TLP **18** (2): 126–166, 2018. © Cambridge University Press 2017
doi:10.1017/S1471068417000497 First published online 4 December 2017

Predicate Pairing for program verification

EMANUELE DE ANGELIS and FABIO FIORAVANTI
DEC, 'G. d'Annunzio' University of Chieti-Pescara, Pescara, Italy
(*e-mails:* emanuele.deangelis@unich.it, fabio.fioravanti@unich.it)

ALBERTO PETTOROSSI
DICII, University of Rome Tor Vergata, Rome, Italy
(*e-mails:* pettorossi@info.uniroma2.it)

MAURIZIO PROIETTI
CNR-IASI, Rome, Italy
(*e-mails:* maurizio.proietti@iasi.cnr.it)

submitted 10 February 2017; accepted 19 October 2017

Abstract

It is well-known that the verification of partial correctness properties of imperative programs can be reduced to the satisfiability problem for constrained Horn clauses (CHCs). However, state-of-the-art solvers for constrained Horn clauses (or CHC solvers) based on *predicate abstraction* are sometimes unable to verify satisfiability because they look for models that are definable in a given class $\mathscr{A}$ of constraints, called $\mathscr{A}$-definable models. We introduce a transformation technique, called *Predicate Pairing*, which is able, in many interesting cases, to transform a set of clauses into an equisatisfiable set whose satisfiability can be proved by finding an $\mathscr{A}$-definable model, and hence can be effectively verified by a state-of-the-art CHC solver. In particular, we prove that, under very general conditions on $\mathscr{A}$, the unfold/fold transformation rules preserve the existence of an $\mathscr{A}$-definable model, that is, if the original clauses have an $\mathscr{A}$-definable model, then the transformed clauses have an $\mathscr{A}$-definable model. The converse does not hold in general, and we provide suitable conditions under which the transformed clauses have an $\mathscr{A}$-definable model *if and only if* the original ones have an $\mathscr{A}$-definable model. Then, we present a strategy, called Predicate Pairing, which guides the application of the transformation rules with the objective of deriving a set of clauses whose satisfiability problem can be solved by looking for $\mathscr{A}$-definable models. The Predicate Pairing (PP) strategy introduces a new predicate defined by the conjunction of two predicates occurring in the original set of clauses, together with a conjunction of constraints. We will show through some examples that an $\mathscr{A}$-definable model may exist for the new predicate even if it does not exist for its defining atomic conjuncts. We will also present some case studies showing that Predicate Pairing plays a crucial role in the verification of *relational properties of programs*, that is, properties relating two programs (such as program equivalence) or two executions of the same program (such as non-interference). Finally, we perform an experimental evaluation of the proposed techniques to assess the effectiveness of Predicate Pairing in increasing the power of CHC solving.

KEYWORDS: program verification, constrained Horn clauses, constraint logic programming, program transformation, relational properties of programs

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

1 Introduction

Constrained Horn clauses (CHCs, for short) have been advocated by many researchers as a suitable logical formalism for the specification and the automated verification of properties of imperative programs (Peralta *et al.* 1998; Albert *et al.* 2007; Podelski and Rybalchenko 2007; Méndez-Lojo *et al.* 2008; Jaffar *et al.* 2009; Rümmer *et al.* 2013; De Angelis *et al.* 2014a; Bjørner *et al.* 2015; Kafle *et al.* 2016). In particular, the problem of showing *partial correctness* properties defined by Hoare triples (Hoare 1969) has a natural translation into the problem of proving the *satisfiability* of a suitable set of CHCs.

Consider, for instance, the C-like program `sum_upto` in Figure 1, which computes the sum of the first `m` non-negative integer numbers:

Suppose, we want to prove the following Hoare triple: $\{\mathtt{m} \geqslant 0\}$ `sum_upto` $\{\mathtt{sum} \geqslant \mathtt{m}\}$. This triple is valid if the following set of clauses, called *verification conditions*, is satisfiable:

(1) $false \leftarrow M > Sum,\ M \geqslant 0,\ R = 0,\ su(M, R, Sum)$
(2) $su(X, R, Sum) \leftarrow X \leqslant 0,\ Sum = R$
(3) $su(X, R, Sum) \leftarrow X > 0,\ R1 = R + X,\ X1 = X - 1,\ su(X1, R1, Sum)$

