

# Layered Fixed Point Logic

Piotr Filipiuk, Flemming Nielson, and Hanne Riis Nielson

DTU Informatics, Richard Petersens Plads, Technical University of Denmark,  
DK-2800 Kongens Lyngby, Denmark  
{pifi,nielson,riis}@imm.dtu.dk

**Abstract.** We present a logic for the specification of static analysis problems that goes beyond the logics traditionally used. Its most prominent feature is the direct support for both inductive computations of behaviors as well as co-inductive specifications of properties. Two main theoretical contributions are a Moore Family result and a parametrized worst case time complexity result. We show that the logic and the associated solver can be used for rapid prototyping and illustrate a wide variety of applications within Static Analysis, Constraint Satisfaction Problems and Model Checking. In all cases the complexity result specializes to the worst case time complexity of the classical methods.

## 1 Introduction

Static analysis [12,20] is a successful approach to the validation of properties of programming languages. It can be seen as a two-phase process where we first transform the analysis problem into a set of constraints that, in the second phase, is solved to produce the analysis result of interest. The constraints may be expressed in a language tailored to the problem at hand, or they may be expressed in a general purpose constraint language such as Datalog [1,5] or ALFP [21].

Model checking [13,2] is an automatic technique for verifying hardware and more recently software systems. Specifications are expressed in modal logic, whereas the system is modeled as a transition system or a Kripke structure. Given a system description the model checking algorithm either proves that the system satisfies the property, or reports a counterexample that violates it.

Constraint Satisfaction Problems (CSPs) [18] are the subject of intense research in both artificial intelligence and operations research. They consist of variables with constraints on them, and many real-world problems can be described as CSPs. A major challenge in constraint programming is to develop efficient generic approaches to solve instances of the CSP.

In this paper we present a logic for specification of analysis problems that goes beyond the logics traditionally used. Its most prominent feature is the direct support for both inductive computations of behaviors as well as co-inductive specifications of properties. At the same time the approach taken falls within the Abstract Interpretation [9,8] framework, thus there always is a unique best solution to the analysis problem considered. We show that the logic and the

associated solver can be used for rapid prototyping and illustrate a wide variety of applications within Static Analysis, Constraint Satisfaction Problems and Model Checking.

One can notice a resemblance of the logic to modal  $\mu$ -calculus [16,13], which is extensively used in various areas of computer science such as e.g computer-aided verification. Its defining feature is the addition of least and greatest fixpoint operators to modal logic; thus it achieves a great increase in expressive power, but at the same time an equally great increase in difficulty of understanding.

The paper is organized as follows. In Section 2 we define the syntax and semantics of LFP. In Section 3 we establish a Moore Family result and estimate the worst case time complexity. In Section 4 we show an application of LFP to Static Analysis. We continue in Section 5 with an application to the Constraint Satisfaction Problem. An application to Model Checking is presented in Section 6. We conclude in Section 7.

## 2 Syntax and Semantics

In this section, we introduce Layered Fixed Point Logic (abbreviated LFP). The LFP formulae are made up of layers. Each layer can either be a *define* formula which corresponds to the inductive definition, or a *constrain* formula corresponding to the co-inductive specification. The following definition introduces the syntax of LFP.

**Definition 1.** *Given a fixed countable set  $\mathcal{X}$  of variables, a non-empty universe  $\mathcal{U}$ , a finite set of function symbols  $\mathcal{F}$ , and a finite alphabet  $\mathcal{R}$  of predicate symbols, we define the set of LFP formulae, *cls*, together with clauses, *cl*, conditions, *cond*, constraints, *con*, definitions, *def*, and terms *u* by the grammar:*

$$\begin{aligned}
u &::= x \mid f(\mathbf{u}) \\
cond &::= R(\mathbf{x}) \mid \neg R(\mathbf{x}) \mid cond_1 \wedge cond_2 \mid cond_1 \vee cond_2 \\
&\quad \mid \exists x : cond \mid \forall x : cond \mid true \mid false \\
def &::= cond \Rightarrow R(\mathbf{u}) \mid \forall x : def \mid def_1 \wedge def_2 \\
con &::= R(\mathbf{u}) \Rightarrow cond \mid \forall x : con \mid con_1 \wedge con_2 \\
cl_i &::= define(def) \mid constrain(con) \\
cls &::= cl_1, \dots, cl_s
\end{aligned}$$

Here  $x \in \mathcal{X}$ ,  $R \in \mathcal{R}$ ,  $f \in \mathcal{F}$  and  $1 \leq i \leq s$ . We say that  $s$  is the order of the LFP formula  $cl_1, \dots, cl_s$ .

We allow to write  $R(\mathbf{u})$  for  $true \Rightarrow R(\mathbf{u})$ ,  $\neg R(\mathbf{u})$  for  $R(\mathbf{u}) \Rightarrow false$  and we abbreviate zero-arity functions  $f()$  as  $f \in \mathcal{U}$ . Occurrences of  $R(\mathbf{x})$  and  $\neg R(\mathbf{x})$  in conditions are called positive and negative queries, respectively. Occurrences of  $R(\mathbf{u})$  on the right hand side of the implication in define formulas are called defined occurrences. Occurrences of  $R(\mathbf{u})$  on the left hand side of the implication in constrain formulas are called constrained occurrences. Defined and constrained occurrences are jointly called assertions.

In order to ensure desirable theoretical and pragmatic properties in the presence of negation, we impose a notion of *stratification* similar to the one in Datalog [1,5]. Intuitively, stratification ensures that a negative query is not performed until the predicate has been fully asserted (defined or constrained). This is important for ensuring that once a condition evaluates to true it will continue to be true even after further assertions of predicates.

**Definition 2.** *The formula  $cl_1, \dots, cl_s$  is stratified if for all  $i = 1, \dots, s$  the following properties hold:*

- *Relations asserted in  $cl_i$  must not be asserted in  $cl_{i+1}, \dots, cl_s$*
- *Relations positively used in  $cl_i$  must not be asserted in  $cl_{i+1}, \dots, cl_s$*
- *Relations negatively used in  $cl_i$  must not be asserted in  $cl_i, \dots, cl_s$*

The function  $rank : \mathcal{R} \rightarrow \{0, \dots, s\}$  is then uniquely defined as

$$rank(R) = \max(\{0\} \cup \{i \mid R \text{ is asserted in } cl_i\})$$

*Example 1.* Using the notion of stratification we can define equality  $eq$  and non-equality  $neq$  predicates as follows

$$define(\forall x : true \Rightarrow eq(x, x)), define(\forall x : \forall y : \neg eq(x, y) \Rightarrow neq(x, y))$$

According to Definition 2 the formula is stratified, since predicate  $eq$  is negatively used only in the layer above the one that defines it.

To specify the semantics of LFP we introduce the interpretations  $\varrho$ ,  $\zeta$  and  $\varsigma$  of predicate symbols, function symbols and variables, respectively. Formally we have

$$\begin{aligned} \varrho &: \prod_k \mathcal{R}_{/k} \rightarrow \mathcal{P}(\mathcal{U}^k) \\ \zeta &: \prod_k \mathcal{F}_{/k} \rightarrow \mathcal{U}^k \rightarrow \mathcal{U} \\ \varsigma &: \mathcal{X} \rightarrow \mathcal{U} \end{aligned}$$

In the above  $\mathcal{R}_{/k}$  stands for a set of predicate symbols of arity  $k$ , then  $\mathcal{R}$  is a disjoint union of  $\mathcal{R}_{/k}$ , hence  $\mathcal{R} = \bigsqcup_k \mathcal{R}_{/k}$ . Similarly  $\mathcal{F}_{/k}$  is a set of function symbols of arity  $k$  and  $\mathcal{F} = \bigsqcup_k \mathcal{F}_{/k}$ . The interpretation of variables is given by  $\llbracket x \rrbracket(\zeta, \varsigma) = \varsigma(x)$ , where  $\varsigma(x)$  is the element from  $\mathcal{U}$  bound to  $x \in \mathcal{X}$ . Furthermore, the interpretation of function terms is defined as  $\llbracket f(\mathbf{u}) \rrbracket(\zeta, \varsigma) = \llbracket f \rrbracket(\zeta, []) (\llbracket \mathbf{u} \rrbracket(\zeta, \varsigma))$ . It is generalized to sequences  $\mathbf{u}$  of terms in a point-wise manner by taking  $\llbracket a \rrbracket(\zeta, \varsigma) = a$  for all  $a \in \mathcal{U}$ , and  $\llbracket (u_1, \dots, u_k) \rrbracket(\zeta, \varsigma) = (\llbracket u_1 \rrbracket(\zeta, \varsigma), \dots, \llbracket u_k \rrbracket(\zeta, \varsigma))$ .

The satisfaction relations for conditions  $cond$ , definitions  $def$  and constrains  $con$  are specified by:

$$(\varrho, \varsigma) \models cond, \quad (\varrho, \zeta, \varsigma) \models def \quad \text{and} \quad (\varrho, \zeta, \varsigma) \models con$$

The formal definition is given in Table 1; here  $\varsigma[x \mapsto a]$  stands for the mapping that is as  $\varsigma$  except that  $x$  is mapped to  $a$ .

