Boundedness of Monadic FO over Acyclic Structures

Stephan Kreutzer[1], Martin Otto[2], and Nicole Schweikardt[1] ⋆

[1] Institut für Informatik, Humboldt Universität zu Berlin,
{kreutzer,schweika}@informatik.hu-berlin.de
[2] Fachbereich Mathematik, Technische Universität Darmstadt,
otto@mathematik.tu-darmstadt.de

Abstract. We study the boundedness problem for monadic least fixed points as a decision problem. While this problem is known to be undecidable in general and even for syntactically very restricted classes of underlying first-order formulae, we here obtain a decidability result for the boundedness issue for monadic fixed points over arbitrary first-order formulae in restriction to acyclic structures.

1 Introduction

The extension of first-order logic by least and greatest fixed points of monotone first-order operators is one of the most natural remedies to some of the obvious limitations of first-order logic when considered as a query language over relational structures. While for instance the very basic graph query concerning reachability of a red node is well known not to be first-order expressible, it possesses a straightforward formalisation in terms of the monadic least fixed point associated with the positive, monotone first-order operator $X \mapsto \{x\colon \mathrm{red}(x) \vee \exists y(Exy \wedge Xy)\}$. In a database context, the query language DATALOG is the extension of positive existential first-order logic by least fixed points. In the context of modal logics, the modal μ-calculus extends basic modal logic by monadic least and greatest fixed points and provides one of the most prominent frameworks for temporal and process logics. In the context of descriptive complexity and finite model theory, the extension of FO by least and greatest fixed points plays a major role. In all these cases it is natural to ask which queries rely on fixed point recursion in an essential way — as opposed to queries which, although they are presented as fixed points, could also be formalised without.

Least fixed points of monotone operators are reached in a (transfinite) iteration of the operator starting from the empty set and taking unions at limit ordinals. The resulting monotone sequence of stages eventually reaches the fixed point. A fixed point process is *bounded* if there is a finite bound on the number of iterations required, uniformly across all input structures. As a decision problem, the *boundedness problem* asks, given a monotone relational operator, whether the least fixed point process for this operator is bounded. As a logical decision problem for first-order formulae, more specifically, we want to decide whether the monotone operator specified by some first-order formula $\varphi(X, x)$ that is positive in X is bounded.

⋆ We gratefully acknowledge our participation in the 2006 Isaac Newton Institute programme on Logic and Algorithms; this opportunity has greatly promoted our collaboration in this research.

The boundedness problem was first studied with a view to query optimisation, in particular for variants of DATALOG. Clearly a bounded fixed point can be eliminated in favour of an explicit unfolding of the iteration to the required depth. By a classical theorem of Barwise and Moschovakis [3], cf. Theorem 2.1 below, boundedness of the fixed point process associated with $\varphi(X, x) \in \mathrm{FO}$ not just implies but is actually equivalent to first-order definability of the fixed point. This reduces the question whether a given fixed point can be eliminated in favour of any "static" first-order definition to the boundedness issue.

The model theoretic link between the procedural behaviour of fixed point recursions and FO-ness provides a source of interest in the boundedness problem that goes far beyond the original motivation from query optimisation. The study of effective criteria for FO-definability of MSO-definable properties has a long tradition in particular over word and tree structures, cf. the handbook chapter [17]. We recall the characterisations by McNaughton–Papert and Schützenberger, and Beauquier–Pin, respectively, in the word case; and the more recent work of Benedikt and Segoufin [4] for trees.

Not surprisingly, the boundedness problem as such is undecidable. In fact one can show that the boundedness problem is at least as hard as the satisfiability problem (for fragments of FO satisfying some very mild closure conditions). But also for many fragments of FO that are decidable for satisfiability, like purely existential or purely universal FO or two-variable first-order logic FO^2, boundedness is known to be undecidable [10, 11, 13, 16]. On the positive side, boundedness is decidable for instance for monadic DATALOG [7] (purely existential positive FO) as well as for its universal counterpart [16], and for the modal fragment of FO [15]. But the decidability region is very narrow; for instance binary DATALOG and monadic DATALOG with inequality are undecidable [10, 11], and also monadic universal FO with equality or with mixed polarities in the static relations [14, 16]. The boundedness problem may thus be viewed as a critical strengthening of the satisfiability issue for fragments of FO; in fact unboundedness of φ precisely corresponds to the satisfiability of the sequence of formulae that express that the n-th stage of φ is non-trivial, for each n.

The undecidability proofs for the boundedness problem of partly very weak fragments of FO suggest that grid-like structures (or structures of unbounded treewidth) play a crucial role in the reduction arguments (in [13, 14, 16], in particular, tiling problems are used). This would explain, for instance, why binary recursion almost inevitably leads to undecidability: here the recursion itself can be used to generate grid-like structures into which other undecidable issues can be coded. For *monadic* recursion, on the other hand, grid-like structure can only reside in the input and not be built up in the fixed point process. It therefore seems conceivable that boundedness of monadic fixed points over certain classes of *tree-like* structures, e.g., structures of bounded treewidth, could be decidable. Such a decidability result, we hope, would provide a uniform framework for most of the known decidability results for monadic boundedness — just as reductions of the tiling problem can provide a rather uniform view of the negative cases, explored in [14].

We therefore want to analyse the boundedness issue from a new perspective, in a way orthogonal to the established approach: rather than looking at fragments of first-order logic, we ask whether decidability can be regained for *all* monadic first-order

fixed points *over suitably restricted classes of structures*. This can be viewed as being part of recent efforts towards developing a model theory for "well-behaved" classes of structures [2, 1, 5].

