The Boundedness Problem for Monadic Universal First-Order Logic

Martin Otto
Fachbereich Mathematik
Technische Universität Darmstadt
64289 Darmstadt, Germany

Abstract

We consider the monadic boundedness problem for least fixed points over FO *formulae as a decision problem: Given a formula $\varphi(X, x)$, positive in X, decide whether there is a uniform finite bound on the least fixed point recursion based on φ. Few fragments of* FO *are known to have a decidable boundedness problem; boundedness is known to be undecidable for many fragments. We here show that monadic boundedness is decidable for purely universal* FO *formulae without equality in which each non-recursive predicate occurs in just one polarity (e.g., only negatively). The restrictions are shown to be essential: waving either the polarity constraint or allowing positive occurrences of equality, the monadic boundedness problem for universal formulae becomes undecidable. The main result is based on a model theoretic analysis involving ideas from modal and guarded logics and a reduction to the monadic second-order theory of trees.*

1 Introduction

Consider a formula $\varphi(X, x) \in$ FO that is positive in the monadic second-order variable X and with a single free first-order variable x. Over corresponding structures $\mathfrak{A}$ such φ defines an operation on subsets, $F^{\mathfrak{A}}_{\varphi}: \mathcal{P}(A) \to \mathcal{P}(A)$, through $P \mapsto \{a \in A : \mathfrak{A} \models \varphi[P, a]\}$. Due to positivity in X, this operation is monotone w.r.t. $\subseteq$, and hence has a least fixed point $(\mu_X \varphi)^{\mathfrak{A}}$, which is also attained as the limit of the monotone sequence of its stages $X^{\alpha}[\mathfrak{A}]$ generated by iteration of $F^{\mathfrak{A}}_{\varphi}$ on $\emptyset$.

When is this recursion uniformly finitely bounded, and hence the least fixed point uniformly definable by a finite iteration of φ? We say that $\varphi(X, x)$ is bounded iff there is $n \in \mathbb{N}$ such that $X^{n+1}[\mathfrak{A}] = X^n[\mathfrak{A}] (= (\mu_X \varphi)^{\mathfrak{A}})$ for all $\mathfrak{A}$. We consider the *boundedness problem* as a decision problem $\mathsf{Bdd}(\mathcal{F})$, for formulae φ from a fixed syntactic fragment $\mathcal{F} \subseteq$ FO.

$\mathsf{Bdd}(\mathcal{F})$ $\boxed{\begin{array}{l}\text{given } \varphi = \varphi(X, x) \in \mathcal{F},\\ \text{decide if } \varphi \text{ is bounded}\end{array}}$

This decision problem has been considered in particular in connection with database query optimisation, with a view to eliminate bounded (i.e., spurious) recursion, e.g., in Datalog programs [1, 13, 9, 11, 12]. Similar concerns can be considered in connection with temporal fixed point logics like the μ-calculus. Because of the model theoretic link between boundedness and FO definability of the least fixed point, there is also a connection with circumscription, which is of interest in artificial intelligence [17, 16].

However, very few fragments of FO are known to have a decidable boundedness problem. In fact, boundedness is known to be undecidable for many fragments. Among the scarce decidability results, the following result of Cosmadakis, Gaifman, Kanellakis, Vardi [9] stands out.

Theorem 1. *Boundedness is decidable for monadic Datalog. I.e.,* Bdd *is decidable for positive existential* FO *formulae in monadic recursion variables X, as well as for systems of such.*

Another decidability result that is important for our considerations is the following from [19].

Theorem 2. $\mathsf{Bdd}(\mathrm{ML})$*, the boundedness problem for the modal fragment of* FO *is decidable. In fact, one can decide for formulae of the μ-calculus whether they are equivalent to plain modal formulae.*

Undecidability results for Bdd abound; we mention some. See [12] for (i), [11] for (ii), and [14] for (iii). For (iv) and (v) compare section 3.2 below.

Theorem 3. $\mathsf{Bdd}(\mathcal{F})$ *is undecidable for the following fragments $\mathcal{F}$ of* FO*:*

(i) *$\mathcal{F}$: positive existential formulae in a binary recursive predicate X.*
(ii) *$\mathcal{F}$: existential formulae in a monadic recursive predicate if negation (or even just inequality) is allowed.*

(iii) $\mathcal{F}$: *formulae in a monadic recursive predicate using just two distinct first-order variables.*

(iv) $\mathcal{F}$: *universal formulae in a monadic recursive predicate, negative in all non-recursive predicates but allowing positive occurrences of equality.*

(v) $\mathcal{F}$: *universal* FO *formulae in a monadic recursive predicate, if some of the non-recursive relations can occur both positively and negatively.*

In a way, therefore, few natural fragments remain that would be candidates for a decidable boundedness problem. Among the most obvious are: the guarded fragment (still open) and the universal fragment of FO (settled here). The present investigation shows that monadic boundedness is decidable for purely universal FO formulae without (positive occurrences of) equality in which each non-recursive predicate occurs in just one polarity. As the boundedness issue is obviously invariant under swapping relations for their complements in the underlying relational structures, we may w.l.o.g. restrict attention to the case where all non-recursive predicates (and equality) only appear negatively, denoted by $\mathcal{F}(\forall, -)$ in the following.

Main Theorem $\mathsf{Bdd}(\mathcal{F}(\forall, -))$ *is decidable. I.e., boundedness is decidable for monadic recursion over purely universal* FO *formulae that are purely negative in all non-recursive predicates and equality (or, by extension, universal formulae in which each predicate letter occurs in just one polarity).*

The restrictions expressed in the theorem are essential. The decidability result obtained provides a counterpart to the one expressed in Theorem 1. Note, however, that the logical "dual" of the decidability result in Theorem 1 would concern boundedness for *greatest* fixed points of universal formulae, rather than least fixed points. Conversely, our main theorem proves, by duality, decidability of the boundedness problem for greatest fixed points of purely positive, purely existential formulae.

Organisation. In the following section, we review the basic definitions and discuss the model theoretic connection between boundedness and definability of the fixed point (Barwise–Moschovakis theorem). Section 3 focuses on boundedness for the universal fragment of FO and introduces a normal form for universal fixed points; we also contrast our decidability result with undecidability proofs for less restricted classes of universal formulae. In section 4 we prepare the reduction argument underlying our main result by collecting the relevant facts about decidability of boundedness for systems in the modal fragment; we also give a new model theoretic argument for decidability there. The main technical contribution in section 5 provides the reduction from the restricted universal case to the modal case (based on an intuitive intermediate stage that is motivated by a formalisation in the guarded fragment). We conclude with some remarks on ramifications and open problems in section 6.

2 Preliminaries

2.1 Boundedness

Consider fixed point recursion in systems for simultaneous least fixed points. (Even when we are ultimately interested in boundedness for single formulae rather than systems, the reductions applied take us to systems.)

