Edinburgh Research Explorer

Bounded Implication for Existential Rules (Extended Abstract)

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Proceedings of the 29th International Workshop on Description Logics

Boundedness for Existential Rules

Existential Rules, Ontology-based Data Access, Boundedness

No Institute Given

Abstract. The property of boundedness in Datalog formalizes whether a set of rules can be equivalently expressed by a non-recursive set of rules. Existential rules extend Datalog to the presence of existential variables in rule heads. In this paper, we introduce and study notions of boundedness for existential rules. We provide a notion of *weak boundedness* and a notion of *strong boundedness* for a rule set, and show that they correspond, respectively, to the notions of first-order rewritability of *atomic queries* and first-order rewritability of *conjunctive queries* over the set. While weak and strong boundedness are in general not equivalent, we show that, for the notable subclasses of Datalog, single-head binary rules, and frontier-guarded rules, the two notions coincide. We finally prove that establishing the boundedness of a rule set is undecidable for single-head binary rules, and is decidable for frontier-guarded rules. These results may have a strong impact on the optimization of reasoning over existential rules.

1   Introduction

The problem   This paper deals with the property of *boundedness* in rule languages. Boundedness is an important notion that formalizes the fact that a rule set Σ can be unfolded into a finite set Σ' of acyclic (i.e., non-recursive) rules such that Σ and Σ' are equivalent on every database: it is therefore a crucial property for optimizing the processing of rules. Such a property has been extensively studied, especially for the Datalog rule language [13,5], and, recently, for Answer Set Programming [22].

In Datalog, the boundedness of a program P can be defined as the existence of an integer k such that, for every database D, the number of iterated applications (in a forward chaining manner) of P to D that are necessary to compute the minimal model of P and D is bounded by k. This definition of boundedness is equivalent to the existence of a finite, non-recursive program that is equivalent to P. Also, it is well-known that a Datalog query is bounded if and only if it is equivalent to a first-order sentence [1,17].

More recently, rule-based languages have been used in the context of *ontology-based data access* [21], which studies the problem of accessing multiple data sources through an ontology that constitutes a virtual, shared and intensional view of the data. In this framework, the main focus is on the problem of answering conjunctive queries over an ontology expressed by a set of rules. Here, one of the most studied properties is the *first-order rewritability of conjunctive queries (CQFO-rewritability)* over an ontology, which corresponds to the above mentioned first-order expressibility in Datalog: an ontology $\mathcal{O}$ is CQFO-rewritable if every conjunctive query q over the ontology can be equivalently rewritten into a first-order query q', i.e., q' is such that, for every database instance $\mathcal{D}$, the evaluation of q over $\mathcal{O}$ and $\mathcal{D}$ coincides with the evaluation of q' over $\mathcal{D}$. Notably, in the case when the ontology is expressed as a set P of Datalog rules, the CQFO-rewritability of P and the boundedness of P are equivalent properties.

Existential rules, which extend Datalog rules to the presence of existentially quantified variables and multiple atoms in rule heads, have been proposed and studied in the last years as a specification language for ontology-based data access [7,2,19]. Several recent studies have focused on the first-order rewritability property for existential rules (e.g., [9,2,11]). On the other hand, the notion of boundedness for existential rules has not been deeply investigated. To our knowledge, one of the most relevant recent approaches to this problem is presented in [2], where the notion of *acyclic graph of rule dependencies (aGRD)* is defined, which corresponds to a form of boundedness for existential rules. Instead, we start our study from a notion of boundedness for existential rules that generalizes the definition of boundedness provided for Datalog to existential rules in a simpler way: we call such a notion *strict boundedness*. However, it can be immediately verified that, for arbitrary sets of existential rules, the above notion of boundedness is much stronger than the first-order rewritability of conjunctive queries. That is, while strict boundedness of a rule set implies its CQFO-rewritability, there exist rule sets that are CQFO-rewritable but are not strictly bounded. Notice that the same property holds for the above mentioned notion of aGRD.

The main goal of this paper is to answer the following question: *is it possible to generalize the notion of boundedness for Datalog to existential rules, in such a way that the correspondence with the notion of first-order rewritability of conjunctive queries is preserved?* Actually, from the forward chaining perspective, such a generalization has been provided by the *bounded derivation depth property (BDDP)* of the *chase* of existential rules [7,14]. However, we would like to characterize this property in terms of equivalent representations of the set of rules, and see how the alternative notion of boundedness as existence of a finite and non-recursive equivalent rule set has to be extended to capture first-order rewritable rules.

Our contribution Our contribution can be summarized as follows.

First, after examining the properties and the limits of the notion of strict boundedness, we define a notion of boundedness for existential rules that weakens strict boundedness by giving up the acyclicity condition: we call such a notion *weak boundedness*. It is based on the idea of looking for a finite representation of all the *single-head* rules (i.e., rules with one atom in the head) that are implied by the initial rule set, and on a notion of equivalence between rule sets that is different from the standard logical equivalence. It turns out (Theorem 3) that such a definition does not have the expected correspondence with the first-order rewritability of conjunctive queries. However, we prove (Theorem 6) its correspondence with the weaker notion of first-order rewritability of *atomic queries*, i.e., conjunctive queries consisting of a single atom.

We thus define a second notion of *strong boundedness* for existential rules. Roughly speaking, such a notion is obtained from weak boundedness by discarding the restriction to single-head rules in the deductive closure of the rule set, and by considering projections of such a deductive closure. We prove (Theorem 4) that this notion has the desired correspondence with the first-order rewritability of conjunctive queries. In turn, this implies the correspondence between strong boundedness and the BDDP [14].