**Table 1.** Semantics of LFP

$(\varrho, \varsigma) \models R(\mathbf{x})$	<u>iff</u> $\llbracket \mathbf{x} \rrbracket([\ ], \varsigma) \in \varrho(R)$
$(\varrho, \varsigma) \models \neg R(\mathbf{x})$	<u>iff</u> $\llbracket \mathbf{x} \rrbracket([\ ], \varsigma) \notin \varrho(R)$
$(\varrho, \varsigma) \models \text{cond}_1 \wedge \text{cond}_2$	<u>iff</u> $(\varrho, \varsigma) \models \text{cond}_1$ and $(\varrho, \varsigma) \models \text{cond}_2$
$(\varrho, \varsigma) \models \text{cond}_1 \vee \text{cond}_2$	<u>iff</u> $(\varrho, \varsigma) \models \text{cond}_1$ or $(\varrho, \varsigma) \models \text{cond}_2$
$(\varrho, \varsigma) \models \exists x : \text{cond}$	<u>iff</u> $(\varrho, \varsigma[x \mapsto a]) \models \text{cond}$ for some $a \in \mathcal{U}$
$(\varrho, \varsigma) \models \forall x : \text{cond}$	<u>iff</u> $(\varrho, \varsigma[x \mapsto a]) \models \text{cond}$ for all $a \in \mathcal{U}$
$(\varrho, \varsigma) \models \text{true}$	<u>iff</u> always
$(\varrho, \varsigma) \models \text{false}$	<u>iff</u> never
$(\varrho, \zeta, \varsigma) \models R(\mathbf{u})$	<u>iff</u> $\llbracket \mathbf{u} \rrbracket(\zeta, \varsigma) \in \varrho(R)$
$(\varrho, \zeta, \varsigma) \models \text{def}_1 \wedge \text{def}_2$	<u>iff</u> $(\varrho, \zeta, \varsigma) \models \text{def}_1$ and $(\varrho, \zeta, \varsigma) \models \text{def}_2$
$(\varrho, \zeta, \varsigma) \models \text{cond} \Rightarrow R(\mathbf{u})$	<u>iff</u> $(\varrho, \zeta, \varsigma) \models R(\mathbf{u})$ whenever $(\varrho, \zeta, \varsigma) \models \text{cond}$
$(\varrho, \zeta, \varsigma) \models \forall x : \text{def}$	<u>iff</u> $(\varrho, \zeta, \varsigma[x \mapsto a]) \models \text{def}$ for all $a \in \mathcal{U}$
$(\varrho, \zeta, \varsigma) \models R(\mathbf{u})$	<u>iff</u> $\llbracket \mathbf{u} \rrbracket(\zeta, \varsigma) \in \varrho(R)$
$(\varrho, \zeta, \varsigma) \models \text{con}_1 \wedge \text{con}_2$	<u>iff</u> $(\varrho, \zeta, \varsigma) \models \text{con}_1$ and $(\varrho, \zeta, \varsigma) \models \text{con}_2$
$(\varrho, \zeta, \varsigma) \models R(\mathbf{u}) \Rightarrow \text{cond}$	<u>iff</u> $(\varrho, \zeta, \varsigma) \models \text{cond}$ whenever $(\varrho, \zeta, \varsigma) \models R(\mathbf{u})$
$(\varrho, \zeta, \varsigma) \models \forall x : \text{con}$	<u>iff</u> $(\varrho, \zeta, \varsigma[x \mapsto a]) \models \text{con}$ for all $a \in \mathcal{U}$
$(\varrho, \zeta, \varsigma) \models cl_1, \dots, cl_s$	<u>iff</u> $(\varrho, \zeta, \varsigma) \models cl_i$ for all $1 \leq i \leq s$

### 3 Optimal Solutions

*Moore Family.* First we establish a Moore family result for LFP, which guarantees that there always is a unique best solution for LFP formulae.

**Definition 3.** A Moore family is a subset  $Y$  of a complete lattice  $L = (L, \sqsubseteq)$  that is closed under greatest lower bounds:  $\forall Y' \subseteq Y : \bigcap Y' \in Y$ .

It follows that a Moore family always contains a least element,  $\bigcap Y$ , and a greatest element,  $\bigcap \emptyset$ , which equals the greatest element,  $\top$ , from  $L$ ; in particular, a Moore family is never empty. The property is also called the model intersection property, since whenever we take a *meet* of a number of models we still get a model.

Let  $\Delta = \{\varrho \mid \varrho : \prod_k \mathcal{R}_{/k} \rightarrow \mathcal{P}(\mathcal{U}^k)\}$  denote the set of interpretations  $\varrho$  of predicate symbols in  $\mathcal{R}$  over  $\mathcal{U}$ . We define a lexicographical ordering  $\sqsubseteq$  defined by  $\varrho_1 \sqsubseteq \varrho_2$  if and only if there is some  $0 \leq j \leq s$ , where  $s$  is the order of the formula, such that the following properties hold:

- (a)  $\varrho_1(R) = \varrho_2(R)$  for all  $R \in \mathcal{R}$  with  $\text{rank}(R) < j$ ,
- (b)  $\varrho_1(R) \subseteq \varrho_2(R)$  for all  $R \in \mathcal{R}$  with  $\text{rank}(R) = j$  and either  $j = 0$  or  $R$  is a *defined* relation,
- (c)  $\varrho_1(R) \supseteq \varrho_2(R)$  for all  $R \in \mathcal{R}$  with  $\text{rank}(R) = j$  and  $R$  is a *constrained* relation,
- (d) either  $j = s$  or  $\varrho_1(R) \neq \varrho_2(R)$  for some relation  $R \in \mathcal{R}$  with  $\text{rank}(R) = j$ .

**Lemma 1.**  $\sqsubseteq$  defines a partial order.

*Proof.* See Appendix A.  $\square$

**Lemma 2.**  $(\Delta, \sqsubseteq)$  is a complete lattice with the greatest lower bound given by

$$(\bigsqcap M)(R) = \begin{cases} \bigcap \{\varrho(R) \mid \varrho \in M_j\} & \text{if } \text{rank}(R) = j \text{ and} \\ & \text{either } j = 0 \text{ or } R \text{ is defined in } cl_j. \\ \bigcup \{\varrho(R) \mid \varrho \in M_j\} & \text{if } \text{rank}(R) = j \text{ and} \\ & R \text{ is constrained in } cl_j. \end{cases}$$

where

$$M_j = \{\varrho \in M \mid \forall R' : \text{rank}(R') < j \Rightarrow (\bigsqcap M)(R') = \varrho(R')\}$$

*Proof.* See Appendix B.  $\square$

Note that  $\bigsqcap M$  is well defined by induction on  $j$  observing that  $M_0 = M$  and  $M_j \subseteq M_{j-1}$ .

**Proposition 1.** Assume  $cls$  is a stratified LFP formula,  $\varsigma_0$  and  $\zeta_0$  are interpretations of the free variables and function symbols in  $cls$ , respectively. Furthermore,  $\varrho_0$  is an interpretation of all relations of rank 0. Then  $\{\varrho \mid (\varrho, \varsigma_0, \varsigma_0) \models cls \wedge \forall R : \text{rank}(R) = 0 \Rightarrow \varrho(R) \supseteq \varrho_0(R)\}$  is a Moore family.

*Proof.* See Appendix C.  $\square$

The result ensures that the approach falls within the framework of Abstract Interpretation [8,9]; hence we can be sure that there always is a single best solution for the analysis problem under consideration, namely the one defined in Proposition 1.

*Complexity.* The least model for LFP formulae guaranteed by Proposition 1 can be computed efficiently as summarized in the following result.

**Proposition 2.** For a finite universe  $\mathcal{U}$ , the best solution  $\varrho$  such that  $\varrho_0 \sqsubseteq \varrho$  of a LFP formula  $cl_1, \dots, cl_s$  (w.r.t. an interpretation of the constant symbols) can be computed in time

$$\mathcal{O}(|\varrho_0| + \sum_{1 \leq i \leq s} |cl_i| |\mathcal{U}|^{k_i})$$

where  $k_i$  is the maximal nesting depth of quantifiers in the  $cl_i$  and  $|\varrho_0|$  is the sum of cardinalities of predicates  $\varrho_0(R)$  of rank 0. We also assume unit time hash table operations (as in [19]).

*Proof.* See Appendix D.  $\square$

For *define* clauses a straightforward method that achieves the above complexity proceeds by instantiating all variables occurring in the input formula in all possible ways. The resulting formula has no free variables thus it can be solved by classical solvers for alternation-free Boolean equation systems [10] in linear time.

In case of *constrain* clauses we first dualize the problem by transforming the co-inductive specification into the inductive one. The transformation increases the size of the input formula by a constant factor. Thereafter, we proceed in the same way as for the *define* clauses.

In addition we need to take into account the number of known facts, which equals to the cardinality of all predicates of rank 0. As a result we get the complexity from Proposition 2.

*The solver.* We developed a state-of-the-art solver for LFP, which is implemented in continuation passing style using Haskell. The solver computes the least model guaranteed by Proposition 1 and has a worst case time complexity as given by Proposition 2. For many clauses it exhibits a running time substantially lower than the worst case time complexity. Indeed, [19] gives a formula estimating the less than worst case time complexity on a given clause.

The solver deals with stratification by computing the relations in increasing order on their rank and therefore the negations present no obstacles. The relations are represented as Ordered Binary Decision Diagrams (OBDDs), which were originally used in hardware verification. OBDDs can efficiently store a large number of states that share many commonalities [4,3], and have already been used in a number of program analyses proving to be very efficient. The algorithm is an extension of the symbolic algorithm presented in [11] and is based on the top-down solving approach of Le Charlier and van Hentenryck [6].

The solver automatically translates LFP formulae into highly efficient OBDD implementations. Since the OBDDs represent sets of tuples, the solver operates on entire relations at a time, rather than individual tuples. The cost of the OBDD operations depends on the size of the OBDD and not the number of tuples in the relation; hence dense relations can be computed efficiently as long as their encoded representations are compact.