These clauses can be obtained in an automatic way from an interpreter of the C-like imperative language we consider and the given Hoare triple by using a technique described in the literature (Peralta *et al.* 1998; Albert *et al.* 2007; Méndez-Lojo *et al.* 2008; De Angelis *et al.* 2014a; De Angelis *et al.* 2017). The predicate $su(M, R, Sum)$, which holds iff $Sum = R + \sum_{x=1}^{M} x$, encodes the operational semantics of the program `sum_upto`. Clause 1 encodes the Hoare triple, stating that if $su(M, R, Sum)$ holds with $M \geqslant 0$ (that is, the precondition $\mathtt{m} \geqslant 0$ holds) and $R = 0$ (that is, `r` is initialized to 0), then $Sum \geqslant M$ (that is, at the end of the execution, the value of the variable `sum` is not smaller than the value of the variable `m`). Clauses 2 and 3 encode the while-loop of the function `f`.

CHCs are syntactically the same as *constraint logic programs* (Jaffar and Maher 1994). However, the term 'CHCs' is mostly used in the field of program verification and, unlike 'constraint logic programs', it is not associated with any operational meaning. Moreover, most of the research on CHCs is devoted to finding a model, expressible in the constraint theory, that proves the *satisfiability* of the clauses, whereas the operational semantics of constraint logic programs is based on a refutation procedure that looks for a proof of the *unsatisfiability* of the clauses. In this respect, the techniques used for finding models of CHCs are closer to the ones proposed for the static analysis of constraint logic programs based on abstract interpretation (Cousot and Cousot 1977; Benoy and King 1997), where the objective is to find an over-approximation of the least model of the program.

The proof of satisfiability of sets of CHCs is supported by *CHC solvers* that have been developed in recent years for various constraint theories, such as (linear or non-linear) integer arithmetic, real (or rational) arithmetic, booleans, integer arrays, lists, heaps, and other data structures (Hoder *et al.* 2011; Grebenshchikov *et al.* 2012; Hojjat *et al.* 2012; McMillan and Rybalchenko 2013; De Angelis *et al.* 2014b; Gurfinkel *et al.* 2015; Kafle *et al.* 2016). However, in general, since the satisfiability of

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

```
int m, sum;
int f(int x) {
   int r = 0;
   while (x > 0) {
      r = r + x;  x--; }
   return r;
   }
void sum_upto() {
   sum = f(m);
   }
```

Fig. 1. Program `sum_upto` computing $\texttt{sum} = \sum_{x=1}^{\texttt{m}} x$.

CHCs is an undecidable problem, CHC solvers may not be able to return conclusive answers.

In order to improve the effectiveness of CHC solvers, several techniques proposed by recent papers perform satisfiability preserving transformations on sets of clauses that, in some cases, derive clauses whose satisfiability is easier to prove (De Angelis *et al.* 2014a; De Angelis *et al.* 2015a; De Angelis *et al.* 2015b; De Angelis *et al.* 2016; Kafle and Gallagher 2017a; Kafle and Gallagher 2017b). These transformations are adaptations to the task of improving the effectiveness of satisfiability checking of earlier techniques, which were developed for improving the efficiency of execution of (constraint) logic programs, such as *query answer transformation*, *specialization* (or *partial deduction*), and *unfold/fold transformations* (Tamaki and Sato 1984; Pettorossi and Proietti 1994; Debray and Ramakrishnan 1994; Etalle and Gabbrielli 1996; Leuschel and Bruynooghe 2002).

In this paper, we further enhance the approach to CHC satisfiability checking based on unfold/fold transformations. Our two main contributions are the following: (1) we prove in a precise mathematical sense that the application of the unfold/fold transformation rules cannot worsen the effectiveness of CHC solvers, and actually these rules are able to enlarge the set of satisfiability problems that can be solved by a given class of CHC solvers; and (2) we provide a specific strategy, called *Predicate Pairing* (PP), for applying the transformation rules with the objective of improving the ability of CHC solvers to prove satisfiability.