As a first significant result in this direction concerning boundedness issues, we here show the decidability of boundedness of monadic first-order fixed points over the class of *acyclic structures*, i.e., structures whose (undirected) Gaifman graph is acyclic (we consider such tree-like structures rather than, say, directed acyclic graphs, since boundedness is undecidable over the class of directed acyclic graphs: the two-dimensional grids that can be used to reduce the tiling problem to monadic boundedness issues can clearly be cast as directed acyclic graphs in this sense). A major goal for the extension of the present paper's approach would consist in a similar decidability result for boundedness of monadic fixed points over classes of bounded treewidth.

2 Preliminaries

We denote first-order structures by German letters $\mathfrak{A}, \mathfrak{B}, \ldots$ and their universes by corresponding Roman letters $A, B, \ldots$. We use $\mathcal{C}, \mathcal{T}, \ldots$ for classes of structures. We always assume that classes of structures are closed under isomorphisms. If $\mathfrak{A}$ is a structure, $U \subseteq A$, and $\psi(X)$ is a formula with a free monadic second-order variable X, we write $(\mathfrak{A}, U) \models \psi$, or $\mathfrak{A} \models \psi[U]$, whichever is more convenient, to denote that the expansion of $\mathfrak{A}$ by $X^{\mathfrak{A}} := U$ satisfies ψ. If $\vartheta(X, x)$ is a formula with a free monadic second-order variable X and a free first-order variable x, we let $\vartheta^{\mathfrak{A}}(U) := \{a \in A : \mathfrak{A} \models \vartheta[U, a]\}$. We omit the index if $\mathfrak{A}$ is understood.

Let $\varphi(X, x)$ be a first-order formula with a free first-order variable x and a free second-order variable X, in which it is positive (i.e., X only occurs within the scope of an even number of negation symbols). On corresponding structures $\mathfrak{A}$, φ defines an operator

$$F_\varphi^{\mathfrak{A}} : \mathcal{P}(A) \longrightarrow \mathcal{P}(A) \quad \text{with} \quad F_\varphi^{\mathfrak{A}}(P) = \varphi^{\mathfrak{A}}(P) \quad \text{for each } P \in \mathcal{P}(A),$$

where $\mathcal{P}(A)$ denotes the power set of A. Due to positivity in X, $F_\varphi^{\mathfrak{A}}$ is monotone and hence has a least fixed point $(\mu X.\varphi)^{\mathfrak{A}}$ which we usually write as $\varphi^\infty(\mathfrak{A})$. The fixed point can also be obtained as the limit of the monotone sequence of stages X^α, with α an ordinal, defined by $X^0 := \emptyset$, $X^{\alpha+1} := \varphi^{\mathfrak{A}}(X^\alpha)$, $X^\lambda := \bigcup_{\beta<\lambda} X^\beta$ for limit ordinals λ. We usually write $\varphi^\alpha(\mathfrak{A})$ to denote the α-th stage X^α.

Note that each finite stage X^n, $n \in \mathbb{N}$, is uniformly first-order definable. We write $\varphi^n(x)$ for the formula that defines the n-th stage of φ, which is obtained inductively by substituting $\varphi^{n-1}(x)$ for each atom Xx in $\varphi(X, x)$, where $\varphi^0(x)$ is meant to be false.

F_φ (or the least fixed point of φ or also just φ) is called *bounded* (over a class of structures $\mathcal{C}$) if there is some $n \in \mathbb{N}$ such that $\varphi^\infty(\mathfrak{A}) = \varphi^n(\mathfrak{A})$ for all $\mathfrak{A}$ (for all $\mathfrak{A} \in \mathcal{C}$).

Recall that a class of structures is *elementary* if it is the class of models of some first-order theory; a class is called *projective* if it is the class of models of some first-order theory in a possibly extended vocabulary (cf., e.g., [12]). The notions of FO-*definability* (and similarly, *projective* FO-*definability*, as well as MSO-*definability*) refer to definability in terms of single sentences rather than possibly infinite theories.

At various places throughout the paper we will use the following classical theorem [3] on boundedness.

Theorem 2.1 (Barwise-Moschovakis) *The following are equivalent for every* FO *formula $\varphi(X, x)$ suitable for positive least fixed-point iteration (also in restriction to any elementary or projective class of structures):*

(i) $\varphi(X, x)$ *is bounded.*
(ii) φ^∞ *is uniformly* FO*-definable.*
(iii) φ^∞ *is* FO*-definable in each structure (non-uniformly).*

We also remark that the Löwenheim–Skolem theorem for FO tells us that a first-order fixed point is bounded (over some elementary or projective class) if, and only if, it is bounded over all countable structures (in that class).

Recall that the *Gaifman graph* $\mathcal{G}(\mathfrak{A})$ of a first-order structure $\mathfrak{A} := (A, \tau^{\mathfrak{A}})$ of signature τ is defined as the undirected graph with vertex set A and an edge between two vertices $a, b \in A$, if $a \neq b$ and there exists an $R \in \tau$ and a tuple $(a_1, \ldots, a_r) \in R^{\mathfrak{A}}$ such that $a, b \in \{a_1, \ldots, a_r\}$.

Definition 2.1. *A structure is called acyclic if its Gaifman graph is acyclic.*
$\mathcal{AC}$ denotes the class of all acyclic structures.

Note that $\mathcal{AC}$ is elementary so that the Barwise-Moschovakis theorem applies to it. For the rest of this paper we work over a fixed finite relational signature of unary and binary relation symbols. The restriction to at most binary signatures is w.l.o.g., as in this paper we only work with acyclic structures and a structure containing a relation R of arity > 2 can only be acyclic if every tuple in R contains at most two distinct elements. Such relations can easily be coded in binary relations.

3 Locality

3.1 Syntactic locality and a positive variant of Gaifman's theorem

In 1981, Gaifman [9] proved that any first-order formula is equivalent to a Boolean combination of basic-local sentences and local formulae. We recall the necessary definitions.