An *admissible system* is given by a tuple of FO formulae $\bar{\varphi} = (\varphi_1, \ldots, \varphi_k)$ in a relational vocabulary τ and monadic second-order variables $\bar{X} = (X_1, \ldots, X_k)$, where each φ_i is of the form

$$\varphi_i(\bar{X}, x) \in \text{FO}, \quad \text{positive in each } X_j.$$

Over any τ-structure $\mathfrak{A}$ associate with this system the monotone operator on $(\mathcal{P}(A))^k$

$$\bar{P} = (P_i)_{1 \leqslant i \leqslant k} \longmapsto \big(\{a \in A \colon \mathfrak{A} \models \varphi_i[\bar{P}, a]\}\big)_{1 \leqslant i \leqslant k}.$$

The simultaneous least fixed point of this operator is denoted $\bar{\varphi}^\infty[\mathfrak{A}]$; its components $(\bar{\varphi}^\infty[\mathfrak{A}])_i$. The least fixed point $\bar{\varphi}^\infty$ is also the limit $\bar{X}^\infty$ of the monotone sequence of stages $\bar{X}^\alpha[\mathfrak{A}] = \bar{X}^\alpha = (X_1^\alpha, \ldots, X_k^\alpha)$ inductively defined over the given $\mathfrak{A}$ according to

$$\begin{array}{rcl} X_i^0 & = & \emptyset \\ X_i^{\alpha+1} & = & \{a \in A \colon \mathfrak{A} \models \varphi_i[\bar{P}, a]\} \\ X_i^\lambda & = & \bigcup_{\alpha<\lambda} X_i^\alpha \text{ for limit } \lambda. \end{array}$$

The *closure ordinal* $\gamma[\bar{\varphi}, \mathfrak{A}]$ of $\bar{\varphi}$ in $\mathfrak{A}$ is the least α such that $\bar{X}^\alpha[\mathfrak{A}] = \bar{\varphi}^\infty[\mathfrak{A}]$.

Observation 4. *Over ω-saturated $\mathfrak{A}$, the closure ordinal for any first-order least fixed point is bounded by ω.*

Note that all finite stages X_i^n, for $n \in \mathbb{N}$, are uniformly FO definable. We write φ^n for the formula that defines the n-th stage of φ, which is obtained inductively by substituting φ^{n-1} for X in $\varphi(X, x)$.

Definition 5. The admissible system $\bar{\varphi} = (\varphi_1, \ldots, \varphi_k)$ is *bounded* if for some $n \in \mathbb{N}$, $\gamma[\bar{\varphi}, \mathfrak{A}] \leqslant n$ for all $\mathfrak{A}$. This notion naturally relativise to restricted classes $\mathcal{C}$ of structures $\mathfrak{A}$; we speak of boundedness of $\bar{\varphi}$ over $\mathcal{C}$ in this sense.

Unboundedness as generalised satisfiability. It is useful to think of unboundedness as a satisfiability issue in the following sense. $\varphi(X, x)$ is unbounded iff $\varphi^{n+1}(x) \wedge \neg\varphi^n(x)$ is satisfiable for all $n \in N$. It is not hard to see that, for fragments satisfying some mild closure properties, $\mathrm{SAT}(\mathcal{F})$ is reducible to $\mathsf{Bdd}(\mathcal{F})$.

Boundedness in finite model theory. For many natural fragments of FO, boundedness (over all structures) coincides with boundedness over the class of all finite structures. This is easily seen for fragments that are closed under positive substitution, negation and conjunction and that have the finite model property. Closure under positive substitution implies that the finite stages X^n in the fixed point generation are themselves definable by formulae φ^n in the given fragment $\mathcal{F}$; closure under conjunction and negation provide the formulae $\varphi^{n+1} \wedge \neg\varphi^n$ within $\mathcal{F}$. For such fragments unboundedness (in infinite structures) implies unboundedness over finite structures, by an appeal to the finite model property. This applies in particular to the modal, guarded and two-variable fragments.

For the existential and universal fragments, one can argue similarly, but invoking the finite model property for the $\exists^*\forall^*$ fragment of FO (the Bernays–Schönfinkel class of the classical decision problem, see [8]) for $\varphi^{n+1} \wedge \neg\varphi^n$.

2.2 The Barwise–Moschovakis Theorem

The following is an adaptation of a well-known classical theorem [5], which links boundedness to FO definability of the fixed point. The non-obvious part, (iii) $\Rightarrow$ (i), is based on a compactness argument using ω-saturated structures $\mathfrak{A}$, applying Observation 4.

Theorem 6 (Barwise-Moschovakis). *For admissible $\bar{\varphi}$, the following are equivalent:*

(i) *$\bar{\varphi}$ is bounded.*
(ii) *$(\bar{\varphi}^\infty)$ is uniformly* FO *definable over all $\mathfrak{A}$.*
(iii) *$(\bar{\varphi}^\infty[\mathfrak{A}])$ is* FO *definable for each $\mathfrak{A}$.*

The statement of the theorem relativises to elementary classes. It also relativises to fragments $\mathcal{F} \subseteq$ FO with the property that the finite stages of admissible systems over $\mathcal{F}$ are themselves $\mathcal{F}$-definable. Natural fragments of this kind are, for instance, the universal or the existential fragments of FO, the k-variable fragments FO^k, the guarded fragment GF and modal logic ML. For the latter, for instance, we have that an admissible system $\bar{\varphi}$ of modal formulae is bounded if, and only if, its least fixed point is definable in modal logic (over each individual Kripke structure).

Status in finite model theory. As discussed in the previous section, the boundedness problem coincides with its finite model theory version for several 'good' fragments, including the universal fragment. For the finite model theory (fmt) version of the Barwise–Moschovakis theorem, however, one also needs to link definability of the fixed point over all finite structures to boundedness over all finite structures, and hence (for a 'good' fragment) to boundedness and therefore definability of the fixed point over all structures. For the existential positive fragment, a strong form of the finite model theory version of the Barwise–Moschovakis theorem was proved by Ajtai and Gurevich: a Datalog query is FO definable if, and only if, the program is bounded. This result has recently been put in a new context, by major new results on the fmt version of the classical preservation theorem that links preservation under homomorphisms to positive existential definability, by Rossmann [21] and Atserias, Dawar and Kolaitis [4]. These fmt versions show that FO definability of the fixed point of a positive existential recursion implies its positive existential definability, due to preservation of the fixed point under homomorphisms in finite structures. By duality, Rossmann's theorem similarly implies that FO definability for a universal equality-free recursion that is negative in all non-recursive predicates implies its universal FO definability, in all finite structures.

For the existential fragment, one can show directly that existential FO definability of the fixed point over all finite structures implies its definability (by the same existential formula) over all structures, again by use of the finite model property of the $\exists^*\forall^*$ fragment. It is unclear, though, whether also for the universal fragment, universal definability of the fixed point in all finite structures implies its definability over all structures.

3 Universal FO

The formulae of the universal fragment of FO are generated from atomic and negated atomic formulae by conjunction, disjunction and universal quantification.

3.1 Moschovakis normal form

The following normal form for universal formulae $\varphi(X, x)$ that are positive in the second-order variable X is an adaptation of the normal form of [18], Chapter 4B.

Proposition 7 (Moschovakis).
Let $\varphi(X, x)$ positive in X and purely universal. Then there is a formula $\tilde{\varphi}(X, x)$ of the form

$$\tilde{\varphi}(X, x) = \forall \mathbf{y}\big(\theta(x, \mathbf{y}) \rightarrow \delta(X, \mathbf{y})\big),$$

where θ is quantifier free and does not contain X, and where δ is a disjunction of atoms Xy_j for some $y_j \in \mathbf{y}$ such that

$$\neg\forall x Xx \models \forall x\big(\varphi(X, x) \leftrightarrow \tilde{\varphi}(X, x)\big).$$

If φ is purely negative in all non-recursive predicates and in $=$, then θ can be chosen purely positive (in all predicates and in $=$).

Proof. We sketch the key steps in the inductive proof.

For $\varphi = Xz$, let $\tilde{\varphi} := \forall y(y = z \to Xy)$.

For $\varphi = R\mathbf{z}$, let $\tilde{\varphi} := \forall y\big(\neg R\mathbf{z} \to Xy\big)$. Here equivalence with φ relies on the existence of elements outside X.

