Deciding the Loosely Guarded Fragment and Querying Its Horn Fragment Using Resolution

Sen Zheng, Renate A. Schmidt
Department of Computer Science, University of Manchester
Oxford Road, Manchester M13 9PL, UK

Abstract

We consider the following query answering problem: Given a Boolean conjunctive query and a theory in the Horn loosely guarded fragment, the aim is to determine whether the query is entailed by the theory. In this paper, we present a resolution decision procedure for the loosely guarded fragment, and use such a procedure to answer Boolean conjunctive queries against the Horn loosely guarded fragment. The Horn loosely guarded fragment subsumes classes of rules that are prevalent in ontology-based query answering, such as Horn \( \mathcal{ALCHOI} \) and guarded existential rules. Additionally, we identify star queries and cloud queries, which using our procedure, can be answered against the loosely guarded fragment.

Introduction

Our motivation of considering query answering problem stems from ontology-based data access (OBDA) systems (Xiao et al. 2018), which have attracted much recent attention in the knowledge representation and database communities. In particular, since the Horn loosely guarded fragment subsumes mainstream rules in OBDA systems, such as Horn \( \mathcal{ALCHOI} \) (Bader et al. 2017) and the guarded existential rules (a.k.a. guarded TGDs) (Cali, Gottlob, and Lukasiewicz 2012), our interest is the development of a practical procedure for answering queries against the Horn loosely guarded fragment (van Benthen 1997).

To retrieve information from OBDA systems, the main querying mechanisms are (Boolean) conjunctive queries. Given a Boolean conjunctive query (BCQ) \( q \), a set \( \Sigma \) of rules and a database \( D \), checking whether \( \Sigma \cup D \models q \) is equivalent to checking whether \( \Sigma \cup \neg D \models \neg q \models \perp \), so that the problem of answering BCQ can be reduced to deciding satisfiability. Such BCQ answering problems can be recast as query containment/entailment/evaluation problems in database research (Baget et al. 2011), constraint satisfaction problems and homomorphism mapping problems in general AI research (Vardi 2000). Although finding answers for queries is also an important problem, Boolean conjunctive query answering is widely studied (Baget et al. 2011; Bárány, Gottlob, and Otto 2010; Calì, Gottlob, and Kifer 2013; Calì, Gottlob, and Lukasiewicz 2012; Glimm et al. 2008; Gottlob, Pieris, and Tendera 2013). In this paper, we particularly focus on an open problem, namely BCQ answering for the Horn loosely guarded fragment.

The complexity of BCQ answering for the guarded fragment is \( 2\text{EXPTIME} \)-complete (Bárány, Gottlob, and Otto 2010), and satisfiability checking for the clique-guarded negation fragment, which subsumes both BCQs and the loosely guarded fragment, is also \( 2\text{EXPTIME} \)-complete (Bárány, ten Cate, and Segoufin 2015). These complexity results show that BCQ answering for the Horn (loosely) guarded fragment is decidable, however, as yet there is no practical (i.e., implementable) procedure.

Let us give a quick review of our settings and explain why we start our investigation for query answering with deciding the loosely guarded fragment (LGF). A loosely guarded quantified formula (van Benthen 1997; Grädel 1999) has the form \( \forall \pi(G_1 \land \ldots \land G_n \rightarrow F) \) where \( G_1, \ldots, G_n \) are atoms that are called guards, \( F \) is a loosely guarded formula where \( i \) all free variables of \( F \) occur in \( G_1 \land \ldots \land G_n \), and \( ii \) the variables in \( G_1, \ldots, G_n \) are either free, or they co-occur with each other in a \( G_i \) among the \( G_1, \ldots, G_n \). The Horn fragment of LGF will be referred to Horn LGF. A Boolean conjunctive query is a first-order formula of the form \( q = \exists \pi \varphi(\pi) \) where \( \varphi \) is a conjunction of atoms containing only constants and variables. One can obtain a query clause \( Q \) by simply negating a BCQ. Hence, \( Q \) is a negative clause containing no compound terms.

To answer BCQs over Horn LGF, we begin with deciding LGF, since it can be observed that \( i \) some loosely guarded formulae are naturally cyclic BCQs (Bernstein and Chiu 1981) because a loosely guarded formula \( F \) allows multiple guards, and all variables in the clausal form of \( F \) (loosely guarded clauses) co-occur with each other in one of the guards. E.g., the loosely guarded formula \( F = \exists xyz(A_1xy \land A_2yz \land A_3xz) \) is a cyclic BCQ. In fact, it shows that the conjunctive queries with the hyper-tree width property are strongly connected to LGF (Gottlob, Leone, and Scarcello 2003); \( ii \) a loosely guarded formula allow variables chaining multiple literals (chained variables), as in \( F \), which can cause term depth increase during reasoning (see Example 1). We see that understanding the handling
of chained variables in LGF helps us handle chained variables in query clauses. Hence, in this paper, we first provide a decision procedure to decide LGF. Subsequently we show that such a procedure can be extended to answer BCQs over Horn LGF and answer restricted BCQs over LGF.

Considering the only existing decision procedure for answering BCQs over (Horn) LGF essentially aims at the theoretical analysis (Bárany, ten Cate, and Segoufin 2015), our focus is on devising a practical decision procedure, so that one can implement our procedure as query engines. We build our procedure using resolution in the framework of (Bachmair and Ganzinger 2001), which is a standard in the area of automated reasoning and provides basis for most first-order logic reasoners such as Spass (Weidenbach et al. 2009), Vampire (Riazanov and Voronkov 2001) and E (Schulz 2013). References for using resolution with refinement as practical decision procedures are (Ganzinger et al. 1998; Ganzinger and de Nivelle 1999; Hustadt and Schmidt 1999; Hustadt and Schmidt 1997; Bachmair, Ganzinger, and Waldmann 1993).

