\mu_0\mu_1 \dots \in M^\omega$ be such that $\sigma_i = \alpha_o(\mu_i, o_i)$ and $\mu_{i+1} = \alpha_u(\mu_i, o_i)$ for all $i \geq 0$. If we write $\mu_i = f_i^{(0)}\sigma_i^{(0)} \dots f_i^{(n_i)}$ and $\mathsf{prop}(\psi, f_I) = g_0\sigma_0 \dots$, then $f_i^{(n_i)} \preceq_0 g_i$ for all $i \geq 0$.*

Proof. We prove this by induction. For $i = 0$ we have $\mu_0 = f_I = g_0$. Now suppose $f_i^{(n_i)} \preceq_0 g_i$. By definition of $\mathsf{prop}(\cdot,\cdot)$ we have that g_{i+1} is the proper σ_i-successor of g_i and $\mathsf{supp}(g_{i+1}) = o_{i+1}$. Assume first that $\mu_i \notin \mathcal{T}_\exists$. Then

$$\mu_{i+1} = \alpha_u(\mu_i, o_i) = \mu_i \cdot \sigma_i \cdot h,$$

where h is the proper σ_i-successor of $f_i^{(n_i)}$ with $\mathsf{supp}(h) = o_{i+1}$. Then, by Lemma 12 (iii) we have $f_{i+1}^{(n_{i+1})} = h \preceq_0 g_{i+1}$.

Now assume $\mu_i \in \mathcal{T}_\exists$, and let $\ell < n_i$ be the index such that $f_i^{(\ell)} \preceq_0 f_i^{(n_i)}$. Then

$$\mu_{i+1} = \alpha_u(\mu_i, o_i) = \left(f_i^{(0)}\sigma_i^{(0)} \dots \sigma_{(i)}^{(\ell-1)} f_i^{(\ell)}\right) \cdot \sigma_i \cdot h$$

where h is the proper σ_i-successor of $f_i^{(\ell)}$ with $\mathsf{supp}(h) = o_{i+1}$. From Lemma 12 item (ii) we have $f_i^{(\ell)} \preceq_0 g_i$, so by Lemma 12 (iii) we have $f_{i+1}^{(n_{i+1})} = h \preceq_0 g_{i+1}$ as required. ☐

We now proceed with the proof of strategy transfer for Eve.

Lemma 17. *Let G be a mean-payoff game with limited observation and let Γ be the associated reachability game. If Eve has a winning strategy in Γ then she has a finite-memory observation-based winning strategy in G.*

Proof. We will show that λ^* described above is a winning strategy for Eve. Let $\psi = o_0\sigma_0 \dots \in \mathsf{Plays}(G)$ be any play consistent with λ^*. That is, there is a sequence $\mu_0\mu_1 \dots M^\omega$ such that $\sigma_i = \alpha_o(\mu_i, o_i)$ and $\mu_{i+1} = \alpha_u(\mu_i, o_i)$ for all $i \geq 0$. We will show that there exists a constant $\beta \geq 0$ such that for all concrete paths $\pi \in \gamma(\psi)$ and all $j \geq 0$, $w(\pi[..j]) \geq \beta$. It follows that $\underline{\mathsf{MP}}(\pi) \geq 0$, and so ψ is winning for Eve.

Let $W = \{f_\ell(q) \mid f_0\sigma_0 \dots f_\ell \in M, q \in Q, \text{ and } f_\ell(q) \neq \bot\}$. Note that W is finite because M and Q are finite, and non-empty because $\mu_0 = f_I$ and $f_I(q_I) = 0 \in W$. Let $\beta = \min W$. As $0 \in W$, we have that $\beta < +\infty$.

As with Lemma 16, let $\mathsf{prop}(\psi, f_I) = g_0\sigma_0 \dots$ and $\mu_i = f_i^{(0)}\sigma_i^{(0)} \dots f_i^{(n_i)}$. Consider an arbitrary $j \in \mathbb{N}$. As $\mathsf{supp}(f_I) = \{q_I\}$ and $f_I(q_I) = 0$, Lemma 11 implies for all $q \in \mathsf{supp}(g_j)$, we have $g_j(q) \neq +\infty$. Hence, for all concrete paths $\pi \in \gamma(\psi)$ we have:

$$\begin{array}{rcll} w(\pi[..j]) & \geq & g_j(\pi[j]) - f_I(\pi[0]) & \text{from Lemma 11} \\ & = & g_j(\pi[j]) & \\ & \geq & f_j^{(n_j)}(\pi[n]) & \text{from Lemma 16} \\ & \geq & \beta & \text{as required.} \end{array}$$

□

5.2 Strategy transfer for Adam

To complete the proof of Theorem 4, we now show how to transfer a winning strategy for Adam from Γ to a winning finite-memory observation-based strategy in G. So let us assume $\lambda : \Pi \times \Sigma \to \Pi$ is a (positional) winning strategy for Adam in Γ. The finite-memory observation-based strategy for Adam is similar to that for Eve in that it perpetually plays λ, returning to a previous position whenever the play reaches $\mathcal{T}_\forall$. However, the proof of correctness is more intricate because we need to handle the $+\infty$ function values.

Formally, the finite-memory observation-based strategy λ^* is given as follows. As before, let $M = \Pi{\downarrow_\lambda}$ and $\mu_0 = f_I$. Given $\mu \in M$, let

$$\mu' = \begin{cases} \text{the proper prefix of } f_0\sigma_0 \ldots f_\ell \text{ such that } f_\ell \preceq_1 f_n & \text{if } \mu \in \mathcal{T}_\forall \\ \mu & \text{otherwise.} \end{cases}$$

Let us write $\mu' = f'_0\sigma'_0 \ldots f'_m$. The output function $\alpha_o : M \times \mathsf{Obs} \times \Sigma \to \mathsf{Obs}$ is defined as: $\alpha_o(\mu, o, \sigma) = \mathsf{supp}(g)$ where $\lambda(\mu', \sigma)) = \mu' \cdot \sigma \cdot g$. The update function $\alpha_u : M \times \mathsf{Obs} \times \Sigma \to M$ is defined as: $\alpha_u(\mu, o, \sigma) = \lambda(\mu', \sigma)$. Note that as the current observation is stored in the memory state, the Obs input to α_o and α_u is redundant.

To show that λ^* is winning for Adam in G we require an analogue to Lemma 16. To be precise, we show that, by following λ^* in G, Adam ensures the belief functions from *ultimately proper function-action sequences* induced by play prefixes consistent with it are $\preceq_r$-larger than his current memory state (for r a function of how many times his memory has been *reset*). Given a finite sequence $\mu_0 \ldots \mu_n \in M^*$ of memory states we denote by $\mathsf{reset}(\mu_0 \ldots \mu_n)$ the number of times the memory is reset along the sequence. That is, $\mathsf{reset}(\mu_0) = 0$, and if $\mu_i \in \mathcal{T}_\forall$ then $\mathsf{reset}(\mu_0 \ldots \mu_{i+1}) = \mathsf{reset}(\mu_0 \ldots \mu_i) + 1$, otherwise $\mathsf{reset}(\mu_0 \ldots \mu_{i+1}) = \mathsf{reset}(\mu_0 \ldots \mu_i)$.

Lemma 18. *Let $\psi = o_0\sigma_0 \cdots \in \mathsf{Plays}(G)$, $\mu_0\mu_1 \cdots \in M^\omega$, $k \in \mathbb{N}$, and $f_0\sigma_0 \cdots \in \mathsf{supp}^{-1}(\psi)$ be such that:*
- *$f_k\sigma_k \cdots = \mathsf{prop}(\psi[k..], f_k)$; and for all $i \geq 0$,*
- *$o_{i+1} = \alpha_o(\mu_i, o_i, \sigma_i)$ and*
- *$\mu_{i+1} = \alpha_u(\mu_i, o_i, \sigma_i)$.*

If we write $\mu_i = g_i^{(0)}\sigma_i^{(0)} \ldots g_i^{(n_i)}$ and $f_k \preceq_r g_k^{(n_k)}$ for some $r \in \mathbb{N}$, then for all $i \geq k$ it holds that $f_i \preceq_{r'_i} g_i^{(n_i)}$ where $r'_i = r + \mathsf{reset}(\mu_k \ldots \mu_i)$.

Proof. We prove this by induction on i. For $i = k$ the result clearly holds. Now suppose $i \geq k$ and $f_i \preceq_{r'_i} g_i^{(n_i)}$ where $r'_i = r + \mathsf{reset}(\mu_k \ldots \mu_i)$. We consider two cases depending on whether $\mu_i \in \mathcal{T}_\forall$. If $\mu_i \notin \mathcal{T}_\forall$ then

$$\mu_{i+1} = \mu_i \cdot \sigma_i \cdot h$$

where h is a σ_i-successor of $g_i^{(n_i)}$ with $\mathsf{supp}(h) = o_{i+1}$. Furthermore, since no reset occurred, we have that $\mathsf{reset}(\mu_0 \ldots \mu_{i+1}) = \mathsf{reset}(\mu_0 \ldots \mu_i)$. Then, by Lemma 12 (iii) we have $f_{i+1} \preceq_{r'_i} h = g_{i+1}^{(n_{i+1})}$ and since

$$r'_i = r + \mathsf{reset}(\mu_0 \ldots \mu_i) = r + \mathsf{reset}(\mu_0 \ldots \mu_{i+1}) = r'_{i+1}$$

the result holds for $i + 1$.

Otherwise if $\mu_i \in \mathcal{T}_\forall$, let $\ell < n_i$ be the index such that $g_i^{(n_i)} \preceq_1 g_i^\ell$. We have that

$$\mu_{i+1} = \left(g_i^{(0)}\sigma_i^{(0)} \ldots \sigma_i^{(\ell-1)}g_i^{(\ell)}\right) \cdot \sigma_i \cdot h$$

where h is a σ_i-successor of $g_i^{(\ell)}$ with $\mathsf{supp}(h) = o_{i+1}$. From Lemma 12 (ii) we have $f_i \preceq_{r_i'+1} h$. So by Lemma 12 (iii) we have $f_{i+1} \preceq_{r_i'+1} h = g_{i+1}^{(n_{i+1})}$, and as

$$r_i' + 1 = r + \mathsf{reset}(\mu_k \dots \mu_i) + 1 = r + \mathsf{reset}(\mu_k \dots \mu_{i+1}) = r_{i+1}'$$

the result holds for $i + 1$. □

We now show how to transfer strategies for Adam.

