The Complexity of Boundedness for Guarded Logics

Michael Benedikt[1], Balder ten Cate[2],
Thomas Colcombet[3], **Michael Vanden Boom**[1]

[1]University of Oxford [2]LogicBlox and UC Santa Cruz [3]Université Paris Diderot

LICS 2015
Kyoto, Japan

Least fixpoint

Consider $\psi(\boldsymbol{y}, Y)$ positive in Y (of arity $m = |\boldsymbol{y}|$).

For all structures $\mathfrak{A}$, the formula ψ induces a monotone operation

$$\begin{aligned} \mathcal{P}(A^m) &\longrightarrow \mathcal{P}(A^m) \\ V &\longmapsto \psi_{\mathfrak{A}}(V) := \{\boldsymbol{a} \in A^m : \mathfrak{A}, \boldsymbol{a}, V \models \psi\} \end{aligned}$$

$\Rightarrow$ there is a unique least fixpoint $[\mathbf{lfp}_{Y,\boldsymbol{y}}.\psi(\boldsymbol{y}, Y)]_{\mathfrak{A}} := \bigcup_\alpha \psi_{\mathfrak{A}}^\alpha$

$$\begin{aligned} \psi_{\mathfrak{A}}^0 &:= \varnothing \\ \psi_{\mathfrak{A}}^{\alpha+1} &:= \psi_{\mathfrak{A}}(\psi_{\mathfrak{A}}^\alpha) \\ \psi_{\mathfrak{A}}^\lambda &:= \bigcup_{\alpha<\lambda} \psi_{\mathfrak{A}}^\alpha \end{aligned}$$

Boundedness problem

Boundedness problem for $\mathcal{L}$

Input: $\psi(\boldsymbol{y}, Y) \in \mathcal{L}$ positive in Y

Question: is there $n \in \mathbb{N}$ s.t. for all structures $\mathfrak{A}$, $\psi^n_{\mathfrak{A}} = \psi^{n+1}_{\mathfrak{A}}$?
(i.e. the least fixpoint is always reached within n iterations)

Boundedness problem

Boundedness problem for $\mathcal{L}$

Input: $\psi(\boldsymbol{y}, Y) \in \mathcal{L}$ positive in Y

Question: is there $n \in \mathbb{N}$ s.t. for all structures $\mathfrak{A}$, $\psi^n_{\mathfrak{A}} = \psi^{n+1}_{\mathfrak{A}}$?
(i.e. the least fixpoint is always reached within n iterations)

$$\psi_1(xy, Y) := Rxy \vee \exists z\,(Rxz \wedge Yzy)$$

Boundedness problem

Boundedness problem for $\mathcal{L}$

Input: $\psi(\boldsymbol{y}, Y) \in \mathcal{L}$ positive in Y

Question: is there $n \in \mathbb{N}$ s.t. for all structures $\mathfrak{A}$, $\psi^n_{\mathfrak{A}} = \psi^{n+1}_{\mathfrak{A}}$?
(i.e. the least fixpoint is always reached within n iterations)

$\psi_1(xy, Y) := Rxy \lor \exists z\,(Rxz \land Yzy)$ unbounded

Boundedness problem

Boundedness problem for $\mathcal{L}$

Input: $\psi(\boldsymbol{y}, Y) \in \mathcal{L}$ positive in Y

Question: is there $n \in \mathbb{N}$ s.t. for all structures $\mathfrak{A}$, $\psi^n_{\mathfrak{A}} = \psi^{n+1}_{\mathfrak{A}}$?
(i.e. the least fixpoint is always reached within n iterations)

$\psi_1(xy, Y) := Rxy \vee \exists z\,(Rxz \wedge Yzy)$ unbounded

$[\mathbf{lfp}_{Y,xy}.\psi_1] \equiv$ transitive closure of R

Boundedness problem

Boundedness problem for $\mathcal{L}$

Input: $\psi(\boldsymbol{y}, Y) \in \mathcal{L}$ positive in Y

Question: is there $n \in \mathbb{N}$ s.t. for all structures $\mathfrak{A}$, $\psi^n_{\mathfrak{A}} = \psi^{n+1}_{\mathfrak{A}}$?
(i.e. the least fixpoint is always reached within n iterations)