In resolution-based reasoning, the main challenges to avoid non-termination are: i) avoiding unlimited growth of the number of literals in resolvents, and ii) avoiding unlimited growth of term depth in resolvents. The former can be tackled by using the property of loosely guarded clauses C: Guards in C contain all variables of C, so that using our inference system, the number of literals cannot increase indefinitely in the resolvents. As for the latter, consider this example:

**Example 1.** Q is a query clause and C_1, C_2 are loosely guarded clauses:

\[ Q = \neg A_1(x, y) \lor \neg A_2(y, z), \]
\[ C_1 = A_1(f(x_1, y_1), x_1) \lor B_1(g(x_1, y_1)) \lor \neg G_1(x_1, y_1), \]
\[ C_2 = A_2(h(x_2, y_2), x_2) \lor \neg G_2(x_2, y_2) \]

Performing resolution among Q, C_1, C_2 derives \( R = B_1(g(h(x_2, y_2), y_1)) \lor \neg G_1(h(x_2, y_2), y_1) \lor \neg G_2(x_2, y_2). \) In R, g(h(x_2, y_2)) is deeper than all the terms in Q, C_1 and C_2. This happens when a query clause Q contains a chained variable y where: i) y needs to be unified with a variable x_1 in C_1 and a non-ground compound term h(x_2, y_2) in C_2 at the same time and, ii) x_1 occurs in a non-ground compound term g(x_1, y_1) in C_1. However, such term depth increase can be avoided if we perform resolution on Q and C_1 first. We introduce the top variable technique in its corresponding section to show how the term depth increase in R can be prevented.

Our resolution decision procedure for LGF is a variation of procedures presented in (de Nivelle and de Rijke 2003) and (Ganzinger and de Nivelle 1999). Like (Ganzinger and de Nivelle 1999), but unlike (de Nivelle and de Rijke 2003), which uses a non-liftable ordering, our procedure uses admissible and liftable orderings with selection, and is consistent with the resolution framework of (Bachmair and Ganzinger 2001). Inspired by the ‘MAXVAR’ technique in (de Nivelle and de Rijke 2003) and the partial hyper-resolution technique from (Ganzinger and de Nivelle 1999), we use the top variable technique to avoid term depth increase. (Ganzinger and de Nivelle 1999) mainly focuses on deciding the guarded fragment, and refers to a manuscript version of (de Nivelle and de Rijke 2003) for technical details about using the ‘MAXVAR’ technique to decide LGF. As for ‘MAXVAR’, unlike (de Nivelle and de Rijke 2003), we use a unification-first approach to identify top variables while (de Nivelle and de Rijke 2003) finds ‘MAXVAR’ by variable depth first, then applies a specific unification algorithm. It turns out that our approach allows top variables (‘MAXVAR’) being easily identified since no specific permutation and unification algorithms are needed. Further, we embed the top variable technique into the framework of (Bachmair and Ganzinger 2001) as a selection function (top selection) with completeness proofs, so that, together by using liftable orderings, one immediately benefits from being able to use the notions of redundancy in that framework. Additionally, we generalise the pre-conditions of ‘MAXVAR’, so that the top technique can be applied to a larger class than LGF (any clauses that satisfy conditions of query pair clauses), including queries, so that one can use the top technique to query other fragments of first-order logic to avoid variable depth growth as well. This makes our procedure applicable for the problems of BCQ answering for Horn LGF, and the star/cloud query answering for LGF.

Without hurting the result of this paper, we discuss variable depth, rather than term depth for the termination result. This holds because a term depth can grow infinitely if and only if a variable depth grows infinitely.

The contributions of this paper are:

- A variation of the resolution-based decision procedure for LGF in (de Nivelle and de Rijke 2003), situated in the framework of (Bachmair and Ganzinger 2001).
- By expanding the top variable technique to query pair clauses, this procedure provides the basis for a practical decision procedure for answering BCQs over Horn LGF.
- We identify star queries and cloud queries so that one can use our procedure to answer these queries against LGF.

**Preliminaries**

Let C, F, P denote pairwise disjoint sets of constant symbols, function symbols and predicate symbols, respectively. The definition of (compound) ground term, atom, literal, clause, expression, argument, unifier, most general unifier (mgu) and simultaneous mgu are defined as usual in automated reasoning (see e.g., (Bachmair and Ganzinger 2001) for details). A literal L is a non-ground compound literal if L contains at least one non-ground compound term. Let \( \pi, A, C \) denote a sequence of variables, a sequence of atoms and a set of clauses, respectively. Let \( \text{var}(t), \text{var}(C) \) and \( \text{var}(A_n) \) denote sets of variables in a term t, a clause C and a sequence of atoms \( A_n \), respectively.

The variable depth of a term t, denoted as vdp(t), is defined as follows: i) if t is ground, then \( vdp(t) = 1 \); and, ii) if t is not ground, then \( vdp(t) = 0 \), and iii) if t is a non-ground compound term \( f(t_1, \ldots, t_n) \), then \( vdp(t) = 1 + \max\{vdp(t_i) \mid 1 \leq i \leq n\} \). A term t is flat if
A term \( t \) is simple if \( vdp(t) \leq 0 \). A flat (simple) atom, literal and clause is an atom, a literal and a clause such that every term in it is flat (simple). We say a term \( t \) is a simple non-ground compound term if \( vdp(t) = 1 \). Assume \( C = \neg A_1 \lor \ldots \lor \neg A_n \lor D \) is a clause where \( \neg A_1 \lor \ldots \lor \neg A_n \) are flat literal. Then \( x \in var(A_1, \ldots, A_n) \) can be: i) a chained variable: \( x \) occurs in two literals \( A_i, A_j \) among \( A_1, \ldots, A_n \), such that \( var(A_i) \nsubseteq var(A_j), var(A_j) \nsubseteq var(A_i) \), and \( x \in var(A_i) \cap var(A_j) \), and ii) an isolated variable: \( x \) is not chained. In Q of Example 1, \( y \) is a chained variable and \( x, z \) are isolated variables. By the length of a clause, we mean the number of literals that occur in a clause, and by the depth of a clause, we mean the deepest variable depth of a clause. In this paper, we assume the input clauses (formulas) are of fixed-length and fixed-width.