Lemma 19. *Let G be a mean-payoff game with limited observation and let Γ be the associated reachability game. If Adam has a winning strategy in Γ then he has a finite-memory observation-based winning strategy in G.*

Proof. We will show that the finite-memory observation-based strategy λ^* constructed above is winning for Adam. Let $\psi = o_0\sigma_0\dots$ be any play consistent with λ^*. That is, there is a sequence $\mu_0\mu_1\dots M^\omega$ such that $o_{i+1} = \alpha_o(\mu_i, o_i, \sigma_i)$ and $\mu_{i+1} = \alpha_u(\mu_i, o_i, \sigma_i)$ for all $i \geq 0$. Let us write $\mu_i = g_i^{(0)}\sigma_i^{(0)}\dots g_i^{(n_i)}$. As M is finite, there exists $\varphi = \dots f^* \in M$ and an infinite set $\mathcal{I} \subseteq \mathbb{N}$ of indices such that for all $i \in \mathcal{I}$ we have $\mu_i = \varphi$. We will show that this implies there exists $\pi \in \gamma(\psi)$ such that $\overline{\mathsf{MP}}(\pi) < 0$. As $\overline{\mathsf{MP}}(\pi) \geq \underline{\mathsf{MP}}(\pi)$ the result follows. For convenience, given $n \in \mathbb{N}$, let $\mathsf{succ}_\mathcal{I}(n) = \min\{i \in \mathcal{I} \mid i > n\}$. Denote by o^* the set $\{q \in \mathsf{supp}(f^*) \mid f^*(q) \neq +\infty\}$. Note that from the definition of $\mathcal{T}_\forall$ it follows that o^* is non-empty.

We will use a function-action sequence to find a concrete path that is winning for Adam. That is, a concrete path where the weights of the prefixes can be identified and seen to be strictly decreasing. Unlike in Lemma 17, the unique proper function-action sequence $\mathsf{prop}(\psi, f_I)$ fulfilled does not hold enough information for us to prove this claim. Indeed, to handle $+\infty$ values, which correspond to irrelevant paths, we require a more complex sequence. Recall that Adam can place $+\infty$ values in a function to tell Eve that the token is **not** in a particular state, *e.g.* $f(q) = +\infty$ signifies q does not hold the token.

The sequence we construct will be *piecewise proper* in the sense that for all $i \in \mathcal{I}$ the sequence will consist of proper successors in the interval $[i, \mathsf{succ}_\mathcal{I}(i))$. When the sequence reaches an element of $\mathcal{I}$ we "reset" the values of the states not in o^* to $+\infty$. More formally, the required sequence, $\mathsf{pw\text{-}prop}_\mathcal{I}(\psi, f_I) = h_0\sigma_0 \dots \in \mathsf{supp}^{-1}(\psi)$, is constructed inductively as follows. Initially, let $h_0 = h_0' = f_I$. For $i \geq 0$, let h_{i+1}' be the proper σ_i-successor of h_i with $\mathsf{supp}(h_{i+1}') = o_{i+1}$. If $i \notin \mathcal{I}$ then $h_i = h_i'$. Otherwise, for any $q \in \mathsf{supp}(h_{i+1}')$ we let

$$h_i'(q) = \begin{cases} +\infty & \text{if } q \notin o \\ h_i'(q) & \text{otherwise.} \end{cases}$$

Observe that, by construction, for all $i \in \mathcal{I}$ and all $q \in o^*$ we have that $h_i(q) = h_i'(q)$.

We now claim that

$$\forall i \in \mathbb{N} : h_i \preceq_{r_i} g_i^{(n_i)},$$

where $r_i = \mathsf{reset}(\mu_0 \dots \mu_i)$. From Lemma 12 (i) and Lemma 18 it follows that we only need to show that for all $i \in \mathcal{I}$ it holds that $h_i \preceq_{r_i} f^*$. Induction and Lemma 18 imply that for all $i \in \mathcal{I}$ we have $h_i' \preceq_{r_i} f^*$. Recall that h_i differs from h_i' only on states where f^* is equal to $+\infty$, we therefore have $h_i \preceq_{r_i} f^*$ as required.

We will now show that there is an infinite concrete path $q_0\sigma_0 \dots \in \gamma(\psi)$ such that $q_i \in o^*$ for all $i \in \mathcal{I}$. To do this we will show for any $i \in \mathcal{I}$ and any $q \in o^*$ there is a concrete path in $\gamma(\psi[i..\mathsf{succ}_\mathcal{I}(i)])$, that ends in q and starts at some state in o^*. The result then follows by induction. Let us fix $i \in \mathcal{I}$, $q \in o^*$, and let $j = \mathsf{succ}_\mathcal{I}(i)$. As $h_j' \preceq_{r_j} f^*$, we have that $h_j'(q) \neq +\infty$. From Lemma 11, there is a concrete path $\pi = q_0\sigma_0 \dots q_n$ from $q_0 \in o_i$ ending at $q_n = q$ such that $h_j'(q) = h_i(q_0) + w(\pi)$. As $h_j'(q) \neq +\infty$ it follows that $h_i(q_0) \neq +\infty$, and as $h_i(q_0) = +\infty$ if and only if $f^*(q_0) = +\infty$, it follows that $q_0 \in o^*$. Note that Lemma 11 implies for all $k \leq |\pi|$ we have

$$w(\pi[..k]) = h_{i+k}'(q_k) - h_i(q_0) = h_{i+k}(q_k) - h_i(q_0).$$

In particular $w(\pi) = h_j(q) - h_i(q_0)$.

Now let $\pi = q_0\sigma_0 \ldots$ be the infinite path implied by the above construction and, for convenience, for $i \in \mathcal{I}$ let $\pi_i = q_i\sigma_i \ldots q_j$ where $j = \mathsf{succ}_{\mathcal{I}}(i)$. To show $\overline{\mathsf{MP}}(\pi) < 0$ we need to show $\limsup_{\ell\to\infty} \frac{1}{\ell}w(\pi[..\ell]) < 0$. To prove this, we will show there exists a constant $\beta < 0$ such that for all sufficiently large ℓ we have $w(\pi[..\ell]) \leq \beta \cdot \ell$.

For convenience, let $i_0 = \min \mathcal{I}$ and let $i_\ell = \max\{i \in \mathcal{I} \mid i \leq \ell\}$. From Lemma 11 and the construction of π_i we have for all $\ell \geq i_0$:

$$\begin{aligned} w(\pi[..\ell]) &= w(\pi[..i_0]) + w(\pi[i_\ell..\ell]) + \sum_{\substack{i\in\mathcal{I}\\ i\leq\ell}} w(\pi_i) \\ &= w(\pi[..i_0]) + (h_\ell(q_\ell) - h_{i_\ell}(q_{i_\ell})) + \sum_{\substack{i\in\mathcal{I}\\ i\leq\ell}} h_{\mathsf{succ}_{\mathcal{I}}(i)}(q_{\mathsf{succ}_{\mathcal{I}}(i)}) - h_i(q_i) \\ &= w(\pi[..i_0]) + h_\ell(q_\ell) - h_{i_0}(q_{i_0}) \\ &\leq w(\pi[..i_0]) + g_\ell^{(n_\ell)}(q_\ell) - r_\ell - h_{i_0}(q_{i_0}) \end{aligned}$$

There are only finitely many values for $g_\ell^{(n_\ell)}(q_\ell)$ and from Lemma 15 we get $r_\ell \geq \lfloor \frac{\ell}{N} \rfloor$. Hence

$$w(\pi[..\ell]) \leq \alpha - \beta' \cdot \ell$$

for constants α and $\beta' > 0$. Thus there exists $\beta < 0$ such that for sufficiently large ℓ we have $w(\pi[..\ell]) \leq \beta \cdot \ell$. Hence $\underline{\mathsf{MP}}(\pi) \leq \overline{\mathsf{MP}}(\pi) < 0$. ☐

6 Forcibly Terminating Games

The reachability game defined in the previous section gives a sufficient condition for determining the winner in an MPG with limited observation. However, as there may be plays where no player wins, such games are not necessarily determined. The first subclass of MPGs with limited observation we investigate is the class of games where the associated reachability game is determined.

Definition 3. An MPG with limited observation is *forcibly terminating* if in the corresponding reachability game Γ either Adam has a winning strategy to reach states in $\mathcal{T}_\forall$ or Eve has a winning strategy to reach states in $\mathcal{T}_\exists$.

It follows immediately from Theorem 4 that finite-memory strategies suffice for both players in forcibly terminating games. Note that an upper bound on the memory required is the number of states in the reachability game restricted to a winning strategy, and this is exponential in N, the bound obtained in Lemma 15.

Theorem 5 (Finite-memory determinacy)**.** *One player always has a winning finite-memory observation-based strategy in a forcibly terminating MPG.*

We now consider the complexity of two natural decision problems associated with forcibly terminating games: the problem of recognizing if an MPG is forcibly terminating and the problem of determining the winner of a forcibly terminating game. Both results follow directly from the fact that we can accurately simulate a Turing Machine with an MPG with limited observation.

Proposition 20. *Let M be a Deterministic Turing Machine. Then there exists an MPG with limited observation G, constructible in polynomial time, such that Eve wins the associated reachability game Γ if and only if M halts in the accept state and Adam wins Γ if and only if M halts in the reject state.*

We will in fact show how to simulate a (deterministic) four-counter machine (4CM). The standard reduction from Turing Machines to 4CMs, via finite state machines with two stacks (see *e.g.* [17]), is readily seen to be constructible in polynomial time.

Counter machines A counter machine (CM) M consists of a finite set of control states S, an initial state $s_I \in S$, a final accepting state $s_A \in S$, a final rejecting state s_R, a set C of integer-valued counters and a finite set δ_M of instructions manipulating the counters. δ_M contains tuples $(s, instr, c, s')$ where $s, s' \in S$ are source and target states respectively, the action $instr \in \{\text{INC}, \text{DEC}\}$ applies to counter $c \in C$. It also contains tuples of the form $(s, 0\text{CHK}, c, s', s^0)$ where s', s^0 are two target states, one of which will be chosen depending on the value of counter c at the moment the instruction is "executed". Without loss of generality we may assume M is deterministic in the sense that for every state $s \in S$ there is exactly one instruction of the form $(s, 0\text{CHK}, \cdot, \cdot, \cdot)$ in δ_M or one of the form $(s, \cdot, \cdot, \cdot)$. We also assume that DEC instructions are always preceded by 0CHK instructions so that counter values never go below 0.

A *configuration* of M is a pair (s, v) where $s \in S$ and $v : C \to \mathbb{N}$ is a valuation of the counters. A *valid run* of M is a finite sequence $(s_0, v_0)\delta_0(s_1, v_1)\delta_1 \ldots\ \delta_{n-1}(s_n, v_n)$ where $\delta_i \in \delta_M$ is either $(s_i, instr_i, c_i, s'_i)$ or $(s_i, instr_i, c_i, s'_i, s^0_i)$ and (s_i, v_i) are configurations of M such that $s_0 = s_I$, $v_0(c) = 0$ for all $c \in C$, and for all $0 \leq i < n$ we have that:

- $v_{i+1}(c) = v_i(c)$ for $c \in C \setminus \{c_i\}$;
- if $instr_i = \text{INC}$ then $v_{i+1}(c_i) = v_i(c_i) + 1$ and $s_{i+1} = s'_i$;
- if $instr_i = \text{DEC}$ then $v_{i+1}(c_i) = v_i(c_i) - 1$ and $s_{i+1} = s'_i$;
- if $instr_i = 0\text{CHK}$ then $v_{i+1}(c_i) = v_i(c_i)$ and if $v_i(c_i) = 0$ we have $s_{i+1} = s^0_i$, otherwise $s_{i+1} = s'_i$.

The run is *accepting* if $s_n = s_A$ and it is *rejecting* if $s_n = s_R$.