Then, we prove (Corollary 1) that both weak and strong boundedness are proper generalizations of the notion of boundedness for Datalog, that is: for Datalog rules, strict, weak and strong boundedness coincide. Moreover, we show that the equivalence between weak and strong boundedness is actually not limited to Datalog, and holds for

two other broad classes of existential rules: *single-head binary* rules (Theorem 7), that is, single-head rules over relations of arity not greater than 2, and *frontier-guarded* [2] rules (Theorem 8). We believe that the equivalence between weak and strong boundedness is a very important property for a set of existential rules. In particular, the above correspondences could be exploited in the optimization of query answering over ontologies expressed by rule sets belonging to the above classes.

Finally, we show (Theorem 9) that checking strong boundedness (or, equivalently, weak boundedness) is undecidable for single-head binary rules, while it is decidable for frontier-guarded rules. These results complement the well-known undecidability of (strict) boundedness for Datalog [13].

Our approach to the study of boundedness for existential rules is inspired by the work in query rewriting for existential rules [2,18,8,15]. In particular, we extend the techniques presented in [2,18] to address the problem of computing an unfolding of a set of existential rules, and the problem of defining an appropriate notion of redundancy between rules. We also use results from [3] to derive the decidability result of weak and strong boundedness for frontier-guarded rules.

2 Preliminaries

We start by recalling the framework of existential rules (see, e.g., [9,2] for more details).

We start from three pairwise disjoint alphabets: $\mathcal{R}$, a countably infinite set of relation names; Δ, a countably infinite set of constants; $\mathcal{X}$, a countably infinite set of variables. We denote by $\bar{a}, \bar{b}, \bar{c}, \ldots$ a tuple, i.e., a (possibly empty) sequence of constants in Δ, and by $\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}, \ldots$ a (possibly empty) sequence of variables in $\mathcal{X}$.

A *relational schema* $\mathbf{S}$ is a finite non-empty set $\{R_1, \ldots, R_n\}$ of relations, where each $R_i \in \mathcal{R}$ and has an associated non-negative integer called the *arity* of the relation and denoted by $Arity(R)$. An *atom* γ over a schema $\mathbf{S}$ is an expression of the form $R(t_1, \ldots, t_k)$, where $R \in \mathbf{S}$ is the relation name, k is the arity of R, and every t_i is called a *term*, and can be either a constant in Δ or a variable in $\mathcal{X}$. If γ has no occurrences of variables we call it a *fact*. $Const(\gamma)$ denotes the set of constants occurring in γ; an analogous notation is used to denote constants occurring in sets of atoms, rules and sets of rules. A *relational instance* $\mathcal{B}$ over $\mathbf{S}$ is a possibly infinite set of facts over $\mathbf{S}$, while a *database* $\mathcal{D}$ over $\mathbf{S}$ is a finite relational instance over $\mathbf{S}$.

An *existential rule*, or simply *rule*, σ over a schema $\mathbf{S}$ is an expression of the form $\forall \bar{\mathbf{x}} \forall \bar{\mathbf{y}} (\Phi(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{a}) \rightarrow \exists \bar{\mathbf{z}} \Psi(\bar{\mathbf{x}}, \bar{\mathbf{z}}, \bar{b}))$, where $\Phi(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{a})$ and $\Psi(\bar{\mathbf{x}}, \bar{\mathbf{z}}, \bar{b})$ are conjunctions of atoms over $\mathbf{S}$ called, respectively, the *body* of σ ($body(\sigma)$) and the *head* of σ ($head(\sigma)$). We use a simplified notation in which we omit the universal quantifiers and represent constants by uppercase letters (e.g., $p(x,y,z), r(y,A) \rightarrow \exists w\ s(x,w)$).

Given a rule σ, we call $\bar{\mathbf{x}}$ the *frontier variables of* σ ($\mathcal{F}(\sigma)$), i.e., the variables occurring both in the head and in the body of σ, and $\bar{\mathbf{z}}$ the *existential head variables* of σ ($\mathcal{E}_H(\sigma)$), i.e., the variables that occur only in the head of σ. We call arity of a rule σ ($Arity(\sigma)$) the number of frontier variables of σ.

Let σ be a rule, $\mathcal{K}$ be a set of constants, and $\eta : \mathcal{F}(\sigma) \rightarrow \mathcal{F}(\sigma) \cup \mathcal{K}$ be a function. The rule $\sigma_s = \eta(\sigma)$ is a called a *specialization* of σ w.r.t. $\mathcal{K}$.

A *single-head rule* rule is a rule with a singleton head atom. A *Datalog* rule is a single-head rule without existential head variables. A *binary rule* is a rule whose atoms

have arity less than or equal to two. A *ternary rule* is a rule whose atoms have arity less than or equal to three. A *frontier-guarded rule* [2] is a rule whose frontier variables appear all together in at least one body atom. We generally focus on sets of rules, and *(sub)classes* (i.e., sets of sets) of rules.

A *first-order query* (FO query) is a FO formula $\varphi(x_1, \ldots, x_n)$, where $x_1, \ldots, x_n$ are called *free* variables, i.e., variables that do not occur in the scope of any quantifier. A *conjunctive query (CQ)* over a schema $\mathbf{S}$ is a special single-head rule over $\mathbf{S}$ of the form $\Phi(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{a}) \rightarrow q(\bar{\mathbf{x}})$, where q is a special relation that occurs only in the head of this rule. We call arity of a CQ q ($Arity(q)$) the arity of the head relation of q. An *atomic query* (AQ) is a CQ with a singleton body atom. A *boolean conjunctive query* (BCQ) is a CQ of zero arity.