A weakly covering term is a compound term \( t \) such that for every non-ground, compound subterm \( s \) of \( t \), it is the case that \( var(s) = var(t) \) (Fermüller et al. 1993). A literal \( L \) is weakly covering if each argument of \( L \) is either a ground term, a variable, or a weakly covering term, such that \( var(t) = var(L) \). A clause \( C \) is weakly covering if each term \( t \) in \( C \) is either a ground term, a variable, or a weakly covering term such that \( var(t) = var(C) \). E.g., the clause \( C_1 = \neg A_1(fxyza, x, y, ga) \cup A_2xyz \) is a weakly covering clause since the only non-ground compound term \( fxyza \) satisfies \( var(fxyza) = var(C_1) \), however, the clause \( C_2 = \neg A_1(gy, y, ga) \cup A_2(hxyz) \) is not weakly covering since \( var(gy) \neq var(C_2) \). Here the notion of weakly covering literals in (Fermüller et al. 1993) is extended to weakly covering clauses; reasons are given in the loosely guarded clauses section.

Recall, the rule set \( \Sigma \) denotes a set of first-order formulæ and the database \( D \) denotes a set of ground atoms. A Boolean conjunctive query is a first-order formula of the form \( q = \exists \phi(x) \) where \( \phi \) is a conjunction of atoms containing only constants and variables. We use the symbol \( Q \) to denote the query clause \( \neg q \), so that we can answer BCQ satisfiability of \( \Sigma \cup D \models \neg q \) by checking whether \( \Sigma \cup D \cup Q \models \bot \).

### The Loosely Guarded Fragment

**Definition 1.** The loosely guarded fragment (LGF) is a fragment of first-order logic without equality and function symbols, defined inductively as follows:

1. \( \top \) and \( \bot \) are in LGF.
2. If \( A \) is an atom, then \( A \) is in LGF.
3. LGF is closed under Boolean combinations.
4. If \( F \in \text{LGF} \) and \( G_1, \ldots, G_n \) are atoms, then a formula \( \forall \varphi(G_1 \land \ldots \land G_n \rightarrow F) \) belongs to LGF (i) if all free variables of \( F \) belong to \( \forall \varphi(G_1, \ldots, G_n) \), and ii) for each variable \( x \in \varphi \) and each variable \( y \in var(G_1, \ldots, G_n) \) where \( x \neq y \) and \( x \) and \( y \) co-occur in a \( G_i \). The negative literals \( \neg G_1, \ldots, \neg G_n \) are called the guards of this formula.

The first-order logic translation of a temporal logic formula \( P \) until \( Q \) is a loosely guarded formula: \( \exists y(Rxy \land Qy \land \forall z((Rxz \land Rzy) \rightarrow Pz)) \), but the transitivity formula \( \forall xyz((Rxy \land Ryz) \rightarrow Rxz) \) is not a loosely guarded formula since \( x \) and \( z \) do not co-occur in a guard.

We use the loosely guarded formula \( F \) in Example 2 to illustrate the clausal normal form transformation for LGF.

**Example 2.** \( \exists y(Rxy \land Qy \land \forall z((Rxz \land Rzy) \rightarrow \exists x Pxy)) \)

### The Resolution Calculus

In this section, we introduce the resolution calculus, which gives us the main termination result of this paper. The inference steps are restricted by an admissible ordering and a selection function, so that the search space can be reduced when a reasoner computes inferences. For more technical details about rules used in this paper, we refer readers to (Bachmair and Ganzinger 2001).

Let \( \succ \) be a strict ordering, called a precedence, on the symbols in the \( C, F \) and \( P \). An ordering \( \succ \) is lift if for all expressions \( E_1 \) and \( E_2 \) and all substitutions \( s, E_1 \succ E_2 \) implies \( E_1(s) \succ E_2(s) \). An ordering \( \succ \) on literals is admissible, if i) it is well-founded and total on ground literals, and lift, ii) \( \neg A \succ A \) for all ground atoms \( A \), and iii) if \( B \succ A \), then \( B \succ \neg A \) for all ground atoms \( A \) and \( B \). A literal \( L \) is \( \succ \)-maximal with respect to a clause \( C \) if for any \( L' \) in \( C \), \( L' \not\succ L \), and \( L \) is strictly \( \succ \)-maximal with respect to a clause \( C \) if for any \( L' \) in \( C \), \( L' \not\sim \succ L \). A selection function \( S \) selects a possibly empty set of occurrences of negative literals in a clause \( C \) with no restriction imposed. Inferences are only performed on eligible literals. A literal \( L \) is eligible in a clause \( C \) if either nothing is selected in the selection function \( S \) and \( L \) is a \( \succ \)-maximal literal with respect to \( C \), or \( L \) is selected by \( S \).

Inferences are computed using the following rules:

- **Deduction:** \( N \) derives \( C \) if \( C \) is either a resolvent or a factor of clauses in the set \( N \).