$\psi_1(xy, Y) := Rxy \vee \exists z\,(Rxz \wedge Yzy)$ unbounded

$[\mathbf{lfp}_{Y,xy}.\psi_1] \equiv$ transitive closure of R

$\psi_2(xy, Y) := Rxy \vee \exists z\,(Yzy)$

Boundedness problem

Boundedness problem for $\mathcal{L}$

Input: $\psi(\boldsymbol{y}, Y) \in \mathcal{L}$ positive in Y

Question: is there $n \in \mathbb{N}$ s.t. for all structures $\mathfrak{A}$, $\psi^n_{\mathfrak{A}} = \psi^{n+1}_{\mathfrak{A}}$?
(i.e. the least fixpoint is always reached within n iterations)

$\psi_1(xy, Y) := Rxy \vee \exists z\,(Rxz \wedge Yzy)$ unbounded

$[\mathbf{lfp}_{Y,xy}.\psi_1] \equiv$ transitive closure of R

$\psi_2(xy, Y) := Rxy \vee \exists z\,(Yzy)$

$[\mathbf{lfp}_{Y,xy}.\psi_2](xy) \equiv Rxy \vee \exists z\,(Rzy)$

Boundedness problem

Boundedness problem for $\mathcal{L}$

Input: $\psi(\boldsymbol{y}, Y) \in \mathcal{L}$ positive in Y

Question: is there $n \in \mathbb{N}$ s.t. for all structures $\mathfrak{A}$, $\psi^n_{\mathfrak{A}} = \psi^{n+1}_{\mathfrak{A}}$?
(i.e. the least fixpoint is always reached within n iterations)

$\psi_1(xy, Y) := Rxy \vee \exists z\,(Rxz \wedge Yzy)$ unbounded

$[\mathbf{lfp}_{Y,xy}.\psi_1] \equiv$ transitive closure of R

$\psi_2(xy, Y) := Rxy \vee \exists z\,(Yzy)$ bounded

$[\mathbf{lfp}_{Y,xy}.\psi_2](xy) \equiv Rxy \vee \exists z\,(Rzy)$

Some prior results

Boundedness is **undecidable** for

- binary predicate in positive existential FO (i.e. Datalog)
  [Hillebrand, Kanellakis, Mairson, Vardi ’95]
- monadic predicate in existential FO with inequalities
  [Gaifman, Mairson, Sagiv, Vardi ’87]
- monadic predicate in FO^2
  [Kolaitis, Otto ’98]

Some prior results

Boundedness is **undecidable** for

- binary predicate in positive existential FO (i.e. Datalog)
  [Hillebrand, Kanellakis, Mairson, Vardi '95]
- monadic predicate in existential FO with inequalities
  [Gaifman, Mairson, Sagiv, Vardi '87]
- monadic predicate in FO^2
  [Kolaitis, Otto '98]

Boundedness is **decidable** for

- monadic predicate in positive existential FO (i.e. monadic Datalog)
  [Cosmadakis, Gaifman, Kanellakis, Vardi '88]
  2EXPTIME
- monadic predicate in modal logic
  [Otto '99]
  EXPTIME
- predicates in "guarded logics"
  [Blumensath, Otto, Weyer '14]
  [Bárány, ten Cate, Otto '12]
  non-elementary upper bound

Some prior results

Boundedness is **undecidable** for

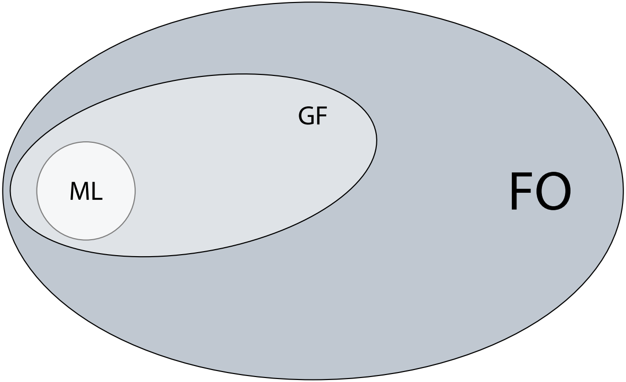
- binary predicate in positive existential FO (i.e. Datalog)
  [Hillebrand, Kanellakis, Mairson, Vardi '95]
- monadic predicate in existential FO with inequalities
  [Gaifman, Mairson, Sagiv, Vardi '87]
- monadic predicate in FO^2
  [Kolaitis, Otto '98]