An interpretation $I = (\Delta^I, \cdot^I)$ is a pair constituted by a non-empty interpretation domain Δ^I and an interpretation function $\cdot^I$, that assigns to each relation $R \in \mathcal{R}$ of arity k, a k-ary relations over Δ^I, and assigns to each constant c an element $c^I \in \Delta^I$ under the unique name assumption (UNA), i.e., such that $c_1^I \neq c_2^I$ for $c_1 \neq c_2$. Obviously, every relational instance $\mathcal{B}$ identifies an interpretation $I^{\mathcal{B}}$ such that there exists an isomorphism between $\mathcal{B}$ and $I^{\mathcal{B}}$.

Given a rule σ and a relational instance $\mathcal{B}$, we say that $\mathcal{B}$ satisfies σ ($\mathcal{B} \models \sigma$) if $I^{\mathcal{B}}$ is a model for σ. If σ is satisfied by every relational instance, we say that σ is *trivial*.

A set Σ of rules *implies* a rule σ (denoted by $\Sigma \models \sigma$) if σ is satisfied in every relational instance that satisfies all the rules in Σ. We say that the set of rule Σ implies the set of rules Σ' ($\Sigma \models \Sigma'$) if $\Sigma \models \sigma'$ for every $\sigma' \in \Sigma'$, and that Σ and Σ' are *equivalent* if both $\Sigma \models \Sigma'$ and $\Sigma' \models \Sigma$. Given a database $\mathcal{D}$ over a schema $\mathbf{S}$, and a set Σ of rules over $\mathbf{S}$, we say that a relational instance $\mathcal{B}$ over $\mathbf{S}$ *satisfies* $(\Sigma, \mathcal{D})$, and we denote it by $\mathcal{B} \models (\Sigma, \mathcal{D})$ if (i) $\mathcal{B} \supseteq \mathcal{D}$, and (ii) $\mathcal{B}$ satisfies every rule $\sigma \in \Sigma$. Moreover, we denote by $Sem(\Sigma, \mathcal{D})$ the (possibly infinite) set of all the relational instances $\mathcal{B}$ over $\mathbf{S}$ such that $\mathcal{B}$ satisfies $(\Sigma, \mathcal{D})$.

Let q be a FO query. We denote by $Ans(q, \mathcal{B})$ the answers to q over $I^{\mathcal{B}}$, i.e., the set of tuples of constants $\bar{c}$ such that $I^{\mathcal{B}}$ satisfies $q(\bar{c})$, where $q(\bar{c})$ is the FO sentence obtained from q by replacing its free variables with the constants $\bar{c}$.

Let q be a CQ. We define the *certain answers* to q over Σ and $\mathcal{D}$, denoted by $Cert(q, \Sigma, \mathcal{D})$, as the set of tuples of constants $\bar{c}$ such that $\bar{c} \in \bigcap_{\mathcal{B} \in Sem(\Sigma, \mathcal{D})} Ans(q, \mathcal{B})$.

The notion of certain answers underlies some key concepts that are crucial for this work. In particular, we recall the definitions of *perfect rewriting* and *FO-rewritability*.

Let Σ be a set of rules, and let q be a query over Σ. A query q' is a *perfect rewriting* of q w.r.t. Σ if, for every database $\mathcal{D}$, $Cert(q, \Sigma, \mathcal{D}) = Ans(q', \mathcal{D})$. Moreover, q is *FO-rewritable* w.r.t. Σ if there exists a FO query q' that is a perfect rewriting of q w.r.t. Σ. Σ is *atomic FO-rewritable* (AFO-rewritable) if every AQ q is FO-rewritable w.r.t. Σ. Σ is *CQFO-rewritable* if every CQ q is FO-rewritable w.r.t. Σ.

3   Boundedness for Existential Rules

We start by providing a first notion of boundedness for existential rules that tries to naturally extend the idea of boundedness as existence of a finite unfolding of the rule set. Let the relation graph of Σ, denoted by $G(\Sigma)$, be the directed graph whose nodes are the relations occurring in Σ, and such that there is an edge $\langle p, p' \rangle$ in the graph iff

there exists a rule $\sigma \in \Sigma$ such that p occurs in $body(\sigma)$ and p' occurs in $head(\sigma)$. We say that Σ is *acyclic* if $G(\Sigma)$ is acyclic.

Definition 1 (Strict boundedness). *A set Σ of rules is* strictly bounded *if it is equivalent to a finite and acyclic set of rules.*

In the case when Σ is a Datalog rule set, this notion corresponds to the well-known notion of boundedness for Datalog [13]. For existential rules, the above definition is too strong to preserve the desired correspondence with the notion of first-order rewritability of queries. Indeed, it is immediate to verify that strict boundedness implies CQFO-rewritability ([2], while CQFO-rewritability does not imply strict boundedness: for instance, the rule set $\{R(x,y) \rightarrow \exists z\ R(y,z)\}$ is CQFO-rewritable, but is not strictly bounded. In particular, the strict boundedness of Σ implies that Σ is a *finite expansion set* [2], i.e., that for every database $\mathcal{D}$, $(\Sigma, \mathcal{D})$ has a finite canonical model, while CQFO-rewritability does not imply such a property.

Therefore, we have to look for weaker notions of boundedness for existential rules. The above example suggests that we have to discard the acyclicity condition from the above definition, replacing it with a different notion of unfoldability of the rules.

Our first attempt is based on the idea of focusing on the single-head rules (since Datalog rules are single-head rules) that are logical consequences of the initial set of rules Σ, and to check whether there exists a finite representation of such logical consequences according to a new notion of equivalence between rule sets that is stronger than the classical one used in Definition 1. We thus start by defining a notion of deductive closure of a set of rules.