## 4 Application to Data Flow Analysis

Datalog has already been used for program analysis in compilers [25,22,23]. In this section we present how the LFP logic can be used to specify analyses that are instances of Bit-Vector Frameworks, which are a special case of the Monotone Frameworks [20,14].

A Monotone Framework consists of (a) a property space that usually is a complete lattice  $L$  satisfying the Ascending Chain Condition, and (b) transfer functions, i.e. monotone functions from  $L$  to  $L$ . The property space is used to represent the data flow information, whereas transfer functions capture the behavior of actions. In the Bit-Vector Framework, the property space is a power

set of some finite set and all transfer functions are of the form  $f_n(x) = (x \setminus kill_n) \cup gen_n$ .

Throughout the section we assume that a program is represented as a control flow graph [15,20], which is a directed graph with one entry node (having no incoming edges) and one exit node (having no outgoing edges), called extremal nodes. The remaining nodes represent statements and have transfer functions associated with them.

*Backward may analyses.* Let us first consider backward may analyses expressed as an instance of the Monotone Frameworks. In the analyses, we require the least sets that solve the equations and we are able to detect properties satisfied by at least one path leading to the given node. The analyses use the reversed edges in the flow graph; hence the data flow information is propagated *against* the flow of the program starting at the exit node. The data flow equations are defined as follows

$$A(n) = \begin{cases} \iota & \text{if } n = n_{exit} \\ \bigcup \{f_n(A(n')) \mid (n, n') \in E\} & \text{otherwise} \end{cases}$$

where  $A(n)$  represents data flow information at the entry to the node  $n$ ,  $E$  is a set of edges in the control flow graph, and  $\iota$  is the initial analysis information. The first case in the above equation, initializes the exit node with the initial analysis information, whereas the second one joins the data flow information from different paths (using the reversed flow). We use  $\bigcup$  since we want to be able to detect properties satisfied by at least one path leading to the given node.

The LFP specification for backward may analyses consists of two conjuncts corresponding to two cases in the data flow equations. Since in case of may analyses we aim at computing the least solution, the specification is defined in terms of a *define* clause. The formula is obtained as

$$define \left( \begin{array}{c} \forall x : \iota(x) \Rightarrow A(n_{exit}, x) \\ \bigwedge_{(s,t) \in E} \forall x : (A(t, x) \wedge \neg kill_s(x)) \vee gen_s(x) \Rightarrow A(s, x) \end{array} \right)$$

The first conjunct initializes the exit node with initial analysis information, denoted by the predicate  $\iota$ . The second one propagates data flow information against the edges in the control flow graph, i.e. whenever we have an edge  $(s, t)$  in the control flow graph, we propagate data flow information from  $t$  to  $s$ , by applying the corresponding transfer function.

Notice that there is no explicit formula for joining analysis information from different paths, as it is the case in the data flow equations, but rather it is done implicitly. Suppose there are two distinct edges  $(s, p)$  and  $(s, q)$  in the flow graph, then we get

$$\begin{aligned} \forall x : \underbrace{(A(p, x) \wedge \neg kill_s(x)) \vee gen_s(x)}_{cond_p(x)} &\Rightarrow A(s, x) \\ \forall x : \underbrace{(A(q, x) \wedge \neg kill_s(x)) \vee gen_s(x)}_{cond_q(x)} &\Rightarrow A(s, x) \end{aligned}$$

which is equivalent to

$$\forall x : \text{cond}_p(x) \vee \text{cond}_q(x) \Rightarrow A(s, x)$$

*Forward must analyses.* Let us now consider the general pattern for defining forward must analyses. Here we require the largest sets that solve the equations and we are able to detect properties satisfied by all paths leading to a given node. The analyses propagate the data flow information along the edges of the flow graph starting at the entry node. The data flow equations are defined as follows

$$A(n) = \begin{cases} \iota & \text{if } n = n_{\text{entry}} \\ \bigcap \{f_n(A(n')) \mid (n', n) \in E\} & \text{otherwise} \end{cases}$$

where  $A(n)$  represents analysis information at the exit from the node  $n$ . Since we require the greatest solution, the greatest lower bound  $\bigcap$  is used to combine information from different paths.

The corresponding LFP specification is obtained as follows

$$\text{constrain} \left( \bigwedge_{(s,t) \in E} \forall x : A(n_{\text{entry}}, x) \Rightarrow \iota(x) \vee (A(s, x) \wedge \neg \text{kill}_t(x)) \vee \text{gen}_t(x) \right)$$

Since we aim at computing the greatest solution, the analysis is given by means of *constrain* clause. The first conjunct initializes the entry node with the initial analysis information, whereas the second one propagates the information along the edges in the control flow graph, i.e. whenever we have an edge  $(s, t)$  in the control flow graph, we propagate data flow information from  $s$  to  $t$ , by applying the corresponding transfer function.

The general patterns for defining forward may and backward must analyses follow similar pattern. In case of forward may analyses the data flow information is propagated along the edges of the flow graph and since we aim at computing the least solution, the analyses are given by means of *define* clauses. Backward must analyses, on the other hand, use reversed edges in the flow graph and are specified using *constrain* clauses.

In order to compute the least solution of the data flow equations, one can use a general iterative algorithm for Monotone Frameworks. The worst case complexity of the algorithm is  $\mathcal{O}(|E|h)$ , where  $|E|$  is the number of edges in the control flow graph, and  $h$  is the height of the underlying lattice [20]. For Bit-Vector Frameworks the lattice is a powerset of a finite set  $\mathcal{U}$ ; hence  $h$  is  $\mathcal{O}(|\mathcal{U}|)$ . This gives the complexity  $\mathcal{O}(|E||\mathcal{U}|)$ .

According to Proposition 2 the worst case time complexity of the LFP specification is  $\mathcal{O}(|\varrho_0| + \sum_{1 \leq i \leq |E|} |\mathcal{U}| |cl_i|)$ . Since the size of the clause  $cl_i$  is constant and the sum of cardinalities of predicates of rank 0 is  $\mathcal{O}(|N|)$  we get  $\mathcal{O}(|N| + |E||\mathcal{U}|)$ . Provided that  $|E| > |N|$  we achieve  $\mathcal{O}(|E||\mathcal{U}|)$  i.e. the same worst case complexity as the standard iterative algorithm.

It is common in the compiler optimization that various analyses are performed at the same time. Since LFP logic has direct support for both least fixed points and greatest fixed points, we can perform both may and must analyses at the same time by splitting the analyses into separate layers.



## 5 Application to Constraint Satisfaction

Arc consistency is a basic technique for solving Constraint Satisfaction Problems (CSP) and has various applications within e.g. Artificial Intelligence. Formally a CSP [18,26] problem can be defined as follows.

**Definition 4.** A Constraint Satisfaction Problem  $(N, D, C)$  consists of a finite set of variables  $N = \{x_1, \dots, x_n\}$ , a set of domains  $D = \{D_1, \dots, D_n\}$ , where  $x_i$  ranges over  $D_i$ , and a set of constraints  $C \subseteq \{c_{ij} \mid i, j \in N\}$ , where each constraint  $c_{ij}$  is a binary relation between variables  $x_i$  and  $x_j$ .

For simplicity we consider binary constraints only. Furthermore, we can represent a CSP problem as a directed graph in the following way.

**Definition 5.** A constraint graph of a CSP problem  $(N, D, C)$  is a directed graph  $G = (V, E)$  where  $V = N$  and  $E = \{(x_i, x_j) \mid c_{ij} \in C\}$ .

Thus vertices of the graph correspond to the variables and an edge in the graph between nodes  $x_i$  and  $x_j$  corresponds to the constraint  $c_{ij} \in C$ .

The arc consistency problem is formally stated in the following definition.

**Definition 6.** Given a CSP  $(N, D, C)$ , an arc  $(x_i, x_j)$  of its constraint graph is arc consistent if and only if  $\forall x \in D_i$ , there exists  $y \in D_j$  such that  $c_{ij}(x, y)$  holds, as well as  $\forall y \in D_j$ , there exists  $x \in D_i$  such that  $c_{ij}(x, y)$  holds. A CSP  $(N, D, C)$  is arc consistent if and only if each arc in its constraint graph is arc consistent.

The basic and widely used arc consistency algorithm is the AC-3 algorithm proposed in 1977 by Mackworth [18]. The complexity of the algorithm is  $O(ed^3)$ , where  $e$  is the number of constraints and  $d$  the size of the largest domain. The algorithm is used in many constraint solvers due to its simplicity and fairly good efficiency [24].

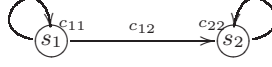
Now we show the LFP specification of the arc consistency problem. A domain of a variable  $x_i$  is represented as a unary relation  $D_i$ , and for each constraint  $c_{ij} \in C$  we have a binary relation  $C_{ij} \subseteq D_i \times D_j$ . Then we obtain

$$\text{constrain} \left( \bigwedge_{c_{ij} \in C} (\forall x : D_i(x) \Rightarrow \exists y : D_j(y) \wedge C_{ij}(x, y)) \wedge (\forall y : D_j(y) \Rightarrow \exists x : D_i(x) \wedge C_{ij}(x, y)) \right)$$

which exactly captures the conditions from Definition 6.

According to the Proposition 2 the above specification gives rise to the worst case complexity  $\mathcal{O}(ed^2)$ . The original AC-3 algorithm was optimized in [26] where it was shown that it achieves the worst case optimal time complexity of  $\mathcal{O}(ed^2)$ . Hence LFP specification is as efficient as the improved version of the AC-3 algorithm.