The universal quantification step is trivial. The interesting inductive steps are for $\wedge$ and $\vee$:
Consider $\varphi = \varphi_1(X, \mathbf{z}) * \varphi_2(X, \mathbf{z})$, where $* \in \{\wedge, \vee\}$. Assume that $\tilde{\varphi}_i = \forall \mathbf{y}^{(i)}\big(\theta_i(\mathbf{x}, \mathbf{y}^{(i)}) \to \delta_i(X, \mathbf{y}^{(i)})\big)$ is as desired for φ_i, w.l.o.g. with disjoint $\mathbf{y}^{(i)}$. Let $\tilde{\varphi}$ be the following, where $\bar{*}$ is the dual of $*$:

$$\forall \mathbf{y}^{(1)} \forall \mathbf{y}^{(2)} \left[\begin{array}{l} \big((\theta_1(\mathbf{z}, \mathbf{y}^{(1)}) \mathbin{\bar{*}} \theta_2(\mathbf{z}, \mathbf{y}^{(2)})\big) \\ \to \big(\delta_1(X, \mathbf{y}^{(1)}) \vee \delta_2(X, \mathbf{y}^{(2)})\big) \end{array} \right].$$

In both cases ($* = \wedge$ or $* = \vee$) it is clear that $\varphi(X, \mathbf{z}) \models \tilde{\varphi}(X, \mathbf{z})$. It remains to establish, that, whenever X is not the full universe, then $\tilde{\varphi}$ implies φ.

Consider first the case of $* = \vee$. We look at a model of ψ, and suppose for instance that $\forall \mathbf{y}^{(1)}\big[\theta_1(\mathbf{y}^{(1)}) \to \delta_1(X, \mathbf{y}^{(1)})\big]$ does not hold true. We therefore find an interpretation for $\mathbf{y}^{(1)}$ such that $\theta_1 \wedge \neg\delta_1$; fixing any such interpretation for $\mathbf{y}^{(1)}$, ψ becomes equivalent to $\forall \mathbf{y}^{(2)}\big[\theta_2(\mathbf{y}^{(2)}) \to \delta_2(X, \mathbf{y}^{(2)})\big]$, as it should.

Consider now the case of $* = \wedge$. Again, we look at a model of ψ, in which there are elements outside X. We need to show that, for instance, this implies that $\theta_1(\mathbf{y}^{(1)}) \to \delta_1(X, \mathbf{y}^{(1)})$. Instantiate the universally quantified $\mathbf{y}^{(2)}$ in ψ by a tuple consisting of components not in X; this makes δ_2 false. Then ψ implies that $\theta_1(\mathbf{y}^{(1)}) \to \delta_1(X, \mathbf{y}^{(1)})$, as desired. ☐

The normal form given in [18] goes further in also eliminating the disjunctions in δ in favour of existential quantification in the prefix; this is not desirable here.

3.2 Limits for decidability

The following is in marked contrast with our main decidability result.

Theorem 8. Bdd *is undecidable for monadic recursion w.r.t. purely universal formulae in which all non-recursive predicates occur only negatively, but equality may occur positively.*

The proof follows a general pattern established in [14] and in [15] for several other undecidable cases of the monadic boundedness problem, via reduction of the tiling problem [7]. A domino system is a finite structure $\mathfrak{D} = (D, R_H, R_V)$ consisting of tile types $d \in D$ with two adjacency relations R_H and R_V that specify when two tile types fit in horizontally or vertically adjacent positions in tiling.

Formally a $\mathfrak{D}$-tiling of $n \times n$ is a homomorphism from the $n \times n$ grid structure $\mathfrak{G}_n$ to $\mathfrak{D}$. Here

$$\begin{array}{rcl} \mathfrak{G}_n & = & (\{0, \ldots, n-1\}^2, H, V) \\ \text{with } H & = & \{((i,j),(i+1,j)) \colon i+1, j < n\}, \\ V & = & \{((i,j),(i,j+1)) \colon i, j+1 < n\}. \end{array}$$

Intuitively, a $\mathfrak{D}$-tiling of $n \times n$ corresponds to a placement of tile types on the nodes of the $n \times n$ grid such that horizontally and vertically adjacent tiles respect the adjacency constraints given by R_H and R_V, respectively. We also refer to a corresponding expansion of $\mathfrak{G}_n$ by colours $(P_d)_{d \in D}$ as a $\mathfrak{D}$-tiling of $n \times n$.

Theorem 9 (Berger). *The following problem is undecidable: Given a tiling system $\mathfrak{D}$, decide whether $\mathfrak{D}$ admits tilings of size n for arbitrarily large $n \in \mathbb{N}$.*

We now want to associate with a given $\mathfrak{D}$ a formula $\varphi_{\mathfrak{D}} = \varphi(X, x)$, positive in X, purely universal and with only negative occurrences of the relations H, V (binary) and P_d (unary) for $d \in D$, such that $\varphi_{\mathfrak{D}}$ is unbounded iff $\mathfrak{D}$ admits arbitrarily large tilings. Consider firstly an auxiliary sentence φ_0 which is the conjunction of the following:

$$\forall y \textstyle\bigwedge_{d \neq d'} \neg(P_d y \wedge P_{d'} y)$$

$$\forall y_1 \forall y_2 \neg\big(H y_1 y_2 \wedge \textstyle\bigvee_{(d,d') \notin R_H} (P_d y_1 \wedge P_{d'} y_2)\big)$$

$$\forall y_1 \exists y_2 \neg\big(V y_1 y_2 \wedge \textstyle\bigvee_{(d,d') \notin R_V} (P_d y_1 \wedge P_{d'} y_2)\big)$$

$$\forall y y_1 y_2 \big((H y y_1 \wedge H y y_2) \to y_1 = y_2\big)$$

$$\forall y y_1 y_2 \big((V y y_1 \wedge V y y_2) \to y_1 = y_2\big)$$

$$\forall y \forall y_1 \forall y_2 \forall z_1 \forall z_2 \left[\begin{array}{r} (H y y_1 \wedge V y_1 z_1 \wedge V y y_2 \wedge H y_2 z_2) \\ \to z_1 = z_2 \end{array} \right]$$

$$\begin{tikzcd} a_{01} \arrow[r, "H"'] & a_{11} \\ a \arrow[u, "V"] \arrow[r, "H"'] & a_{10} \arrow[u, "V"'] \end{tikzcd}$$

$\varphi = \varphi_{\mathfrak{D}}$ will be of the form $\varphi(X, x) = \varphi_0 \wedge \varphi_1(X, x)$ with static part φ_0 as given above. Clearly φ is unbounded iff φ_1 is unbounded in models of φ_0. Models of φ_0 are H-V-graphs in which H and V are the graphs of partial functions that commute where defined, and in which each node carries at most one colour P_d in such a way that the colours of adjacent nodes respect the tiling constraints.

φ_1 is set up so as to ensure that (in models of φ_0) the recursion on $\varphi_1(X, x)$ has closure ordinal n on $\mathfrak{D}$-tilings of $\mathfrak{G}_n$, and that any structure on which φ_1 has a non-trivial $(n+1)$-st stage, embeds $\mathfrak{G}_n$ and hence establishes the existence of a $\mathfrak{D}$-tiling of $n \times n$.