Boundedness is **decidable** for

- monadic predicate in positive existential FO (i.e. monadic Datalog)
  [Cosmadakis, Gaifman, Kanellakis, Vardi '88]
  2EXPTIME
- monadic predicate in modal logic
  [Otto '99]
  EXPTIME
- predicates in "guarded logics"
  [Blumensath, Otto, Weyer '14]
  [Bárány, ten Cate, Otto '12]
  non-elementary upper bound

our contribution:
elementary upper bound
(or better)

Guarded logics

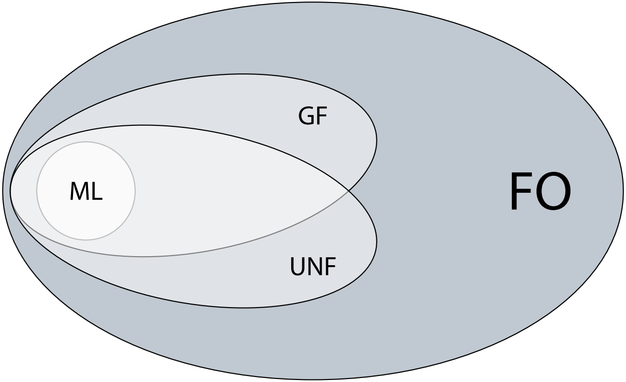


constrain quantification

$\exists \boldsymbol{x}(G(\boldsymbol{xy}) \wedge \psi(\boldsymbol{xy}))$
$\forall \boldsymbol{x}(G(\boldsymbol{xy}) \rightarrow \psi(\boldsymbol{xy}))$

[Andréka, van Benthem, Németi '95-'98]

Guarded logics



constrain
quantification

$\exists \boldsymbol{x}(G(\boldsymbol{xy}) \wedge \psi(\boldsymbol{xy}))$
$\forall \boldsymbol{x}(G(\boldsymbol{xy}) \rightarrow \psi(\boldsymbol{xy}))$

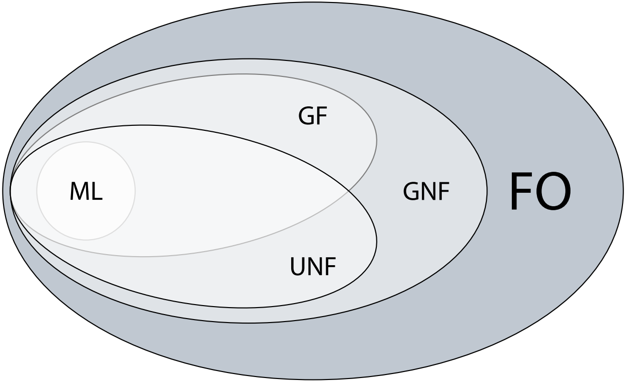
[Andréka, van Benthem, Németi ’95-’98]

constrain
negation

$\exists \boldsymbol{x}(\psi(\boldsymbol{xy}))$
$\neg\psi(x)$

[ten Cate, Segoufin ’11]

Guarded logics



constrain quantification

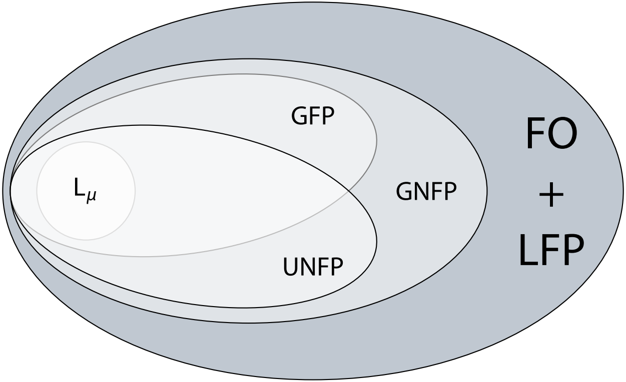
$\exists \boldsymbol{x}(G(\boldsymbol{xy}) \wedge \psi(\boldsymbol{xy}))$
$\forall \boldsymbol{x}(G(\boldsymbol{xy}) \rightarrow \psi(\boldsymbol{xy}))$