*Example 2.* As an example let us consider the following problem. Assume we have two processes  $P_1$  and  $P_2$  that need to be finished before 8 time units have elapsed. The process  $P_1$  is required to run for 3 or 4 time units, the process  $P_2$



**Fig. 1.** Arc consistency.

is required to run for precisely 2 time units, and  $P_2$  should start at the exact moment when  $P_1$  finishes.

The problem can be defined as an instance of CSP  $(N, D, C)$  where  $N = \{s_1, s_2\}$  denoting the starting times of the corresponding process. Since both processes need to be completed before 8 time units have elapsed we have  $D_1 = D_2 = \{0, \dots, 8\}$ . Moreover, we have the following constraints  $C = \{c_{12} = (3 \leq s_2 - s_1 \leq 4), c_{11} = (0 \leq s_1 \leq 4), c_{22} = (0 \leq s_2 \leq 6)\}$ . We can represent the above CSP problem as a constraint graph depicted in Figure 1. Furthermore it can be specified as the following LFP formulae

$$\begin{aligned} & \text{define } (\bigwedge_{0 \leq x \leq 4} C_1(x) \wedge \bigwedge_{0 \leq y \leq 6} C_2(y) \wedge \bigwedge_{3 \leq z \leq 4} C_{12}(z)), \\ & \text{constrain } \left( \begin{array}{l} (\forall x : D_1(x) \Rightarrow \exists y : D_2(y) \wedge C_{12}(y - x)) \wedge \\ (\forall y : D_2(y) \Rightarrow \exists x : D_1(x) \wedge C_{12}(y - x)) \end{array} \right) \end{aligned}$$

where we write  $y - x$  for a function  $f_{sub}(y, x)$ .

## 6 Application to Model Checking

This section is concerned with the application of the LFP logic to the model checking problem [2]. In particular we show how LFP can be used to specify a prototype model checker for a special purpose modal logic of interest. Here we illustrate the approach on the familiar case of Computation Tree Logic (CTL) [7]. Throughout this section, we assume that  $TS$  is finite and has no terminal states.

CTL distinguishes between state formulae and path formulae. CTL state formulae over the set  $AP$  of atomic propositions are formed according to the following grammar

$$\Phi ::= \text{true} \mid a \mid \Phi_1 \wedge \Phi_2 \mid \neg\Phi \mid \mathbf{E}\varphi \mid \mathbf{A}\varphi$$

where  $a \in AP$  and  $\varphi$  is a path formula. CTL path formulae are formed according to the following grammar

$$\varphi ::= \mathbf{X}\Phi \mid \Phi_1 \mathbf{U}\Phi_2 \mid \mathbf{G}\Phi$$

where  $\Phi$ ,  $\Phi_1$  and  $\Phi_2$  are state formulae. The satisfaction relation  $\models$  is defined for state formula by

$$\begin{aligned} s \models \text{true} & \quad \text{iff } \text{true} \\ s \models a & \quad \text{iff } a \in L(s) \\ s \models \neg\Phi & \quad \text{iff } \text{not } s \models \Phi \\ s \models \Phi_1 \wedge \Phi_2 & \quad \text{iff } s \models \Phi_1 \text{ and } s \models \Phi_2 \\ s \models \mathbf{E}\varphi & \quad \text{iff } \pi \models \varphi \text{ for some } \pi \in \text{Paths}(s) \\ s \models \mathbf{A}\varphi & \quad \text{iff } \pi \models \varphi \text{ for all } \pi \in \text{Paths}(s) \end{aligned}$$

where  $Paths(s)$  denote the set of maximal path fragments  $\pi$  starting in  $s$ . The satisfaction relation  $\models$  for path formulae is defined by

$$\begin{aligned} \pi &\models \mathbf{X}\Phi && \text{iff } \pi[1] \models \Phi \\ \pi &\models \Phi_1 \mathbf{U} \Phi_2 && \text{iff } \exists j \geq 0 : (\pi[j] \models \Phi_2 \wedge (\forall 0 \leq k < j : \pi[k] \models \Phi_1)) \\ \pi &\models \mathbf{G}\Phi && \text{iff } \forall j \geq 0 : \pi[j] \models \Phi \end{aligned}$$

where for path  $\pi = s_0 s_1 \dots$  and an integer  $i \geq 0$ ,  $\pi[i]$  denotes the  $(i+1)$ th state of  $\pi$ , i.e.  $\pi[i] = s_i$ .

The CTL model checking amounts to a recursive computation of the set  $Sat(\Phi)$  of all states satisfying  $\Phi$ , which is sometimes referred to as *global* model checking. The algorithm boils down to a bottom-up traversal of the abstract syntax tree of the CTL formula  $\Phi$ . The nodes of the abstract syntax tree correspond to the sub-formulae of  $\Phi$ , and leaves are either a constant *true* or an atomic proposition  $a \in AP$ .

**Table 2.** LFP specification of satisfaction sets

$$\begin{aligned} &define(\forall s : Sat_{true}(s)) \\ &define(\forall s : L_a(s) \Rightarrow Sat_a(s)) \\ &define(\forall s : Sat_{\Phi_1}(s) \wedge Sat_{\Phi_2}(s) \Rightarrow Sat_{\Phi_1 \wedge \Phi_2}(s)) \\ &define(\forall s : \neg Sat_{\Phi}(s) \Rightarrow Sat_{\neg \Phi}(s)) \\ \\ &define(\forall s : (\exists s' : T(s, s') \wedge Sat_{\Phi}(s')) \Rightarrow Sat_{\mathbf{EX}\Phi}(s)) \\ \\ &define(\forall s : (\forall s' : \neg T(s, s') \vee Sat_{\Phi}(s')) \Rightarrow Sat_{\mathbf{AX}\Phi}(s)) \\ \\ &define \left( \begin{aligned} &(\forall s : Sat_{\Phi_2}(s) \Rightarrow Sat_{\mathbf{E}[\Phi_1 \mathbf{U} \Phi_2]}(s)) \wedge \\ &(\forall s : Sat_{\Phi_1}(s) \wedge (\exists s' : T(s, s') \wedge Sat_{\mathbf{E}[\Phi_1 \mathbf{U} \Phi_2]}(s')) \Rightarrow Sat_{\mathbf{E}[\Phi_1 \mathbf{U} \Phi_2]}(s)) \end{aligned} \right) \\ \\ &define \left( \begin{aligned} &(\forall s : Sat_{\Phi_2}(s) \Rightarrow Sat_{\mathbf{A}[\Phi_1 \mathbf{U} \Phi_2]}(s)) \wedge \\ &(\forall s : Sat_{\Phi_1}(s) \wedge (\forall s' : \neg T(s, s') \vee Sat_{\mathbf{A}[\Phi_1 \mathbf{U} \Phi_2]}(s')) \Rightarrow Sat_{\mathbf{A}[\Phi_1 \mathbf{U} \Phi_2]}(s)) \end{aligned} \right) \\ \\ &constrain \left( \begin{aligned} &(\forall s : Sat_{\mathbf{EG}\Phi}(s) \Rightarrow Sat_{\Phi}(s)) \wedge \\ &(\forall s : Sat_{\mathbf{EG}\Phi}(s) \Rightarrow (\exists s' : T(s, s') \wedge Sat_{\mathbf{EG}\Phi}(s'))) \end{aligned} \right) \\ \\ &constrain \left( \begin{aligned} &(\forall s : Sat_{\mathbf{AG}\Phi}(s) \Rightarrow Sat_{\Phi}(s)) \wedge \\ &(\forall s : Sat_{\mathbf{AG}\Phi}(s) \Rightarrow (\forall s' : \neg T(s, s') \vee Sat_{\mathbf{AG}\Phi}(s'))) \end{aligned} \right) \end{aligned}$$

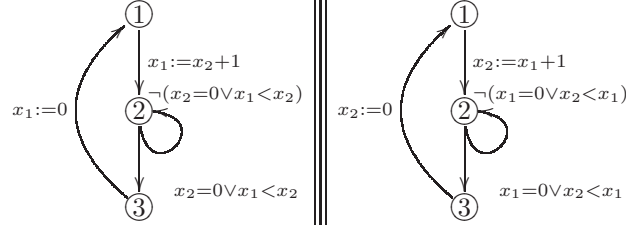
Now let us consider the LFP specification, where for each formula  $\Phi$  we define a relation  $Sat_{\Phi} \subseteq S$  characterizing states where  $\Phi$  hold. The specification is defined in Table 2. The clause for *true* is straightforward and says that *true* holds in all states. The clause for an atomic proposition  $a$  expresses that a state satisfies  $a$  whenever it is in  $L_a$ , where we assume that we have a predicate  $L_a \subseteq S$