The formalisation of φ_1 uses the following abbreviations (corresponding to positive quantifier-free definitions):

$$\begin{array}{rcl} \dot{H} & := & \{(x,y)\colon Hxy \wedge \bigvee_{d\in D} P_d y\}, \\ \dot{V} & := & \{(x,y)\colon Vxy \wedge \bigvee_{d\in D} P_d y\}. \end{array}$$

$$\begin{array}{rl} \text{Put } \varphi_1(X,x) := & \forall y(\dot{H}xy \to Xy) \\ \vee & \forall y(\dot{V}xy \to Xy) \\ \vee & \forall y \forall z\big((\dot{H}xy \wedge \dot{V}yz) \to Xz\big) \\ \vee & \forall y \forall z\big((\dot{V}xy \wedge \dot{H}yz) \to Xz\big). \end{array}$$

Obviously $\varphi_{\mathfrak{D}} := \varphi_0 \wedge \varphi_1(X,x)$ is purely negative in H, V and the P_d, but uses equality positively (in φ_0).

It is obvious that $\gamma[\varphi_{\mathfrak{D}}, \mathfrak{A}] = n$ for a $\mathfrak{D}$-tiling $\mathfrak{A} = (\mathfrak{G}_n, (P_d))$ of $n \times n$. It remains to show that conversely, unboundedness of $\varphi_{\mathfrak{D}}$ implies that $\mathfrak{D}$ has tilings of arbitrarily large $n \times n$ grids.

Lemma 10. *Let $\mathfrak{A} \models \varphi_0$, $X^i[\mathfrak{A}]$ the stages of $\varphi_{\mathfrak{D}}$ on $\mathfrak{A}$. Let $\dot{\mathfrak{A}}$ be $\mathfrak{A}$ with H and V replaced by $\dot{H}$ and $\dot{V}$, respectively. Let $a \notin X^n[\mathfrak{A}]$. Then there is a homomorphism $h\colon \mathfrak{G}_{n+1} \to \dot{\mathfrak{A}}$ with $h(0,0) = a$. It follows that there is a $\mathfrak{D}$-tiling of $\mathfrak{G}_n$ if $\gamma[\varphi_{\mathfrak{D}}, \mathfrak{A}] > n$.*

Proof. The main claim is proved by induction. Note that $a \notin X^n[\mathfrak{A}]$ implies that a has $\dot{H}$-, $\dot{V}$- and $(\dot{V} \circ \dot{H})$- successors a_{10}, a_{01} and a_{11} that are not in $X^{n-1}[\mathfrak{A}]$. By the inductive hypothesis there are homomorphisms $h_i\colon \mathfrak{G}_n \to \dot{\mathfrak{A}}$ with $h_1(0,0) = a_{10}$, $h_2(0,0) = a_{01}$ and $h_3(0,0) = a_{11}$.

φ_0 forces $h_1(0,1)$ (the V-successor of a_{10}) to be the same as $h_2(1,0)$ (the H-successor of a_{01}) and $h_3(0,0)$. Similarly, $h_1(i,j+1) = h_2(i+1,j) = h_3(i,j)$ for all i,j where defined. But this implies that these three homomorphisms can be combined consistently to obtain the desired homomorphism h by putting $h(0,0) := a$, $h(i+1,j) := h_1(i,j)$, $h(i,j+1) := h_2(i,j)$ and $h(i+1,j+1) := h_3(i,j)$ (whichever righthand sides are defined). □

Instead of equality one can introduce an extra binary relation $\approx$ which, by purely universal axioms, can be forced to be a congruence w.r.t. H, V and the P_d. However, these axioms will necessarily involve $\approx$ and the predicates H, V, P_d in both polarities. In other words, we may trade the restriction on polarities for avoidance of equality.

Theorem 11. Bdd *is undecidable for monadic recursion w.r.t. purely universal formulae without equality.*

4 Boundedness in modal logic

We regard modal logic as a fragment $\mathrm{ML} \subseteq \mathrm{FO}$ in the usual way. In a vocabulary σ consisting of binary relations E_i and unary predicates P_j, the formulae of $\mathrm{ML}[\sigma]$ are generated as the closure of the atomic formulae P_j under booleans and modal quantification $\langle E_i \rangle$ and $[E_i]$. In FO terms, $(\langle E_i \rangle \varphi)(x)$ is $\exists y\big(E_i xy \wedge \varphi(y)\big)$, and dually $([E_i]\varphi)(x)$ is $\forall y\big(E_i xy \to \varphi(y)\big)$.

We consider admissible systems in a tuple of monadic recursive predicates $\bar{X}$. The modal version of the Barwise–Moschovakis theorem provides the means to decide boundedness, ultimately via reduction to the MSO theory of the ω-branching tree, which is decidable by Rabin's well known theorem. We outline the chain of model theoretic arguments underlying this reduction.

Due to bisimulation invariance and the Löwenheim–Skolem theorem, it suffices to consider the behaviour of systems in ML over countable tree structures.

Let σ be a relational type consisting of binary predicates E_i (for labelled edges) and unary predicates P_j (for colouring or labelling vertices). A *tree structure* of type σ is a rooted directed tree w.r.t. $E := \bigcup E_i$ with disjoint E_i (no multiple edge labels).

For a tree structure $\mathfrak{A}, \lambda$ with root λ and $\ell \geqslant 1$ we let $A{\restriction}\ell \subseteq A$ denote the initial segment consisting of all nodes up to depth ℓ. Some related terminology:

(a) $A' \subseteq A$ is called *initial* in $\mathfrak{A}, \lambda$ if it contains the root λ and is connected.

(b) $A' \subseteq A$ is called *bounded* if $A' \subseteq A{\restriction}\ell$ for some $\ell \in \mathbb{N}$.

(c) $A' \subseteq A$ is called *path-finite* if it contains no infinite paths.

Note that, while properties (a) and (c) are MSO definable, property (b) is not.

Definition 12. Call a formula $\psi(x) \in \mathrm{MSO}$ *tree-local* if there is some $\ell \in \mathbb{N}$ such that for all countable tree structures $\mathfrak{A}$ and for all initial subsets $A' \supseteq A{\restriction}\ell$:

$$\mathfrak{A}, \lambda \models \psi \quad \Leftrightarrow \quad \mathfrak{A}{\restriction}A', \lambda \models \psi.$$

The following is a straightforward consequence of the combination of the modal variant of Theorem 6 and the expressive completeness of ML for bisimulation invariant properties that are ℓ-local for some ℓ [6, 20].

Lemma 13. *The following are equivalent for admissible systems $\bar{\varphi} \in \mathrm{ML}$:*

(i) *$\bar{\varphi}$ bounded.*

(ii) *$\bar{\varphi}^\infty$ tree-local.*

(iii) *$\bar{\varphi}^\infty$ ML-definable.*

Theorem 14. Bdd(ML) *is decidable via reduction to the* MSO *theory of the ω-branching tree $\mathfrak{T}_\omega$. Similarly for boundedness of modal systems over* MSO *definable classes of trees.*

The proof of correctness for the intended reduction involves a regular version of König's lemma that allows us to "apply König's lemma" in the context of the ω-branching tree, its infinite branching notwithstanding.

A *regular* expansion of $\mathfrak{T}_\omega$ is one that realises only finitely many isomorphism types of subtrees.