Proof of Proposition 20. Given a 4CM $M = (S, s_I, s_A, s_R, C, \delta_M)$, we now show how to construct an MPG with limited observation G in which Eve wins the associated reachability game Γ if and only if M has an accepting run, and Adam wins Γ if and only if M has a rejecting run. Plays in G correspond to executions of M. As we will see, the tricky part is to make sure that zero-check instructions are faithfully simulated by one of the players. Initially, both players will be allowed to declare how many instructions the machine needs to execute in order to reach an accepting or rejecting state. Either player can bail out of this initial "pumping phase" and become the *Simulator*. The Simulator is then responsible for the faithful simulation of M and the opponent will be monitoring the simulation and punish him if the simulation is not executed correctly. Let us now go into the details.

Control structure The control structure of the machine M is encoded in the observations of our game. To be precise, to each state of M, there will correspond at most three observations in the game. We require two copies of each such observation since, in order to punish Adam or Eve (whoever plays the role of Simulator), existential and universal gadgets have to be set up in a different manner. For technical reasons that will be made clear below, we also need two additional observations. Formally, the observation set in our game contain observations $\{b^+, b^0, b^-\}$, $\{a^+, a^-\}$, and $\{q_I\}$, which do not correspond to instructions from the 4CM but they are used in gadgets that will make sure that zero tests are faithfully executed.

Counter values The values of counters will be encoded using the weights of transitions that reach designated states. We will associate to each observation (so to each state in the 4CM) two states for each counter: c_i^+ and c_i^-, for $i \in \{1, 2, 3, 4\}$. Intuitively, an abstract path, corresponding to the simulation of a run of the machine, will encode the value of counter i, at each step, as the weight of the shortest suffix from the initial pumping gadget to c_i^+.

Start of the construction The mean-payoff game with limited observation $G = (Q, q_I, \Sigma, \Delta, w, \mathsf{Obs})$ starts in observation $\{q_I\}$. For now, we will describe the transitions of G on symbols $\Sigma' \subset \Sigma$ which allow Eve to "declare" the value of a counter as being zero or non-zero, as well as "bailing" from certain gadgets in the game. More formally, let $\Sigma' := \{z, \overline{z}, \texttt{bail}\}$. The transition relation Δ contains σ-transitions (for all $\sigma \in \Sigma'$) from q_I to b^+, b^0, b^-. This observation represents the pumping phase of the simulation. From here each player will be allowed to declare how many steps they require to reach a halting state that will accept or reject.

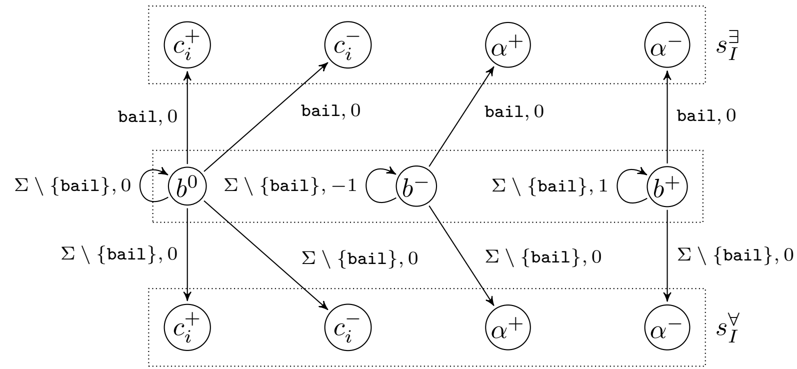

Figure 3: Initial pumping gadget for the 4CM simulation

If Adam bails, we go to the initial instruction of M on the universal side of the construction ($s_I^\forall$), if Eve does so then we go to the analogue in the existential side ($s_I^\exists$). Σ' contains a symbol `bail` which represents Eve choosing to leave the gadget and try simulating an accepting run of M, that is $\Delta \ni (b^+, \texttt{bail}, (s_I^\exists, \alpha^-))$, $(b^-, \texttt{bail}, (s_I^\exists, \alpha^+))$, and $(b^0, \texttt{bail}, (s_I^\exists, c))$ for $c \in \{c_i^+, c_i^- \mid i = 1, \dots, 4\}$. For all other actions in Σ', self-loops are added on the states b^+, b^0, b^- with weights $+1, 0, -1$ respectively. Meanwhile, Adam is able to exit the gadget at any moment—via non-deterministic transitions $(b^+, \sigma, (s_I^\forall, \alpha^-))$, $(b^-, \sigma, (s_I^\forall, \alpha^+))$, $(b^0, \sigma, (s_I^\forall, c))$ where $c \in \{c_i^+, c_i^-\}$ for all i and $\sigma \in \Sigma' \setminus \{\texttt{bail}\}$—to the universal side of the construction, *i.e.* he will try to simulate a rejecting run of the machine. Bailing transitions (transitions going to states $(s_I^\exists, \cdot)$ or $(s_I^\forall, \cdot)$) have weight 0.

Note that after these initial transitions the (simulated) value of all the counters is 0. Indeed, this corresponds to the beginning of a simulation of M starting from configuration (s_I, v) where $v(c) = 0$ for all $c \in C$.

Counter increments & decrements Let us now explain how Eve simulates increments of counter values using this encoding (decrements are treated similarly). The gadget we explain below actually works the same in both sides of the construction, *i.e.* the universal and existential gadgets for increments and decrements are identical. For that, consider Figure 4, the upper part of the figure is related to the state s of M, while the bottom part is related to the state s' of M, and assume that $(s, \text{INC}, c_i, s') \in \delta_M$.

As can be seen in the figure, the observation related to the instruction s contains the states c_i^+, c_i^-. These states are used for the encoding of the value of counter c_i. The additional states α^+, α^- are used to encode the number of steps in the simulation (again one positive ending in α^+ and one negative encoding in α^-). Now, let us consider the transitions of the gadget. The increment of the counter c_i from state s to state s' is encoded using the weights on the transitions that go from the observation s to the observation s'. As you can see, the weight on the edge between the copy of state c_i^+ of observation s to the copy of this state in observation s' is equal to $+1$, while the weight on the edge between the copy of state c_i^- of observation s to the copy of this state in observation s' is equal to -1. As you can see from the figure, when going from state s to state s', we also increment the additional counter that keeps track of the number of steps in the simulation of M. As the machine is deterministic there is no choice for Eve in observation s, since only an increment can be executed, this is why, regardless of the action chosen from Σ', the same transition is taken.

Existential zero checks Now, let us turn to the gadget of Figure 5, that is used to simulate zero-check instructions. We first focus on the case in which it is the duty of Eve to reveal if the counter has value zero or not, by forcibly choosing the next letter to play in $\{z, \overline{z}\} \subset \Sigma'$. In the observation that corresponds to the state s of M, Eve decides to declare that the counter c_i is equal to zero (by issuing z) or not (by issuing $\overline{z}$), then Adam resolves non-determinism as follows. If Eve does not cheat then Adam should let the simulation continue to either s^0 or s' depending on Eve's choice (the figure only depicts the branching to s^0,

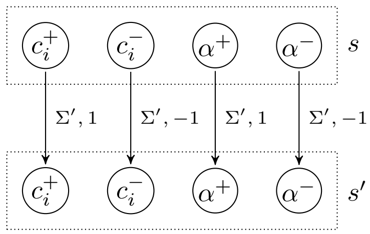

Figure 4: Observation gadget for (s, INC, c_i, s') instruction. For $(s, q) \in Q$ only the q component is shown.

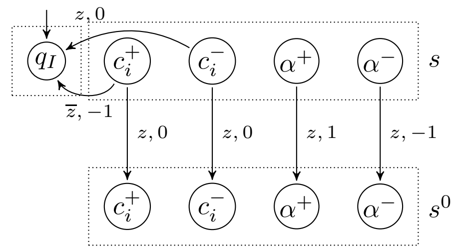

Figure 5: Existential observation gadget for $(s, 0\text{CHK}, c_i, s', s^0)$ instruction. Transitions to s' observation not shown.

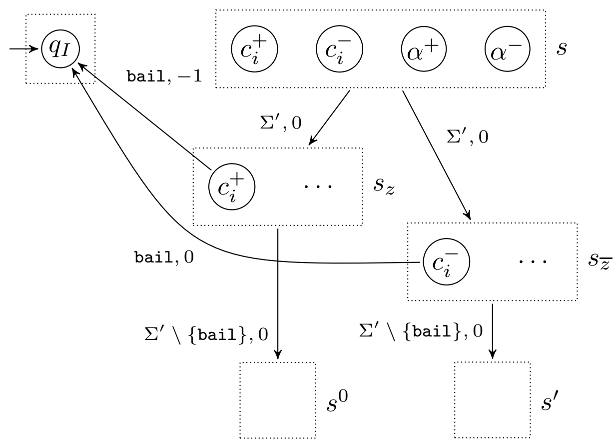

Figure 6: Universal observation gadget for $(s, 0\text{CHK}, c_i, s', s^0)$ instruction. Transitions to s', s^0 observations are weighted as with the existential observation gadget.

the branching to s' is similar). Now if Eve has cheated, then Adam should have a way to retaliate: we allow him to do so by branching to observation $\{q_I\}$ from state (s, c_i^-) with weight 0 in case z has been issued and the counter c_i is not equal to zero and with weight -1 in case $\overline{z}$ has been issued and the counter c_i is equal to zero. It should be clear that in both cases Adam closes a bad abstract cycle.

Universal zero checks A similar trick is used for the gadget from Figure 6, where Adam is forced to simulate a truthful zero check or lose Γ. Since Adam can control non-determinism and not the action chosen, we have transitions going from $(s, \cdot)$ to states in both $(s_z, \cdot)$ and $(s_{\overline{z}})$ with weight 0 and all actions in Σ'. Eve is then allowed to branch back to q_I as follows. If Adam does not cheat, then Eve will play any action in $\Sigma' \setminus \{\texttt{bail}\}$ and transitions, with weights similar to those used in the the existential check gadget, will take the play from $(s_z, \cdot)$ to $(s^0, \cdot)$ and from $(s_{\overline{z}}, \cdot)$ to $(s', \cdot)$. Now if Adam has cheated by taking the play to $(s_z, \cdot)$ when c_i was not zero, then Eve—by playing `bail`—can go from (s_z, c_i^+) to the initial observation with weight -1 and close a good abstract cycle. (Recall that a good abstract cycle is **good for Eve**: by repeatedly closing good cycles, Eve can win the MPG. Thus, closing a good cycle is our way of punishing Adam.) If Adam cheated by taking the play to $(s_{\overline{z}}, \cdot)$ when c_i was indeed zero, Eve can go (with the same action) from $(s_{\overline{z}}, c_i^-)$ to the initial observation with weight 0 again and close a good abstract cycle. Indeed, Adam can escape the zero-check gadget by choosing a non-proper successor. We will shortly explain why this is not a viable option for him.

Stopping Simulator It should be clear also from the gadgets, that the opponent of Simulator has no incentive to interrupt the simulation if there is no cheat. Doing so is actually beneficial to Simulator, who can get a function-action sequence which makes him win Γ.

Finally, Δ also contains self-loops at all $(s_A, \cdot)$ with all Σ' and with 0 weights and at all $(s_R, \cdot)$ with all Σ' and with -1 weights. Thus, if the play reaches the observation representing state s_F or s_R from M then Simulator will be able to force function-action sequences which allow him to win Γ.