for each  $a \in AP$ . The clause for  $\Phi_1 \wedge \Phi_2$  captures that a state satisfies  $\Phi_1 \wedge \Phi_2$  whenever it satisfies both  $\Phi_1$  and  $\Phi_2$ . Similarly a state satisfies  $\neg\Phi$  if it does not satisfy  $\Phi$ . The formula for  $\mathbf{EX}\Phi$  captures that a state  $s$  satisfies  $\mathbf{EX}\Phi$ , if there is a transition to state  $s'$  such that  $s'$  satisfies  $\Phi$ . The formula for  $\mathbf{AX}\Phi$  expresses that a state  $s$  satisfies  $\mathbf{AX}\Phi$  if for all states  $s'$ : either there is no transition from  $s$  to  $s'$ , or otherwise  $s'$  satisfies  $\Phi$ . The formula for  $\mathbf{E}[\Phi_1 \mathbf{U} \Phi_2]$  captures two possibilities. If a state satisfies  $\Phi_2$  then it also satisfies  $\mathbf{E}[\Phi_1 \mathbf{U} \Phi_2]$ . Alternatively if the state  $s$  satisfies  $\Phi_1$  and there is a transition to a state satisfying  $\mathbf{E}[\Phi_1 \mathbf{U} \Phi_2]$  then  $s$  also satisfies  $\mathbf{E}[\Phi_1 \mathbf{U} \Phi_2]$ . The formula  $\mathbf{A}[\Phi_1 \mathbf{U} \Phi_2]$  also captures two cases. If a state satisfies  $\Phi_2$  then it also satisfies  $\mathbf{A}[\Phi_1 \mathbf{U} \Phi_2]$ . Alternatively state  $s$  satisfies  $\mathbf{A}[\Phi_1 \mathbf{U} \Phi_2]$  if it satisfies  $\Phi_1$  and for all states  $s'$  either there is no transition from  $s$  to  $s'$  or  $\mathbf{A}[\Phi_1 \mathbf{U} \Phi_2]$  is valid in  $s'$ . Let us now consider the formula for  $\mathbf{EG}\Phi$ . Since the set of states satisfying  $\mathbf{EG}\Phi$  is defined as a largest set satisfying the semantics of  $\mathbf{EG}\Phi$ , the property is defined by means of constrain clause. The first conjunct expresses that whenever a state satisfies  $\mathbf{EG}\Phi$  it also satisfies  $\Phi$ . The second conjunct says that if a state satisfies  $\mathbf{EG}\Phi$  then there exists a transition to a state  $s'$  such that  $s'$  satisfies  $\mathbf{EG}\Phi$ . Finally let us consider the formula for  $\mathbf{AG}\Phi$ , which is also defined in terms of constrain clause and distinguishes between two cases. In the first one whenever a state satisfies  $\mathbf{AG}\Phi$ , it also satisfies  $\Phi$ . Alternatively, if a state  $s$  satisfies  $\mathbf{AG}\Phi$  then for all states  $s'$ : either there is no transition from  $s$  to  $s'$  or otherwise  $s'$  satisfies  $\mathbf{AG}\Phi$ .

The generation of clauses for  $Sat_\Phi$  is performed in the postorder traversal over  $\Phi$ ; hence the clauses defining sub-formulas of  $\Phi$  are defined in the lower layers. It is important to note that the specification in Table 2 is both correct and precise. It follows that an implementation of the given specification of CTL by means of the LFP solver constitutes a model checker for CTL.

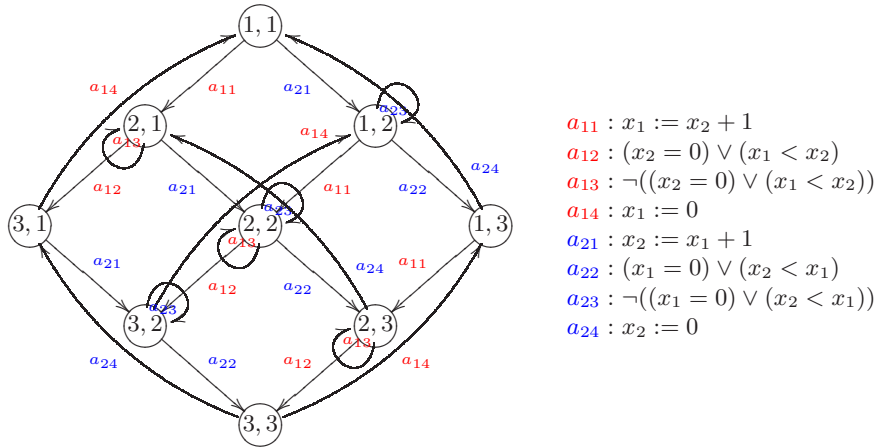
We may estimate the worst case time complexity of model checking performed using LFP. Consider a CTL formula  $\Phi$  of size  $|\Phi|$ ; it is immediate that the LFP clause has size  $\mathcal{O}(|\Phi|)$ , and the nesting depth is at most 2. According to Proposition 2 the worst case time complexity of the LFP specification is  $\mathcal{O}(|S| + |S|^2|\Phi|)$ , where  $|S|$  is the number of states in the transition system. Using a more refined reasoning than that of Proposition 2 we obtain  $\mathcal{O}(|S| + |T||\Phi|)$ , where  $|T|$  is the number of transitions in the transition system. It is due to the fact that the "double quantifications" over states in Table 2 really correspond to traversing all possible transitions rather than all pairs of states. Thus our LFP model checking algorithm has the same worst case complexity as classical model checking algorithms [2].

*Example 3.* As an example let us consider the Bakery mutual exclusion algorithm [17]. Although the algorithm is designed for an arbitrary number of processes, we consider the simpler setting with two processes. Let  $P_1$  and  $P_2$  be the two processes, and  $x_1$  and  $x_2$  be two shared variables both initialized to 0. We can represent the algorithm as an interleaving of two program graphs [2], which are directed graphs where actions label the edges rather than the nodes. The algorithm is as follows



The variables  $x_1$  and  $x_2$  are used to resolve the conflict when both processes want to enter the critical section. When  $x_i$  is equal to zero, the process  $P_i$  is not in the critical section and does not attempt to enter it — the other one can safely proceed to the critical section. Otherwise, if both shared variables are non-zero, the process with smaller “ticket” (i.e. value of the corresponding variable) can enter the critical section. This reasoning is captured by the conditions of busy-waiting loops. When a process wants to enter the critical section, it simply takes the next “ticket” hence giving priority to the other process.

From the algorithm above, we can obtain a program graph corresponding to the interleaving of the two processes, which is depicted in Figure 2.



**Fig. 2.** Interleaved program graph.

The CTL formulation of the mutual exclusion property is  $AG\neg(crit_1 \wedge crit_2)$ , which states that along all paths globally it is never the case that  $crit_1$  and  $crit_2$  hold at the same time.

As already mentioned, in order to specify the problem we proceed bottom up by specifying formulae for the sub problems. After a bit of simplification we

obtain the following LFP clauses

$$\begin{aligned} & \text{define}(\forall s : L_{crit_1}(s) \wedge L_{crit_2}(s) \Rightarrow Sat_{crit}(s)), \\ & \text{constrain} \left( \begin{aligned} & (\forall s : Sat_{AG(\neg crit)}(s) \Rightarrow \neg Sat_{crit}(s)) \wedge \\ & (\forall s : Sat_{AG(\neg crit)}(s) \Rightarrow (\forall s' : \neg T(s, s') \vee Sat_{AG(\neg crit)}(s'))) \end{aligned} \right) \end{aligned}$$

where relation  $L_{crit_1}$  (respectively  $L_{crit_2}$ ) characterizes states in the interleaved program graph that correspond to process  $P_1$  (respectively  $P_2$ ) being in the critical section. Furthermore, the  $AG$  modality is defined by means of a constrain clause. The first conjunct expresses that whenever a state satisfies a mutual exclusion property  $AG(\neg crit)$  it does not satisfy  $crit$ . The second one states that if a state satisfies a mutual exclusion property then all successors do as well, i.e. for an arbitrary state, it is either not a successor or else satisfies the mutual exclusion property.

## 7 Conclusions

In the paper we introduced the Layered Fixed Point Logic, which is a suitable formalism for the specification of analysis problems. Its most prominent feature is the direct support for both inductive as well as co-inductive specifications of properties.

We established a Moore Family result that guarantees that there always is a best solution for the LFP formulae. More generally this ensures that the approach taken falls within the general Abstract Interpretation framework. Other theoretical contribution is the parametrized worst case time complexity result, which provide a simple characterization of the running time of the LFP programs.

We developed a state-of-the-art solving algorithm for LFP, which is a continuation passing style algorithm based on OBDD representations of relations. The solver achieves the best known theoretical complexity bounds, and for many clauses exhibit a running time substantially lower than the worst case time complexity.

We showed that the logic and the associated solver can be used for rapid prototyping by presenting applications within Static Analysis, Constraint Satisfaction Problems and Model Checking. In all cases the complexity result specializes to the worst case time complexity of classical results.

## References

1. Apt, K.R., Blair, H.A., Walker, A.: Towards a theory of declarative knowledge. In: Foundations of Deductive Databases and Logic Programming., pp. 89–148. Morgan Kaufmann (1988)
2. Baier, C., Katoen, J.P.: Principles of Model Checking (Representation and Mind Series). The MIT Press (2008)
3. Bryant, R.E.: Graph-based algorithms for boolean function manipulation. IEEE Trans. Computers 35(8), 677–691 (1986)