- Factors and resolvents are derived using:
  - **Ordered factoring:** \( C \lor A_1 \lor A_2 \) derives \( (C \lor A_1) \sigma \), where \( \sigma \) is the mgu of \( A_1 \) and \( A_2 \), and ii) no literal is selected in \( C \), and iii) \( A_1 \sigma \) is \( \succ \)-maximal with respect to \( C \).
  - **Ordered resolution with selection:** \( \neg A_1 \lor \ldots \lor \neg A_n \lor D \lor B_1 \lor D_1 \lor \ldots \lor B_n \lor D_n \lor (D_1 \lor \ldots \lor D_n) \sigma \) where i) either \( A_1 \lor \ldots \lor \neg A_n \) are selected in \( D \), or \( n = 1 \), no literal is selected, and \( A_1 \sigma \) is \( \succ \)-maximal with respect to \( D \), and ii) no literal is selected in \( D_1 \lor \ldots \lor D_n \) and \( B_1 \lor \ldots \lor B_n \), respectively, and iii) \( \sigma \) is a simultaneous mgu such that \( A_1 \sigma = B_1 \sigma, \ldots, A_n \sigma = B_n \sigma \), and iv) \( \neg A_1 \lor \ldots \lor \neg A_n \lor D, B_1 \lor D_1, \ldots, B_n \lor D_n \) are pairwise variable-disjoint.

In ordered factoring and ordered resolution, maximality is computed using a-posteriori application of the mgu \( \sigma \). This means the maximal literal is determined after application of \( \sigma \), derived from the unification algorithm applied to the premises of a rule. If the maximal literal is determined before the application of \( \sigma \), we call this a-priori application.

Redundancy is eliminated using:

- **Deletion:** \( N, C \) derives \( N \) if \( C \) is a tautology, or \( N \) contains a variant of \( C \), or \( N \) contains a condensed form of \( C \).

The ‘Deletion’ rule is the only rule used to eliminate redundancy, and turns out to be sufficient for the termination result. Since we employ an admissible ordering with a selection function as resolution refinement in accordance with
the framework of (Bachmair and Ganzinger 2001), we can also use more sophisticated simplification rules and redundancy elimination of that framework, e.g., subsumption deletion and forward/backward subsumption.

A ground clause \( C \) is redundant with respect to \( N \) if there are ground instances \( C_1 \sigma, \ldots, C_n \sigma \) of clauses in \( N \) such that \( C_1 \sigma, \ldots, C_n \sigma \models C \) and for each \( i \), \( C \succ C_i \sigma \). A non-ground clause \( C \) is redundant with respect to \( N \) if every ground instance of \( C \) is redundant with respect to \( N \). A set of clauses \( N \) is saturated up to redundancy (with respect to ordered resolution and selection) if any inference from non-redundant premises in \( N \) is redundant in \( N \) (Bachmair and Ganzinger 2001).

### The Decision Procedure

Now we can discuss the resolution procedure for LGF.

#### Clausal Normal Form Translation LGF-Trans

The clausal normal form transformation we use is similar to the one in (de Nivelle and de Rijke 2003) and (Ganzinger and de Nivelle 1999), but i) free variables are assumed to be existentially quantified since the focus is on satisfiability checking, and ii) prenex normal form and outer Skolemisation (Nonnengart and Weidenbach 2001) are used. Though outer Skolemisation may introduce Skolem functions of higher arities than inner/standard Skolemisation, outer Skolemisation turns out to be critical to guarantee that output clauses have the weakly covering property.

We use \( \text{LGF-Trans} \) to denote the clausal normal form transformation below. Using \( F \) in Example 2, one can obtain a set of loosely guarded clauses via the following steps:

i) Add existential quantifiers to all free variables in \( F \):
\[
\exists x (Rxy \land Qy \land \forall z ((Rxz \land Rzy) \rightarrow \exists x Pxy)).
\]

ii) Rewrite \( \rightarrow \) and \( \leftrightarrow \) using conjunction, disjunction and negation, and transform \( F \) into negation normal form, obtaining the formula \( F_{nnf} \):
\[
\exists y (Rxy \land Qy \land \forall z (\neg Rxz \lor \neg Rzy \lor \exists x Pxy)).
\]

iii) Apply optimised structural transformation to \( F_{nnf} \), that introduces fresh predicate symbols \( (Q_1) \) for universally quantified subformulae \( \forall z (\neg Rxz \lor \neg Rzy \lor \exists x Pxy) \), obtaining the formula \( F_{str} \):
\[
\exists y (Rxy \land Qy \land Q_1xy) \land \forall x (\neg Q_1xy \lor \forall z (\neg Rxz \lor \neg Rzy \lor \exists x Pxy)).
\]

iv) Find \( \exists y \forall w x \forall z f (Rxy \land Qy \land Q_1xy) \land \neg Q_1wz \lor \neg Rxw \lor \neg Rwv \lor Pxv) \) as the prenex normal form of \( F_{str} \), and apply outer Skolemisation: if \( \forall x \) is the subsequence of all universal quantifiers of the \( \exists y \)-prefix of subformula \( \exists y \varphi \) of \( \varphi \), then \( \varphi[y/f(x)] \) is the outer Skolemisation of \( \exists y \varphi \). Skolem terms \( a, b, f, z, y \) are introduced, obtaining \( F_{sko} \):
\[
Rab \land Qb \land Q_1ab \land \forall x (\neg Q_1xy \lor \neg Rxz \lor \neg Rzy \lor P(fxyz, y)).
\]

v) Drop all universal quantifiers and transform \( F_{sko} \) into conjunctive normal form, obtaining loosely guarded clauses:
\[
Rab, Qb, Q_1ab, \neg Q_1xy \lor \neg Rxz \lor \neg Rzy \lor P(fxyz, y)
\]

### Loosely Guarded Clauses

We now describe loosely guarded clauses and their properties.

#### Definition 2

A loosely guarded clause (LGC) \( C \) is a clause satisfying the following conditions:

1. \( C \) is simple and weakly covering, and
2. if \( C \) is non-ground, then there is a set of negative literals \( \neg G_1, \ldots, \neg G_n \) in \( C \) that are flat. Then \( \neg G_1, \ldots, \neg G_n \) are called the guards of \( C \), such that each pair of variables in \( C \) co-occur in at least one of the guards.

We can immediately see that a ground clause is an LGC.