The basic idea behind the first contribution is as follows. Similarly to what is introduced in a paper by Bjørner *et al.* (2015), we consider the notion of the $\mathscr{A}$*-definable* model, which is a model definable in a class $\mathscr{A}$ of first-order formulas. Typically, CHC solvers (and, in particular, the solvers based on *predicate abstraction*) look for models in specific classes, such as *linear* (*integer or real*) *arithmetic* formulas, or *quantifier-free array* formulas. While satisfiability is undecidable and not semidecidable, the existence of an $\mathscr{A}$-definable model is semidecidable, as long as the validity problem for the formulas in $\mathscr{A}$ is decidable, and hence solvers that find an $\mathscr{A}$-definable model whenever it exists, can indeed be constructed. We prove that, under very general conditions on $\mathscr{A}$, the unfold/fold rules preserve the existence of an $\mathscr{A}$-definable model, that is, if the original clauses have an $\mathscr{A}$-definable model, then also the transformed clauses have an $\mathscr{A}$-definable model. The converse does not hold: there are cases where the original clauses have no $\mathscr{A}$-definable model, while

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

the transformed clauses have an $\mathscr{A}$-definable model. In this sense, the application of the unfold/fold rules may improve the effectiveness of a CHC solver that works by searching for $\mathscr{A}$-definable models, because the solver may be able to find an $\mathscr{A}$-definable model after the transformation in cases where there was no such a model before the transformation.

We also provide less general conditions under which the transformed clauses have an $\mathscr{A}$-definable model *if and only if* the original ones have an $\mathscr{A}$-definable model. These conditions prevent the introduction of new predicates that have recursive definitions in terms of the old predicates occurring in the original clauses. Thus, the source of the improvement of the effectiveness of the CHC solver due to the unfold/fold transformations is the introduction of one or more new predicates and the derivation of new (mutually) recursive definitions for these predicates.

The second contribution of our paper is related to the fact that, due to the already mentioned undecidability limitations, there is no universal algorithm that, starting from a set of clauses, applies the unfold/fold rules, and derives a set of clauses such that, if it is satisfiable, then it has an $\mathscr{A}$-definable model, for some theory $\mathscr{A}$ whose validity problem is decidable. Therefore, it should not be unexpected that the PP strategy we propose for guiding the use of the unfold/fold transformation rules is based on heuristics. We show that this strategy is capable, in many significant cases, of transforming sets of clauses into new, equisatisfiable sets of clauses, whose satisfiability problem can be solved by constructing $\mathscr{A}$-definable models, while the original sets have no $\mathscr{A}$-definable models. PP introduces a new predicate defined by the conjunction of two predicates together with a conjunction of constraints. We will explain through examples why an $\mathscr{A}$-definable model may exist for a conjunction of predicates, even if it does not exist for the atomic conjuncts in isolation. (Obviously, by a repeated application of PP, we may introduce new predicates corresponding to the conjunction of more than two old predicates.) Thus, PP can be viewed as an extension to CHCs of techniques for transforming logic programs, such as the *tupling* unfold/fold strategy (Pettorossi and Proietti 1994) and *conjunctive partial deduction* (De Schreye *et al.* 1999).

We will show that constraint-based reasoning is essential for guiding the introduction of the suitable pairs of predicates during the transformation process. Moreover, we will show that PP works well for solving many satisfiability problems that arise from the field of imperative program verification. In particular, PP is a crucial technique for verifying *relational program properties* (Barthe *et al.* 2011), that is, properties relating two programs (such as program equivalence) or two executions of the same program (such as non-interference).

The paper is structured as follows. In Section 2, we recall the basic notions concerning CHCs and we define the notion of an $\mathscr{A}$-definable model. In Section 3, we prove our results concerning the preservation of $\mathscr{A}$-definable models when using the unfold/fold transformation rules. In Section 4, we present the PP strategy and in Section 5 we show some examples of its application for verifying relational program properties. In Section 6, we report the results obtained by our implementation of that strategy by using the VeriMAP transformation system (De Angelis *et al.*

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

2014b). Finally, in Section 7, we discuss the related work in the fields of program transformation and verification.

2 Constrained Horn clauses

In this section, we recall the basic definitions concerning CHCs and their satisfiability, and we introduce the notion of an $\mathscr{A}$*-definable model*.

Let $\mathscr{L}$ be a first-order language with equality and $Pred_u \subseteq \mathscr{L}$ be a set of predicate symbols, called the *user-defined* predicate symbols. Let $\mathscr{C}$ be a set of formulas of $\mathscr{L} \setminus Pred_u$, called the set of *constraints*. We assume that: (1) *true*, *false*, and equalities between terms belong to $\mathscr{C}$, and (2) $\mathscr{C}$ is closed under conjunction.