Making Adam play properly We will now explain the idea behind observation gadget $\{a^+, a^-\}$. Note that Adam could break Eve's simulation of an accepting run by declaring the value of functions from Γ, which are actually our means of encoding the values of the counters, to be $+\infty$ (or at least some subset of the values of the functions). We describe how we obtain the final set of actions for the constructed game and mention the required transitions from every observation in the game so that it is not in the interest of Adam to do the latter.

Denote by (o, q_i) the i-th state in observation o. Observe that in our construction we need at most 10 states per observation: two copies of every counter state and two additional step counters. The full action set in the game is defined as follows $\Sigma := \Sigma' \cup \{q_i \mid 0 \leq i < 10\} \cup \{\texttt{ex}\}$. For every observation o in G we add the transitions $((o, q_i), q_i, a^-), ((o, q_j), q_i, a^+)$ for all q_i, q_j in o where $q_i \neq q_j$. To finish, we add the self-loops (a^+, σ, a^+) and (a^-, σ, a^-), with weights $+1$ and -1 respectively, as part of Δ for all $\sigma \in \Sigma$. Clearly, Adam cannot choose anything other than proper σ-successors in Γ or he gives Eve enough information for her to win the game. Indeed, if the game is currently at observation o and Adam has chosen a non-proper successor, then Eve knows some state $q_i \in o$ does not currently hold the token. Hence, she can play action q_i and be sure to reach the state a^+ where she will win the game.

Bound on the length of the simulation All that remains is to show how we allow the opponent of Simulator to stop the simulated run of M in case Simulator exhausts the number of instructions he initially declared would be used to accept or reject. The ex action is used in transitions $((o, \alpha^+), \texttt{ex}, q_I) \in \Delta$ for all observation gadgets o in the universal side of the construction. This allows Eve to stop Adam (who is playing Simulator) in case he tries to simulate more steps than he said were required for M to reject. (Indeed, Adam may instead choose to move to another observation besides $\{q_I\}$ on ex, but then he reveals to Eve that some state q_i in the following observation cannot hold the token, and she will then play $q_i \in \Sigma$ to win from there.) Similarly, in the existential part of the construction, we add a transition $((o, \alpha^-), \sigma, q_I)$ for all observation gadgets o and all $\sigma \in \Sigma$, which lets Adam stop Eve's simulation if she tries to cheat in the same way.

Finally, to have the game be limited observation we let all missing σ-transitions on the existential (resp. universal) side of the simulation go to a sink state in which Adam (Eve) wins.

Correctness Now, let us prove the correctness of the overall construction. Assume that M has an accepting or rejecting run. Then, Simulator, by simulating faithfully the run of M has an observation-based strategy that allows him to force abstract paths which induce good or bad abstract cycles depending on who is simulating. Clearly, in this case even if the opponent decides to interrupt the simulation M at a zero-check gadget, he will only be helping Simulator.

If M has no accepting or rejecting run, then by simulating the machine faithfully, Simulator will be generating cycles in the control state of the machine and such abstract paths are "mixed" because of concrete paths between corresponding α^-, α^+ states. Cheating does not help him either since after the opponent catches him cheating and restarts the simulation of the machine (by returning to the initial observation), the corresponding path is losing for him. □

It follows from Proposition 20 that determining if a given limited-observation MPG is forcibly terminating is as hard as determining if a given 4CM halts (by accepting or rejecting). As the latter problem is known to be undecidable, we obtain the following result.

Theorem 6 (Class membership)**.** *Let G be an MPG with limited observation. Determining if G is forcibly terminating is undecidable.*

Although determining if Eve wins a forcibly terminating MPG is decidable, Proposition 20 implies the problem is extremely hard. Let R denote the class of all decision problems solvable by a Turing machine. We say a problem is R-complete under polynomial reductions if it is decidable and if any decidable problem reduces to it in polynomial time.

Theorem 7 (Winner determination)**.** *Let G be a forcibly terminating MPG. Determining if Eve wins G is* R*-complete under polynomial reductions.*

Proof. For decidability, Lemma 15 implies that an alternating Turing Machine simulating a play on Γ will terminate. Regarding hardness, we will argue that any decidable problem reduces to winner determination of forcibly terminating MPGs via our 4CM simulation. Indeed, it is known that any given Turing machine can be simulated by a 4CM of polynomial size with respect to the size of the original Turing machine. Also, given a decidable problem, we know there exists a Turing machine which, given any instance of the problem, always halts and outputs a positive or negative answer. We construct, from the Turing machine and a given instance of the problem, the corresponding 4CM and, in turn, the corresponding limited-observation MPG as in the proof of Proposition 20. Since the original Turing machine always halts, the game is guaranteed to be forcibly terminating. Now, it should be clear that in the constructed game Eve wins if and only if the 4CM accepts if and only if the Turing machine accepts the instance. As both the construction of the CM and the game are feasible in polynomial time, the result follows. □

6.1 Modifications for Theorem 1

To prove Theorem 1 we reduce from the non-termination problem for 2CMs using a construction similar to the one used in the proof of Proposition 20. Given a 2CM M, we construct a game G as in the proof of Proposition 20, with the following adjustments:

- We only consider the universal side of the simulation, but allow both players to exit the initial pumping phase into it;
- The observation corresponding to the accept state of M is a sink state winning for Adam;
- The α^- states are replaced with β states which have transitions to other β states of weight 0 except in one case specified below;
- The pumping gadget has self loops of weights $0, 0, -1$ and the transition from b^+ to β has weight -1 if Eve exits and weight 0 if Adam exits;
- The `ex` transition also goes from β states to q_I.

Suppose the counter machine halts in N steps. The observation-based winning strategy for Adam is as follows. Exit the pumping gadget after N steps and faithfully simulate the counter machine. Suppose Eve can beat this strategy. If she allows a faithful simulation for N steps then Adam reaches a sink state and wins, so Eve must play `ex` within N steps of the simulation. Let us consider each cycle of at most $2N$ steps. If she waits for Adam to exit the pumping gadget then the number of steps in the simulation is less than the number of steps in the pumping gadget, so a negative cycle is closed. On the other hand if she exits the pumping gadget before N steps then the cycle through the β states has negative weight. In both cases, a negative cycle is closed in at most $2N$ steps, so the limit average is bounded above by $-\frac{1}{2N}$. Thus this strategy is winning for Adam.

Now suppose the counter machine does not halt. The (infinite memory) observation-based winning strategy for Eve is as follows. For increasing n, exit the pumping gadget after n steps and faithfully simulate (*i.e.* call any, and only, cheats of Adam) the counter machine for n steps. Then play `ex` and increase n. Cheating in the simulation does not benefit Adam, so we can assume Adam faithfully simulates the counter machine. Likewise, if Eve always waits until the number of steps in the simulation exceeds the number of steps in the pumping gadget, then there is no benefit for Adam to exit the pumping gadget. However if the play proceeds as Eve intends then the weight of the path through the α^+ states is non-negative and although the weight through the β states is negative, the limit average is 0. Thus the strategy is winning for Eve. □

7 Forcibly First Abstract Cycle Games

In this section and the next we consider restrictions of forcibly terminating games in order to find sub-classes with more efficient algorithmic bounds. The negative algorithmic results from the previous section largely arise from the fact that the abstract cycles required to determine the winner are not necessarily simple cycles. Our first restriction of forcibly terminating games is the restriction of the abstract cycle-forming game to simple cycles.

More precisely, let $G = (Q, q_I, \Sigma, \Delta, w, \mathsf{Obs})$ be an MPG with limited observation and $\Gamma = (\Pi, \Sigma, f_I, \delta, \mathcal{T}_\exists, \mathcal{T}_\forall)$ be the associated reachability game. Define $\Pi' \subseteq \Pi$ as the set of all sequences $f_0\sigma_0 f_1\sigma_1 \ldots f_n \in \Pi$ such that $\mathsf{supp}(f_i) \neq \mathsf{supp}(f_j)$ for all $0 \leq i < j < n$ and denote by Γ' the reachability game $(\Pi', \Sigma, f_I, \delta', \mathcal{T}'_\exists, \mathcal{T}'_\forall)$ where δ' is δ restricted to Π', $\mathcal{T}'_\exists = \mathcal{T}_\exists \cap \Pi'$, and $\mathcal{T}'_\forall = \mathcal{T}_\forall \cap \Pi'$.

Definition 4. An MPG with limited observation is *forcibly first abstract cycle* (or forcibly FAC) if in the associated reachability game Γ' either Adam has a winning strategy to reach states in $\mathcal{T}'_\forall$ or Eve has a winning strategy to reach states in $\mathcal{T}'_\exists$.

One immediate consequence of the restriction to simple abstract cycles is that the bound in Lemma 15 is at most $|\mathsf{Obs}|$. In particular an alternating Turing Machine can, in linear time, simulate a play of the reachability game and decide which player, if any, has a winning strategy. Hence the problems of deciding if a given MPG with partial observation is forcibly FAC and deciding the winner of a forcibly FAC game are both solvable in PSPACE. The next results show that there is a matching lower bound for both these problems.

Theorem 8 (Class membership)**.** *Let G be an MPG with limited observation. Determining if G is forcibly FAC is* PSPACE*-complete.*

Proof. For PSPACE membership we observe that a linear bounded alternating Turing Machine can decide whether one of the players can force to reach $\mathcal{T}'_\exists$ or $\mathcal{T}'_\forall$ in Γ'. To show hardness we use a reduction from the True Quantified Boolean Formula (TQBF) problem. Given a *fully quantified* Boolean formula $\Psi = \exists x_0 \forall x_1 \ldots \mathcal{Q} x_{n-1}(\Phi)$, where $\mathcal{Q} \in \{\exists, \forall\}$ and Φ is a Boolean formula expressed in *conjunctive normal form* (CNF), the TQBF problem asks whether Ψ is true or false. The TQBF problem is known to be PSPACE-complete [22].

This problem is often rephrased as a game between Adam and Eve. In this game the two players alternate choosing values for each x_i from Φ. Eve wins if the resulting evaluation of Φ is true while Adam wins if it is false. We simulate such a game with the use of "diamond" gadgets that allow Eve to choose a value for existentially quantified variables by letting her choose the next observation. Similarly, the same gadget—except for the labels on the transitions, which are completely non-deterministic in the following case—allow Adam to choose values for variables that are universally quantified.

We construct a game $G_\Psi = (Q, q_I, \Sigma, \Delta, w, \mathsf{Obs})$ in which there are no concrete negative cycles, hence it follows from Lemma 13 that there are no bad cycles. The game will thus be forcibly FAC if and only if Eve is able to force good cycles. If Eve is unable to prove the QBF is true, Adam will be able to avoid such plays. For this purpose, the "diamond" gadgets employed have two states per observation. This will allow two disjoint concrete paths to go from the initial state q_I through the whole arena and form a simple abstract cycle that is either good or not good depending on where the cycle started from.

Concretely, let x_1 be a universally quantified variable from Ψ. We add a gadget to G_Ψ consisting of eight states grouped into four observations: $\{b_0^-, b_0^0\}$, $\{\overline{x_1}, \overline{z_1}\}$, $\{x_1, z_1\}$, $\{b_1^-, b_1^0\}$. We also add the following transitions:

- from b_0^- to $\overline{x_1}$ and x_1, b_0^0 to $\overline{z_1}$ and z_1, with all Σ and weight 0;
- from $\overline{x_1}$ and x_1 to b_1^-, $\overline{z_1}$ and z_1 to b_1^0, with all Σ and the first two with weight -1 while the last two have weight 0.