[Andréka, van Benthem, Németi '95-'98]

constrain negation

$\exists \boldsymbol{x}(\psi(\boldsymbol{xy}))$
$G(\boldsymbol{xy}) \wedge \neg\psi(\boldsymbol{xy})$

[ten Cate, Segoufin '11]
[Bárány, ten Cate, Segoufin '11]



constrain quantification

$\exists \boldsymbol{x}(G(\boldsymbol{xy}) \wedge \psi(\boldsymbol{xy}))$
$\forall \boldsymbol{x}(G(\boldsymbol{xy}) \rightarrow \psi(\boldsymbol{xy}))$

[Andréka, van Benthem, Németi '95-'98]

constrain negation

$\exists \boldsymbol{x}(\psi(\boldsymbol{xy}))$
$G(\boldsymbol{xy}) \wedge \neg\psi(\boldsymbol{xy})$

[ten Cate, Segoufin '11]
[Bárány, ten Cate, Segoufin '11]

Guarded logics

Guarded logics are expressive. For instance, GNFP captures:

- mu-calculus, even with backwards modalities;
- positive existential FO (i.e. unions of conjunctive queries);
- description logics including $\mathcal{ALC}$, $\mathcal{ALCHIO}$, $\mathcal{ELI}$;
- monadic Datalog.

Guarded logics

Guarded logics are expressive. For instance, GNFP captures:

- mu-calculus, even with backwards modalities;
- positive existential FO (i.e. unions of conjunctive queries);
- description logics including $\mathcal{ALC}$, $\mathcal{ALCHIO}$, $\mathcal{ELI}$;
- monadic Datalog.

Guarded logics have many nice model theoretic properties.

- GF, UNF, and GNF have finite models.
- GFP, UNFP, and GNFP have tree-like models
  (models of bounded tree-width).

Guarded logics

Guarded logics are expressive. For instance, GNFP captures:

- mu-calculus, even with backwards modalities;
- positive existential FO (i.e. unions of conjunctive queries);
- description logics including $\mathcal{ALC}$, $\mathcal{ALCHIO}$, $\mathcal{ELI}$;
- monadic Datalog.

Guarded logics have many nice model theoretic properties.

- GF, UNF, and GNF have finite models.
- GFP, UNFP, and GNFP have tree-like models (models of bounded tree-width).

Guarded logics have nice computational properties.

- Satisfiability is decidable, and is 2EXPTIME-complete (even EXPTIME-complete for fixed-width GFP).

Boundedness for guarded logics

(We say $\psi(\boldsymbol{x})$ is **answer-guarded** if it is of the form $G(\boldsymbol{x}) \wedge \psi'(\boldsymbol{x})$.)

Corollary to tree-like model property

For ψ in GFP or answer-guarded GNFP:
ψ is bounded over all structures iff ψ is bounded over tree-like structures.

Boundedness for guarded logics

(We say $\psi(\boldsymbol{x})$ is **answer-guarded** if it is of the form $G(\boldsymbol{x}) \wedge \psi'(\boldsymbol{x})$.)

Corollary to tree-like model property

For ψ in GFP or answer-guarded GNFP:
ψ is bounded over all structures iff ψ is bounded over tree-like structures.

⇒ amenable to techniques using tree automata

Boundedness for guarded logics

(We say $\psi(\boldsymbol{x})$ is **answer-guarded** if it is of the form $G(\boldsymbol{x}) \wedge \psi'(\boldsymbol{x})$.)

Corollary to tree-like model property

For ψ in GFP or answer-guarded GNFP:
ψ is bounded over all structures iff ψ is bounded over tree-like structures.

$\Rightarrow$ amenable to techniques using tree automata

Logic-automata connection utilized in Blumensath et al. '14
but only yields non-elementary complexity since their proof goes via MSO.

Boundedness for guarded logics

(We say $\psi(\boldsymbol{x})$ is **answer-guarded** if it is of the form $G(\boldsymbol{x}) \wedge \psi'(\boldsymbol{x})$.)