An *atom* is an atomic formula of the form $p(X_1, \ldots, X_m)$, where p is a predicate symbol in $Pred_u$ and $X_1, \ldots, X_m$ are distinct variables. Let *Atom* be the set of all atoms. A *definite constrained clause* is an implication of the form $c \wedge G \rightarrow H$ whose premise (or *body*) is the conjunction of a constraint c and a (possibly empty) conjunction G of $n\,(\geqslant 0)$ atoms $A_1 \wedge \ldots \wedge A_n$, and whose conclusion (or *head*) H is an atom. A *constrained goal* (or simply, a *goal*) is an implication of the form $c \wedge G \rightarrow \mathit{false}$, where c is a constraint and G is a conjunction of atoms. A CHC (or simply, a *clause*) is either a definite constrained clause or a constrained goal. A set of CHCs is said to be a *CHC set*. A CHC Cl (or a set P of clauses) is said to be '*over* $\mathscr{C}$' in case we want to stress that the constraints occurring in clause Cl (or in the set P of clauses) are taken from the set $\mathscr{C}$ of constraints. A clause $c \wedge G \rightarrow H$ is said to be *linear* if G consists of at most one atom, and *non-linear* otherwise.

We will often use the logic programming syntax and we write $H \leftarrow c, A_1, \ldots, A_n$, instead of $c \wedge A_1 \wedge \ldots \wedge A_n \rightarrow H$. We will also feel free to write non-variable terms as arguments of atoms. Thus, the clause $p(\ldots, t, \ldots) \leftarrow c, G$ should be viewed as a shorthand for $p(\ldots, X, \ldots) \leftarrow X = t, c, G$, where X is a variable not occurring elsewhere in the clause, and likewise, $H \leftarrow c, G_1, p(\ldots, t, \ldots), G_2$ should be viewed as a shorthand for $H \leftarrow X = t, c, G_1, p(\ldots, X, \ldots), G_2$.

Given a formula $\varphi \in \mathscr{L}$, we denote by $\exists(\varphi)$ its *existential closure* and by $\forall(\varphi)$ its *universal closure*. By $vars(\varphi)$ and $Fvars(\varphi)$, we denote the set of variables and the set of the free variables, respectively, occurring in φ.

For the notions of an *interpretation* and a *model* of a first-order formula, we will use the standard notions and notations (Mendelson 1997). We fix a *canonical interpretation* $\mathbb{D}$ of the symbols in $\mathscr{L} \setminus Pred_u$. A $\mathbb{D}$-*interpretation* is an interpretation of $\mathscr{L}$ that for all symbols occurring in $\mathscr{L} \setminus Pred_u$, agrees with $\mathbb{D}$. If U is the universe of $\mathbb{D}$, then a $\mathbb{D}$-interpretation $\mathbb{I}$ can be identified with the set of atoms:

$$\{p(a_1, \ldots, a_m) \mid (a_1, \ldots, a_m) \in U^m \text{ and } p^{\mathbb{I}}(a_1, \ldots, a_m) \text{ holds in } \mathbb{I}\}$$

where $p^{\mathbb{I}}$ denotes the m-ary relation, which is the interpretation of p in $\mathbb{I}$. Given any set F of formulas, a $\mathbb{D}$-interpretation $\mathbb{M}$ is a $\mathbb{D}$-*model* of F, written $\mathbb{M} \models F$, if, for all formulas $\varphi \in F$, $\mathbb{M} \models \varphi$ holds. F is $\mathbb{D}$-*satisfiable* if it has a $\mathbb{D}$-model. We will feel free to say *satisfiable*, instead of $\mathbb{D}$-*satisfiable*, when the interpretation $\mathbb{D}$ is clear from the context. We write $\mathbb{D} \models F$ if, for every $\mathbb{D}$-interpretation $\mathbb{M}$, $\mathbb{M} \models F$ holds.

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

A set P of definite constrained clauses is $\mathbb{D}$-satisfiable and has a *least* (with respect to set inclusion) $\mathbb{D}$-model, denoted $lm(P)$ (Jaffar and Maher 1994). Thus, if P is any set of CHCs and Q is the subset of the constrained goals in P, then P is $\mathbb{D}$-satisfiable if and only if $lm(P \setminus Q) \models Q$.