Figure 7 shows the universal "diamond" gadget just described. The observation $\{\overline{x_1}, \overline{z_1}\}$ corresponds to the variable being given a false valuation, whereas the $\{x_1, z_1\}$ observation models a true valuation having been picked. Observe that the choice of the next observation from $\{b_0^-, b_0^0\}$ is completely non-deterministic, *i.e.* Adam chooses the valuation for this variable.

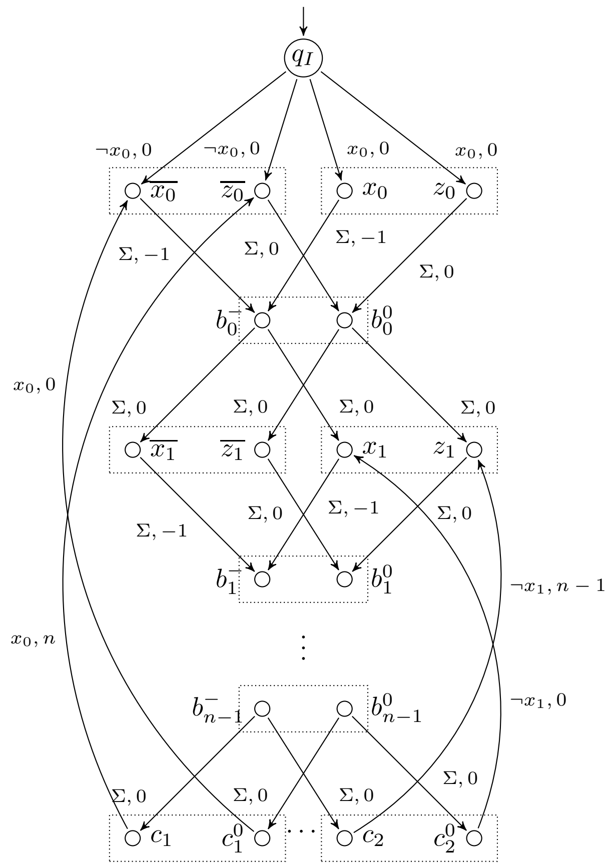

Figure 7: Corresponding game for QBF $\exists x_0 \forall x_1 \ldots (\neg x_0) \wedge (x_1) \ldots$

For existentially quantified variables, the first set of transitions from the gadget is slightly different. Let x_i be an existentially quantified variable in Ψ, then the upper part of the gadget includes transitions from b_i^- to $\overline{x_i}$ and from b_i^0 to $\overline{z_i}$ with action symbol $\neg x_i$ and weight 0; as well as transitions from b_i^- to x_i and from b_i^0 to z_i with action symbol x_i and weight 0.

A play in G_Ψ traverses gadgets for all the variables from the QBF and eventually gets to the observation $\{b_{n-1}^-, b_{n-1}^0\}$ where the assignment of values for every variable has been simulated. At this point we want to check whether the valuation of the variables makes Φ true. We do so by allowing Adam to choose the next observation (corresponding to one of the clauses from the CNF formula Φ) and letting Eve choose a variable from the clause (which might be negated). Let x_i (resp. $\overline{x_i}$) be the variable chosen by Eve, in G_Ψ the next observation will correspond to closing a good abstract cycle if and only if the chosen valuation of the variables for Ψ assigns to x_i a true (false) value. For this part of the construction we have $2m$ states grouped in m observations, where m is the number of clauses in the formula. The lower part of figure 7 shows the clause observations we just described.

Denote by $\{c_i, c_i^0\}$ the observation associated to clause c_i. The game has transitions from c_i to x_i (or $\overline{x_i}$) and from c_i^0 to z_i ($\overline{z_i}$) with action symbol x_i ($\neg x_i$) and weight $n - i$ for the first, 0 for the latter, if and only if the clause c_i includes the (negated) variable x_i.[3]

After Eve and Adam have chosen values for all variables (and the game reaches observation $\{b_{n-1}^-, b_{n-1}^0\}$)

[3]All missing transitions for G_Ψ to be complete go to a dummy state with a negative and 0-valued non-deterministic transitions.

there are two concrete paths corresponding with the current play: one with payoff 0 and one with payoff $-n$. When Adam has chosen a clause and Eve chooses a variable x_i from the clause, the next observation is reached with both concrete paths having payoffs 0 and $-i$. Observe, however, that if we consider the suffix of said concrete paths starting from $\{x_i, z_i\}$ or $\{\overline{x_i}, \overline{z_i}\}$—depending on which valuation the players chose—both payoffs are 0. Indeed, if the observation was previously visited, *i.e.* Eve has proven the clause to be true, then a good cycle is closed. On the other hand, if the observation has not been visited previously, then Eve has no choice but to keep playing and the play thus reaches observation $\{b_i^-, b_i^0\}$. We note that traversing the lower part of our "diamond" gadgets results in a *mixed* payoff of -1 and 0 and since $\{b_i^-, b_i^0\}$ has already been visited, a cycle is closed that is not good. To summarize, either a good cycle is closed when moving from $\{b_{n-1}^-, b_{n-1}^0\}$ to $\{x_i, z_i\}$ (or, respectively, $\{\overline{x_i}, \overline{z_i}\}$) if the latter observation had been visited before; or a bad cycle is closed on the next step when moving to $\{b_i^-, b_i^0\}$.

Therefore, if Ψ is true then Eve has a strategy to make the first cycle closed be a good one, so G_Ψ is forcibly FAC. Conversely, if Ψ is false then Adam has a strategy to make the first cycle formed be not good (mixed, in fact). Hence G_Ψ is not forcibly FAC. ☐

We can slightly modify the above construction in such a way that if the game does not finish when the play returns to a variable then Adam can close a bad cycle (instead of just being able to force a mixed cycle). This results in a forcibly FAC game that Eve wins if and only if the formula is satisfied. Hence,

Theorem 9 (Winner determination)**.** *Let G be a forcibly FAC MPG. Determining if Eve wins G is* PSPACE-*complete.*

Proof. We describe the modifications required to the construction used in the proof of Theorem 8.

First, we augment every observation with $2n$ states corresponding to variables from Φ and their negation (say, y_i and $\overline{y_i}$ for $0 \leq i < n$).

We then add transitions from every new state y_i ($\overline{y_i}$) to its counterpart in the next observation so as to form $2n$ new disjoint cycles going from q_I through the whole construction. (Note that, up to this point $\mathsf{Plays}(G_\Psi)$ remains unchanged. That is, the set of abstract paths in the game constructed for the proof of Theorem 8 is the same as the set of abstract paths in the present game.) These new transitions all have weight zero except for a few exceptions:

- the transition corresponding to the lower part of the gadget which represents the variable itself, *i.e.* the transition from augmented observation $\{x_i, z_i, \dots\}$ to $\{b_i^-, b_i^0, \dots\}$ (resp. $\{\overline{x_i}, \overline{z_i}, \dots\}$ to $\{b_i^-, b_i^0, \dots\}$) has weight of $+1$ for the y_i-transition ($\overline{y_i}$-transition);
- outgoing transitions from clause observations have weight -1 on the y_i-transition going to the x_i-gadget—*i.e.* if we let y_j be one of the new states in the clause observation and y_j' the corresponding state in the x_i-gadget, then $w(y_j, \sigma, y_j') = -1$; and
- at every $\{x_i, z_i, \dots\}$ and $\{\overline{x_i}, \overline{z_i}, \dots\}$ augmented observation, Adam is allowed to resolve non-determinism by going back to q_I—*i.e.* in these observations we add a transition from y_i and $\overline{y_i}$, respectively, back to the initial state.

Let us argue that the game is forcibly FAC and that the QBF instance is true if and only if Eve wins the reachability game associated with the constructed MPG. When the play reaches $\{x_i, z_i, \dots\}$ (or $\{\overline{x_i}, \overline{z_i}, \dots\}$) after Eve and Adam choose values for all the variables and after she has chosen a variable from a clause given by Adam, then the concrete path ending at y_i (resp. $\overline{y}_i$) has weight 0 if the observation was previously visited, and weight -1 if it was not. The concrete paths ending at all the other new states have weight 0 or $+1$ depending on the choices made by the players. Concrete paths ending at x_i and z_i states are as before (mixed if the observation has not been witnessed, and good otherwise). Thus if the observation was previously visited, then the cycle closed is good as before. If the observation was not previously visited, then Adam can now choose to play to q_I from y_i ($\overline{y_i}$) and close a bad cycle (of weight -1). Note that if Adam chooses to play to q_I before the clause gadgets are reached then he will only be closing good cycles. Following the same argument as before, if Ψ is true then Eve has a winning strategy and if Ψ is false then Adam has a winning strategy. So G_Ψ is forcibly FAC and Eve wins if and only if Ψ is true. ☐

It also follows from the $|\mathsf{Obs}|$ upper bound on plays in Γ' that there is an exponential upper bound on the memory required for a winning strategy for either player. Furthermore, we can show this bound is tight—the games constructed in the proof of Theorem 9 can be used to show that there are forcibly FAC games that require exponential memory for winning strategies.

Theorem 10 (Exponential memory determinacy)**.** *One player always has a winning observation-based strategy with exponential memory in a forcibly FAC MPG. Further, for any $n \in \mathbb{N}$ there exists a forcibly FAC MPG, of size polynomial in n, such that any winning observation-based strategy has memory at least 2^n.*

Proof. For the upper bound we observe that plays in Γ' are bounded in length by $|\mathsf{Obs}|$. It follows that the strategy constructed in Theorem 4 has memory at most $|\Sigma|^{|\mathsf{Obs}|}$.

For the lower bound, consider the forcibly FAC game G_n constructed in the proof of Theorem 9 for the formula

$$\varphi_n = \forall x_1 \forall x_2 \ldots \forall x_n \exists y_1 \ldots \exists y_n . \bigwedge_{i=1}^{n} (x_i \vee \neg y_i) \wedge (\neg x_i \vee y_i).$$

As φ_n is satisfied, Eve wins G_n. Now consider any observation-based strategy for Eve with memory $< 2^n$. As there are 2^n possible assignments for the values of $x_1, \ldots x_n$ it follows there are at least two different assignments of values such that Eve makes the same choices in the game. Suppose these two assignments differ at x_i and assume w.l.o.g. that Eve's choice is at $(n+i)$-th gadget to play to y_i. Then Adam can win the game by choosing values for the universal variables that correspond to the assignment which sets x_i to false, and then playing to the clause $(x_i, \vee \neg y_i)$. Thus any winning observation-based strategy for Eve must have size at least 2^n.

In a similar way the game defined by the formula $\neg\varphi_n$ is won by Adam, but any winning observation-based strategy must have size at least 2^n. ☐

8 First Abstract Cycle Games

We now consider a structural restriction that guarantees Γ' is determined. Recall that to any limited-observation MPG $G = (Q, q_I, \Sigma, \Delta, w, \mathsf{Obs})$ we associate a reachability game $\Gamma = (\Pi, \Sigma, f_I, \delta, \mathcal{T}_\exists, \mathcal{T}_\forall)$ and that Γ' is the restriction of Γ to simple function-action sequences (with respect to the supports). That is, Π' is the set of all sequences $f_0 \sigma_0 f_1 \sigma_1 \ldots f_n \in \Pi$ such that $\mathsf{supp}(f_i) \neq \mathsf{supp}(f_j)$ for all $0 \leq i < j < n$ and the other components of Γ' are the corresponding restrictions of Γ to Π'.