Corollary to tree-like model property

For ψ in GFP or answer-guarded GNFP:
ψ is bounded over all structures iff ψ is bounded over tree-like structures.

$\Rightarrow$ amenable to techniques using tree automata

Logic-automata connection utilized in Blumensath et al. '14
but only yields non-elementary complexity since their proof goes via MSO.

Our strategy: construct automata for boundedness problem directly.

Boundedness for guarded logics

(We say $\psi(\boldsymbol{x})$ is **answer-guarded** if it is of the form $G(\boldsymbol{x}) \wedge \psi'(\boldsymbol{x})$.)

Corollary to tree-like model property

For ψ in GFP or answer-guarded GNFP:
ψ is bounded over all structures iff ψ is bounded over tree-like structures.

$\Rightarrow$ amenable to techniques using tree automata

Logic-automata connection utilized in Blumensath et al. '14
but only yields non-elementary complexity since their proof goes via MSO.

Our strategy: construct cost automata for boundedness problem directly.

Cost automata

Cost automaton $\mathcal{A}$
classical automaton + finite set of counters with operations $\mathtt{i}$, $\mathtt{r}$, and ε

Cost automata

Cost automaton $\mathcal{A}$
classical automaton + finite set of counters with operations `i`, `r`, and ε

Semantics $[\![\mathcal{A}]\!]$: trees $\to \mathbb{N} \cup \{\infty\}$

$$[\![\mathcal{A}]\!](t) := \min \{n : \exists \text{ run } \rho \text{ of } \mathcal{A} \text{ on } t \text{ such that } \rho \text{ satisfies the acceptance condition and keeps counters below } n\}$$

Cost automata

Cost automaton $\mathcal{A}$
classical automaton + finite set of counters with operations $\mathtt{i}$, $\mathtt{r}$, and ε

Semantics $[\![\mathcal{A}]\!]$: trees $\to \mathbb{N} \cup \{\infty\}$

$$[\![\mathcal{A}]\!](t) := \min \{n : \exists \text{ run } \rho \text{ of } \mathcal{A} \text{ on } t \text{ such that } \rho \text{ satisfies the acceptance condition and keeps counters below } n\}$$

Theorem

For all $\psi \in \mathrm{GNFP}[\sigma]$, we can construct a 2-way cost automaton $\mathcal{A}_\psi$ such that

ψ is bounded
iff $\exists\, n \in \mathbb{N}$ such that $\forall$ trees t, $[\![\mathcal{A}_\psi]\!](t) \leq n$.

Boundedness for cost automata

Boundedness problem for cost automata

Input: cost automaton $\mathcal{B}$

Question: is there $n \in \mathbb{N}$ such that for all trees t, $[\![\mathcal{B}]\!](t) \leq n$?

Boundedness for cost automata

Boundedness problem for cost automata

Input: cost automaton $\mathcal{B}$

Question: is there $n \in \mathbb{N}$ such that for all trees t, $[\![\mathcal{B}]\!](t) \leq n$?

Decidability of boundedness is not known in general for cost automata over infinite trees...

Boundedness for cost automata

Boundedness problem for cost automata

Input: cost automaton $\mathcal{B}$

Question: is there $n \in \mathbb{N}$ such that for all trees t, $[\![\mathcal{B}]\!](t) \leq n$?

Decidability of boundedness is not known in general for cost automata over infinite trees...

...but we are interested in special types of cost automata:
1 counter that is only incremented or left unchanged (never reset).

Theorem

For some special types of 2-way cost automata, the boundedness problem is decidable in elementary time.

Summary of results

Theorem

Boundedness is decidable in elementary time for answer-guarded GNF and GF.

Summary of results

Theorem

Boundedness is decidable in elementary time for answer-guarded GNF and GF.

Using unpublished results of Colcombet, this can be improved to 2EXPTIME, and elementary bound can be extended to answer-guarded GNFP and GFP.

Summary of results

Theorem

Boundedness is decidable in elementary time for answer-guarded GNF and GF.

Using unpublished results of Colcombet, this can be improved to 2EXPTIME, and elementary bound can be extended to answer-guarded GNFP and GFP.

This yields elementary time algorithms for:

- deciding boundedness for some Datalog-like languages

Summary of results

Theorem

Boundedness is decidable in elementary time for answer-guarded GNF and GF.