Definition 2 (Single-head closure of a set of rules). *Let Σ be a set of rules over a signature **S**. We define the SH-closure of Σ as the set $\Sigma^{\star s} = \{\sigma \mid \sigma$ is a single-head rule over **S** and $Const(\Sigma)$, and $\Sigma \models \sigma\}$.*

Then, we define a notion of *redundancy* between two rules.

Definition 3 (Redundancy of a rule). *Given two rules $\sigma : \Phi(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{a}) \rightarrow \exists\bar{\mathbf{z}}\ \Psi(\bar{\mathbf{x}}, \bar{\mathbf{z}}, \bar{b})$, and $\sigma' : \Phi'(\bar{\mathbf{x}}', \bar{\mathbf{y}}', \bar{c}) \rightarrow \exists\bar{\mathbf{z}}'\ \Psi(\bar{\mathbf{x}}', \bar{\mathbf{z}}', \bar{d})$, we say that σ is redundant with respect to σ' if there exists a specialization $\eta(\sigma') = \sigma'_s$ of σ' w.r.t. $Const(\sigma)$, and a bijective function $\epsilon : \mathcal{F}(\sigma'_s) \rightarrow \mathcal{F}(\sigma)$ such that the following FO sentences are valid:*

$$\forall\bar{\mathbf{x}}\forall\bar{\mathbf{y}}\ body(\sigma)(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{a}) \rightarrow \exists\bar{\mathbf{y}}'\ \epsilon(body(\sigma'_s))(\bar{\mathbf{x}}, \bar{\mathbf{y}}', \bar{e})$$
$$\forall\bar{\mathbf{x}}'\forall\bar{\mathbf{z}}'\ head(\sigma'_s)(\bar{\mathbf{x}}', \bar{\mathbf{z}}', \bar{d}) \rightarrow \exists\bar{\mathbf{z}}\ \epsilon^-(head(\sigma))(\bar{\mathbf{x}}', \bar{\mathbf{z}}, \bar{f})$$

where $\bar{e} = \bar{c} \cup \bar{a} \cup \bar{b}$ and $\bar{f} = \bar{d} \cup \bar{a} \cup \bar{b}$.

Example 1. Let $\sigma_1 : R(x,y) \rightarrow \exists z\ S(x,y), Q(z,y)$, and $\sigma_2 : R(x',x'), P(x',z') \rightarrow S(x',x')$. It is easy to verify that σ_2 is redundant w.r.t. σ_1. Indeed, let $\sigma_{1_s} = \eta(\sigma_1) : R(x,x) \rightarrow \exists z\ S(x,x), Q(z,x)$ be a specialization of σ_1, where $\eta = \{x \rightarrow x, y \rightarrow x\}$. Moreover, let $\epsilon = \{x \rightarrow x'\}$. Then we have that the following FO sentences are both valid:

$$\forall x'\forall z'\ R(x',x'), P(x',z') \rightarrow R(x',x')$$
$$\forall x, \forall z\ S(x,x), Q(z,x) \rightarrow S(x,x)$$

We remark that redundancy between two rules corresponds to a weakened notion of logical implication between such rules. Indeed, if σ is redundant with respect to σ' then $\{\sigma'\} \models \sigma$, while in general the converse does not hold. As a simple example, consider $\sigma : R(x,y), R(y,z), R(z,w) \to R(x,w)$ and $\sigma' : R(x,y), R(y,z) \to R(x,z)$: in this case, $\{\sigma'\} \models \sigma$ but σ is not redundant w.r.t. σ'. Notice also that the implication between two rules is an undecidable problem ([2], Theorem 8), while it can be easily verified that checking redundancy between two rules is NP-complete.

The notion of redundancy allows us to define the desired stronger notion of equivalence. Given two sets of rules Σ, Σ', we say that Σ' *R-entails* Σ if, for each non-trivial rule $\sigma \in \Sigma$ there exists a rule $\sigma' \in \Sigma'$ such that σ is redundant w.r.t. σ'. Moreover, we say that Σ and Σ' are *R-equivalent* if both Σ' R-entails Σ and Σ R-entails Σ'.

Definition 4 (Weak boundedness). *A set Σ of rules is* weakly bounded *if $\Sigma^{\star s}$ is R-equivalent to a finite set of rules.*

Let Σ be a set of rules. A *cover* of Σ is a minimal subset $\Sigma_c \subseteq \Sigma$ such that for each non-trivial rule $\sigma \in \Sigma$ there exists a rule $\sigma' \in \Sigma_c$ such that σ is redundant w.r.t. σ'. It is immediate to verify that Σ is weakly bounded iff there exists a finite cover of $\Sigma^{\star s}$.

We now focus on the properties of weak boundedness. We first prove that the strict boundedness implies the weak one.

Theorem 1. *Let Σ be a rule set. If Σ is strictly-bounded, then Σ is weakly bounded.*

Proof sketch. Suppose Σ is strictly bounded. Then, there exists a finite and acyclic rule set Σ' that is equivalent to Σ. Moreover, from Definition 4 it follows that Σ is weakly bounded if and only if Σ' is weakly bounded. We now prove that Σ' is weakly bounded. To do so, we make use of a structure, called the *SH-forest* of Σ' and obtained by applying a special form of resolution between the rules of Σ' in a breadth-first fashion and by excluding, at each level of the expansion, the rules that are redundant w.r.t. the rules belonging to the structure. Such an expansion technique for sets of rules is inspired by (and is an extension of) the conjunctive query rewriting technique for existential rules presented in [18]. The key properties of the SH-forest of a set of rules are: (1) it represents a superset of the cover of the SH-closure of the original set; (2) it is finite if and only if the cover is finite. It can easily be shown that, since Σ' is finite and acyclic, the SH-forest of Σ' is finite as well, which by the above property (1) implies the thesis. $\quad\square$

On the other hand, it is immediate to verify that, for arbitrary set of rules, weak boundedness does not always imply strict boundedness: e.g., the rule set $\{R(x,y) \to \exists z\, R(y,z)\}$ mentioned above is weakly bounded but not strictly bounded.