Many CHC solvers based on *predicate abstraction* (Bjørner *et al.* 2015) try to check the $\mathbb{D}$-satisfiability of a set of CHCs by looking for the existence of $\mathbb{D}$-models that are definable by formulas belonging to a given set $\mathscr{A}$, which is a subset of the set $\mathscr{C}$ of constraints. This restriction when looking for models may significantly ease the satisfiability test, as shown by the following example.

Example 1
Let us assume that $\mathscr{C}$ is the set of *linear integer arithmetic* (LIA) constraints, that is, equalities ($=$) and inequalities ($>$) between linear polynomials with integer coefficients and integer-valued variables, closed with respect to conjunction and disjunction. We also use the symbols $\geqslant$, $\leqslant$, $<$, and $\neq$ with the usual definitions in terms of $=$ and $>$. Let $\mathbb{Z}$ denote the usual interpretation of integer arithmetic.

Now, let us consider clauses 1–3 listed in the Introduction. The satisfiability of these clauses can be proved by looking for models that are definable by constraints φ in the subset of LIA, which we call 2VAR, defined by the following grammar:

$$\varphi ::= true \mid false \mid X > 0 \mid X = 0 \mid X > Y \mid X = Y \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2$$

where X and Y are variables. Thus, 2VAR is the set of linear integer constraints constructed from arithmetic comparisons between at most two variables, and it is a subset of the *Octagons* domain often considered in the field of abstract interpretation (Miné 2006). A 2VAR-definable model of clauses 1–3 is given by interpreting the predicate $su(M, R, Sum)$ as the set of triples satisfying the following constraint in 2VAR:

$$(Sum \geqslant M \wedge Sum \geqslant R) \ \vee \ R < 0,$$

In order to show that the above interpretation indeed defines a $\mathbb{Z}$-model, we replace the instances of $su(M, R, Sum)$ by the corresponding instances of the formula $(Sum \geqslant M \wedge Sum \geqslant R) \vee R < 0$ in clauses 1–3, and we check that the resulting implications hold in $\mathbb{Z}$.

This Example 1 motivates the introduction of the notion of an $\mathscr{A}$-definable model, which is a generalization of the one presented in the literature (Bjørner *et al.* 2015), where $\mathscr{A}$ coincides with $\mathscr{C}$.

Definition 1
Let $\mathscr{A} \subseteq \mathscr{C}$ be a set of formulas of $\mathscr{L} \setminus Pred_u$ such that: (1) *true*, *false*, and equalities between terms belong to $\mathscr{A}$, and (2) $\mathscr{A}$ is closed under conjunction. Let $\mathbb{D}$ be the canonical interpretation of the symbols in $\mathscr{L} \setminus Pred_u$. We denote by $\mathscr{A}^{\exists\vee}$ the set of formulas $\{\exists X_1 \ldots \exists X_m(\varphi_1 \vee \ldots \vee \varphi_n) \mid m \geqslant 0, n > 0,$ and for $i = 1, \ldots, n, \ \varphi_i \in \mathscr{A}\}$. A *symbolic interpretation* is a function $\Sigma : Atom \cup \{false\} \longrightarrow \mathscr{A}^{\exists\vee}$ such that $\Sigma(false) = false$ and, for every $A \in Atom$, (1) $Fvars(\Sigma(A)) \subseteq Fvars(A)$, and (2) for every renaming substitution ρ for A (Lloyd 1987), $\Sigma(A\rho) = \Sigma(A)\rho$. We extend Σ to conjunctions of atoms by stating that $\Sigma(A_1 \wedge \ldots \wedge A_n) = \Sigma(A_1) \wedge \ldots \wedge \Sigma(A_n)$. Given a set P of CHCs

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

```
int n, sqr;
int g(int y, int k) {
  int s = 0;
  while (y > 0) {
    s = s + k;   y--; }
  return s;
  }
void square() {
  sqr = g(n,n);
  }
```

Fig. 2. Program `square` computing $\mathtt{sqr} = \mathtt{n}^2$.

over $\mathscr{C}$, a symbolic interpretation Σ is an $\mathscr{A}$-definable model of P, written $\Sigma \models P$, if for every clause $H \leftarrow c, A_1, \ldots, A_n$ in P, $\mathbb{D} \models \forall(c \wedge \Sigma(A_1 \wedge \ldots \wedge A_n) \rightarrow \Sigma(H))$ holds.