Definition 5. An MPG with limited observation is a *first abstract cycle game* (FAC) if in the associated reachability game Γ' all leaves are in $\mathcal{T}'_\forall \cup \mathcal{T}'_\exists$.

Intuitively, in an FAC game G all simple abstract cycles (that can be formed) are either good or bad. Since Γ' is a full-observation finite reachability game, G is determined. Thus, by Theorem 4, we get that in every FAC one of the two players has a winning finite-memory observation-based strategy. However, we can show an even stronger result holds: one of them has a winning positional observation-based strategy.

Theorem 11 (Positional determinacy)**.** *One player always has a positional winning observation-based strategy in an FAC MPG.*

Proof. It follows then from Corollary 14 that any cyclic permutation of a good cycle is also good and any cyclic permutation of a bad cycle is also bad. Together with Lemma 13, this implies the abstract cycle-forming games associated with FAC games can be seen to satisfy the following three assumptions: (1) A play stops as soon as an abstract cycle is formed; (2) The winning condition and its complement are preserved under cyclic permutations; and (3) The winning condition and its complement are preserved under interleavings. These assumptions were shown in [1] to be sufficient for winning positional strategies to exist in any game.[4] ☐

[4]These conditions supersede those of [4] which were shown in [1] to be insufficient for positional strategies.

As we can check in polynomial time if a positional observation-based strategy is winning in an FAC MPG, we immediately have:

Theorem 12 (Winner determination). *Let G be an FAC MPG. Determining if Eve wins G is in* NP∩coNP.

A path in Γ' to a leaf not in $\mathcal{T}'_\forall \cup \mathcal{T}'_\exists$ provides a short certificate to show that an MPG with limited observation is not FAC. Thus deciding if an MPG is FAC is in coNP. A matching lower bound can be obtained using a reduction from the complement of the Hamiltonian cycle problem.

Theorem 13 (Class membership). *Let G be an MPG with limited observation. Determining if G is FAC is* coNP*-complete.*

Proof. For coNP membership, one can guess a large enough simple abstract cycle ψ and (in polynomial time with respect to Q) check that it is neither good nor bad. To show coNP-hardness we use a reduction from the complement of the Hamiltonian Cycle problem.

Given graph $\mathcal{G} = (V, E)$ where V is the set of vertices and $E \subseteq V \times V$ the set of edges. We construct a directed weighted graph with limited observation $G = (Q, q_I, \Sigma, \Delta, w, \mathsf{Obs})$ where:

- $Q = V \cup \{q_I, q_+, q_-\}$;
- $\mathsf{Obs} = \{\{v\} \mid v \in V\} \cup \{\{q_-, q_+\}, \{q_I\}\}$;
- $\Sigma = V \cup \{\tau\}$;
- Δ contains transitions (u, v, v) such that $(u, v) \in E$ and self-loops (u, v', u) for all $(u, v') \notin E$, transitions (with all σ) from q_I to both q_+ and q_- and from these last two to all states $v \in V$, as well as τ-transitions from every state $v \in V$ to q_+ and q_-;
- w is such that all outgoing transitions from q_+ and q_- have weight $1 - |V|$, (u, v, v) transitions where $(u, v) \in E$ have weight $+1$, τ-transitions to q_- from states $v \in V$ have weight -1 and all other transitions have weight 0.

Notice that the only non-deterministic transitions in G are those incident on and outgoing from the states q_+, q_-. Clearly, the only way for a simple abstract cycle to be not good and not bad (thus making G not FAC) is if there is a path from $\{q_-, q_+\} \in \mathsf{Obs}$ that traverses $|V|$ unique observations and ends with a τ-transition back at $\{q_-, q_+\}$. Such a path corresponds to a Hamiltonian cycle in $\mathcal{G}$. If there is no Hamiltonian cycle in $\mathcal{G}$ then for any play π in G, a bad cycle will be formed (hence, G is FAC). □

9 MPGs with Partial Observation

In the introduction it was mentioned that an MPG with partial observation can be transformed into an MPG with limited observation. This translation allows us to extend the notions of FAC and forcibly FAC games to the larger class of MPGs with partial observation. In this section we will investigate the resulting algorithmic effect of this translation on the decision problems we have been considering.

The idea behind the translation is to take subsets of the observations and restrict transitions to those that satisfy the limited-observation requirements. More formally, given an MPG with partial observation $G = (Q, \Sigma, \Delta, q_I, w, \mathsf{Obs})$ we construct an MPG with limited observation $G' = (Q', \Sigma, \Delta', q'_I, w', \mathsf{Obs}')$ where:

- $Q' = \{(q, K) \in Q \times 2^Q \mid q \in K \text{ and } K \subseteq o \in \mathsf{Obs}\}$,
- $q'_I = (q_I, \{q_I\})$,
- $\mathsf{Obs}' = \big\{\{(q, K) \mid q \in K\} \mid K \subseteq o \text{ for some } o \in \mathsf{Obs}\big\}$,
- Δ' contains the transitions $((q, K), \sigma, (q', K'))$ such that $(q, \sigma, q') \in \Delta$ and $K' = post_\sigma(K) \cap o$ for some $o \in \mathsf{Obs}$, and
- $w'((q, K), \sigma, (q', K')) = w(q, \sigma, q')$ for all $((q, K), \sigma, (q', K')) \in \Delta'$.

It is folklore to show that this *knowledge-based* subset construction (also known as a belief construction) preserves winning strategies for Eve. The terms belief and knowledge are used to denote a state from any variation of the classic "Reif construction" [21] to turn a game with partial observation into a game with full observation. Other names for similar constructions include "knowledge-based subset construction" (see *e.g.* [9]). In this case the resulting game is not one with full observation but one with limited observation.

Theorem 14 (Equivalence)**.** *Let G be an MPG with partial observation and G' be the corresponding MPG with limited observation as constructed above. Eve has a winning observation-based strategy in G if and only if she has a winning observation-based strategy in G'.*

The result above is shown by proving that winning observation-based strategies for Eve transfer between G and G'. It is worth noting that an observation-based strategy for Eve in G can directly be used by her in G'. Conversely, for her to use a strategy from G' in G she must keep in memory the knowledge-based subset construction herself. Hence,

Theorem 15 (Memory requirements)**.** *Let G be a partial-observation MPG and G' be the corresponding limited-observation MPG. If a player has a finite-memory observation-based winning strategy in G', then (s)he has a finite-memory observation-based winning strategy in G' which requires exponentially more memory (on the size of G).*

We say an MPG with partial observation is *(forcibly) first belief cycle*, or FBC, if the corresponding MPG with limited observation is (forcibly) FAC.

10 FBC and Forcibly FBC MPGs

Our first observation is that FBC MPGs generalize the class of *visible weight games* studied in [9]. An MPG with partial observation is considered a visible weights game if its weight function satisfies the condition that all σ-transitions between any pair of observations have the same weight. We base some of our results for FBC and forcibly FBC games on lower bounds established for problems on visible weights games.

Lemma 21. *Let G be a visible weights MPG with partial observation. Then G is FBC.*

We now turn to the decision problems we have been investigating throughout the paper. Given the exponential blow-up in the construction of the game of limited observation, it is not surprising that there is a corresponding exponential increase in the complexity of the class membership problem.

Theorem 16 (Class membership)**.** *Let G be an MPG with partial observation. Determining if G is FBC is* coNEXPTime*-complete and determining if G is forcibly FBC is in* EXPSpace *and* NEXPTime*-hard.*

Membership of the relevant classes is straightforward, they follow directly from the upper bounds for MPGs with limited observation and the (at worst) exponential blow-up in the translation from games of partial observation to games of limited observation. For hardness, we prove first the result for FBC games and comment on the changes necessary for the construction to yield the result for forcibly FBC games.

Lemma 22. *Let G be an MPG with partial observation. Determining if G is FBC is* coNEXPTime*-hard.*

Proof. We reduce from the complement of the succinct Hamilton cycle problem: Given a Boolean circuit C with $2N$ inputs, does the graph on 2^N nodes with edge relation encoded by C have a Hamiltonian cycle? This problem is known to be NEXPTime-complete [19].

The idea is to simulate a traversal of the succinct graph in our MPG: if we make 2^N valid steps without revisiting a vertex of the succinct graph then that guarantees a Hamiltonian cycle. To do this, we start with a transition of weight -2^N and add 1 to all paths every time we make a valid transition. We include a pair of transitions back to the initial state with weights 0 and -1 and ensure this is the only transition that can be taken that results in paths of different weight. The resulting game then has a mixed lasso if and only if we can make 2^N valid transitions. If we encode the succinct graph vertex in the knowledge set then the definition of an FAC game will give us an automatic check if we revisit a vertex. In fact, we store several pieces of information in the knowledge sets of the observations: the current (succinct) graph vertex, the potential successor, and the evaluation of the edge-transition circuit up to a point.

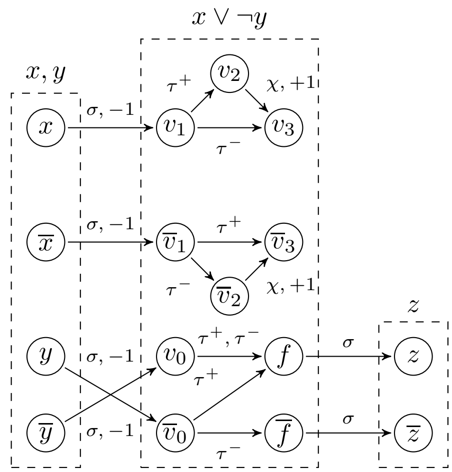

Figure 8: This is the partial-observation gadget to simulate $x \vee \neg y = z$. Inside the gate gadget we also have on all states: self-loops with weight $+1$ on χ, and 0 on $\tau^{\pm}$; 0-weight transitions to a sink on external actions. (Zero-weights have been omitted for clarity.)

Simulating gates The crucial technical trick used in our reduction is the construction of an observation gadget which simulates a logical gate. Figure 8 depicts the gadget corresponding to $z = x \vee \neg y$. We assume we have a knowledge set K that is a subset of the states from the leftmost observation. Further, we assume K induces a valid valuation of x and y, *i.e.* $x \in K$ if and only if $\overline{x} \notin K$ and similarly for y. Denote by $K \models x \vee \neg y$ the fact that the valuation of x and y makes the formula true. We also assume all concrete paths arriving at states in K have the same weight. Now, by playing σ, Eve reaches the $x \vee \neg y$ observation where she can play internal actions τ^-, τ^+, χ. We claim observation $x \vee \neg y$ allows concrete plays to reach the z observation with weight 0 without creating a non-mixed belief lasso if and only if $K \models x \vee \neg y$. The main idea is that Eve declares the truth value of x using τ^+ if it is true and τ^- otherwise, she then plays χ to cancel the -1 weight seen upon entering the observation. For example, if $K = \{x, y\}$, Eve plays σ and enters the gate observation with knowledge set $\{v_1, \overline{v}_0\}$. Then, Eve plays τ^+ and one concrete path moves from v_1 to v_2, the other from $\overline{v}_0$ to f; Eve then plays χ and concrete paths reach v_3 and f, both with weight 0; finally, she plays σ, and a concrete play reaches a sink or a concrete play reaches z (as expected since $K \models x \vee \neg y$). Crucially, the sequence of internal transitions on $\tau^{\pm}\chi$ induces a sequence of three distinct knowledge sets if and only if she declared the correct value of x. Otherwise, a lasso is formed.