Lemma 15. *The following are equivalent for initial subsets D in regular expansions $(\mathfrak{T}_\omega, D)$:*

(i) *D is path-finite.*

(ii) *D is bounded.*

Proof. For (i) $\Rightarrow$ (ii) one shows that one can inductively choose an infinite path within an unbounded D. Starting from the root, select a successor node in D such that D is unbounded in restriction to the subtree rooted in that node. This is always possible, as there are only finitely many isomorphism types of subtrees rooted in the available successor nodes. So D cannot be bounded in each of those without being bounded in the father node. $\Box$

Let

$$\psi(Z) := \forall \bar{I}\big(\psi_0(\bar{I}) \to \bigwedge_i (\varphi(\bar{I})_i^\infty \leftrightarrow \varphi(\bar{I}{\restriction}Z)_i^\infty)\big)$$

be an MSO formula that says (of an initial subset Z of $\mathfrak{T}_\omega$) that for all interpretations of countable tree structures of type suitable for $\bar{\varphi}$, the evaluation of the fixed point produces the same outcome at the root whether $\bar{\varphi}$ is evaluated in the whole structure or in the initial substructure induced on Z. Here $\psi_0(\bar{I})$ collects the obvious FO conditions on unary predicates $\bar{I}$ to encode an interpretation of a countable tree structure of the appropriate type within $\mathfrak{T}_\omega$ in the natural way. The intended interpretations are such that all nodes at depth ℓ in the interpreted tree are represented by depth ℓ nodes in $\mathfrak{T}_\omega$. This compatibility with depth is crucial for the reduction; it is for this reason that we work with the ω-branching tree rather than the binary tree.

Lemma 16. *The following are equivalent:*

(i) *$\bar{\varphi}$ bounded.*

(ii) *for some $\ell \in \mathbb{N}$ and for $D = T_\omega{\restriction}\ell$: $\mathfrak{T}_\omega, \lambda \models \psi[D]$.*

(iii) *$\mathfrak{T}_\omega, \lambda \models \exists Z\big(Z$ initial and path-finite $\wedge\, \psi(Z)\big)$.*

(iv) *there is some regular expansion $(\mathfrak{T}_\omega, D)$ of $\mathfrak{T}_\omega$ with an initial, path-finite D, such that $\mathfrak{T}_\omega, \lambda \models \psi[D]$.*

Note that (iii) is decidable, as part of the MSO theory of $\mathfrak{T}_\omega$, thus proving the theorem.

Proof. (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) are clear. (iii) $\Rightarrow$ (iv) is a well known fact about MSO: any MSO formula satisfiable in some tree model has a regular tree model. (iv) $\Rightarrow$ (i) follows with Lemma 15. $\Box$

5 From modal to universal boundedness

Consider boundedness for an equality-free universal formula $\varphi(X, x)$, positive in X, and with all other predicates of one polarity only. W.l.o.g. assume that φ is in Moschovakis normal form (see Proposition 7) and purely negative in all predicates other than X.

$$\begin{aligned} \varphi(X, x) &= \forall \mathbf{y}\Big(\big(\textstyle\bigvee_i \theta_i(x, \mathbf{y})\big) \to \delta(X, \mathbf{y})\Big) \\ &\equiv \textstyle\bigwedge_{i=1}^k \forall \mathbf{y}\big(\theta_i(x, \mathbf{y})\big) \to \delta(X, \mathbf{y})\big), \end{aligned}$$

the θ_i conjunctions of relational atoms (and not involving X), δ a disjunction over X-atoms in variables $\mathbf{y}$.

Intuitively we treat the θ_i as if they were relational guards (in the sense of the guarded fragment, [3]). To this end, we use new relation symbols R_i, whose intended interpretations are the relations defined by θ_i. We may also assume w.l.o.g. that all these R_i have the same arity $m + 1$ where m is the arity of $\mathbf{y}$. Let τ be the original vocabulary of φ, τ^θ its extension by these R_i.

With τ-structures $\mathfrak{A}$ we associate corresponding expansions to τ^θ-structures $\mathfrak{A}^\theta := (\mathfrak{A}, \bar{R})$ by R_i defined by θ_i according to

$$R_i^{\mathfrak{A}} := \{\mathbf{a} = (a_0, a_1, \ldots, a_m) \in A^{m+1} : \mathfrak{A} \models \theta_i[\mathbf{a}]\}.$$

Over these expansions

$$\varphi(X, x) \equiv \textstyle\bigwedge_{i=1}^k \forall \mathbf{y}\big(R_i x \mathbf{y} \to \delta(X, \mathbf{y})\big).$$

5.1 Unfoldings and tree representations

The following technique of switching between tree representations and relational structures is similar to the one used in [10].

Let the vocabulary of φ be τ, and τ^θ its extension by the R_i as introduced above. For tree representations of τ^θ structures we use a vocabulary $\sigma[\tau^\theta]$ consisting of the following unary and binary predicates:

– for each τ^θ-structure $\mathfrak{B}$ of size up to $m + 1$ and an enumeration of these elements as $\mathbf{b} = (b_0, \ldots, b_m)$, a unary predicate P_β associated with the isomorphism type β of $(\mathfrak{B}, \mathbf{b})$ (or, equivalently, the quantifier-free type of $\mathbf{b}$ in $\mathfrak{B}$);

– for each $\rho \subseteq \{0, \ldots, m\} \times \{0, \ldots, m\}$ that is the graph of a partial function, a binary relation E_ρ. We include $\rho = \emptyset$.

With $\mathfrak{A}^\theta$ and a suitable tuple $\mathbf{a} \in A^{m+1}$ we associate

- a transition system $K(\mathfrak{A}^\theta)$ of type $\sigma[\tau^\theta]$,
  with distinguished node $\lambda = \mathbf{a}$,
- the tree unfolding $T(\mathfrak{A}^\theta, \mathbf{a})$ of $K(\mathfrak{A}^\theta)$,
  which is a $\sigma[\tau^\theta]$ tree structure with root $\lambda = \mathbf{a}$.

Conversely, we associate with every $\sigma[\tau^\theta]$ tree structure T satisfying some weak FO definable consistency conditions outlined below:

- a tree-like τ^θ-structure $\mathfrak{B}(T)$,
  of which T is a tree decomposition.

In particular, the tree unfolding of $\mathfrak{A}^\theta$ from $\mathbf{a}$:

- $(\mathfrak{A}^\theta)^* := \mathfrak{B}(T(\mathfrak{A}^\theta, \mathbf{a}))$.

The transition system: The nodes of $K(\mathfrak{A}^\theta)$ are those $(m+1)$-tuples of $\mathfrak{A}^\theta$ that are in one of the relations R_i (i.e., that satisfy one of the θ_i), plus all $(m+1)$-tuples of the form $(a, \dots, a)$ representing single elements. A node $\mathbf{a}$ is in P_β, if β is the quantifier-free type of $\mathbf{a}$ in $\mathfrak{A}^\theta$. Transitions are labelled with labels ρ to indicate identities between components of different tuples. $(\mathbf{a}, \mathbf{a}') \in E_\rho$ for $\rho = \{(i,j) \colon a_i = a'_j\}$. Note that $\rho = \emptyset$ is admitted, and applies to any two nodes corresponding to disjoint tuples; between any two nodes of $K(\mathfrak{A}^\theta)$ there is a transition.

The trees: $T(\mathfrak{A}^\theta, \mathbf{a})$ is the usual bisimilar unfolding of $K(\mathfrak{A}^\theta)$ into a tree structure, rooted in node $\mathbf{a}$.

Let $\mathcal{C}^* = \{T(\mathfrak{A}^\theta, \mathbf{a}) \colon \mathfrak{A}$ a τ-structure $\}$ be the class of all $\sigma[\tau^\theta]$ trees obtained in this manner.