4. Bryant, R.E.: Symbolic boolean manipulation with ordered binary-decision diagrams. *ACM Comput. Surv.* 24(3), 293–318 (1992)
5. Chandra, A.K., Harel, D.: Computable queries for relational data bases (preliminary report). In: *STOC*. pp. 309–318 (1979)
6. Charlier, B.L., Hentenryck, P.V.: A universal top-down fixpoint algorithm. Tech. rep., CS-92-25, Brown University (1992)
7. Clarke, E.M., Emerson, E.A.: Design and synthesis of synchronization skeletons using branching-time temporal logic. In: *Logic of Programs*. pp. 52–71 (1981)
8. Cousot, P., Cousot, R.: Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fixpoints. In: *POPL*. pp. 238–252 (1977)
9. Cousot, P., Cousot, R.: Systematic design of program analysis frameworks. In: *POPL*. pp. 269–282 (1979)
10. Dowling, W.F., Gallier, J.H.: Linear-time algorithms for testing the satisfiability of propositional horn formulae. *J. Log. Program.* 1(3), 267–284 (1984)
11. Filipiuk, P., Nielson, H.R., Nielson, F.: Explicit versus symbolic algorithms for solving ALFP constraints. *Electr. Notes Theor. Comput. Sci.* 267(2), 15–28 (2010)
12. Hecht, M.S.: *Flow Analysis of Computer Programs*. North Holland (1977)
13. (Jr.), E.M.C., Grumberg, O., Peled, D.A.: *Model Checking*. MIT Press (1999)
14. Kam, J.B., Ullman, J.D.: Monotone data flow analysis frameworks. *Acta Inf.* 7, 305–317 (1977)
15. Kildall, G.A.: A unified approach to global program optimization. In: *POPL*. pp. 194–206 (1973)
16. Kozen, D.: Results on the propositional  $\mu$ -calculus. *Theor. Comput. Sci.* 27, 333–354 (1983)
17. Lamport, L.: A new solution of Dijkstra’s concurrent programming problem. *Commun. ACM* 17(8), 453–455 (1974)
18. Mackworth, A.K.: Consistency in networks of relations. *Artif. Intell.* 8(1), 99–118 (1977)
19. McAllester, D.A.: On the complexity analysis of static analyses. *J. ACM* 49(4), 512–537 (2002)
20. Nielson, F., Nielson, H.R., Hankin, C.: *Principles of Program Analysis*. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1999)
21. Nielson, F., Seidl, H., Nielson, H.R.: A Succinct Solver for ALFP. *Nord. J. Comput.* 9(4), 335–372 (2002)
22. Reps, T.W.: Demand interprocedural program analysis using logic databases. In: *Workshop on Programming with Logic Databases (Book)*, ILPS. pp. 163–196 (1993)
23. Ullman, J.D.: Bottom-up beats top-down for datalog. In: *PODS*. pp. 140–149 (1989)
24. Wallace, R.J.: Why AC-3 is almost always better than AC4 for establishing arc consistency in cps. In: *IJCAI*. pp. 239–247 (1993)
25. Whaley, J., Lam, M.S.: Cloning-based context-sensitive pointer alias analysis using binary decision diagrams. In: *PLDI*. pp. 131–144 (2004)
26. Zhang, Y., Yap, R.H.C.: Making AC-3 an optimal algorithm. In: *IJCAI*. pp. 316–321 (2001)

These appendices are not intended for publication and references to them will be removed in the final version.

## A Proof of Lemma 1

*Proof.* **Reflexivity**  $\forall \varrho \in \Delta : \varrho \sqsubseteq \varrho$ .

To show that  $\varrho \sqsubseteq \varrho$  let us take  $j = s$ . If  $\text{rank}(R) < j$  then  $\varrho(R) = \varrho(R)$  as required. Otherwise if  $\text{rank}(R) = j$  and either  $R$  is a defined relation or  $j = 0$ , then from  $\varrho(R) = \varrho(R)$  we get  $\varrho(R) \subseteq \varrho(R)$ . The last case is when  $\text{rank}(R) = j$  and  $R$  is a constrained relation. Then from  $\varrho(R) = \varrho(R)$  we get  $\varrho(R) \supseteq \varrho(R)$ . Thus we get the required  $\varrho \sqsubseteq \varrho$ .

**Transitivity**  $\forall \varrho_1, \varrho_2, \varrho_3 \in \Delta : \varrho_1 \sqsubseteq \varrho_2 \wedge \varrho_2 \sqsubseteq \varrho_3 \Rightarrow \varrho_1 \sqsubseteq \varrho_3$ .

Let us assume that  $\varrho_1 \sqsubseteq \varrho_2 \wedge \varrho_2 \sqsubseteq \varrho_3$ . From  $\varrho_i \sqsubseteq \varrho_{i+1}$  we have  $j_i$  such that conditions (a)–(d) are fulfilled for  $i = 1, 2$ . Let us take  $j$  to be the minimum of  $j_1$  and  $j_2$ . Now we need to verify that conditions (a)–(d) hold for  $j$ . If  $\text{rank}(R) < j$  we have  $\varrho_1(R) = \varrho_2(R)$  and  $\varrho_2(R) = \varrho_3(R)$ . It follows that  $\varrho_1(R) = \varrho_3(R)$ , hence (a) holds. Now let us assume that  $\text{rank}(R) = j$  and either  $R$  is a defined relation or  $j = 0$ . We have  $\varrho_1(R) \subseteq \varrho_2(R)$  and  $\varrho_2(R) \subseteq \varrho_3(R)$  and from transitivity of  $\subseteq$  we get  $\varrho_1(R) \subseteq \varrho_3(R)$ , which gives (b). Alternatively  $\text{rank}(R) = j$  and  $R$  is a constrained relation. We have  $\varrho_1(R) \supseteq \varrho_2(R)$  and  $\varrho_2(R) \supseteq \varrho_3(R)$  and from transitivity of  $\supseteq$  we get  $\varrho_1(R) \supseteq \varrho_3(R)$ , thus (c) holds. Let us now assume that  $j \neq s$ , hence  $\varrho_i(R) \neq \varrho_{i+1}(R)$  for some  $R \in \mathcal{R}$  and  $i = 1, 2$ . Without loss of generality let us assume that  $\varrho_1(R) \neq \varrho_2(R)$ . In case  $R$  is a defined relation we have  $\varrho_1(R) \subsetneq \varrho_2(R)$  and  $\varrho_2(R) \subseteq \varrho_3(R)$ , hence  $\varrho_1(R) \neq \varrho_3(R)$ . Similarly in case  $R$  is a constrained relation we have  $\varrho_1(R) \supsetneq \varrho_2(R)$  and  $\varrho_2(R) \supseteq \varrho_3(R)$ . Hence  $\varrho_1(R) \neq \varrho_3(R)$ , and (d) holds.

**Anti-symmetry**  $\forall \varrho_1, \varrho_2 \in \Delta : \varrho_1 \sqsubseteq \varrho_2 \wedge \varrho_2 \sqsubseteq \varrho_1 \Rightarrow \varrho_1 = \varrho_2$ .

Let us assume  $\varrho_1 \sqsubseteq \varrho_2$  and  $\varrho_2 \sqsubseteq \varrho_1$ . Let  $j$  be minimal such that  $\text{rank}(R) = j$  and  $\varrho_1(R) \neq \varrho_2(R)$  for some  $R \in \mathcal{R}$ . If  $j = 0$  or  $R$  is a defined relation, then we have  $\varrho_1(R) \subseteq \varrho_2(R)$  and  $\varrho_2(R) \subseteq \varrho_1(R)$ . Hence  $\varrho_1(R) = \varrho_2(R)$  which is a contradiction. Similarly if  $R$  is a constrained relation we have  $\varrho_1(R) \supseteq \varrho_2(R)$  and  $\varrho_2(R) \supseteq \varrho_1(R)$ . It follows that  $\varrho_1(R) = \varrho_2(R)$ , which again is a contradiction. Thus it must be the case that  $\varrho_1(R) = \varrho_2(R)$  for all  $R \in \mathcal{R}$ .  $\square$

## B Proof of Lemma 2

*Proof.* First we prove that  $\sqcap M$  is a lower bound of  $M$ ; that is  $\sqcap M \sqsubseteq \varrho$  for all  $\varrho \in M$ . Let  $j$  be maximum such that  $\varrho \in M_j$ ; since  $M = M_0$  and  $M_j \supseteq M_{j+1}$  clearly such  $j$  exists. From definition of  $M_j$  it follows that  $(\sqcap M)(R) = \varrho(R)$  for all  $R$  with  $\text{rank}(R) < j$ ; hence (a) holds.

If  $\text{rank}(R) = j$  and either  $R$  is a defined relation or  $j = 0$  we have  $(\sqcap M)(R) = \bigcap \{\varrho'(R) \mid \varrho' \in M_j\} \subseteq \varrho(R)$  showing that (b) holds.

Similarly, if  $R$  is a constrained relation with  $\text{rank}(R) = j$  we have  $(\sqcap M)(R) = \bigcup \{\varrho'(R) \mid \varrho' \in M_j\} \supseteq \varrho(R)$  showing that (c) holds.



Finally let us assume that  $j \neq s$ ; we need to show that there is some  $R$  with  $\text{rank}(R) = j$  such that  $(\prod M)(R) \neq \varrho(R)$ . Since we know that  $j$  is maximum such that  $\varrho \in M_j$ , it follows that  $\varrho \notin M_{j+1}$ , hence there is a relation  $R$  with  $\text{rank}(R) = j$  such that  $(\prod M)(R) \neq \varrho(R)$ ; thus (d) holds.

Now we need to show that  $\prod M$  is the greatest lower bound. Let us assume that  $\varrho' \sqsubseteq \varrho$  for all  $\varrho \in M$ , and let us show that  $\varrho' \sqsubseteq \prod M$ . If  $\varrho' = \prod M$  the result holds vacuously, hence let us assume  $\varrho' \neq \prod M$ . Then there exists a minimal  $j$  such that  $(\prod M)(R) \neq \varrho'(R)$  for some  $R$  with  $\text{rank}(R) = j$ . Let us first consider  $R$  such that  $\text{rank}(R) < j$ . By our choice of  $j$  we have  $(\prod M)(R) = \varrho'(R)$  hence (a) holds.

Next assume that  $\text{rank}(R) = j$  and either  $R$  is a defined relation of  $j = 0$ . Then  $\varrho' \sqsubseteq \varrho$  for all  $\varrho \in M_j$ . It follows that  $\varrho'(R) \subseteq \varrho(R)$  for all  $\varrho \in M_j$ . Thus we have  $\varrho'(R) \subseteq \bigcap \{\varrho(R) \mid \varrho \in M_j\}$ . Since  $(\prod M)(R) = \bigcap \{\varrho(R) \mid \varrho \in M_j\}$ , we have  $\varrho'(R) \subseteq (\prod M)(R)$  which proves (b).