#### Proposition 1

Using LGF-Trans, every loosely guarded formula can be transformed into a set of LGCs.

The class of LGCs strictly subsumes LGF since function symbols are allowed. If using an admissible ordering in which function symbols have higher precedence than other symbols, then non-ground compound terms in an LGC \( C \) are always larger than variables in \( C \) due to the weakly covering property. E.g., with a lexicographic path ordering \( \succ_{\text{lpw}} \) (Dershovitz 1982), considering an LGC \( C = \neg A_1xy \lor \neg A_2yz \lor \neg A_3xz \lor D(fxyz) \) and an arbitrary substitution \( \sigma \), \( D \) is \( \succ_{\text{lpw}} \)-maximal with respect to \( C \) if \( D \sigma \) is \( \succ_{\text{lpw}} \)-maximal with respect to \( C \sigma \) since \( \text{var}(D) = \text{var}(C) \).

An obvious property of a weakly covering clause \( C \) is that all non-ground terms and literals in \( C \) are also weakly covering. Formally stated as:

#### Lemma 1

Assume \( C \) is a weakly covering clause containing a non-ground compound literal \( L \) and \( \sigma \) is an arbitrary substitution. Using any admissible ordering \( \succ \) with a precedence that function symbols are larger than other symbols, \( L \) is \( \succ \)-maximal with respect to \( C \) if \( L \sigma \) is \( \succ \)-maximal with respect to \( C \sigma \).

#### The Top Variable Technique

Before discussing the resolution calculus for LGCs, we introduce the top variable technique, as a variation of the ‘MAXVAR’ technique in (de Nivelle and de Rijke 2003). The top variable technique is a look-ahead approach to prevent variable depth increase in the resolvents: Suppose we have a set of clauses that can lead to variable depth increase in the resolvent. Using the top variable technique, we first identify clauses that lead to the potentially deepest terms, and then perform resolution on those clauses first. Next we perform inference on the rest of the clauses. In such a manner of performing resolution, we show that no variable depth increase occurs in the resolvents.
Definition 3 (Query Pair Clauses). Let $\bar{A}_n$, $\bar{B}_n$ be a sequence of atoms $A_1, \ldots, A_n$ and a sequence of weakly covering atoms $B_1, \ldots, B_n$, respectively. $(\bar{A}_n, \bar{B}_n)$ is a query pair if they satisfy these conditions:

1. $\bar{A}_n$ is flat and non-ground, and $\bar{B}_n$ is simple.
2. Each $B_i \in \bar{B}_n$ either is a non-ground compound literal or is a ground literal.
3. $\text{var}(\bar{A}_n) \cap \text{var}(\bar{B}_n) = \emptyset$, and $B_1, \ldots, B_n$ are pairwise variable disjoint.
4. There exists an mgu (simultaneous mgu if $n > 1$) $\sigma$ such that for each $A_i \in \bar{A}_n$, $B_i \in \bar{B}_n$, $A_i\sigma = B_i\sigma$.

Let $(\bar{A}_n, \bar{B}_n)$ be a query pair. Query pair clauses for a query pair $(\bar{A}_n, \bar{B}_n)$ is a set of clauses:

$$C = \neg A_1 \lor \ldots \lor \neg A_n \lor D,$$

where $D$ is a flat clause and $D_1, \ldots, D_n$ are simple clauses.

From now on, we also use mgu to denote the simultaneous mgu. To find top variables in $\bar{A}_n$ in a query pair, one needs to find the mgu between $A_n$ and $B_n$ to identify the variable orderings over $\text{var}(\bar{A}_n)$.

Definition 4 (Variable Ordering). Let $(\bar{A}_n, \bar{B}_n)$ be a query pair and let an mgu $\sigma$ satisfy Condition 4 in Definition 3.

By $x \succ y$, we denote a variable ordering over $\text{var}(\bar{A}_n)$, which is defined by: for $x, y \in \text{var}(\bar{A}_n)$, $x \succ y$ if $\text{vdp}(x\sigma) > \text{vdp}(y\sigma)$, or $x = y$ if $\text{vdp}(x\sigma) = \text{vdp}(y\sigma)$.

Using the notion of variable orderings, we define top variables, and show the existence of top variables in query pairs:

Definition 5 (Top variable). Given a query pair $(\bar{A}_n, \bar{B}_n)$, a variable $x \in \text{var}(\bar{A}_n)$ is a top variable iff for each $y \in \text{var}(\bar{A}_n)$, $x \succ y$ or $x = y$.

Proposition 2. Let $(\bar{A}_n, \bar{B}_n)$ be a query pair. Then at least one of the variables in $\bar{A}_n$ is a top variable.

The idea behind the top variable technique is finding the potentially deepest term of query pair clauses. To realise it, we first apply the unification algorithm, then make the literals in the main premise containing the potentially deepest terms eligible literals. In Example 1, the mgu $\sigma = \{x/f(xz_2y_2, y_1), y/yz_2y_2, z/xx, z/yz_2y_2\}$, thus applying resolution among $Q, C_1, C_2$ derives $B_1(g(xz_2y_2, y_1))$, in which the first argument is deeper than all terms in $Q, C_1, C_2$. Now we use top variables to find the deepest terms. First we find top variables in $Q$: since $\text{vdp}(x\sigma) > \text{vdp}(y\sigma) > \text{vdp}(z\sigma)$, $x \succ y \succ z$. Since $x$ is the top variable (potentially the deepest term), we make the literal $A_1$ eligible since $x$ only occurs in $A_1$, then applying resolution on clauses $Q, C_1$, deriving the resolvent $C_3 = \neg A_2(x, z) \lor B_1(g(xz_2y_1) \lor \neg A_1(x_1, y_1))$. Though $C_3$ is not weakly covering, there is no variable depth increase in $C_3$. Let $(\bar{A}_n, \bar{B}_n)$ be a query pair, and assume query pair clauses $C, C_1, \ldots, C_n$ such that $C = \neg A_1 \lor \ldots \lor \neg A_n \lor D$, as the main premise, $C_1 = B_1 \lor D_1, \ldots, C_n = B_n \lor D_n$ as the side premises, where $D$ is flat and $D_1, \ldots, D_n$ are simple. Assume $A_1, \ldots, A_t$ is a sequence of atoms containing top variables and the respective counterparts are $B_1, \ldots, B_t$, which occur in $C_1, \ldots, C_t$, respectively, and $\sigma$ is the mgu such that $A_i\sigma = B_i\sigma$ where $1 \leq i \leq t \leq n$. We denote $\text{Res}$ as an application of resolution among the clauses $C, C_1, \ldots, C_t$:

$$\text{Res}(B_1 \lor D_1, \ldots, B_t \lor D_t, \neg A_1 \lor \ldots \lor \neg A_n \lor D) = \neg A_t \lor \ldots \lor \neg A_n \lor D.$$

Given two expressions $A(t, \ldots, u)$ and $B(u, \ldots, v)$, we say $t$ matches $u$ if the argument position of $t$ in $A$ is the same as the argument position of $u$ in $B$. We show how top variables in $\text{var}(\bar{A}_n) \cap \text{var}(\bar{B}_n)$ the result is stated as:

Lemma 3. In an application of $\text{Res}$,

1. a top variable matches either a ground term or a non-ground compound term, and
2. a non-ground compound term matches a top variable.

Based on the matching in Lemma 3, we now show properties of mgu in $\text{Res}$:

Lemma 4. In an application of $\text{Res}$, these conditions hold:

1. The mgu assigns to top variables either simple non-ground compound terms or ground terms.
2. The mgu assigns to non-top variables in the main premise either variables or ground terms.
3. The mgu assigns to variables in the side premises either variables or ground terms.

Using Lemma 4, we give Theorem 1, which says that, for query pair clauses, only resolving literals that contain the potentially deepest terms does not lead to variable depth growth in the resolvents.

Theorem 1. In an application of $\text{Res}$, no variable depth growth occurs in the resolvents of a set of query pair clauses.

Resolution Refinement LGC-Refine

Now we formally describe the orderings and selection as refinement to decide LGF. One can use any admissible ordering that satisfies the conditions in LGC-Refine. Here a lexicographic path ordering $\succ_{lpo}$ (Dershowitz 1982) is used.

Definition 6 (LGC-Refine). Let LGC-Refine denote the refinement: A lexicographic path ordering $\succ_{lpo}$ based on a precedence $f > a > p$ for $f \in F$, $a \in C$ and $p \in P$, and a selection function such that the following conditions hold:

1. If a clause contains negative non-ground compound literals, then at least one of these literals is selected.
2. If a clause contains no negative non-ground compound literal, but there are positive non-ground compound literals, then the maximality principle with respect to $\succ_{lpo}$ is applied to determine the eligible literals.
3. If a clause contains no non-ground compound literals, select all the negative literals containing top variables.

We use top selection to denote selection based on the top variable technique. Condition 3 in LGC-Refine implies that top selection is imposed not only to guards. E.g., although $\neg B(x)$ is not a guard in $\neg A(x, y) \lor \neg B(x)$, top selection would select both $A$ and $B$ if $x$ is a top variable.
The Resolution Calculus LGC-Res

Now we discuss how resolution with LGC-Refine performs over LGCs. We use the notation LGC-Res to denote the calculus consisting of the following: the ‘Deduction’ rule, ordered factoring and ordered resolution with selection refined by LGC-Refine, and the ‘Deletion’ rule. When applying LGC-Res, the ‘Deletion’ rule and the ‘Deduction’ rule are used whenever they are applicable. As usual, we assume the input clauses (after condensation and modulo variable renaming) are a finite set of fixed LGCs.

First we discuss the ordered resolution rule. We use Res' to denote the resolution rule when one of the premises satisfies Condition 3 in LGC-Refine.

Let $C = \lnot A_1 \lor \ldots \lor \lnot A_n \lor D$ be a flat LGC, as the main premise, and $C_i = B_i \lor D_i$ be a set of LGCs, as the side premises. Let $A_{i(n)}, B_i$ denote $A_1, \ldots, A_{i(n)}$ and $B_1, \ldots, B_t$ where $1 \leq t \leq n$, respectively. Using LGC-Refine, Res' is performed as:

$$
(\text{LGC-Res})_{C_i = B_i \lor D_i} :\sigma \left [ \frac{B_1 \lor D_1 \lor \ldots \lor \lnot A_1 \lor \ldots \lor \lnot A_i \lor \ldots \lor \lnot A_n \lor D}{(D_1 \lor \ldots \lor D_t \lor \lnot A_{i+1} \lor \ldots \lor \lnot A_n \lor D)} \right ]^\sigma
$$

where i) $C$ is non-ground and $D$ is positive, ii) each $A_i \in A_{i(t)}$ contains at least one top variable and each $B_i \in B_i(t)$ is strictly $\sigma$-maximal with respect to $C_i$, respectively, and $\sigma$ is the mu such that $A_i^\sigma = B_i^\sigma$ where $1 \leq i \leq t$, iii) $C, C_1, \ldots, C_t$ are pairwise variable disjoint.

Since the premises in Res' (LGCs) satisfy conditions of query pair clauses, we can inherit results of Res. The particularities in Res' are: i) all premises are weakly covering clauses rather than only literals are weakly covering, and ii) each premise contains a set of guards.