Then, we prove that, for Datalog rules, strict and weak boundedness coincide (and hence that weak boundedness is a proper generalization of the notion of boundedness for Datalog).

Theorem 2. *Let Σ be a Datalog rule set. Σ is strictly-bounded iff Σ is weakly bounded.*

Proof sketch. One direction of the theorem follows from Theorem 1. For the other direction, suppose Σ is not strictly-bounded. This implies that there exists an infinite set of acyclic rules Σ' such that $\Sigma' \subseteq \Sigma^{\star s}$ and for every pair of rules σ, σ' in Σ', σ is not redundant w.r.t. σ'. Since Σ' is acyclic, there exists a numbering N of the relations

occurring in Σ' such that, for every $\sigma \in \Sigma'$ and for every relation R occurring in the body of σ'', $N(R) < N(S)$, where S is the relation occurring in the head of σ''. Now, it is easy to see that, from the definition of redundancy, for every cyclic rule σ'' (i.e., for every rule having at least a relation R occurring in the body such that $N(R) \geq N(S)$ where S is the relation occurring in the head of σ'') and for every rule $\sigma \in \Sigma'$, σ is not redundant w.r.t. σ''. This implies that Σ' must belong to every set of rules that is R-equivalent to $\Sigma^{\star s}$, which in turn implies the thesis. $\qquad\square$

However, it turns out that weak boundedness is not equivalent to CQFO-rewritability.

Theorem 3. *Weak boundedness does not imply CQFO-rewritability for both: (i) single-head ternary set of rules; and (ii) binary set of rules.*

Proof sketch. (i) We prove the statement by exhibiting a counterexample. Let Σ be the following set of rules:

$$\sigma_1 : T_1(y, z_1, z_2), T_2(z, z_1, z_2), P(x, z_1) \to R(x, y, z)$$
$$\sigma_2 : T_1(y, z_1, z_2), T_2(z, z_1, z_2), P(x, z_1) \to S(x, y, z)$$
$$\sigma_3 : R(x, z_1, z_2), P'(z_1, z_2, y) \to \exists z \ T_1(z, x, y)$$
$$\sigma_4 : S(x, z_1, z_2), P'(z_1, z_2, y) \to \exists z \ T_2(z, x, y)$$
$$\sigma_5 : R(z_1, x, y) \to \exists z \ P'(x, y, z)$$
$$\sigma_6 : S(z_1, x, y) \to \exists z \ P'(x, y, z)$$
$$\sigma_7 : P(x, z_1) \to \exists z, w \ R(x, z, w)$$
$$\sigma_8 : P(x, z_1) \to \exists z, w \ S(x, z, w)$$

It is possible to verify that Σ is weakly bounded, i.e., that the there exists a finite cover of the SH-closure of Σ. For the above Σ, it can be shown that its SH-forest (see the proof of Theorem 1) is finite; consequently, Σ is weakly bounded. On the other hand, we show that there exists a CQ $q : R(x, y, z), S(x, y, z) \to Q(x)$ that is not FO-rewritable w.r.t. Σ, which proves that the set is not CQFO-rewritable.

(ii) If we allow for multiple-head binary rules, then every arbitrary set of existential rules can be transformed into a set of binary rules through the well-known reification technique (see, e.g., [10]), which represents n-ary relations through auxiliary binary relations. For instance, the rule $\{R(x, y, z), S(y, v) \to T(x, y, v)\}$ can be transformed by reification into the rule

$$\{R_1(w, x), R_2(w, y), R_3(w, z), S_1(w', y), S_2(w', v) \to \exists w'' \ T_1(w'', x), T_2(w'', y), T_3(w'', v)\}$$

Now, it is easy to verify that the reification Σ_r of a rule set Σ over a schema $\mathbf{S}$ is such that, for every query q over $\mathbf{S}$, if q is not FO-rewritable w.r.t. Σ then q_r (i.e, the reification of q) is not FO-rewritable w.r.t. Σ_r. Therefore, if we now apply the reification to the rule set Σ defined at the above point (i), obtaining the set Σ_r, it follows that Σ_r is not CQFO-rewritable. Moreover, by extending the proof of point (i), it is possible to show that the SH-forest of Σ_r is finite, which implies that Σ_r is weakly bounded. $\qquad\square$

The above theorem shows that, while strict boundedness is stronger than CQFO-rewritability, weak boundedness is weaker than CQFO-rewritability. We thus now look for a notion of boundedness that lies in the middle between the two previous notions. To do so, we need to discard the restriction to single-head rules, i.e., we have to consider the generalized deductive closure of Σ.

Definition 5 (Closure of a set of rules). *Let Σ be a set of rules over a signature* **S**. *We call* closure of Σ *the set* $\Sigma^{\star} = \{\sigma \mid \sigma$ *is a rule over* **S** *and* $Const(\Sigma)$, *and* $\Sigma \models \sigma\}$.