Let $\mathfrak{A} := (A, \tau^{\mathfrak{A}})$ be a first-order structure of signature τ. The *distance* $d^{\mathfrak{A}}(a, b)$ between two elements $a, b \in A$ is defined as the length of the shortest path in the Gaifman graph $\mathcal{G}(\mathfrak{A})$ connecting a and b. For $r \geq 0$ and $a \in A$ we define the *r-neighbourhood* of a in $\mathfrak{A}$ as $N_r^{\mathfrak{A}}(a) := \{b \in A : d^{\mathfrak{A}}(a, b) \leq r\}$. It is easily seen that for any $r \geq 0$ there is a first-order formula $\delta_r(x, y) \in \text{FO}[\tau]$ such that $\mathfrak{A} \models \delta_r[a, b]$ iff $d^{\mathfrak{A}}(a, b) \leq r$, for all τ-structures $\mathfrak{A}$ and all $a, b \in A$. For notational convenience we write $d(x, y) \leq r$ for $\delta_r(x, y)$ and $d(x, y) > r$ for $\neg\delta_r(x, y)$.

If $\varphi(x)$ is a first-order formula, then $\varphi^{N_r(x)}(x)$ is the formula obtained from φ by relativising all quantifiers in φ to the r-neighbourhood of x, i.e. replacing $\forall y\psi$ by $\forall y(d(x, y) \leq r \to \psi)$ and $\exists y\psi$ by $\exists y(d(x, y) \leq r \wedge \psi)$. A formula $\psi(x)$ of the form $\varphi^{N_r(x)}(x)$ is called *r-local*. A formula $\psi(x)$ is *local*, or *local in x*, if it is r-local for some $r \geq 0$.

Theorem 3.1 (Gaifman [9]) *Every first-order formula $\varphi(x)$ is equivalent to a Boolean combination of local formulae $\chi(x)$, and basic local sentences, i.e., sentences of the form*

$$\exists x_1 \ldots \exists x_k \Big(\bigwedge_{1 \leq i < j \leq k} d(x_i, x_j) > 2r \ \wedge \ \bigwedge_{1 \leq i \leq k} \vartheta(x_i) \Big)$$

for suitable $r, k > 0$ and an r-local formula $\vartheta(x)$.

The following theorem establishes a variant of Gaifman's locality theorem for first-order formulae $\varphi(X, x)$ which are positive in the monadic second-order variable X. The proof is an adaptation of the proof of the analogous statement for sentences as it appears in [8]. Due to space limitations, we refrain from giving the proof here and refer the reader to the full version of the paper.

Theorem 3.2 *Every formula $\varphi(X, x) \in$ FO that is positive in the monadic second-order variable X is logically equivalent to a finite disjunction of formulae $\varphi_i(X, x) \wedge \psi_i(X)$, where the $\varphi_i(X, x)$ are local in x and the ψ_i are positive in X and conjunctions of (possibly negated) basic local sentences. Furthermore, for every formula $\varphi(X, x) \in$ FO that is positive in X, we can effectively compute a finite disjunction of formulae $\varphi_i(X, x) \wedge \psi_i(X)$ that is equivalent to φ over $\mathcal{AC}$.*

In what follows we shall actually not even rely on the basic local nature of the X-positive sentential components $\psi_i(X)$.

Note that the $\varphi_i(X, x)$ are local in x but not necessarily positive in X. The following example demonstrates that the theorem fails if in addition we require the $\varphi_i(X, x)$ to be positive in X. Let $\tau := \{E, P, X\}$, where E is binary and P and X are unary, and consider the formula $\varphi(X, x) := Px \wedge \exists y(x \neq y \wedge Xy \wedge Py)$. Suppose that $\varphi(X, x)$ is equivalent to a formula $\psi := \bigvee_{i=1}^{k} \big(\varphi_i(X, x) \wedge \psi_i(X)\big)$, where the φ_i are positive in X and local in x and the ψ_i are positive in X. Let $\mathfrak{A} := (\{a, b\}, \tau^{\mathfrak{A}})$ with $E^{\mathfrak{A}} := \emptyset$ and $P^{\mathfrak{A}} := \{a, b\}$ and $X^{\mathfrak{A}} := \{b\}$. Clearly, $\mathfrak{A} \models \varphi[a]$ and therefore there exists an $i \in \{1, \ldots, k\}$ such that $\mathfrak{A} \models (\varphi_i \wedge \psi_i)[a]$. The l-neighbourhoods of a and b are distinct for all l and their $\{E, P\}$-reducts are isomorphic. As $\mathfrak{A} \models \varphi_i[a]$ and φ_i is positive in X it follows that $\mathfrak{A} \models \varphi_i[b]$. Hence, $\mathfrak{A} \models (\varphi_i \wedge \psi_i)[b]$ but $\mathfrak{A} \not\models \varphi[b]$ contradicting the equivalence of φ and ψ.

3.2 Locality of queries

Definition 3.3 *Let $\mathcal{C}$ be a class of τ-structures.*

(i) *A* monadic query *Q on $\mathcal{C}$ is a mapping which assigns to each $\mathfrak{A} \in \mathcal{C}$ a set $Q^{\mathfrak{A}} \subseteq A$ so that for all isomorphisms $\pi : \mathfrak{A} \cong \mathcal{B}$ and all $a \in A$, $a \in Q^{\mathfrak{A}} \iff \pi(a) \in Q^{\mathfrak{B}}$.*

(ii) *A monadic query Q is* MSO*-definable, if there is an* MSO*-formula $\varphi(x) \in \mathrm{MSO}[\tau]$ such that $Q^{\mathfrak{A}} = \{a \in A : \mathfrak{A} \models \varphi[a]\}$ for all $\mathfrak{A} \in \mathcal{C}$.*

(iii) *Let $l \in \mathbb{N}$. Q is l-*local *over $\mathcal{C}$, if for all countable structures $\mathfrak{A}, \mathfrak{B} \in \mathcal{C}$ and $a \in A, b \in B$ such that $(\mathfrak{A}|_{N_l^{\mathfrak{A}}(a)}, a) \cong (\mathfrak{B}|_{N_l^{\mathfrak{B}}(b)}, b)$: $a \in Q^{\mathfrak{A}} \iff b \in Q^{\mathfrak{B}}$.*
Here, $\mathfrak{A}|_{N_l^{\mathfrak{A}}(a)}$ denotes the restriction of $\mathfrak{A}$ to the l-neighbourhood of a.