We now describe the construction in detail.

Simplifying assumptions Let us assume inputs of the circuit C are labelled $x_1, \ldots x_{2N}$ and that it has k gates $G_1, \ldots, G_k$ numbered in an order that respects the circuit graph, so G_j has inputs from $\{x_i, \neg x_i : 1 \leq i < 2N + j\}$ where, for convenience, x_{2N+i} indicates the output of gate G_i. We may assume each gate has two inputs and (as we are allowing negated inputs) we may assume we only have AND and OR gates.

Construction description The game consists of two *external* actions primarily for transitions between observations: σ (solid lines) and σ' (dotted lines); and a number of *internal* actions denoted with τ and χ for transitions primarily within observations (not shown). The numbers in parentheses indicate the maximum weight that can be added to the total with internal transitions, and the edge weights indicate the weight of *all* transitions between observations.

Our game proceeds in several stages:

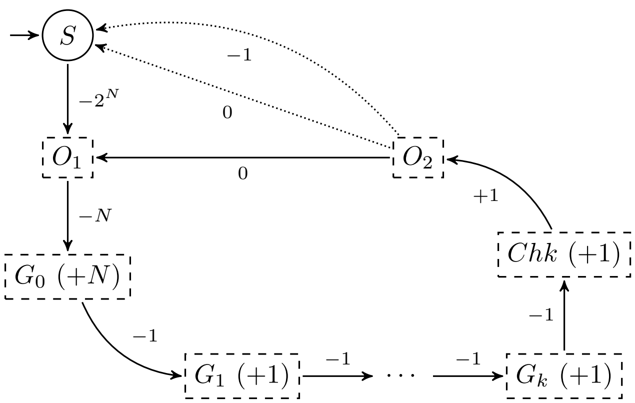

Figure 9: Overall structure of the game for succinct Hamiltonian cycle

1. The transition from S to O_1 sets the initial (succinct) vertex (stored in a subset of the states of O_1) and initializes the vertex counter to -2^N.
2. Internal transitions in G_0 select the next vertex, the transition from O_1 to G_0 initializes this procedure.
3. For $i > 0$, internal transitions in G_i evaluate gate i, incoming transitions initialize this by passing on the previous evaluations (including the current and next vertices).
4. Internal transitions in Chk test if the circuit evaluates to 1.
5. The next succinct vertex (chosen in G_0) is passed to O_2, where there is an implicit check that this vertex has not been visited before, and the counter is incremented.
6. The play can return to S, generating a mixed lasso if and only if the vertex counter is 0, *i.e.* 2^N vertices have been correctly visited, or return to O_1 with a new current succinct vertex.

The weights on the incoming transitions to an observation are designed to impose a penalty that can only be nullified if the correct sequence of internal transitions is taken. We observe that if there is a penalty that is not nullified then the game can never enter a mixed lasso (as the vertex counter will still be negative when a vertex is necessarily revisited). The overall (*i.e.* observation-level) structure of the game is shown in Figure 9.

We now describe the structure of the observations.

Observation O_1 It contains $2N$ states: $\{x_i, \overline{x}_i \mid 1 \leq i \leq N\}$. For convenience we will use the same labels across different observations, using observation membership to distinguish them. There are σ-transitions from S to $\{x_i \mid 1 \leq i \leq N\}$ with weight -2^N.

Observation O_2 It contains $2N+1$ states: $\{x_i, \overline{x}_i \mid 1 \leq i \leq N\} \cup \{\perp\}$. There are σ-transitions from each state in O_2 other than $\perp$ to its corresponding state in O_1 with weight 0. There is a σ'-transition from each state in O_2 other than $\perp$ to S with weight 0, and a σ'-transition from $\perp$ to S with weight -1.

Observation G_0 It contains $5N$ states: $\{x_i, \overline{x}_i \mid 1 \leq i \leq 2N\} \cup \{y_i \mid N < i \leq 2N\}$. There is a σ-transition from each state in O_1 to its corresponding state in G_0 of weight $-N$ and in addition, σ-transitions from every state in O_1 to $\{y_i \mid N < i \leq 2N\}$ also of weight $-N$. For $N < j \leq 2N$ there is a τ_j^+ transition of weight 1 from y_j to x_j and a τ_j^- transition of weight 1 from y_j to $\overline{x}_j$. For all states in G_0 other than y_j there is a τ_j^+ and τ_j^- loop of weight 1. Figure 10 shows the construction.

Observations G_j **(for** $j > 0$**)** Observation gadgets for the logical gates follow the idea laid out earlier. The observation corresponding to gate j contains $4N+2j+8$ states: $\{x_i, \overline{x}_i \mid 1 \leq i \leq 2N+j\} \cup \{v_m, \overline{v}_m \mid 0 \leq$

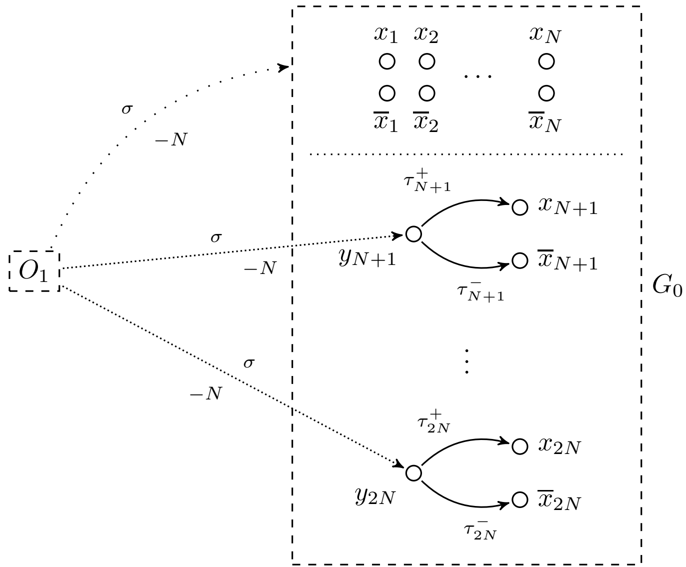

Figure 10: Gadget for G_0

$m \leq 3\}$. Recall gate j has inputs from $\{x_i, \overline{x}_i \mid 1 \leq i < 2N + j\}$. Suppose these inputs are $y_l \in \{x_l, \overline{x}_l\}$ and $y_r \in \{x_r, \overline{x}_r\}$, and for convenience let $\overline{y}_l$ and $\overline{y}_r$ denote the other member of the pair (*i.e.* the complement of the input). We have a σ-transition of weight -1 from $\{x_i, \overline{x}_i \mid 1 \leq i < 2N + j\} \subseteq G_{j-1}$ to the corresponding vertex in G_j. In addition we have σ-transitions of weight -1 from $y_l, \overline{y}_l, y_r, \overline{y}_r \in G_{j-1}$ to $v_0, \overline{v}_0, v_1, \overline{v}_1 \in G_j$ respectively. We have the following internal transitions:

- τ^+ (weight 0): v_1 to v_2, $\overline{v}_1$ to $\overline{v}_3$, v_0 to x_{2N+j}, $\overline{v}_0$ to $\overline{x}_{2N+j}$ if gate j is an AND gate, $\overline{v}_0$ to x_{2N+j} if it is an OR gate,
- τ^- (weight 0): v_1 to v_3, $\overline{v}_1$ to $\overline{v}_2$, $\overline{v}_0$ to $\overline{x}_{2N+j}$, v_0 to $\overline{x}_{2N+j}$ if gate j is an AND gate, v_0 to x_{2N+j} if it is an OR gate,
- χ (weight 1): v_2 to v_3, $\overline{v}_2$ to $\overline{v}_3$.

For all states in G_j we have $\tau^\pm, \chi$-loops with the same weights as above (*i.e.* χ loops have weight 1, $\tau^\pm$ loops have weight 0). Bit states (*i.e.* x_{N+1}, $\overline{x}_{2N}$) transition to the next observation on external actions.

Figure 11 shows an example of the construction of G_j for the gate $x_l \wedge \neg x_r$.

Observation Chk This last observation gadget contains $4N + 2$ states: $\{x_i, \overline{x}_i \mid 1 \leq i \leq 2N\} \cup \{y, z\}$. There is a σ-transition of weight -1 from $\{x_i, \overline{x}_i \mid 1 \leq i \leq 2N\} \subseteq G_k$ to their corresponding states in Chk, and a σ-transition of weight -1 from $x_{2N+k} \in G_k$ to y. There is a χ-transition of weight 1 from y to z and for all other states in Chk there is a χ-loop of weight 1. There is a σ-transition of weight 1 from all states in Chk to $\perp \in O_2$ and for $N < i \leq 2N$ there is a σ-transition of weight 1 from $x_i \in Chk$ to $x_{i-N} \in O_2$ and from $\overline{x}_i \in Chk$ to $\overline{x}_{i-N} \in O_2$.

Correctness of the construction We present a similar argument to that given for the proof of Theorem 8. Recall the game's initial transition is weighted -2^N. Further, note that internal transitions in all observations can only lead to reaching a good or bad sink or reaching the next observation gadget (while nullifying the incoming -1 weight). Hence, completing 2^N full simulations of the circuit is the only way of not forming a bad cycle and reaching observation O_2 with all concrete paths having weight 0. From there, a mixed cycle can be formed by going back to S. The latter thus holds if and only if the graph encoded succinctly by the given circuit has a Hamiltonian cycle. □

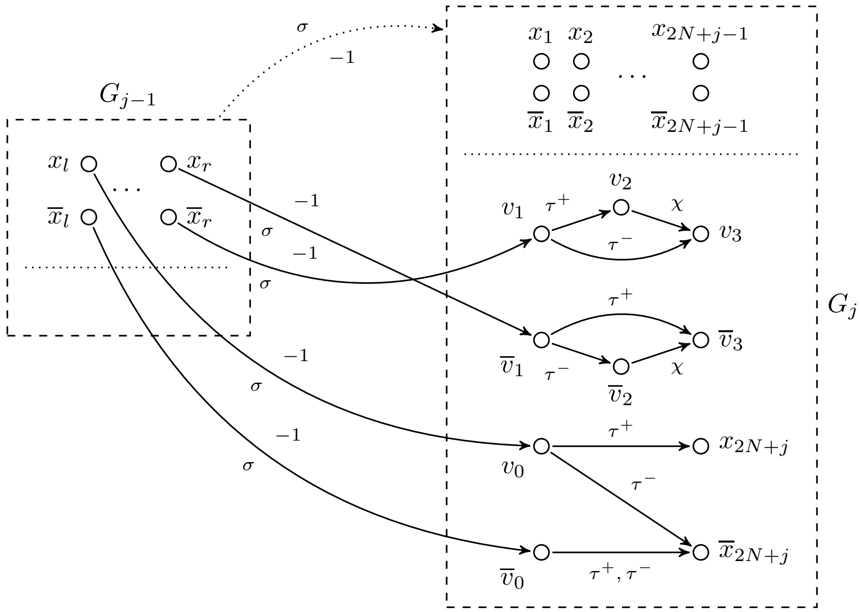

Figure 11: Gadget for gate $x_l \wedge \neg x_r$ (self-loops not shown)

Based on the construction used to prove the above result, we will now show hardness for forcibly FBC class membership.