All $T \in \mathcal{C}^*$ satisfy some local and FO definable consistency criteria Cons related to the node labels and transition labels. These enforce, for $(u, u') \in E_\rho$, $u \in P_\beta$, $u' \in P_{\beta'}$, the obvious local compatibility conditions between the quantifier-free types β and β' w.r.t. the identification encoded in ρ. For instance, for ρ such that $i_1, \dots, i_k \in \mathrm{dom}(\rho)$, the following is in Cons:

$$\forall y_1 \forall y_2 \big(E_\rho y_1 y_2 \rightarrow \bigvee \Phi(y_1, y_2)\big),$$

where Φ is the set of all formulae $P_\beta y_1 \wedge P_{\beta'} y_2$ for pairs β, β' for which $\rho \colon \beta\restriction\{i_1, \dots, i_k\} \simeq \beta'\restriction\{\rho(i_1), \dots, \rho(i_k)\}$.

The class $\mathcal{C}^*$, however, is not even MSO definable. A crucial property of $T \in \mathcal{C}^*$ is their homogeneity which stems from the fact that in $K(\mathfrak{A}^\theta)$ every node is reachable from every other one. For the tree unfolding this implies that the sibling subtrees at any two nodes are set-wise isomorphic.

Relational structures from trees: With every tree $T \models$ Cons associate a relational structure $\mathfrak{B}(T)$ as follows. Each node $u \in P_\beta$ of T gives rise to a relational structure $\mathfrak{B}_u$ of size up to $(m+1)$ according to the isomorphism type encoded in β. $\mathfrak{B}(T)$ is obtained from the disjoint union of the $\mathfrak{B}_u$ for $u \in T$, with identifications according to ρ between $\mathfrak{B}_u$ and $\mathfrak{B}_{u'}$ for $(u, u') \in E_\rho$. More formally, let $(u, i) \in T \times \{0, \dots, m\}$ be the element of the disjoint union of the $\mathfrak{B}_u$ that corresponds to the i-th element of $\mathfrak{B}_u$. Let then $\approx$ be the symmetric and transitive closure of the binary relation consisting of the pairs $((u,i),(u',i'))$ for $(u,u') \in E_\rho$ and $(i,i') \in \rho$. The universe of $\mathfrak{B}(T)$ is the quotient of the disjoint union of the $\mathfrak{B}_u$ w.r.t. $\approx$. Cons is such that $\approx$ is compatible with the relational information in the individual $\mathfrak{B}_u$.

It is important to note that the interpretation of the relations R_i over $\mathfrak{B}(T)$ is *not* in general the one that would be defined by the θ_i over the τ-reduct of $\mathfrak{B}(T)$. In general one only has that

$$(*) \quad \mathfrak{B}(T) \models \forall \mathbf{z}(R_i \mathbf{z} \rightarrow \theta_i(\mathbf{z})),$$

due to positivity of the θ_i and the construction of $\mathfrak{B}(T)$.

Remark: T provides a tree decomposition of width m of $\mathfrak{B}(T)$; the patches of this tree decomposition are guarded in the extended vocabulary τ^θ.

For a tree unfolding $(\mathfrak{A}^\theta)^* := \mathfrak{B}(T(\mathfrak{A}^\theta, \mathbf{a}))$ let $\pi \colon (\mathfrak{A}^\theta)^* \rightarrow \mathfrak{A}^\theta$ be the natural projection.

We denote as $\mathfrak{A}^*$ the τ-reduct of $(\mathfrak{A}^\theta)^*$.

Lemma 17. *$\pi \colon (\mathfrak{A}^\theta)^* \rightarrow \mathfrak{A}^\theta$ preserves the stages of the fixed point generation for φ in the sense that for all α: $X^\alpha[\mathfrak{A}^*] = \pi^{-1}\big(X^\alpha[\mathfrak{A}]\big)$.*

Proof. The claim is clear for $\alpha = 0$; also limit stages are trivial. So consider a successor step. Let $\pi(c^*) = c$. We show that

$$\mathfrak{A}^* \models \varphi[X^\alpha[\mathfrak{A}^*], c^*] \quad \text{iff} \quad \mathfrak{A} \models \varphi[X^\alpha[\mathfrak{A}], c],$$

assuming that $X^\alpha[\mathfrak{A}^*] = \pi^{-1}\big(X^\alpha[\mathfrak{A}]\big)$.

Let $\mathfrak{A}^* \models \varphi[X^\alpha, c^*]$ and let $\mathfrak{A} \models \theta_i[c, \mathbf{c}]$. By construction of $\mathfrak{A}^*$ there is $\mathbf{c}^* \in \pi^{-1}(\mathbf{c})$ such that $\mathfrak{A}^* \models \theta_i[c^*, \mathbf{c}^*]$. So $\mathfrak{A}^* \models \delta[X^\alpha, \mathbf{c}^*]$ and hence $\mathfrak{A} \models \delta[X^\alpha, \mathbf{c}]$. It follows that $\mathfrak{A} \models \varphi[X^\alpha, c]$.

Conversely, let $\mathfrak{A} \models \varphi[X^\alpha, c]$ and $\mathfrak{A}^* \models \theta_i[c^*, \mathbf{c}^*]$. Then $\mathfrak{A} \models \theta_i[c, \mathbf{c}]$ for $\mathbf{c} := \pi(\mathbf{c}^*)$, as π is a homomorphism. It follows that $\mathfrak{A} \models \delta[X^\alpha, \mathbf{c}]$ and thus $\mathfrak{A}^* \models \delta[X^\alpha, \mathbf{c}^*]$ by the inductive hypothesis. So $\mathfrak{A}^* \models \varphi[X^\alpha, c^*]$. □

5.2 A translation into modal logic

We want to capture the generation of the least fixed point of φ over $\mathfrak{A}^*$ in terms of the associated tree $T(\mathfrak{A}^\theta, \mathbf{a})$. A modal system $\bar{\varphi}(\bar{X})$ for $\bar{X} = (X_0, \dots, X_m)$ can serve this purpose. A system rather than recursion in a single predicate variable X is necessary, as each node in the tree stands for an $(m+1)$-tuple. The intended meaning of $u \in X_j^\alpha$ will be that the j-th component of this tuple is in X^α. Correspondingly, $\bar{\varphi}$ is of the form $\bar{\varphi}(\bar{X}) = \big(\varphi_0(\bar{X}), \dots, \varphi_m(\bar{X})\big)$.

Recall that $\varphi(X, x)$ is the conjunction of the following, for $1 \leqslant i \leqslant k$:

$$\forall y_1 \dots \forall y_m \big(\theta_i(x, y_1, \dots, y_m) \rightarrow \delta(X, y_1, \dots, y_m)\big).$$

$\varphi_j(\bar{X})$ has, for each $1 \leqslant i \leqslant k$ and each quantifier-free τ^θ-type $\beta = \beta(x_0, \dots, x_m)$ such that $\beta \models \theta_i$ the conjunct

$$\bigwedge_{\rho(j)=0} [E_\rho]\big(P_\beta \rightarrow \delta(\bar{X})\big)$$

where $\delta(\bar{X})$ is the disjunction of those X_s for which Xy_s is a disjunct in $\delta(X, \mathbf{y})$.

The following correspondence between φ and $\bar{\varphi}$ is based on a correspondence between the individual stages, using Lemma 17.

Recall that $\mathcal{C}^*$ is the class of tree structures $T(\mathfrak{A}^\theta, \mathbf{a})$ that represent unfoldings $\mathfrak{A}^*$ of τ-structures $\mathfrak{A}$. With an node u of $T(\mathfrak{A}^\theta, \mathbf{a})$ we associate the tuple $\mathbf{b} = \pi(u) \in A^{m+1}$, which gives rise to $\mathbf{b}$ as an element of $K(\mathfrak{A}^\theta)$ and which is the last node in the path u; we also denote the j-th component b_j of this tuple $\mathbf{b}$ by $\pi_j(u)$, thus regarding π_j as a map $\pi_j \colon T(\mathfrak{A}^\theta, \mathbf{a}) \to A$ for $j \leqslant m$.