First we show that using Res', every resolvent is simple:

**Corollary 1.** In an application of Res', the resolvents of a set of LGCs are simple clauses.

To show the resolvents in Res' are LGCs, we need to discuss some unique properties in Res' comparing to Res:

**Lemma 5.** In an application of Res', if we use notions from Res', and let $x$ be a top variable in $A_1, \ldots, A_s$ ($s \leq t$). Then
1. $\text{var}(A_1, \ldots, A_s) = \text{var}(C)$, and
2. $\text{var}(x^\sigma) = \text{var}(C^\sigma)$, and
3. $\text{var}(x^\sigma) = \text{var}(y^\sigma)$ if $x, y$ are distinct top variables.

Now we can show the resolvents in Res' have are indeed LGCs: they are weakly covering and contain a set of guards.

**Lemma 6.** In an application of Res', the resolvents of a set of LGCs are LGCs.

It remains to consider other possibilities in LGC-Res. In particular, we discuss situations that are not covered by Res' such that there is no premise satisfying Condition 3 in LGC-Refine: the negative premise satisfies Condition 1 in LGC-Refine or is ground. This is the case when the ordered resolution with selection is naturally reduced to a binary case.

**Lemma 7.** In an application of LGC-Res, the factors of LGCs are LGCs, and the resolvents of LGCs are LGCs.

Now we can show the main result of this section:

**Theorem 2.** Given a set of LGCs, using LGC-Res, all inferred clauses are LGCs.

So far we have shown that using LGC-Res, the resolvents of LGCs are LGCs. Since an LGC is a simple clause, there is no variable depth increase during the inference. We still need to consider that using LGC-Res, the length of the resolvents cannot be infinitely long:

**Lemma 8.** In an application of LGC-Res, the number of variables in derived clauses is no more than the number of variables of one of the premises of these derived clauses.

**Refutational Completeness and Termination**

This section we give the refutational completeness and termination results of applications of LGC-Res over LGCs. For refutational completeness result, we particularly show that the top selection used in LGC-Refine is compatible within the framework of (Bachmair and Ganzinger 2001).

**Theorem 3** (Refutational Completeness). Let $N$ be a set of clauses that are saturated up to redundancy under LGC-Res, then $N$ is unsatisfiable iff $N$ contains the empty clause.

Let LGF-Res denote the combination of the clausal transformation LGF-Trans and the resolution calculus LGC-Res with refinement LGC-Refine. Now we give the first main result of this paper:

**Theorem 4.** LGF-Res decides LGF.

**Querying Horn LGF and LGF**

In this section, we aim to check whether $\Sigma \cup \mathcal{D} \cup Q \models \bot$ where $\Sigma$ are formulae in (Horn) LGF, $\mathcal{D}$ is a set of ground atoms and $Q$ is a query clause. Because $\Sigma$ and $\mathcal{D}$ can be transformed into (Horn) LGCs using LGF-Trans, the aim now is to check whether $\bar{C} \cup Q \models \bot$ where $\mathcal{C}$ is a set of (Horn) LGCs and $Q$ is a query clause. In particular, we assume $Q$ is a fixed query clause. We show that when either $Q$ is restricted to star/cloud queries, or $\Sigma$ is restricted to Horn LGF, our procedure guarantees termination.

Since a query clause $Q$ contains no non-ground compound terms, $Q$ satisfies Condition 3 in LGC-Refine, thus no particular new refinement for $Q$ is needed. Hence we still can use LGC-Refine as refinement for inference rules. However, LGC-Res does not contain a rule to compute resolvents of a query clause and a set of LGCs. Using LGC-Refine, a query clause $Q$ and a set of LGCs $C_1, \ldots, C_n$ satisfy conditions of query pair clauses, thus we apply Res to compute resolvents of $Q$ and $C_1, \ldots, C_n$. We use Query-Res to denote LGF-Res and Res using refinement LGC-Refine. Since no positive factoring can be applied to query clauses, we only discuss how resolution is performed on query clauses.

According to Theorem 1, the following result holds:

**Corollary 2.** In the application of Res, there is no variable depth growth in the resolvents of a query clause and a set of LGCs, thus the resolvents are simple clauses.

**Querying Horn LGF**

It follows from Theorem 2 that using Query-Res, the resolvents of a set of LGCs are LGCs. We now discuss the resolvent of a query clause and a set of LGCs. Corollary 2 shows
that a resolvent $R$ of a query clause and a set of LGCs is a simple clause. However, this simple clause $R$ can be neither an LGC nor a query clause. In Example 1, one can obtain $C_3 = \neg A_2(x_1,z) \lor B_1(gx_1y_1) \lor \neg G_1(x_1,y_1)$ using Res with LGF-Refine. Although $C_3$ is simple, $C_3$ is neither an LGC (not weakly covering), nor a query clause (not flat).

We observe that by disallowing multiple positive non-ground compound literals in LGCs, using Res, the resolvent $R$ of a query clause and Horn LGCs is a negative clause containing no non-ground compound term, thus $R$ is a query clause. If we change $C_1$ in Example 1 to a Horn LGC $C'_1 = A_1(fx_1y_1,x_1) \lor \neg G_1(x_1,y_1)$, then by applying Res with refinement LGF-Refine, since $x$ is a top variable, resolution between $Q,C'_1$ derives $C'_3 = \neg A_2(x_1,z) \lor \neg G_1(x_1,y_1)$, which is a query clause. According to LGF-Refine, further inference between $C_2$ and $C'_3$ requires another positive premise that contains a positive literal $G_1$ that is either ground or contains non-ground compound terms. Hence, no inference is performed between $C_2$ and $C'_3$.

We show that Query-Res can decide $C \cup Q \models \top$ where $C$ are formulae in Horn LGF and $Q$ is a query clause. Notice that ground atoms $D$ are immediately in Horn LGF. A Horn loosely guarded clause (Horn LGC) is an LGC that contains at most one positive literal. The Horn loosely guarded fragment (Horn LGF) is a subset of LGF that can be transformed into a set of Horn LGCs using LGF-Trans.