However, extending Definition 4 based on the deductive closure $\Sigma^{\star}$ is not an easy task: if we just replace $\Sigma^{\star s}$ with $\Sigma^{\star}$ in Definition 4, we end up with a meaningless notion. Indeed, it is immediate to verify that even one of the simplest non-trivial (and CQFO-rewritable) rule sets, i.e., $\Sigma = \{r(x, y) \rightarrow s(x, y)\}$ is such that $\Sigma^{\star}$ does not have any finite R-equivalent set (observe that the infinite set of rules

$$\{R(x_0, x_1), \dots R(x_{i-1,x_i}) \rightarrow S(x_0, x_1), \dots S(x_{i-1,x_i}) \mid i \geq 1\}$$

is contained in $\Sigma^{\star}$, and no such rule is redundant with respect to another rule in the set).

To overcome the problem, we introduce a notion of *projection* of a set of rules with respect to a given rule, based on a notion of compatibility between two rule heads.

Let σ, σ' be two rules. We say that σ' is *head-unifiable* w.r.t. σ if there exists a homomorphism $\mu : \mathcal{F}(\sigma) \rightarrow \mathcal{F}(\sigma') \cup Const(\sigma')$ and an isomorphism $\epsilon : \mathcal{E}_H(\sigma) \rightarrow \mathcal{E}_H(\sigma')$ such that $head(\mu(\epsilon(\sigma))) = head(\sigma')$.

Example 2. Let $\sigma : R(x, y) \rightarrow \exists z\ S(x, z)$, and let $\sigma' : Q(C, x), P(x, y) \rightarrow \exists w\ S(C, w)$. Now, σ' is head-unifiable with σ, because there exists a homomorphism $\mu = \{x \rightarrow C\}$ and a isomorphism $\epsilon = \{z \leftrightarrow w\}$ between $\mathcal{E}_H(\sigma)$ and $\mathcal{E}_H(\sigma')$ such that $head(\mu(\epsilon(\sigma)) = head(\sigma')$. On the contrary, $\sigma'' : Q(C, x), P(x, y) \rightarrow \exists w\ S(w, C)$ is not head-unifiable with σ, because $x \in \mathcal{F}(\sigma)$, while $w \in \mathcal{E}_H(\sigma'')$.

Let Σ be a set of rules, and let σ be a rule. We define the *projection* of Σ with respect to σ, as the set $\Pi_\sigma(\Sigma) = \{\sigma' \in \Sigma \mid \sigma'$ is head-unifiable with $\sigma\}$.

We are now ready to define a new notion of boundedness by replacing $\Sigma^{\star s}$, in Definition 4, with all the possible projections of $\Sigma^{\star}$.

Definition 6 (Strong boundedness). *A set Σ of rules over a schema* **S** *is* strongly bounded *if, for each rule σ over* **S**, $\Pi_\sigma(\Sigma^{\star})$ *is R-equivalent to a finite set of rules.*

We now prove the central property of strong boundedness, i.e., the desired equivalence with CQFO-rewritability.

Theorem 4. *Let Σ be a set of rules. Σ is strongly bounded if and only if Σ is CQFO-rewritable.*

Proof sketch. First, we show that if there exists a CQ q that is not FO-rewritable w.r.t. Σ, then Σ can not be strongly bounded. We use again the notion of SH-forest, and, in particular, we show that the SH-forest of the set of rules $\Sigma \cup \{q\}$ has an infinite branch, which, in turn, implies that $\Pi_q(\Sigma^{\star s})$ has an infinite cover, and contradicts the assumption that Σ is strongly bounded. For the other direction, we show that if Σ is not strongly bounded, then the CQ $q(\sigma)$, defined as $head(\sigma) \rightarrow q(\mathcal{F}(\sigma))$, can not be CQFO-rewritable, where σ is a rule such that $\Pi_\sigma(\Sigma^{\star s})$ has an infinite cover. In particular, we show that Σ is strongly bounded if and only if $\Sigma \cup \{q(\sigma)\}$ is weakly bounded, and that the SH-forest of $\Sigma \cup \{q(\sigma)\}$ has an infinite branch, thus the thesis follows. $\square$

Then, from Definition 6, it immediately follows that strong boundedness implies weak boundedness. Moreover, the following property is an immediate consequence of the previous theorem and of the fact that strict boundedness implies CQFO-rewritability:

Theorem 5. *Let Σ be a set of rules. If Σ is strictly bounded then Σ is strongly bounded.*

This theorem and Theorem 2) imply that strong boundedness is a proper generalization of the notion of boundedness for Datalog. Consequently:

Corollary 1. *For the Datalog class of rules, strict boundedness, strong boundedness and weak boundedness coincide.*

Of course, from Theorem 3 and Theorem 4 it follows that, in general, weak boundedness does not imply strong boundedness.

Finally, we prove the equivalence between the notion of weak boundedness and the notion of AFO-rewritability.