Lemma 18. *Consider the stages $(X_j^\alpha)_{j \leqslant m}$ of the fixed point generation for system $\bar{\varphi}$ over $T(\mathfrak{A}^\theta, \mathbf{a})$ and the stages X^α in the generation of the fixed point for φ over $\mathfrak{A}$. Then, for all α, the (X_j^α) represent X^α:*

$$X_j^\alpha = \left\{ u \in T(\mathfrak{A}^\theta, \mathbf{a}) \colon \pi_j(u) \in X^\alpha \right\}.$$

It follows that φ is bounded iff $\bar{\varphi}$ is bounded over $\mathcal{C}^*$.

Proof. As $\bar{\varphi}$ is modal, and since $T(\mathfrak{A}^\theta, \mathbf{a}) = T(K(\mathfrak{A}^\theta), \mathbf{a})$ is bisimilar to $K(\mathfrak{A}^\theta)$, we may prove the correspondence between φ on $\mathfrak{A}$ and $\bar{\varphi}$ on $K(\mathfrak{A}^\theta)$, where is becomes:

$$X_j^\alpha = \left\{ \mathbf{b} \in K(\mathfrak{A}^\theta) \colon b_j \in X^\alpha \right\}.$$

We write K for $K = K(\mathfrak{A}^\theta)$ for the rest of this proof.

The claim for the stages $(X_j^\alpha)_{j \leqslant m}$ of $\bar{\varphi}$ over K and the stages X^α of φ over $\mathfrak{A}$ is proved by induction on α. Consider the successor step from α to $\alpha + 1$ and assume the claim for α.

Let $\mathbf{b} \in X_j^{\alpha+1}$, i.e., $K, \mathbf{b} \models \varphi_j[\bar{X}^\alpha]$. To show that $b_j \in X^{\alpha+1}$ in $\mathfrak{A}$, we need to establish that, for all $\mathbf{b}' = (b_j, b_1', \ldots, b_m')$ such that $\mathfrak{A} \models \theta_i[\mathbf{b}']$ also $\mathfrak{A}, \mathbf{b}' \models \delta[X^\alpha]$. $\mathfrak{A} \models \theta_i[\mathbf{b}']$ implies that $\mathbf{b}'$ is a node of K, contained in some P_β for which $\beta \models \theta_i$, and $(\mathbf{b}, \mathbf{b}') \in E_\rho$ for some ρ with $\rho(j) = 0$. For these β and ρ, therefore, φ_j has the conjunct $[E_\rho](P_\beta \to \delta(\bar{X}))$, and $K, \mathbf{b} \models \varphi_j[\bar{X}^\alpha]$ implies that $K, \mathbf{b}' \models \delta[\bar{X}^\alpha]$. It follows that $\mathfrak{A}, \mathbf{b}' \models \delta[X^\alpha]$.

Conversely, let $\mathbf{b} \in K$ and $b_j \in X^{\alpha+1}$. Then $\mathfrak{A} \models \varphi[X^\alpha, b_j]$ implies that $\mathbf{b} \in X_j^{\alpha+1}$ as follows. Consider a conjunct $[E_\rho](P_\beta \to \delta(\bar{X}))$ in φ_j, where $\beta \models \theta_i$. Then $(\mathbf{b}, \mathbf{b}') \in E_\rho$ implies that $b_j = b_0'$, and $\mathbf{b}' \in P_\beta$ implies that $\mathfrak{A} \models \theta_i[\mathbf{b}']$. Therefore, $\mathfrak{A} \models \varphi[X^\alpha, b_j]$ implies, as $\mathfrak{A} \models \theta_i[\mathbf{b}']$, that $\mathfrak{A} \models \delta[X^\alpha, \mathbf{b}']$. Therefore $K, \mathbf{b}' \models \delta[\bar{X}^\alpha]$. We thus get $K, \mathbf{b} \models \varphi_j[\bar{X}^\alpha]$ and $\mathbf{b} \in X_J^{\alpha+1}$. ☐

This reduction, however, does not yield decidability directly as the class $\mathcal{C}^*$ of relevant tree structures is *not* MSO definable. We must therefore extend the class $\mathcal{C}^*$ of actual tree representations to a wider class $\mathcal{C} \supseteq \mathcal{C}^*$ that is MSO definable and still supports a correspondence as in the lemma. The lemma implies

$$\varphi \text{ unbounded} \quad \Rightarrow \quad \bar{\varphi} \text{ unbounded over } \mathcal{C}$$

for any $\mathcal{C} \supseteq \mathcal{C}^*$. Crucial for the choice of $\mathcal{C}$ is that it is tight enough to make the converse true as well.

5.3 Monotonicity and admissible trees

Consider an arbitrary tree $T \models \mathrm{Cons}$ and the associated τ-structure $\mathfrak{B}(T)$ represented by T. It is clear from the construction of $\mathfrak{B}(T)$ that there is a translation from MSO over $\mathfrak{B}(T)$ into MSO over T. Essentially this amounts to a translation $\psi \mapsto \hat{\psi}$ such that $\mathfrak{B}(T) \models \psi \;\Leftrightarrow\; T \models \hat{\psi}$. Some care has to be taken w.r.t. first- and second-order free variables, though. Consider for instance a formula $\psi(X, x)$. An instantiation by $P \subseteq \mathfrak{B}(T)$ is represented in T by a tuple $\bar{P} = (P_0, \ldots, P_m)$ of subsets of T, where

$$P_j = \{u \in T \colon u_j \in P\} \subseteq T,$$

u_j the j-th component of the tuple represented in $\mathfrak{B}_u$. Similarly, w.r.t. its first-order variable x, ψ translates into a tuple of formulae $\hat{\psi}_j$ such that $\hat{\psi}_j$ expresses of $u \in T$ whether ψ is true of u_j.

So in this case (with just one free variable of each type), $\hat{\psi}$ is of the form $\big(\hat{\psi}_j(\bar{X}, y)\big)_{0 \leqslant j \leqslant m}$ and such that for all j:

$$\mathfrak{B}(T) \models \psi[P, u_j] \quad \text{iff} \quad T \models \hat{\psi}_j[\bar{P}, u].$$

The availability of MSO quantification in this translation is essential even for the translation of FO formulae ψ, since the set of nodes of T in which a given element of $\mathfrak{B}(T)$ is represented can only be defined in MSO. This is because the same element can be carried along (E_ρ)-paths of arbitrary lengths. For instance, in order to check whether $\psi(x) = \exists y Rxy$ is true of some x, which is represented as the i-th component of $u \in T$, $\hat{\psi}_i$ needs to involve checking all nodes u' reachable from u on paths labelled $\rho_1 \ldots \rho_k$ such that $i \in \mathrm{dom}(\rho_k \circ \cdots \circ \rho_1)$.