(iv) *Q is* local*, if it is l-local for some $l \in \mathbb{N}$.*

Note that locality for queries is a purely semantic property, as opposed to syntactic locality of formulae discussed above. We refer to just *countable* structures in (iii), as this will allow us to use interpretations of the relevant structures in the ω-branching tree; for our boundedness concerns we may always restrict attention to countable structures, by the Löwenheim–Skolem theorem (cf., Section 2).

Lemma 3.4 *Let $\varphi(x) \in$ MSO and let Q be the query defined by φ over a class $\mathcal{C} \subseteq \mathcal{AC}$. If Q is r-local for some $r \in \mathbb{N}$, then Q is* FO*-definable over $\mathcal{C}$, in fact even by a local first-order formula.*

In other words, a local query Q is MSO*-definable if, and only if, it is* FO*-definable.*

Proof (Sketch). The proof is obtained on the basis of a decomposition argument for MSO-types and the observation that MSO-formulas of quantifier rank q can count multiplicities of types only up to a threshold $g(q)$. This can be used to translate r-local MSO-formulae into FO-formulae by induction on the locality radius r. □

4 Boundedness over acyclic structures

The goal of this section is to prove the following theorem, the main result of the paper. Recall Definition 2.1 for the class of acyclic structures $\mathcal{AC}$.

Theorem 4.1 *Let $\mathcal{C} \subseteq \mathcal{AC}$ be an* FO*-definable subclass of $\mathcal{AC}$ and let $\varphi(X, x)$ be a first-order formula positive in the monadic second-variable X. Then boundedness of φ in $\mathcal{C}$ is decidable.*

The main ingredients for the proof are the following. In Section 4.2, we use Theorem 3.2 to show that deciding boundedness for arbitrary first-order formulae $\varphi(X, x)$, positive in X, can be decomposed into a sequence of boundedness tests for purely local formulae $\varphi_i(X, x)$ in certain projectively FO-definable subclasses $\mathcal{C} \subseteq \mathcal{AC}$.

To decide boundedness for local formulae $\varphi_i(X, x)$, we show that this can further be reduced to deciding whether the global relation $\varphi^\infty(x)$ defined by $\varphi_i(X, x)$ is local. Locality of MSO-definable queries in MSO-definable subclasses of $\mathcal{AC}$ can be decided via reduction to the MSO-theory of the ω-branching tree and Rabin's theorem, see Section 4.1.

4.1 Locality testing for MSO-queries

Theorem 4.2 *Let $\mathcal{C} \subseteq \mathcal{AC}$ be an* MSO*-definable subclass of $\mathcal{AC}$ and let Q be a query that is* MSO*-definable over $\mathcal{C}$. Then locality of Q over $\mathcal{C}$ is decidable. The decision procedure is uniform in the* MSO*-formula defining the subclass $\mathcal{C}$.*

The remainder of this section is devoted to the proof of the theorem, which is based on a reduction to Rabin's following classical result.

Theorem 4.3 (Rabin) *The* MSO*-theory of the infinite ω-branching tree is decidable.*

We denote the infinite ω-branching tree as $\mathfrak{T}_\omega$, its root as λ. The first step towards the proof of Theorem 4.2 is to show that there is a uniform MSO-interpretation that interprets structures $\mathfrak{A} \in \mathcal{C}$ in colourings of $\mathfrak{T}_\omega$ in a suitable way. We can then use the decidability of the MSO-theory of $\mathfrak{T}_\omega$ to check if the query is local. This method follows ideas from [15, 16] and in particular allows us to capture locality (boundedness) in the framework of MSO over trees through a regular analogue of König's lemma that is available in $\mathfrak{T}_\omega$.

Let $\mathcal{C}$ be an MSO-definable subclass of $\mathcal{AC}$ over a signature σ. Given a σ-structure $\mathfrak{A} \in \mathcal{AC}$ and an element $a \in A$ we encode $\mathfrak{A}, a$ in a colouring of $\mathfrak{T}_\omega$ by a suitable set $\tau := \tau(\sigma)$ of colours: every tree in the Gaifman graph $\mathcal{G}(\mathfrak{A})$ is coded in a subtree rooted at a successor of the root λ so that the distinguished element a is encoded by a direct successor of λ and marked by a special colour. We refrain from giving details here as the encoding is straightforward. It is important, however, that the encoding preserves the distance between elements of the structure within each component of $\mathcal{G}(\mathfrak{A})$.

The interpretation allows us to translate MSO-formulae over structures from $\mathcal{AC}$ to MSO-formulae over encodings in $\mathfrak{T}_\omega$. In particular, there is an MSO-sentence ψ that is true in a τ-expansion $\mathfrak{T}$ of $\mathfrak{T}_\omega$ if, and only if, $\mathfrak{T}$ encodes some structure $\mathfrak{A} \in \mathcal{C}$.