Note that the symbolic interpretation Σ of an atom is independent of the variable names occurring in that atom, and hence, for each predicate symbol p, the formula $\Sigma(p(X_1, \ldots, X_m))$ is unique up to variable renaming. Note also that the definition of a symbolic interpretation is essentially equivalent to that given by Kafle and Gallagher, who define an interpretation as a set of *constrained facts* of the form $p(X_1, \ldots, X_m) \leftarrow c$, where c is a constraint in $\mathscr{A}$ (Kafle and Gallagher 2017b). Indeed, the set of constraints in the bodies of the constrained facts with the same head predicate can be represented as a disjunction of those constraints, and the variables occurring in the body of a constrained fact and not in its head are implicitly existentially quantified.

Clearly, if P has an $\mathscr{A}$-definable model, then P is $\mathbb{D}$-satisfiable. In general, the converse does not hold, as shown by the following example.

Example 2

Let us continue Example 1, where the sets of constraints $\mathscr{C}$ and $\mathscr{A}$ are LIA and 2VAR, respectively. Let us consider the program in Figure 2, which computes the square of a non-negative integer `n` by summing up `n` times the value of `n`.

Clauses 4–6 below express the following property, which relates program `sum_upto` and program `square`: if the value of `m` is equal to the value of `n` before the execution of the programs `sum_upto` and `square` and they both terminate, then at the end of their execution the value of `sqr` is not smaller than the value of `sum`. Note that, since the programs `sum_upto` and `square` have disjoint sets of variables, the order of their execution is immaterial.

(4) $false \leftarrow Sum > Sqr,\ M \geqslant 0,\ M = N,\ N = Y,\ R0 = 0,\ S0 = 0,$
$su(M, R0, Sum),\ sq(N, Y, S0, Sqr)$

(5) $sq(K, Y, S0, S) \leftarrow Y \leqslant 0,\ S = S0$

(6) $sq(K, Y, S0, S) \leftarrow Y > 0,\ Y1 = Y - 1,\ S1 = S0 + K,\ sq(K, Y1, S1, S)$

For $Y \geqslant 0$, the atom $sq(K, Y, S0, S)$ holds iff $S = S0 + (K \times Y)$. Properties like the one between programs `sum_upto` and `square` are called *relational properties* (Barthe *et al.* 2011). Similarly to the verification conditions for partial correctness properties, the clauses for relational properties, also called verification conditions,

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

can be automatically generated from the formal specification of those properties and the operational semantics of the programming language (De Angelis *et al.* 2016). For brevity, we do not give here the details of that generation process, which is inessential for understanding the techniques presented in this paper.

Clauses 2–6 are CHCs over LIA and they are $\mathbb{Z}$-satisfiable. Indeed, in the least $\mathbb{Z}$-model of clauses 2, 3, 5, 6, for all integers M, Sum, and Sqr, with $M \geqslant 0$, if $su(M, 0, Sum)$ and $sq(M, M, 0, Sqr)$ hold, then $Sum \leqslant Sqr$ holds. However, clauses 2–6 do not admit a 2VAR-definable model. Indeed, no constraint of the form $Sum \leqslant X$, for any variable X, is a consequence of $su(M, 0, Sum)$, and hence we cannot infer $Sum \leqslant Sqr$, independently of the constraints that are consequences of $sq(M, M, 0, Sqr)$. Actually, it is not difficult to see that a similar limitation holds even if we look for a LIA-definable model, rather than a 2VAR-definable model. Indeed, in order to infer the constraint $Sum \leqslant Sqr$ one should discover quadratic relations, such as $Sum = M \times (M-1)/2$ and $Sqr = M \times M$, starting from $su(M, 0, Sum)$ and $sq(M, M, 0, Sqr)$, respectively, and these relations cannot be expressed by linear arithmetic constraints.

3 Transformation rules and preservation of $\mathscr{A}$-definable models

Let $\mathscr{C}$ be a set of constraints and $\mathscr{A} \subseteq \mathscr{C}$. A *transformation sequence over* $\mathscr{C}$ is a sequence of CHC sets $P_0, P_1, \ldots, P_n$ over $\mathscr{C}$, where, for $i = 0, \ldots, n-1$, P_{i+1} is derived from P_i by applying one of the following rules R1–R4.