Lemma 23. *Let G be an MPG with partial observation. Determining if G is forcibly FBC is* NEXPTIME-*hard.*

Proof. Suppose we make the following adjustments to the construction given in the proof of Lemma 22:

- Change the weights of incoming transitions to G_i ($i > 0$) to -5 and the weights of all internal τ-transitions to 1,
- Change the weight of the σ'-transition from $\bot \in O_2$ to S to 0,
- Add a new state $\bot$ to all observations other than S (and O_2),
- Add a σ-transition of weight 2^N from S to $\bot \in O_1$, and
- Whenever there is a transition of weight w from $x_i \in o$ to $x_j \in o'$ (o, o' and i, j possibly the same) add a transition of weight $-w$ from $\bot \in o$ to $\bot \in o'$.

Then the only possible non-mixed lasso in the resulting graph[5] is one that would correspond to a successful traversal of a Hamiltonian cycle. Eve can force the play to this cycle if and only if the succinct graph has a Hamiltonian cycle. ☐

Somewhat surprisingly, for the winner determination problem we have an EXPTIMEalgorithm matching the EXPTIME-hardness lower bound from games with visible weights. This is in contrast to the class membership problem in which an exponential increase in complexity occurs when moving from limited to partial observation.

Theorem 17 (Winner determination)**.** *Let G be a forcibly FBC MPG. Determining if Eve wins G is* EXPTIME-*complete.*

Proof. The lower bound follows from the fact that forcibly FBC games are a generalization of visible weights games (see Lemma 21), shown to be EXPTIME-complete in [9]. For the upper bound, rather than working on the reachability game Γ' associated to G', which is doubly-exponential in the size of G, we instead reduce

[5]We assume dead-ends go to a dummy state with a single mixed self-loop.

the problem of determining the winner to that of solving a safety game which is only exponential in the size of G.

Given an MPG with partial observation $G = (Q, q_I, \Sigma, \Delta, w, \mathsf{Obs})$, let $G' = (Q', q'_I, \Sigma, \Delta', w', \mathsf{Obs}')$ be its limited-observation version and Γ' be the finite reachability game, as defined in Section 7, constructed for G' (not for G!). Let $\mathcal{E} = [-1, 2W|\mathsf{Obs}'|] \cup \{\bot\}$ where $W = \max\{|w(e)| : e \in \Delta\}$, and let $\mathcal{B}' \subseteq \mathcal{B}$ be the set of functions $f : Q \to \mathcal{E}$.

The safety game will be played on $\mathcal{B}'$ with the transitions defined by σ-successors. The idea is that a given position $f \in \mathcal{B}'$ of the safety game corresponds to being in an observation of G', namely $\mathsf{supp}(f)$. The functions additionally keep track of the minimal weight of all concrete paths ending in states from $\mathsf{supp}(f)$. However, they do so only up to the point where a belief cycle is formed. Since W is the biggest absolute weight in G and in G', and the length of any simple belief path is bounded by $|\mathsf{Obs}'|$, it suffices to keep track of weights from $\mathcal{E}$.

Formally, the safety game is $\mathcal{S}_G = (\mathcal{B}', f'_I, \Sigma, \Delta_{succ}, \mathcal{F}'_{neg})$ where $f'_I(q_I) = W|\mathsf{Obs}'|$ and $f'_I(q) = \bot$ for all other $q \in Q$; $(f, \sigma, f') \in \Delta_{succ}$ if f' is a proper σ-successor of f where we let

$$a + b = \begin{cases} \bot & \text{if } a = \bot \text{ or } b = \bot, \\ -1 & \text{if } a = -1,\ b = -1, \text{ or } a + b < 0, \text{ and} \\ \min\{a + b, 2W|Q|\} & \text{otherwise.} \end{cases}$$

$\mathcal{F}'_{neg}$ is the set of all functions $f \in \mathcal{F}'$ such that $f(q) = -1$ for some $q \in \mathsf{supp}(f)$. The game is played similar to the reachability game Γ, *i.e.* Eve chooses an action σ and Adam resolves non-determinism by selecting a proper σ-successor. In this case, however, Eve's goal is to avoid visiting any function in $\mathcal{F}'_{neg}$.

In this safety game (just like in the weighted unfolding) the non-negative integer values of f give a lower bound for the minimum weights of the concrete paths ending in the given state (see Lemma 11). More formally, since obtaining a -1 weight means that henceforth the weight stays -1, we have that if $f(q) \neq \bot$ and $f(q) \geq 0$ then the minimum weight over all concrete paths starting at q_I and ending at q is at least $f(q) + W|\mathsf{Obs}'|$. We do not have equality because of the max applied after each sum. If $f(q) = -1$ then there is a concrete path of weight at most $-W|\mathsf{Obs}'| - 1$, because $f_I(q_I) = W|\mathsf{Obs}'|$. As the winner of a forcibly FAC game can be resolved in at most $|\mathsf{Obs}'|$ transitions it turns out that this is sufficient information to determine the winner.

The above observation that non-negative values give lower bounds for concrete paths ending at the given state implies that if Eve has a strategy to always avoid $\mathcal{F}'_{neg}$ then $\liminf_{n\to\infty} \frac{\pi[..n]}{n} \geq 0$ for all concrete paths π consistent with the play. That is, if Eve has a winning strategy in $\mathcal{S}_G$ then she has a winning strategy in G.

Now suppose Eve has a winning strategy in G. It follows from the determinacy of forcibly FAC games and Theorem 4 that she has a winning strategy λ in Γ'. Let λ^* be the translation of λ to G' as per Theorem 4, and let M denote the set of memory states required for λ^*. Clearly λ^* induces a strategy in $\mathcal{S}_G$. We claim this induced strategy is winning in $\mathcal{S}_G$. Let $\varrho = f_0\sigma_0 \ldots$ be any play in $\mathcal{S}_G$ consistent with λ^*, and let $\mu_i = g_i^{(0)} \ldots g_i^{(n_i)}$ denote the i-th memory state obtained in the generation of ϱ (as in Lemma 16). Then, with a slight adjustment to the proof of Lemma 16 to account for function values not exceeding $2W|\mathsf{Obs}'|$ we have for all i and all q:

$$\begin{aligned} f_i(q) - W|\mathsf{Obs}'| &\geq g_i^{(n_i)}(q) \\ &= \min\{w(\pi) \mid \pi \in \gamma(\mathsf{supp}(\mu_i)) \text{ and } \pi \text{ ends at } q\}^6 \\ &\geq -W|\mathsf{Obs}'| \end{aligned}$$

because $|\mu_i| \leq |\mathsf{Obs}'|$ from the definition of Γ'. Thus $f_i(q) \geq 0$ for all i. Hence ϱ does not reach $\mathcal{F}'_{neg}$ and is winning for Eve. Thus λ^* is a winning strategy for Eve.

[6] The step follows from Lemma 11.

So to determine the winner of G, it suffices to determine the winner of $\mathcal{S}_G$. This is just the complement of alternating reachability, known to be decidable in polynomial time (see *e.g.* [18]). As

$$|\mathcal{S}_G| = O(|\mathcal{F}'|^2) = O\left((2W|\mathsf{Obs}'| + 1)^{|Q|}\right) = 2^{O(|Q|^2)},$$

determining the winner of $\mathcal{S}_G$, and hence G, is in EXPTIME. □

Corollary 24. *Let G be an FBC MPG. Determining if Eve wins G is* EXPTIME*-complete.*

References

[1] B. Aminof and S. Rubin. First cycle games. In F. Mogavero, A. Murano, and M. Y. Vardi, editors, *SR*, volume 146 of *EPTCS*, pages 91–96, 2014.

[2] D. Berwanger, K. Chatterjee, L. Doyen, T. A. Henzinger, and S. Raje. Strategy construction for parity games with imperfect information. In *CONCUR*, pages 325–339, 2008.

[3] D. Berwanger and L. Doyen. On the power of imperfect information. In *FSTTCS*, pages 73–82, 2008.

[4] H. Björklund, S. Sandberg, and S. Vorobyov. Memoryless determinacy of parity and mean payoff games: a simple proof. *TCS*, 310(1):365–378, 2004.

[5] K. Chatterjee and L. Doyen. The complexity of partial-observation parity games. In *LPAR*, pages 1–14. Springer, 2010.

[6] K. Chatterjee and L. Doyen. Partial-observation stochastic games: How to win when belief fails. In *LICS*, pages 175–184. IEEE, 2012.

[7] K. Chatterjee, L. Doyen, and T. A. Henzinger. Quantitative languages. In *CSL*, pages 385–400, 2008.

[8] K. Chatterjee, L. Doyen, T. A. Henzinger, and J.-F. Raskin. Generalized mean-payoff and energy games. In *FSTTCS*, pages 505–516, 2010.

[9] A. Degorre, L. Doyen, R. Gentilini, J.-F. Raskin, and S. Toruńczyk. Energy and mean-payoff games with imperfect information. In *CSL*, pages 260–274, 2010.

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

[11] P. Hunter, G. A. Pérez, and J.-F. Raskin. Mean-payoff games with partial-observation - (extended abstract). In *RP*, pages 163–175, 2014.

[12] M. Jurdziński. Deciding the winner in parity games is in UP ∩ coUP. *IPL*, 68(3):119–124, 1998.

[13] A. S. Kechris. *Classical descriptive set theory*, volume 156. Springer, 1995.

[14] O. Kupferman and M. Y. Vardi. Synthesis with incomplete informatio. *Advances in Temporal Logic*, 16:109–127, 2000.

[15] D. A. Martin. Borel determinacy. *The annals of Mathematics*, 102(2):363–371, 1975.

[16] D. A. Martin and J. R. Steel. Projective determinacy. *Proceedings of the National Academy of Sciences of the United States of America*, 85(18):6582, 1988.

[17] M. L. Minsky. *Computation: Finite and Infinite Machines*. Prentice-Hall, 1967.

[18] C. H. Papadimitriou. *Computational complexity*. John Wiley and Sons Ltd., 2003.

[19] C. H. Papadimitriou and M. Yannakakis. A note on succinct representations of graphs. *Information and Control*, 71(3):181 – 185, 1986.

[20] D. Perrin and J.-E. Pin. *Infinite words: automata, semigroups, logic and games*, volume 141. Academic Press, 2004.

[21] J. H. Reif. The complexity of two-player games of incomplete information. *Journal of Computer and System Sciences*, 29(2):274–301, 1984.

[22] L. J. Stockmeyer and A. R. Meyer. Word problems requiring exponential time. In *STOC*, pages 1–9. ACM, 1973.

[23] U. Zwick and M. Paterson. The complexity of mean payoff games on graphs. *TCS*, 158(1):343–359, 1996.