Let $\varphi(x)$ be an MSO-formula defining a query Q over $\mathcal{C}$ (which again can be translated to a corresponding formula $\varphi'(x)$ over τ-expansions of $\mathfrak{T}_\omega$). For an acyclic structure $\mathfrak{A}$ with distinguished element $a \in A$ and $r \geq 1$ we let $\mathfrak{A}|_r := \mathfrak{A}|_{N_r^{\mathfrak{A}}(a)}$ denote the initial segment of all nodes up to depth r. A set $A' \subseteq A$ is *initial*, if it contains a and is connected (in the Gaifman graph). A' is called *local*, if $A' \subseteq \mathfrak{A}|_r$ for some r. It is *path-finite* if it contains no infinite paths. Note that while it is MSO-definable that a subset of $\mathcal{T}_\omega$ is initial or path-finite, locality is not (this follows from König's lemma together with the fact that every MSO-formula that is satisfiable in $\mathfrak{T}_\omega$ is also satisfiable in a finitely branching tree).

A *regular* expansion of $\mathfrak{T}_\omega$ is one that realises only finitely many isomorphism types of subtrees. The following regular analogue of König's lemma is proved in [16].

Lemma 4.4 *An initial subset D in a regular expansion $(\mathfrak{T}_\omega, D)$ of $\mathfrak{T}_\omega$ is path-finite if, and only if, it is local.*

Let $\varphi_{local}(Z)$ be an MSO-formula that says of an inital subset Z of $\mathfrak{T}_\omega$ rooted at some immediate successor z of the root λ, that whenever encodings of two structures $\mathfrak{A}, \mathfrak{B} \in \mathcal{C}$ in $\mathfrak{T}_\omega$, whose distinguished elements $a \in A$ and $b \in B$ are represented by z, agree on Z, then $\mathfrak{A} \models \varphi[a]$ if, and only if, $\mathfrak{B} \models \varphi[b]$. Clearly, this is MSO-definable. Analogous to the reasoning in [16], we can show the following lemma.

Lemma 4.5 *Let λ be the root of $\mathfrak{T}_\omega$. The following are equivalent:*

(i) *$\varphi(x)$ is local*
(ii) *for some $r \in \mathbb{N}$ and for $D = \mathfrak{T}_\omega|_r$: $\mathfrak{T}_\omega \models \varphi_{local}[D]$.*
(iii) *$\mathfrak{T}_\omega \models \exists Z\big(Z$ is initial and path-finite $\wedge\ \varphi_{local}(Z)\big)$.*
(iv) *there is a regular expansion $(\mathfrak{T}_\omega, D)$ of $\mathfrak{T}_\omega$ with an initial, path-finite D such that $\mathfrak{T}_\omega \models \varphi_{local}[D]$.*

Proof. $(i) \implies (ii) \implies (iii)$ are obvious. $(iii) \implies (iv)$ is a well-known fact about MSO. For $(iv) \implies (i)$ use Lemma 4.4. □

By Rabin's theorem, part (iii) of the previous lemma, and hence locality of φ, is decidable. This completes the proof of Theorem 4.2.

4.2 Boundedness of arbitrary monadic FO-formulae

We show that the boundedness problem for arbitrary monadic first-order formulae over acyclic structures can be reduced to the locality test for MSO-formulae as provided by Theorem 4.2.

Let $\varphi(X, x) \in \text{FO}$ be a formula, positive in X. By Theorem 3.2, φ is equivalent to $\bigvee_i(\varphi_i(X, x) \wedge \psi_i(X))$, where the $\varphi_i(X, x)$ are local in x and the ψ_i are positive in X. To simplify the presentation, we only consider the case where

$$\varphi \equiv \big(\varphi_1(X, x) \wedge \psi_1(X)\big) \vee \big(\varphi_2(X, x) \wedge \psi_2(X)\big).$$

We can treat the ψ_i as *guards*, that enable or disable the contribution of φ_i to the fixed-point induction, in the sense that certain ψ_i may be true already at the initial stage (and stay true for the whole induction process) whereas others may only become true at later stages. In this case, the corresponding φ_i only contribute to the induction process beginning with these stages and can be neglected before.

Inductions over φ can therefore be decomposed into a sequence of *phases*. Let $\mathfrak{A}$ be a structure. We distinguish between several cases:

(0) $\mathfrak{A} \not\models \psi_1[\emptyset]$ and $\mathfrak{A} \not\models \psi_2[\emptyset]$. Then $\varphi^\infty(\mathfrak{A}) = \emptyset$, so we can ignore this case.

(1) $\mathfrak{A} \models \psi_1[\emptyset]$ and $\mathfrak{A} \models \psi_2[\emptyset]$. In this case, the induction on φ in $\mathfrak{A}$ is equivalent to the induction on $(\varphi_1 \vee \varphi_2)$, which is purely local.

(2) $\mathfrak{A} \models \psi_1[\emptyset]$ and $\mathfrak{A} \not\models \psi_2[\varphi^\infty(\mathfrak{A})]$. In this case, the induction on φ in $\mathfrak{A}$ is equivalent to the induction on φ_1, as the guard ψ_2 for φ_2 will never become true.

(3) $\mathfrak{A} \models \psi_1[\emptyset]$ and $\mathfrak{A} \models \psi_2[\varphi^\infty(\mathfrak{A})]$. Over $\mathfrak{A}$, the induction has two phases. It starts with an induction on φ_1. As soon as $\psi_2(X^\alpha)$ is satisfied for a stage α, the induction continues on $\varphi_1 \vee \varphi_2$.

There are two more cases, just like (2) and (3) but with the roles of φ_1/ψ_1 and φ_2/ψ_2 exchanged (we suppress this duplication in the following).