Using Res, when query pair clauses are query clauses and a set of Horn LGCs, the resolvents are query clauses:

**Lemma 9.** In an application of Res, the resolvents of Horn LGCs and a query clause are query clauses.

It turns out the proof of Lemma 9 does not require the premises to be guarded. Thus we can generalise Lemma 9 to a result such that the resolvents of a query clause and a set of simple, weakly covering Horn clauses are query clauses.

Since a Horn LGC contains at most one positive literal, ordered factoring cannot be applied. Now we show other possibilities of applying resolution in Query-Res:

**Lemma 10.** Using Query-Res, the resolvents of Horn LGCs are Horn LGCs.

So far we showed that query clauses and Horn LGCs are closed under the inference system Query-Res, and the variable depth of derived clauses does not increase. Now we consider the length of derived clauses:

**Lemma 11.** In an application of Query-Res, given a finite set of fixed Horn LGCs and fixed query clauses, all derived clauses are fixed query clauses.

Now we give the second main result of this paper:

**Theorem 5.** Query-Res decides the problem of the BCQ answering for Horn LGF.

**Restricted Queries for LGF**

In this section, we answer loosely guarded queries, star queries and cloud queries over LGF. Theorem 4 implies that if a query clause $Q$ is expressible in LGCs, then using LCF-Res, one can immediately answer $Q$ over LGF. E.g., one can answer a loosely guarded query $\exists xyz(\text{Postgrad}(x) \land \text{citedBy}(x,y) \land \text{citedBy}(y,z) \land \text{citedBy}(z,x))$ over LGF using LCF-Res. This result is formally stated as:

**Corollary 3.** LCF-Res decides the loosely guarded query answering problem for LGF.

Another observation from Example 1 is: The top variable $x$ does not occur with all other variables in $Q$. If $y$, which occurs with all other variables in $Q$, is a top variable, then using Res, the resolvent among $Q,C_1,C_2$ is an LGC. E.g., if we change $C_1$ to $C'_1 = A_1(x_1,fx_1y_1) \lor B_1(gx_1y_1) \lor \neg G_1(x_1,y_1)$, to make $y$ a top variable, then using Res, the resolvent of $Q,C'_1,C_2$ is an LGC $B_1(gx_1y_1) \lor \neg G_1(x_1,y_1) \lor \neg G_2(x_1,y_1)$. This observation motivates our definition of star queries and cloud queries, which both guarantee the co-occurrence property between top variables and all other variables in the query clause.

Before discussing star/cloud queries, we first give the definition of partners of a variable. The partner of a variable $x$ in a query $Q \text{par}(x,Q)$ is a set of variables that co-occur with $x$ in an atom of $Q$, and $\text{par}(\pi,Q)$ is interpreted as $\text{par}(x_1,Q) \cup \ldots \cup \text{par}(x_n,Q)$ where $x_i$ is in an atom of $Q$. E.g., Let a query $Q = \exists xyz(A_1xy \land A_2yz)$. Then $\text{par}(x,Q) = \{y\}$ and $\text{par}(x,z,Q) = \text{par}(x,Q) \cup \text{par}(z,Q) = \{y\}$.

**Definition 7** (Star Query). A BCQ $Q$ is a star query if $Q$ contains a top variable $x$ such that $\text{par}(x,Q) = \text{var}(Q)$.

The notion of star query strictly extends that of loosely guarded query, since only one top variable need to occur with all other variables. E.g., a query clause $Q = \exists xyz(A_1xy \land A_2yz)$ is not a loosely guarded query since $x$ and $z$ do not co-occur in any literal in $Q$, but if $y$ is a top variable, $Q$ is a star query.

**Definition 8** (Cloud Query). A BCQ $Q$ is a cloud query if $Q$ satisfies these conditions:

1. $Q$ contains chained variables $V$ that are top variables.
2. Each pair in $V$ co-occur in an atom of $Q$.
3. $\text{par}(V,Q) = \text{var}(Q)$.

The notion of cloud query is a further extension of that of a star query since a top variable does not need to occur with all other variables. An example of a cloud, but not star query is $Q = \exists xyzuv(A_1xy \land A_2yz \land A_3zuw \land A_4v)$ if $y,z$ are top variables, $Q$ is a cloud query because that $V = \{y,z\}$, $y,z$ co-occur in $A_2$, and $\text{par}(V,Q) = \text{var}(Q)$.

Unlike loosely guarded queries, star queries and cloud queries vary depending on whether the top variables co-occur with all other variables.

Now we show that the resolvents of a set of LGCs and a star/cloud query are LGCs:

**Lemma 12.** Using Res with refinement LGF-Refine, the resolvents of a set of LGCs and a star/cloud query are LGCs.

Following the same idea of Lemma 8, we can show that the resolvents of a set of LGCs and a star/cloud query cannot have more variables than one of their side premises. Now we can state the third result of this paper:

**Theorem 6.** Query-Res decides the problem of the loosely guarded query and star/cloud query answering for LGF.
Conclusion and Future Work

In this paper, we have presented, as far as we know, the first practical decision procedures for answering BCQs over Horn LGF. Inspired by the ‘MAXVAR’ notion from (de Nivelle and de Rijke 2003), we used the top variable technique to handle chained variables in query clauses. Based on this top variable technique, we showed the method LGF-Res decides LGF, the method Query-Res answers BCQs over Horn LGF, and it answers loosely guarded queries, star queries and cloud queries over LGF. This shows that Query-Res provides essentials for the implementation of query answering over guard-related fragments as an extension for existing first-order logic reasoners.

Using Query-Res, an issue of answering BCQs against the whole of LGF is that: If the top variable is an isolated variable in a query clause $Q$, then the resolvents of $Q$ and a set of LGCs is neither a query clause nor an LGC (Example 1).

Our next step is extending Query-Res so that we can use it to answer BCQs against LGF, for which as yet there is no practical procedure.

Acknowledgements

We very thank the anonymous reviewers for their useful comments.

References