Lemma 19. *There is an effective translation of $\varphi(X, x)$ into $(\hat{\varphi}_j(\bar{X}, x))_{0 \leqslant j \leqslant m}$ such that for all $P \subseteq \mathfrak{B}(T)$, represented by $\bar{P}$ over T, and for $0 \leqslant j \leqslant m$:*

$$\mathfrak{B}(T) \models \varphi[P, u_j] \quad \textit{iff} \quad T \models \hat{\varphi}_j[\bar{P}, u].$$

Similarly, for the fixed points:

$$\mathfrak{B}(T) \models \varphi^\infty[u_j] \quad \textit{iff} \quad T \models (\hat{\varphi})_j^\infty[u].$$

Due to (negative) monotonicity of φ in the non-recursive predicates, the evaluation of the fixed point w.r.t. $\bar{\varphi}$ in T will in general result in an overestimate for the fixed point of φ over $\mathfrak{B}(T)$. More generally, for any $P \subseteq \mathfrak{B}(T)$ and its representation $\bar{P}$ over T, by $(*)$ in section 5.1,

$$(**) \quad T \models \forall x \big(\hat{\varphi}_i(\bar{X}, x) \to (\bar{\varphi})_i(\bar{X}, x)\big).$$

Definition 20. Call a tree T *admissible for* φ if T is consistent (cf. local FO conditions Cons) and $\hat{\varphi}_j^\infty = (\bar{\varphi})_j^\infty$ on T for $j = 0, \dots, m$. Let $\mathcal{C}_\varphi$ be the class of trees that are admissible for φ.

Observation 21. *$\mathcal{C}_\varphi$ is* MSO *definable.*

Lemma 22. *φ is bounded iff $\bar{\varphi}$ is bounded over $\mathcal{C}_\varphi$.*

Proof. $\mathcal{C}_\varphi \supseteq \mathcal{C}^*$ follows from Lemma 18. The tree structure $T(\mathfrak{A}^\theta, \mathbf{a})$ underlying $\mathfrak{A}^*$ provides a faithful representation of the fixed point generation of φ (over $\mathfrak{A}$ or $\mathfrak{A}^*$, see Lemma 17) in terms of $\bar{\varphi}$. It follows that $\bar{\varphi}$ is unbounded over $\mathcal{C}_\varphi$ if φ is unbounded.

Consider φ that is bounded by $n \in \mathbb{N}$. By the definition of $\mathcal{C}_\varphi$, the results of the fixed points w.r.t. $\bar{\varphi}$ and the faithful translation $\hat{\varphi}$ of φ on $T \in \mathcal{C}_\varphi$ are the same. By monotonicity $(**)$, $\bar{\varphi}$ can only *over*estimate individual stages, which means that $\gamma[\bar{\varphi}, T] \leqslant \gamma[\hat{\varphi}, T] = \gamma[\varphi, \mathfrak{B}(T)] \leqslant n$. So $\bar{\varphi}$ is bounded over $\mathcal{C}_\varphi$. □

This finishes the proof of the main theorem, since boundedness of the modal system $\bar{\varphi}$ over the MSO definable class $\mathcal{C}_\varphi$ is decidable by Theorem 14.

6 Summary

Decidability of $\mathsf{Bdd}(\mathcal{F}(\forall, -))$ has been shown by model theoretic methods, involving a reduction to the boundedness problem for modal systems over some suitable MSO definable class of trees. Both the modal system and the class over which its boundedness is checked depend on the input formula $\varphi(X, x) \in \mathcal{F}(\forall, -)$, and the ultimate target in the reduction is a satisfiability issue in the MSO theory of the ω-branching tree.

As a result of this approach we cannot expect to extract good complexity bounds. The complexity of $\mathsf{Bdd}(\mathcal{F}(\forall, -))$ remains to be determined.

Applicability of the reduction seems to depend, in a rather subtle way, on the very special monotonicity properties involved in universal quantifications in which the only role that the (exclusively negative) occurrences of nonrecursive predicates play is that of restricting the scope of the universal quantifiers, as in universal guarded quantification.

Among the obvious ramifications, therefore, that can be treated in an analogous fashion are systems of $\mathcal{F}(\forall, -)$ formulae in monadic recursive predicates, and universal formulae of the guarded fragment in arbitrary arities.

It remains open whether binary purely universal recursion – with otherwise the same constraints on polarities and equality – has a decidable boundedness problem.

The present investigation covers recursion in monadic recursive predicates (or higher arity but then guarded universal rather than arbitrary universal quantification would seem to be essential).

In the wider picture, major open issues concern the boundedness problem for the full guarded fragment, as well as the monadic boundedness problem for larger fragments or all of FO on the class of structures of tree-width k, for fixed k. In both cases, suitable reductions to the MSO theory of trees could provide a route to decidability proofs.

References

[1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases*. Addison-Wesley, 1995.

[2] M. Ajtai and Y. Gurevich. Datalog vs. first-order logic. *Journal of Computer and System Sciences*, 49, pages 562–588, 1994.

[3] H. Andréka, J. van Benthem, and I. Németi. Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic*, 27, pages 217–274, 1998.

[4] A. Atserias, A. Dawar, and P. Kolaitis. On preservation under homomorphisms and unions of conjunctive queries. In *Proc. 23rd Symp. on Principles of Database Systems*, pages 319–329, 2004.

[5] J. Barwise and Y. Moschovakis. Global inductive definability. *Journal of Symbolic Logic*, 43(3), pages 521–534, 1978.

[6] J. van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, Naples, 1985.

[7] R. Berger. *The Undecidability of the Domino Problem*, volume 66 of *Memoirs of the AMS*. American Mathematical Society, 1966.

[8] E. Börger, E. Grädel, and Y. Gurevich. *The Classical Decision Problem*. Springer-Verlag, 1997.

[9] S. Cosmadakis, H. Gaifman, P. Kanellakis, and M. Vardi. Decidable optimization problems for database logic programs. In *Proc. 20th ACM Symp. on Theory of Computing*, pages 477–490, 1988.

[10] E. Grädel, C. Hirsch, and M. Otto. Back and forth between guarded and modal logics. *ACM Transactions on Computational Logic*, 3, pages 418–463, 2002.

[11] H. Gaifman, H. Mairson, Y. Sagiv, and M. Vardi. Undecidable optimization problems for database logic problems. *Journal of the ACM*, 40(3), pages 683–713, 1993.

[12] G. Hillebrand, P. Kanellakis, H. Mairson, and M. Vardi. Undecidable boundedness problems for datalog programs. *Journal of Logic Programming*, 25(2), pages 163–190, 1995.

[13] P. Kanellakis. *Elements of Relational Database Theory*. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 1073–1157. MIT Press/Elsevier, 1990.

[14] P. Kolaitis and M. Otto. On the boundedness problem for two-variable first-order logic. In *Proc. 13th IEEE Symp. on Logic in Computer Science*, pages 513–524, 1998.

[15] P. Kolaitis and M. Otto. On the boundedness problem for fragments of first-order logic. *unpublished draft*.

[16] P. Kolaitis and C. Papadimitriou. Some computational aspects of circumscription. *Journal of the ACM*, 37, pages 1–14, 1990.

[17] J. McCarthy. Circumscription - a form of nonmonotonic reasoning. *Artificial Intelligence*, 13, pages 27–39, 1980.

[18] Y. Moschovakis. *Elementary Induction on Abstract Structures*. North Holland, 1974.

[19] M. Otto. Eliminating recursion in the μ-calculus. In *Proc. 16th Symp. on Theoretical Aspects of Computer Science*, volume 1563 of LNCS, pages 531–540, 1999.

[20] M. Otto. Modal and guarded characterisation theorems over finite transition systems. *Annals of Pure and Applied Logic*, 130, pages 173–205, 2004.

[21] B. Rossmann. Existential positive types and preservation under homomorphisms. In *Proc. 20th IEEE Symp. on Logic in Computer*, pages 467–476, 2005.