To illustrate the possible interplay between the various phases, let us consider formulae in positive Gaifman form of the kind

$$\varphi(X, x) = \psi_0 \wedge \big[Px \vee \big(\varphi_1(X, x) \wedge \exists x(Xx \wedge Q_1x)\big) \vee \big(\varphi_2(X, x) \wedge \exists x(Xx \wedge Q_2x)\big)\big],$$

with local formulae $\varphi_i(X, x)$, and with extra unary predicates P and Q_i, which serve to initialise the fixed point process and to trigger phases driven by φ_1, φ_2 or their combination $\varphi_1 \vee \varphi_2$. Static side conditions in ψ_0 and the local formulae φ_i may be such that, for instance, the induction starts on φ_1 (e.g., ψ_0 may force $Q_2 \cap P = \emptyset$), and that the induction on φ_1 alone would be unbounded while the induction on φ_2 alone would be bounded, but such that the unbounded induction on φ_1 always triggers the induction on φ_2 (due to Q_2-elements in the 17-th stage of φ_1 in any model of ψ_0 in which φ_1 is unbounded, say). The interplay between φ_1 and φ_2 could now still be such that the overall process is bounded or unbounded. It should be clear from this naive analysis of a very simple family of examples that the full variety of phase patterns needs to be taken

into account and that several boundedness issues for purely local processes determine the overall boundedness in a non-trivial manner.

Let us proceed with the proof of Theorem 4.1. Considering the possible sources for unboundedness of φ, it is clear that φ is unbounded if, and only if, at least one of the following applies:

Case 1: $\varphi_1 \vee \varphi_2$ is unbounded over $\mathcal{C}_1 := \{\mathfrak{A} \in \mathcal{AC} : \mathfrak{A} \models (\psi_1 \wedge \psi_2)[\emptyset]\}$.
Case 2: φ_1 is unbounded over $\mathcal{C}_2 := \{\mathfrak{A} \in \mathcal{AC} : \mathfrak{A} \models \psi_1[\emptyset]$ and $\mathfrak{A} \not\models \psi_2[\varphi^\infty(\mathfrak{A})]\}$.
Case 3: **a)** $\varphi_1 \vee \psi_2$ is unbounded over $\mathcal{C}_3 := \{\mathfrak{A} \in \mathcal{AC} : \mathfrak{A} \models \psi_1[\emptyset]\}$, or
b) $\varphi_1 \vee \psi_2$ is bounded over $\mathcal{C}_3$, with some bound $n \in \mathbb{N}$, and $\varphi_1 \vee \varphi_2$ is unbounded over $\mathcal{C}_4 := \{\mathfrak{A} \in \mathcal{AC} : \mathfrak{A} \models \psi_1[\emptyset]$ and $\mathfrak{A} \models \psi_2[\varphi^n(\mathfrak{A})]\}$.
Case 4: like 2 and 3 with the roles of φ_1/ψ_1 and φ_2/ψ_2 exchanged.

We further reduce the number of the cases that need to be considered.

Lemma 4.6 (1) *φ_1 is unbounded over $\mathcal{C}_2$ if, and only if, φ_1 is unbounded over $\mathcal{C}_2' := \{\mathfrak{A} \in \mathcal{AC} : (\mathfrak{A}, \emptyset) \models \psi_1$ and $(\mathfrak{A}, P) \not\models \psi_2$ for some fixed point P of $\varphi_1\}$.*
(2) *If $\varphi_1 \vee \psi_2$ is unbounded over $\mathcal{C}_3$, then φ_1 is unbounded over $\mathcal{C}_2$.*
(3) *If $\varphi_1 \vee \psi_2$ is bounded over $\mathcal{C}_3$, then φ is unbounded over $\mathcal{C}_3$ if, and only if, $\varphi' := \varphi_1 \vee \varphi_2 \vee Px$ is unbounded over $\mathcal{C}_3' := \{(\mathfrak{A}, P) : \mathfrak{A} \in \mathcal{AC}, \mathfrak{A} \models \psi_1[\emptyset], \mathfrak{A} \models \psi_2[P], P \subseteq \varphi^\infty[\mathfrak{A}]\}$.*

Note that $\mathcal{C}_2'$ is projectively FO-definable, as the fact that P is a fixed point of φ_1 is FO-definable with a new relation for P. Both $\mathcal{C}_3'$ and $\mathcal{C}_2'$ are clearly MSO-definable.

Proof. (1) is trivial. For (2), assume $\varphi_1 \vee \psi_2$ is unbounded over $\mathcal{C}_3$ but φ_1 is bounded over $\mathcal{C}_2$. Consider the theories $T_0 := \{\exists x(\varphi^{n+1}(x) \wedge \neg\varphi^n(x)) : n \in \mathbb{N}\}$ and

$$T := T_0 \cup \{\neg\psi_2(\varphi^n) \colon n \in \mathbb{N}\} \cup \text{ (axiomatisation of } \mathcal{C}_3\text{)}.$$

T is consistent, as $\varphi_1 \vee \psi_2$ is unbounded over $\mathcal{C}_3$. Note that T speaks about φ, not φ_1, but $\varphi^n \equiv \varphi_1^n$ in models of T as T implies ψ_1 and $\neg\psi_2(\varphi^n)$ for all $n \in \mathbb{N}$. By assumption, φ is bounded over $\mathcal{C}_2$ and hence T_0 is not satisfiable in $\mathcal{C}_2$. Therefore,

$$\text{every model of } T \text{ satisfies } \psi_2(\varphi^\infty), \qquad (*)$$

as $\neg\psi_2(\varphi^\infty)$ is the defining condition of the subclass $\mathcal{C}_2$ of $\mathcal{C}_3$. But in every model $\mathfrak{A}$ of $\psi_2(\varphi^\infty)$, $(\varphi \vee \psi_2)^\infty(\mathfrak{A}) = (\varphi_1 \vee \psi_2)^\infty(\mathfrak{A}) = A$ and with (4.2), therefore, $(\varphi \vee \psi_2)^\infty(\mathfrak{A}) = (\varphi_1 \vee \psi_2)^\infty(\mathfrak{A}) = A$ in every model $\mathfrak{A}$ of T. Furthermore, the equivalence between the first and the second induction is even stage-by-stage.