Now assume  $\text{rank}(R) = j$  and  $R$  is a constrained relation. We have that  $\varrho' \sqsubseteq \varrho$  for all  $\varrho \in M_j$ . Since  $R$  is a constrained relation it follows that  $\varrho'(R) \supseteq \varrho(R)$  for all  $\varrho \in M_j$ . Thus we have  $\varrho'(R) \supseteq \bigcup \{\varrho(R) \mid \varrho \in M_j\}$ . Since  $(\prod M)(R) = \bigcup \{\varrho(R) \mid \varrho \in M_j\}$ , we have  $\varrho'(R) \supseteq (\prod M)(R)$  which proves (c).

Finally since we assumed that  $(\prod M)(R) \neq \varrho'(R)$  for some  $R$  with  $\text{rank}(R) = j$ , it follows that (d) holds. Thus we proved that  $\varrho' \sqsubseteq \prod M$ .  $\square$

## C Proof of Proposition 1

In order to prove Proposition 1 we first state and prove two auxiliary lemmas.

**Definition 7.** We introduce an ordering  $\subseteq_{/j}$  defined by  $\varrho_1 \subseteq_{/j} \varrho_2$  if and only if

- $\forall R : \text{rank}(R) < j \Rightarrow \varrho_1(R) = \varrho_2(R)$
- $\forall R : \text{rank}(R) = j \Rightarrow \varrho_1(R) \subseteq \varrho_2(R)$

**Lemma 3.** Assume a condition *cond* occurs in  $cl_j$ , and let  $\varsigma$  be a valuation of free variables in *cond*. If  $\varrho_1 \subseteq_{/j} \varrho_2$  and  $(\varrho_1, \varsigma) \models \text{cond}$  then  $(\varrho_2, \varsigma) \models \text{cond}$ .

*Proof.* We proceed by induction on  $j$  and in each case perform a structural induction on the form of the condition *cond* occurring in  $cl_j$ .

**Case:** *cond* =  $R(\mathbf{x})$

Assume  $\varrho_1 \subseteq_{/j} \varrho_2$  and

$$(\varrho_1, \varsigma) \models R(\mathbf{x})$$

From Table 1 it follows that

$$\llbracket \mathbf{x} \rrbracket([], \varsigma) \in \varrho_1(R)$$

Depending of the rank of  $R$  we have two sub-cases.

(1) Let  $\text{rank}(R) < j$ , then from Definition 7 we know that  $\varrho_1(R) = \varrho_2(R)$  and hence

$$\llbracket \mathbf{x} \rrbracket([], \varsigma) \in \varrho_2(R)$$

Which according to Table 1 is equivalent to

$$(\varrho_2, \varsigma) \models R(\mathbf{x})$$

(2) Let us now assume  $\text{rank}(R) = j$ , then from Definition 7 we know that  $\varrho_1(R) \subseteq \varrho_2(R)$  and hence

$$\llbracket \mathbf{x} \rrbracket([\ ], \varsigma) \in \varrho_2(R)$$

which is equivalent to

$$(\varrho_2, \varsigma) \models R(\mathbf{x})$$

and finishes the case.

**Case:**  $\text{cond} = \neg R(\mathbf{x})$

Assume  $\varrho_1 \subseteq_{/j} \varrho_2$  and

$$(\varrho_1, \varsigma) \models \neg R(\mathbf{x})$$

From Table 1 it follows that

$$\llbracket \mathbf{x} \rrbracket([\ ], \varsigma) \notin \varrho_1(R)$$

Since  $\text{rank}(R) < j$ , then from Definition 7 we have  $\varrho_1(R) = \varrho_2(R)$  and hence

$$\llbracket \mathbf{x} \rrbracket([\ ], \varsigma) \notin \varrho_2(R)$$

Which according to Table 1 is equivalent to

$$(\varrho_2, \varsigma) \models \neg R(\mathbf{x})$$

**Case:**  $\text{cond} = \text{cond}_1 \wedge \text{cond}_2$

Assume  $\varrho_1 \subseteq_{/j} \varrho_2$  and

$$(\varrho_1, \varsigma) \models \text{cond}_1 \wedge \text{cond}_2$$

From Table 1 it follows that

$$(\varrho_1, \varsigma) \models \text{cond}_1 \text{ and } (\varrho_1, \varsigma) \models \text{cond}_2$$

The induction hypothesis gives

$$(\varrho_2, \varsigma) \models \text{cond}_1 \text{ and } (\varrho_2, \varsigma) \models \text{cond}_2$$

Hence we have

$$(\varrho_2, \varsigma) \models \text{cond}_1 \wedge \text{cond}_2$$

**Case:**  $\text{cond} = \text{cond}_1 \vee \text{cond}_2$

Assume  $\varrho_1 \subseteq_{/j} \varrho_2$  and

$$(\varrho_1, \varsigma) \models \text{cond}_1 \vee \text{cond}_2$$

From Table 1 it follows that

$$(\varrho_1, \varsigma) \models \text{cond}_1 \text{ or } (\varrho_1, \varsigma) \models \text{cond}_2$$

The induction hypothesis gives

$$(\varrho_2, \varsigma) \models \text{cond}_1 \text{ or } (\varrho_2, \varsigma) \models \text{cond}_2$$

Hence we have

$$(\varrho_2, \varsigma) \models \text{cond}_1 \vee \text{cond}_2$$

**Case:**  $\text{cond} = \exists x : \text{cond}'$

Assume  $\varrho_1 \subseteq_{/j} \varrho_2$  and

$$(\varrho_1, \varsigma) \models \exists x : \text{cond}'$$

From Table 1 it follows that

$$\exists a \in \mathcal{U} : (\varrho_1, \varsigma[x \mapsto a]) \models \text{cond}'$$

The induction hypothesis gives

$$\exists a \in \mathcal{U} : (\varrho_2, \varsigma[x \mapsto a]) \models \text{cond}'$$

Hence from Table 1 we have

$$(\varrho_2, \varsigma) \models \exists x : \text{cond}'$$

**Case:**  $\text{cond} = \forall x : \text{cond}'$

Assume  $\varrho_1 \subseteq_{/j} \varrho_2$  and

$$(\varrho_1, \varsigma) \models \forall x : \text{cond}'$$

From Table 1 it follows that

$$\forall a \in \mathcal{U} : (\varrho_1, \varsigma[x \mapsto a]) \models \text{cond}'$$

The induction hypothesis gives

$$\forall a \in \mathcal{U} : (\varrho_2, \varsigma[x \mapsto a]) \models \text{cond}'$$

Hence from Table 1 we have

$$(\varrho_2, \varsigma) \models \forall x : \text{cond}'$$

□

**Lemma 4.** *If  $\varrho = \sqcap M$  and  $(\varrho', \zeta, \varsigma) \models cl_j$  for all  $\varrho' \in M$  then  $(\varrho, \zeta, \varsigma) \models cl_j$ .*

*Proof.* We proceed by induction on  $j$  and in each case perform a structural induction on the form of the clause  $cl$  occurring in  $cl_j$ .

**Case:**  $cl_j = \text{define}(\text{cond} \Rightarrow R(\mathbf{u}))$

Assume

$$\forall \varrho' \in M : (\varrho', \zeta, \varsigma) \models \text{cond} \Rightarrow R(\mathbf{u}) \quad (1)$$

Let us also assume

$$(\varrho, \varsigma) \models \text{cond}$$

Since  $\varrho = \bigcap M$  we know that

$$\forall \varrho' \in M : \varrho \subseteq \varrho' \quad (2)$$

Let  $R'$  occur in *cond*. We have two possibilities; either  $\text{rank}(R') = j$  and  $R'$  is a defined relation, then from (2) it follows that  $\varrho(R') \subseteq \varrho'(R')$ . Alternatively  $\text{rank}(R') < j$  and from (2) it follows that  $\varrho(R') = \varrho'(R')$ . Hence from Definition 7 we have that  $\varrho \subseteq_{/j} \varrho'$ . Thus from Lemma 3 it follows that

$$\forall \varrho' \in M : (\varrho', \varsigma) \models \text{cond}$$

Hence from (1) we have

$$\forall \varrho' \in M : (\varrho', \zeta, \varsigma) \models R(\mathbf{u})$$

Which from Table 1 is equivalent to

$$\forall \varrho' \in M : \llbracket \mathbf{u} \rrbracket(\zeta, \varsigma) \in \varrho'(R)$$

It follows that

$$\llbracket \mathbf{u} \rrbracket(\zeta, \varsigma) \in \bigcup \{ \varrho'(R) \mid \varrho' \in M \} = \varrho(R)$$

Which from Table 1 is equivalent to

$$(\varrho, \zeta, \varsigma) \models R(\mathbf{u})$$

and finishes the case.