Theorem 6. *Let Σ be a set of rules. Σ is weakly bounded if and only if Σ is AFO-rewritable.*

Proof sketch. The proof is similar to the proof of Theorem 4: first, we show that if there exists an AQ q that is not FO-rewritable w.r.t. Σ, then Σ can not be weakly bounded. Towards this aim, we refer to the SH-forest of a set of rules mentioned in the previous proofs, and we use a property of the forest that guarantees that if a set of rules is weakly bounded, then the set $\Sigma \cup \{\sigma\}$ is still weakly bounded if σ is a linear rule, i.e., a rule with a singleton body atom. Notice that an AQ is a special case of linear rule. In the other direction, we show that if Σ is not weakly bounded, i.e., the above mentioned SH-forest has an infinite branch, then the AQ $q(\sigma)$, defined as $head(\sigma) \to q(\mathcal{F}(\sigma))$, can not be AFO-rewritable, where the rule σ is the source of such a branch and q is a relation not occurring in Σ. More precisely, we show that Σ is weakly bounded if and only if $\Sigma \cup \{q(\sigma)\}$ is weakly bounded, thus the thesis follows. $\square$

4 Weak vs. Strong Boundedness

Theorem 3 shows that, in general, the notions of weak and strong boundedness are not equivalent. However, in this section we are interested in finding subclasses of rules for which the two notions coincide. We believe that identifying notable classes of rules for which the two notions of boundedness are equivalent would be very important. For instance, this property could be exploited in the optimization of query answering over existential rules. Indeed, the above equivalence corresponds to the equivalence between AFO-rewritability and CQFO-rewritability for a set of rules. A possible interpretation of this result from the query processing perspective is that, for such a rule set, the join of atomic expressions (as is done in conjunctive queries) can not produce non-FO-rewritable queries: this strongly suggests that such joins (and hence, conjunctive queries) can be processed in an easier way than in the general case. This aspect, however, is outside the scope of the present paper (see also the conclusions).

The first subclass we identify for which strong and weak boundedness coincide is the class of single-head binary rules.

Theorem 7. *Let Σ be a set of single-head binary rules. Σ is weakly bounded if and only if Σ is strongly bounded.*

Proof sketch. We refer again to the SH-forest of Σ. In particular, we classify the edges of such structure based on their relationship with the atoms of the rule that is the source of the branch to which they belong. We distinguish between the case in which the edge depends on a single atom (single-ancestor edge), and the case in which it depends on more than one atom (multiple-ancestor edge). Moreover, we use two crucial properties: if Σ is weakly bounded, then (i) the SH-forest of $\Sigma \cup \{\sigma\}$ need be infinite at least for a rule σ in order for Σ not to be strongly bounded, and (ii) the SH-forest of $\Sigma \cup \{\sigma\}$ has an infinite branch originating in σ if and only there is an infinite number of multiple-ancestor edges along such branch. We suse such properties to show that, if Σ is weakly bounded, then there exists no rule σ such that the SH-forest of $\Sigma \cup \{\sigma\}$ is infinite. We use the fact that a multiple-ancestor edge necessarily applies a rule with an existential head variable, and thus can propagate at most one frontier variable. This structural property, in turn, implies that there exists a bound on the number of non-frontier variables in the body of rules occurring along the branches of the forest, and thus that the branches can not grow in an unbounded way. □

Notice that, in the above theorem, the restriction to single-head rules is essential (cf. Theorem 3). The second subclass that we identify is the class of frontier-guarded existential rules.

Theorem 8. *Let Σ be a set of frontier-guarded rules. Then Σ is weakly bounded if and only if Σ is strongly bounded.*

Proof sketch. For single-head frontier-guarded rules, the proof is analogous to the proof of Theorem 7, i.e., we show that if Σ is weakly bounded, then there exists no rule σ such that the SH-forest of $\Sigma \cup \{\sigma\}$ is infinite, due to a structural property. In fact, the syntactic restriction of frontier-guarded rules is such that, after each application of a rule, all the variables that have become equal as an effect of the application must appear together in the atom corresponding to the guard of the rule that is applied. As a result of this property, the number of existential body variables that can be introduced in each rule that is derived along the branches of the SH-forest is bounded by the maximum arity of the relations in Σ, thus the branches can not grow in an unbounded way. Then, we extends the proof to arbitrary frontier-guarded rules. We recall that every rule can be suitably rewritten via a LOGSPACE transformation [6] that uses auxiliary relations and produces a set of single-head rules that is equivalent to the original rule w.r.t. query answering. We show that such a transformation also preserves the properties of weak boundedness and strong boundedness. □

As an immediate consequence of Theorem 7, Theorem 8, Theorem 2, Theorem 4, and Theorem 6, we obtain:

Corollary 2. *Let Σ be a set of single-head binary rules, frontier-guarded rules, or Datalog rules. Then Σ is AFO-rewritable if and only if Σ is CQFO-rewritable.*

We now turn our attention to the decidability of the problem of checking the forms of boundedness presented in this paper. First, the undecidability of strict boundedness for Datalog [13] immediately implies, by Corollary 1, that both strong and weak boundedness for arbitrary set of existential rules are in general undecidable properties.

We now focus on checking strong (i.e., weak) boundedness for single-head binary rules and frontier-guarded rules.

Theorem 9. *The strong boundedness (and weak boundedness) of a set Σ of rules is: (1) undecidable for single-head binary rules; (2) decidable for frontier-guarded rules.*

Proof sketch. The first statement directly follows from Theorem 2.1 in [16], which states the undecidability of the (strict) boundedness problem for linear binary Datalog programs. In fact, since linear binary Datalog programs identify a subclass of single-head binary rules, by Corollary 1, the decidability of weak boundedness would contradict the undecidability results in [16]. The second statement follows from Theorem 6 and Theorem 8 together with results in [4,3]. In particular, we prove the decidability of the weak boundedness by presenting a technique for deciding the AFO-rewritability of a set of single-head frontier-guarded rules based on the fact that atomic queries are answer guarded queries, that answer guarded queries over a set of frontier-guarded rules can be rewritten into a frontier-guarded Datalog program [3], and that frontier-guarded Datalog programs are a subclass of GN-Datalog programs, for which boundedness is decidable [4]. Such a technique can be extended to arbitrary frontier-guarded rules as in the proof of Theorem 8. □

We finally remark that Property 2 of the above theorem can alternatively be derived from Theorem 8, Theorem 6 and from an observation reported in the conclusions of [5]. Notice also that the decidability of the CQFO-rewritability property for the description logic $\mathcal{ELHI}$ (which corresponds to a fragment of frontier-guarded rules) was shown in [20]. Such a result was extended to sets of binary guarded existential rules in [12]. Thus, Property 2 of the above theorem (together with Theorem 4) can be seen as a generalization of the above results to the whole class of frontier-guarded rules.