Hence, in the model class of T, the fixed point $(\varphi \vee \psi_2)^\infty$ is trivially first-order definable – it is defined by *true*. Therefore, by the Barwise-Moschovakis Theorem, $(\varphi \vee \psi_2)$ is bounded in models of T, contradicting T_0.

(3) Note that the phase in the generation of φ^∞ over some $\mathfrak{A} \in \mathcal{C}_3$ that is driven by $\varphi_1 \vee \varphi_2$ is stage-by-stage equivalent to the fixed point generation of $(\varphi')^\infty$ over $(\mathfrak{A}, \varphi^n(\mathfrak{A})) \in \mathcal{C}_3'$, for the minimal n such that $\mathfrak{A} \models \psi_2[\varphi^n(\mathfrak{A})]$. Therefore boundedness of φ over $\mathcal{C}_3$ trivially implies boundedness of φ' over $\mathcal{C}_3'$. Conversely, by decomposition into corresponding phases, boundedness of $\varphi_1 \vee \psi_2$ and of φ' over $\mathcal{C}_3'$ together imply that φ is bounded over $\mathcal{C}_3$. □

The desired decision procedure for boundedness of φ is built on a sequence of applications of decision procedures that detect (un)boundedness in the form of (non)locality. Consider the individual boundedness issues in cases 1–3 above in the light of their reformulation in Lemma 4.6:

- Case 1 corresponds to a boundedness test for a purely local fixed point: $\varphi_1 \vee \varphi_2$ is purely local, and stage-by-stage equivalent with φ over the FO-definable subclass $\mathcal{C}_1$ of $\mathcal{AC}$. As $\mathcal{C}_1$ is elementary, by the Barwise-Moschovakis Theorem, boundedness is equivalent to FO-definability of the fixed point. Further, as a bounded fixed point over a purely local formula is local itself, boundedness of $\varphi_1 \vee \varphi_2$ is equivalent to *local* FO-definability. Finally, as $\mathcal{C}_1$ and the fixed point are MSO-definable, local FO-definability of the fixed point is decidable according to Theorem 4.2 and Lemma 3.4.
- Case 2 is similarly reduced to a boundedness test for a purely local fixed point over the projective and MSO-definable class $\mathcal{C}_2'$ in part 1 of the lemma.
- Case 3a) is in fact subsumed by case 2, according to part 2 of the lemma.

For what remains (case 3b) we may restrict attention to the situation where $\varphi_1 \vee \psi_2$ is bounded over $\mathcal{C}_3$. The corresponding decision issue obtained in part 3 of Lemma 4.6 is further reduced to a locality issue that can be decided according to Theorem 4.2, through condition (iii) in the following.

Claim 4.7 *Let $\varphi_1 \vee \psi_2$ be bounded over $\mathcal{C}_3$. Then the following are equivalent:*
(i) *φ is bounded over $\mathcal{C}_3$.*
(ii) *φ' is bounded over $\mathcal{C}_3'$.*
(iii) *$(\varphi')^\infty$ is local over $\mathcal{C}_3'$.*

Proof. The equivalence between (i) and (ii) was dealt with in part 3 of Lemma 4.6. As φ' is purely local, boundedness clearly implies locality for the fixed point, i.e., (ii) $\Rightarrow$ (iii). Assume now (iii). Obviously, $(\varphi')^\infty$ is MSO-definable and therefore FO-definable due to locality, by Lemma 3.4. As $\varphi_1 \vee \psi_2$ is bounded over $\mathcal{C}_3$, there is some n such that for $\mathfrak{A} \in \mathcal{C}_3$, $\mathfrak{A} \models \psi_2[\varphi^\infty(\mathfrak{A})] \iff \mathfrak{A} \models \psi_2[\varphi^n(\mathfrak{A})]$. Further, φ' is stage-by-stage equivalent to $\varphi \vee Px$ over $\mathcal{C}_3'$ and hence $(\varphi \vee Px)^\infty$ is FO-definable over the elementary subclass $\mathcal{C}_3'' := \{(\mathfrak{A}, P) : \mathfrak{A} \in \mathcal{AC}, \mathfrak{A} \models \psi_1[\emptyset], \mathfrak{A} \models \psi_2[P], P = \varphi^n(\mathfrak{A})\} \subseteq \mathcal{C}_3'$. By the Barwise-Moschovakis theorem, therefore, $\varphi \vee Px$ is bounded over $\mathcal{C}_3''$, which implies that φ is bounded over $\mathcal{C}_3$, since also $\varphi_1 \vee \psi_2$ is bounded. □

It follows that testing for unboundedness of $\varphi_1 \vee \varphi_2$ over $\mathcal{C}_1$;
else, for unboundedness of[3] φ_1 over $\mathcal{C}_2'$;
else, for unboundedness of[3] φ' over $\mathcal{C}_3'$,

fails if, and only if, φ is bounded over $\mathcal{AC}$. In the given format, each one of these unboundedness tests can be realised as a (non)locality test, which is effective through Theorem 4.2. This cascade of locality tests can be adapted to the (more complex) phase pattern of a formula in positive Gaifman form with more than two disjuncts. We also

[3] and the mirror symmetric case with φ_1/ψ_1 exchanged for φ_2/ψ_2

point out that all the arguments used relativise to FO-definable subclasses of $\mathcal{AC}$. This concludes the proof of Theorem 4.1. □

Theorem 4.1 yields a decision procedure for the boundedness problem of FO on acyclic structures. However, the running time of the procedure grows non-elementarily with the formula size. There is no hope for an algorithm whose running time is bounded by an elementary function. As the first-order satisfiability problem over $\mathcal{AC}$ is reducible to the boundedness problem over $\mathcal{AC}$, the non-elementary lower bound for satisfiability on $\mathcal{AC}$ established in [6] gives us a non-elementary lower bound here.