**Case:**  $cl_j = \text{define}(def_1 \wedge def_2)$

Assume

$$\forall \varrho' \in M : (\varrho', \zeta, \varsigma) \models def_1 \wedge def_2$$

From Table 1 we have that for all  $\varrho' \in M$

$$(\varrho', \zeta, \varsigma) \models def_1 \text{ and } (\varrho', \zeta, \varsigma) \models def_2$$

The induction hypothesis gives

$$(\varrho, \zeta, \varsigma) \models def_1 \text{ and } (\varrho, \zeta, \varsigma) \models def_2$$

Hence from Table 1 we have

$$(\varrho, \zeta, \varsigma) \models def_1 \wedge def_2$$

**Case:**  $cl_j = \text{define}(\forall x : def)$

Assume

$$\forall \varrho' \in M : (\varrho', \zeta, \varsigma) \models \forall x : def \quad (3)$$

From Table 1 we have that

$$\varrho' \in M : \forall a \in \mathcal{U} : (\varrho', \zeta, \varsigma[x \mapsto a]) \models def$$

Thus

$$\forall a \in \mathcal{U} : \varrho' \in M : (\varrho', \zeta, \varsigma[x \mapsto a]) \models \text{def}$$

The induction hypothesis gives

$$\forall a \in \mathcal{U} : (\varrho, \zeta, \varsigma[x \mapsto a]) \models \text{def}$$

Hence from Table 1 we have

$$(\varrho, \zeta, \varsigma) \models \forall x : \text{def}$$

**Case:**  $cl_j = \text{constrain}(R(\mathbf{u}) \Rightarrow \text{cond})$

Assume

$$\forall \varrho' \in M : (\varrho', \zeta, \varsigma) \models R(\mathbf{u}) \Rightarrow \text{cond} \quad (4)$$

Let us also assume

$$(\varrho, \zeta, \varsigma) \models R(\mathbf{u})$$

From Table 1 it follows that

$$\llbracket \mathbf{u} \rrbracket(\zeta, \varsigma) \in \bigcup \{ \varrho'(R) \mid \varrho' \in M \}$$

Thus there is some  $\varrho' \in M$  such that

$$\llbracket \mathbf{u} \rrbracket(\zeta, \varsigma) \in \varrho'(R)$$

From (4) it follows that

$$(\varrho', \varsigma) \models \text{cond}$$

Since  $\varrho = \bigcap M$  we know that

$$\forall \varrho' \in M : \varrho \sqsubseteq \varrho' \quad (5)$$

Let  $R'$  occur in  $\text{cond}$ . We have two possibilities; either  $\text{rank}(R') = j$  and  $R'$  is a constrained relation, then from (5) it follows that  $\varrho(R') \supseteq \varrho'(R')$ . Alternatively  $\text{rank}(R') < j$  and from (5) it follows that  $\varrho(R') = \varrho'(R')$ . Hence from Definition 7 we have that  $\varrho' \sqsubseteq_{/j} \varrho$ . Thus from Lemma 3 it follows that

$$(\varrho, \varsigma) \models \text{cond}$$

which finishes the case.

**Case:**  $cl_j = \text{constrain}(\text{con}_1 \wedge \text{con}_2)$

Assume

$$\forall \varrho' \in M : (\varrho', \zeta, \varsigma) \models \text{con}_1 \wedge \text{con}_2$$

From Table 1 we have that for all  $\varrho' \in M$

$$(\varrho', \zeta, \varsigma) \models \text{con}_1 \text{ and } (\varrho', \zeta, \varsigma) \models \text{con}_2$$

The induction hypothesis gives

$$(\varrho, \zeta, \varsigma) \models \text{con}_1 \text{ and } (\varrho, \zeta, \varsigma) \models \text{con}_2$$

Hence from Table 1 we have

$$(\varrho, \zeta, \varsigma) \models \text{con}_1 \wedge \text{con}_2$$

**Case:**  $cl_j = \text{constrain}(\forall x : \text{con})$

Assume

$$\forall \varrho' \in M : (\varrho', \zeta, \varsigma) \models \forall x : \text{con} \quad (6)$$

From Table 1 we have that

$$\varrho' \in M : \forall a \in \mathcal{U} : (\varrho', \zeta, \varsigma[x \mapsto a]) \models \text{con}$$

Thus

$$\forall a \in \mathcal{U} : \varrho' \in M : (\varrho', \zeta, \varsigma[x \mapsto a]) \models \text{con}$$

The induction hypothesis gives

$$\forall a \in \mathcal{U} : (\varrho, \zeta, \varsigma[x \mapsto a]) \models \text{con}$$

Hence from Table 1 we have

$$(\varrho, \zeta, \varsigma) \models \forall x : \text{con}$$

□

**Proposition 1:** Assume  $cls$  is a stratified LFP formula,  $\varsigma_0$  and  $\zeta_0$  are interpretations of the free variables and function symbols in  $cls$ , respectively. Furthermore,  $\varrho_0$  is an interpretation of all relations of rank 0. Then  $\{\varrho \mid (\varrho, \zeta_0, \varsigma_0) \models cls \wedge \forall R : \text{rank}(R) = 0 \Rightarrow \varrho(R) \supseteq \varrho_0(R)\}$  is a Moore family.

*Proof.* The result follows from Lemma 4. □

## D Proof of Proposition 2

**Proposition 2:** For a finite universe  $\mathcal{U}$ , the best solution  $\varrho$  such that  $\varrho_0 \sqsubseteq \varrho$  of a LFP formula  $cl_1, \dots, cl_s$  (w.r.t. an interpretation of the constant symbols) can be computed in time

$$\mathcal{O}(|\varrho_0| + \sum_{1 \leq i \leq s} |cl_i| |\mathcal{U}|^{k_i})$$

where  $k_i$  is the maximal nesting depth of quantifiers in the  $cl_i$  and  $|\varrho_0|$  is the sum of cardinalities of predicates  $\varrho_0(R)$  of rank 0. We also assume unit time hash table operations (as in [19]).

*Proof.* Let  $cl_i$  be a clause corresponding to the  $i$ -th layer. Since  $cl_i$  can be either a define clause, or a constrain clause, we have two cases.

Let us first assume that  $cl_i = \text{define}(\text{def})$ ; the proof proceed in three phases. First we transform  $\text{def}$  to  $\text{def}'$  by replacing every universal quantification  $\forall x : \text{def}_{cl}$  by the conjunction of all  $|\mathcal{U}|$  possible instantiations of  $\text{def}_{cl}$ , every existential quantification  $\exists x : \text{cond}$  by the disjunction of all  $|\mathcal{U}|$  possible instantiations of

$cond$  and every universal quantification  $\forall x : cond$  by the conjunction of all  $|\mathcal{U}|$  possible instantiations of  $cond$ . The resulting clause  $def'$  is logically equivalent to  $def$  and has size

$$\mathcal{O}(|\mathcal{U}|^k |def|) \quad (7)$$

where  $k$  is the maximal nesting depth of quantifiers in  $def$ . Furthermore,  $def'$  is *boolean*, which means that there are no variables or quantifiers and all literals are viewed as nullary predicates.

In the second phase we transform the formula  $def'$ , being the result of the first phase, into a sequence of formulas  $def'' = def'_1, \dots, def'_l$  as follows. We first replace all top-level conjunctions in  $def'$  with  $\text{,}$ . Then we successively replace each formula by a sequence of simpler ones using the following rewrite rule

$$cond_1 \vee cond_2 \Rightarrow R(\mathbf{u}) \mapsto cond_1 \Rightarrow Q_{new}, cond_2 \Rightarrow Q_{new}, Q_{new} \Rightarrow R(\mathbf{u})$$

where  $Q_{new}$  is a fresh nullary predicate that is generated for each application of the rule. The transformation is completed as soon as no replacement can be done. The conjunction of the resulting define clauses is logically equivalent to  $def'$ .

To show that this process terminates and that the size of  $def''$  is at most a constant times the size of the input formula  $def'$ , we assign a cost to the formulae. Let us define the cost of a sequence of clauses as the sum of costs of all occurrences of predicate symbols and operators (excluding  $\text{,}$ ). In general, the cost of a symbol or operator is 1 except disjunction that counts 6. Then the above rule decreases the cost from  $k + 7$  to  $k + 6$ , for suitable value of  $k$ . Since the cost of the initial sequence is at most 6 times the size of  $def$ , only a linear number of rewrite steps can be performed. Since each step increases the size at most by a constant, we conclude that the  $def''$  has increased just by a constant factor. Consequently, when applying this transformation to  $def'$ , we obtain a boolean formula without sharing of size as in (7).

The third phase solves the system that is a result of phase two, which can be done in linear time by the classical techniques of e.g. [10].

Let us now assume that the  $cl_i = constrain(con)$ . We begin by transforming  $con$  into a logically equivalent (modulo fresh predicates) *define* clause. The transformation is done by function  $f_i$  defined as

$$f_i(constrain(con)) = define(g(con)), define(h_i(con))$$

$$\begin{aligned} g(\forall x : con) &= \forall x : g(con) \\ g(con_1 \wedge con_2) &= g(con_1) \wedge g(con_2) \\ g(R(\mathbf{u}) \Rightarrow cond) &= (\neg cond[R^G(\mathbf{u})/\neg R(\mathbf{u})] \Rightarrow R^G(\mathbf{u})) \end{aligned}$$

$$\begin{aligned} h_i(\forall x : con) &= \forall x : h_i(con) \\ h_i(con_1 \wedge con_2) &= h_i(con_1) \wedge h_i(con_2) \\ h_i(R(\mathbf{u}) \Rightarrow cond) &= let\ cond' = cond[true/(R'(\mathbf{v}) \mid rank(R') = i)]\ in \\ &\quad cond' \wedge \neg R^G(\mathbf{u}) \Rightarrow R(\mathbf{u}) \end{aligned}$$

where  $R^c$  is a new predicate corresponding to the complement of  $R$ . The size of the formula increases by a number of constraint predicates; hence the size of the input formula is increased by a constant factor. Then the proof proceeds as in case of *define* clause.

The three phases of the transformation result in the sequence of define clauses of size

$$\mathcal{O}\left(\sum_{1 \leq i \leq s} |cl_i| |\mathcal{U}|^{k_i}\right)$$

which can then be solved in linear time. We also need to take into account the size of the initial knowledge i.e. the cardinality of all predicates of rank 0; thus the overall worst case complexity is

$$\mathcal{O}(|\varrho_0| + \sum_{1 \leq i \leq s} |cl_i| |\mathcal{U}|^{k_i})$$

□