Using unpublished results of Colcombet, this can be improved to 2EXPTIME, and elementary bound can be extended to answer-guarded GNFP and GFP.

This yields elementary time algorithms for:

- deciding boundedness for some Datalog-like languages
- deciding FO-rewritability of $[\mathbf{lfp}_{Y,\boldsymbol{y}}.\psi](\boldsymbol{y})$ for ψ in answer-guarded GNF or GF (using [Bárány, ten Cate, Otto '12])

Summary of results

Theorem

Boundedness is decidable in elementary time for answer-guarded GNF and GF.

Using unpublished results of Colcombet, this can be improved to 2EXPTIME, and elementary bound can be extended to answer-guarded GNFP and GFP.

This yields elementary time algorithms for:

- deciding boundedness for some Datalog-like languages
- deciding FO-rewritability of $[\mathbf{lfp}_{Y,\boldsymbol{y}}.\psi](\boldsymbol{y})$ for ψ in answer-guarded GNF or GF (using [Bárány, ten Cate, Otto '12])
- deciding FO-rewritability of CQs over guarded and frontier-guarded TGDs (using [Bárány, Benedikt, ten Cate '13])

Boundedness is decidable
in elementary time for guarded logics.

Contributions

- General translation from GNFP to automata that can be used for satisfiability testing and boundedness questions.
- Finer analysis of complexity of some cost automata constructions.

Bringing cost capabilities to guarded logics

Syntax of cGNFP$[\sigma]$

$\varphi ::= \cdots \mid [\mathbf{lfp}^{N}_{Y,\boldsymbol{y}}.G(\boldsymbol{y}) \wedge \varphi(\boldsymbol{y}, Y, \boldsymbol{Z})](\boldsymbol{x})$ for φ positive in Y

where $\mathbf{lfp}^{N}$ operators only appear positively in the formula.

Bringing cost capabilities to guarded logics

Syntax of cGNFP$[\sigma]$

φ ::= $\cdots$ | $[\mathbf{lfp}^{N}_{Y,\boldsymbol{y}}.G(\boldsymbol{y}) \wedge \varphi(\boldsymbol{y}, Y, \boldsymbol{Z})](\boldsymbol{x})$ for φ positive in Y

where $\mathbf{lfp}^{N}$ operators only appear positively in the formula.

Semantics $[\![\varphi]\!] : \sigma$-structures $\to \mathbb{N} \cup \{\infty\}$

$$[\![\varphi]\!](\mathfrak{A}) := \min \left\{ n \in \mathbb{N} : \mathfrak{A} \text{ satisfies } \varphi \text{ when } [\mathbf{lfp}^{N}_{Y,\boldsymbol{y}}.\psi] \text{ replaced by } \psi^{n} \right\}$$

where $\psi^{0} := \bot$ and $\psi^{n} := \psi[\psi^{n-1}/Y]$

Bringing cost capabilities to guarded logics

Syntax of cGNFP$[\sigma]$

$\varphi \ ::= \ \cdots \ | \ [\mathbf{lfp}^{N}_{Y,\boldsymbol{y}}.G(\boldsymbol{y}) \wedge \varphi(\boldsymbol{y}, Y, \boldsymbol{Z})](\boldsymbol{x})$ for φ positive in Y

where $\mathbf{lfp}^{N}$ operators only appear positively in the formula.

Semantics $[\![\varphi]\!] : \sigma$-structures $\to \mathbb{N} \cup \{\infty\}$

$[\![\varphi]\!](\mathfrak{A}) := \min \left\{ n \in \mathbb{N} : \mathfrak{A} \text{ satisfies } \varphi \text{ when } [\mathbf{lfp}^{N}_{Y,\boldsymbol{y}}.\psi] \text{ replaced by } \psi^{n} \right\}$

where $\psi^{0} := \bot$ and $\psi^{n} := \psi[\psi^{n-1}/Y]$

Example

$$\varphi(y) := [\mathbf{lfp}^{N}_{Y,y}.Sy \vee \exists z(Ryz \wedge Yz)](y)$$

$[\![\varphi]\!](\mathfrak{A}, a) :=$ minimum length of R-chain to reach S from a