5 Outlook

Given the decidability of the boundedness problem on acyclic structures, it is natural to ask whether, for instance, the result also holds for the class of trees or extends to the class of structures of treewidth at most k, for fixed k. We collect a few observations to indicate that the results and the methods used here are indeed rather sensitive to the underlying class of structures.

Remark 5.1 (a) *Boundedness over trees (connected acyclic structures) does not imply boundedness over all acyclic structures; similarly, boundedness over finite acyclic structures does not imply boundedness over all acyclic structures.*
(b) *The Barwise-Moschovakis theorem is well known to fail over the class of all finite structures. It also fails over the class of finite acyclic structures, over the class of trees and over the class of finite trees.*
(c) *The connection between locality and* FO*-definability of Lemma 3.4. fails over simple* FO*-definable classes of graphs of bounded treewidth.*

Proof. For (a) consider, over acyclic structures or over trees, a reachability query (in itself unbounded) modified by a side condition that renders the process trivial unless there are at least two nodes of degree 0 (impossible over trees), or unless there is precisely one vertex of degree 1 (impossible in *finite* acyclic structures).

For (b) consider over finite forests the fixed point process that evaluates to the well-founded part; while this process is unbounded, it evaluates to the full vertex set. Over (finite or infinite) trees, the unbounded fixed point based on $\mathrm{red}(x) \vee \exists y(Exy \wedge Xy) \vee \exists y(\mathrm{root}(y) \wedge Xy)$ is in FO (equivalent to $\exists x\ \mathrm{red}(x)$).

For (c) consider connectivity over the FO-definable class of graphs that are disjoint unions of cycles and two-way infinite successor chains (treewidth 2). Clearly this query is in MSO but neither local nor FO. Joining all the nodes in such graphs to one new central vertex, one obtains an FO-definable class of graphs of treewidth 3, which is FO-bi-interpretable with the original class. As all structures in this class have diameter 2, any query is trivially local here. □

Clearly, these observations do not imply that boundedness is undecidable over corresponding classes. However, we need to develop new tools to show decidability. This is part of ongoing work.

As pointed out above, classes of bounded treewidth in particular provide an interesting wider framework for the analysis of the boundedness problem, because of their

relevance for other known decidable cases. Here, arguments based on compactness and the Barwise-Moschovakis theorem are still available, and so is in principle the phase analysis of Section 4.2. However, as indicated in (c) above, the direct link with locality is lost (at a more technical level, distances and locality are not preserved in the passage between the structures themselves and their tree representations). The search for alternative methods to decide FO-definability or boundedness in this context remains of particular interest.

References

1. A. Atserias, A. Dawar, and M. Grohe. Preservation under extensions on well-behaved finite structures. In *Proc. ICALP'05*, volume 3580 of *Springer LNCS*, pages 1437–1449, 2005.
2. A. Atserias, A. Dawar, and P. G. Kolaitis. On preservation under homomorphisms and unions of conjunctive queries. *Journal of the ACM*, 53(2):208–237, 2006.
3. J. Barwise and Y. Moschovakis. Global inductive definability. *Journal of Symbolic Logic*, 43:521–534, 1978.
4. M. Benedikt and L. Segoufin. Regular tree languages definable in FO and $\mathrm{FO}_{\mathrm{mod}}$. In *Proc. STACS'05*, volume 3404 of *Springer LNCS*, pages 327–339, 2005. Also see manuscript of journal paper.
5. M. Benedikt and L. Segoufin. Towards a characterization of order-invariant queries over tame structures. In *Proc. CSL'05*, volume 3634 of *Springer LNCS*, pages 276–291, 2005.
6. K. Compton and C. Ward Henson. A uniform method for proving lower bounds on the computational complexity of logical theories. In S. Abramsky, D. M. Gabbay, and T. Maibaum, editors, *Handbook of Logic in Computer Science*, volume 5: Logic and Algebraic Methods, pages 129–216. Oxford University Press, 2000.
7. S. Cosmadakis, H. Gaifman, P. Kanellakis, and M. Vardi. Decidable optimization problems for database logic programs. In *Proc. STOC'88*, pages 477–490, 1988.
8. A. Dawar, M. Grohe, S. Kreutzer, and N. Schweikardt. Approximation schemes for first-order definable optimisation problems. In *Proc. LICS'06*, pages 411–420, 2006.
9. H. Gaifman. On local and non-local properties. In J. Stern, editor, *Herbrand Symposium, Logic Colloquium '81*, pages 105–135. North Holland, 1982.
10. H. Gaifman, H. Mairson, Y. Sagiv, and M. Vardi. Undecidable optimization problems for database logic problems. *Journal of the ACM*, 40:683–713, 1993.
11. G. Hillebrand, P. Kanellakis, H. Mairson, and M. Vardi. Undecidable boundedness problems for datalog programs. *Journal of Logic Programming*, 25:163–190, 1995.
12. W. Hodges. *Model Theory*. Cambridge University Press, 1993.
13. P. Kolaitis and M. Otto. On the boundedness problem for two-variable first-order logic. In *Proc. LICS'98*, pages 513–524, 1998.
14. P. Kolaitis and M. Otto. On the boundedness problem for fragments of first-order logic: undecidability results. Unpublished draft, 1999.
15. M. Otto. Eliminating recursion in the μ-calculus. In *Proc. STACS'99*, volume 1563 of *Springer LNCS*, pages 531–540, 1999.
16. M. Otto. The boundedness problem for monadic universal first-order logic. In *Proc. LICS'06*, pages 37–46, 2006.
17. W. Thomas. Languages, automata, and logic. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Language Theory*, volume 3, pages 389–455. Springer, 1997.