5 Conclusions

The present work raises several interesting questions and can be extended in different directions. First, we believe that the results presented in this paper may have significant consequences for query answering and query rewriting over existential rules. In particular, for weakly bounded and frontier-guarded sets of rules, an optimized rewriting algorithm for conjunctive queries may be defined, based on the techniques used to prove Theorem 8.

Then, it would be very interesting to further explore the relationship between the strong and weak boundedness properties (that is, between CQFO-rewritability and AFO-rewritability). More precisely, we wonder whether it is possible to find further classes of rules that extend the classes identified in the present paper, i.e., single-head binary rules and frontier-guarded rules. Moreover, for classes of rules for which we are not able to decide weak boundedness, we aim at defining techniques that allow for identifying sufficient conditions for such a property.

Finally, this work leaves open some complexity issues. In particular, we showed that strong boundedness is decidable for sets of frontier-guarded rules. The exact complexity of the decision problem, however, remains open, and we would like to identify more precise bounds in the future.

References

1. Miklós Ajtai and Yuri Gurevich. Datalog vs first-order logic. *J. Comput. Syst. Sci.*, 49(3):562–588, 1994.
2. Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. On rules with existential variables: Walking the decidability line. *Artificial Intelligence*, 175(9–10):1620–1654, 2011.
3. Vince Bárány, Michael Benedikt, and Balder ten Cate. Rewriting guarded negation queries. In *Mathematical Foundations of Computer Science 2013 - 38th International Symposium, MFCS 2013*, pages 98–110, 2013.
4. Vince Bárány, Balder ten Cate, and Martin Otto. Queries with guarded negation. *Proc. VLDB Endow.*, 5(11):1328–1339, July 2012.
5. Michael Benedikt, Balder ten Cate, Thomas Colcombet, and Michael Vanden Boom. The complexity of boundedness for guarded logics. In *30th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2015, Kyoto, Japan, July 6-10, 2015*, pages 293–304, 2015.
6. Andrea Calì, Georg Gottlob, and Michael Kifer. Taming the infinite chase: Query answering under expressive relational constraints. *J. of Artificial Intelligence Research*, 48:115–174, 2013.
7. Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *Semantic Web J.*, 14:57–83, 2012.
8. Andrea Calì, Georg Gottlob, and Andreas Pieris. Advanced processing for ontological queries. *Proc. of the VLDB Endowment*, 3(1):554–565, 2010.
9. Andrea Calì, Georg Gottlob, and Andreas Pieris. Towards more expressive ontology languages: The query answering problem. *Artificial Intelligence*, 193:87–128, 2012.
10. Diego Calvanese and Giuseppe De Giacomo. Expressive description logics. In Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors, *The Description Logic Handbook: Theory, Implementation and Applications*, chapter 5, pages 178–218. Cambridge University Press, 2003.
11. Cristina Civili and Riccardo Rosati. A broad class of first-order rewritable tuple-generating dependencies. In *Proc. of the 2nd Datalog 2.0 Workshop*, 2012.
12. Cristina Civili and Riccardo Rosati. On the first-order rewritability of conjunctive queries over binary guarded existential rules (extended abstract). In *Proceedings of the 30th Italian Conference on Computational Logic, Genova, Italy, July 1-3, 2015.*, 2015.
13. Haim Gaifman, Harry G. Mairson, Yehoshua Sagiv, and Moshe Y. Vardi. Undecidable optimization problems for database logic programs. *J. of the ACM*, 40(3):683–713, 1993.
14. Georg Gottlob, Stanislav Kikot, Roman Kontchakov, Vladimir V. Podolskii, Thomas Schwentick, and Michael Zakharyaschev. The price of query rewriting in ontology-based data access. *Artif. Intell.*, 213:42–59, 2014.
15. Georg Gottlob, Giorgio Orsi, and Andreas Pieris. Ontological queries: Rewriting and optimization. In *Proc. of the 27th IEEE Int. Conf. on Data Engineering (ICDE 2011)*, pages 2–13, 2011.
16. Gerd G Hillebrand, Paris C Kanellakis, Harry G Mairson, and Moshe Y Vardi. Undecidable boundedness problems for datalog programs. *The Journal of Logic Programming*, 25(2):163 – 190, 1995.
17. Phokion G. Kolaitis and Christos H. Papadimitriou. Some computational aspects of circumscription. *J. ACM*, 37(1):1–14, 1990.
18. Mélanie König, Michel Leclère, Marie-Laure Mugnier, and Michaël Thomazo. Sound, complete and minimal ucq-rewriting for existential rules. *Semantic Web*, 6(5):451–475, 2015.
19. Markus Krötzsch and Sebastian Rudolph. Extending decidable existential rules by joining acyclicity and guardedness. In *Proc. of the 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI 2011)*, pages 963–968, 2011.

20. Carsten Lutz and Frank Wolter. Non-uniform data complexity of query answering in description logics. In *Proc. of the 24th Int. Workshop on Description Logic (DL 2011)*, 2011.
21. Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. *J. on Data Semantics*, X:133–173, 2008.
22. Heng Zhang and Yan Zhang. First-order expressibility and boundedness of disjunctive logic programs. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, 2013.