Let $Defs_i$ denote the set of all the clauses, called *definitions*, introduced by rule R1 during the construction of the transformation sequence $P_0, P_1, \ldots, P_i$. Thus, $Defs_0 = \emptyset$.

(R1) *Definition.* We introduce a clause D: $newp(X_1, \ldots, X_k) \leftarrow c, G$, where: (1) $newp$ is a predicate symbol in $Pred_u$ not occurring in the sequence $P_0, P_1, \ldots, P_i$, (2) $c \in \mathscr{A}$, (3) G is a non-empty conjunction of atoms whose predicate symbols occur in P_0, and (4) $X_1, \ldots, X_k$ are distinct variables occurring free in (c, G). Then, we derive the new set $P_{i+1} = P_i \cup \{D\}$ and $Defs_{i+1} = Defs_i \cup \{D\}$.

(R2) *Unfolding.* Let C: $H \leftarrow c, G_1, p(X_1, \ldots, X_k), G_2$ be a clause in P_i. Let

$$\{p(X_1, \ldots, X_k) \leftarrow c_j, B_j \mid j = 1, \ldots, m\}$$

be the (possibly empty) set of clauses in P_i whose head predicate is p. Without loss of generality, we assume that, for $j = 1, \ldots, m$, $vars(c_j, B_j) \cap vars(C) \subseteq \{X_1, \ldots, X_k\}$. By *unfolding the atom* $p(X_1, \ldots, X_k)$ *in* C *using* P_i we derive the new set $P_{i+1} = (P_i \setminus \{C\}) \cup \{H \leftarrow c, c_j, G_1, B_j, G_2 \mid j = 1, \ldots, m\}$.

(R3) *Folding.* Let C: $H \leftarrow c, G_1, Q, G_2$ be a clause in P_i, where Q is a non-empty conjunction of atoms, and let D: $K \leftarrow d, B$ be (a variant of) a clause in $Defs_i$ with $vars(C) \cap vars(D) = \emptyset$. Suppose that there exist a substitution ϑ and a constraint e such that: (1) $Q = B\vartheta$, (2) $\mathbb{D} \models \forall (c \leftrightarrow (e \wedge d\vartheta))$, and (3) for every variable $X \in vars(d, B) \setminus vars(K)$, the following conditi3.1) $X\vartheta$ is a variable not occurring in $\{H, c, G_1, G_2\}$, and (3.2) $X\vartheta$ does not occur in the term $Y\vartheta$, for any variable Y occurring in (d, B) and different from X. By *folding* C *using the definition* D, we derive clause E: $H \leftarrow e, G_1, K\vartheta, G_2$. In this case,

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

we also say that E is derived *by folding Q in C*. We derive the new set $P_{i+1} = (P_i \setminus \{C\}) \cup \{E\}$.

(R4) *Constraint Replacement*. Let us consider a subset of P_i of the form $\{(H \leftarrow c_1, G), \ldots, (H \leftarrow c_k, G)\}$. Suppose that, for some constraints $d_1, \ldots, d_m$,

$$\mathbb{D} \models \forall\, (\exists Y_1 \ldots \exists Y_r\ (c_1 \vee \ldots \vee c_k) \leftrightarrow \exists Z_1 \ldots \exists Z_s\ (d_1 \vee \ldots \vee d_m))$$

where $\{Y_1, \ldots, Y_r\} = Fvars(c_1 \vee \ldots \vee c_k) \setminus vars(\{H, G\})$ and $\{Z_1, \ldots, Z_s\} = Fvars(d_1 \vee \ldots \vee d_m) \setminus vars(\{H, G\})$. Then, we derive the new set $P_{i+1} = (P_i \setminus \{(H \leftarrow c_1, G), \ldots, (H \leftarrow c_k, G)\}) \cup \{(H \leftarrow d_1, G), \ldots, (H \leftarrow d_m, G)\}$.

Note that rule R4 enables the deletion of a clause with an inconsistent constraint in its body. Indeed, if c_1 is unsatisfiable, then $\mathbb{D} \models \forall\, (c_1 \leftrightarrow d_1 \vee \ldots \vee d_m)$ with $m = 0$.

The following result (Etalle and Gabbrielli 1996) shows that the transformation rules R1–R4 derive sets of clauses that are equivalent with respect to the least $\mathbb{D}$-model.

Theorem 1 (*Equivalence with respect to the Least $\mathbb{D}$-Model*)
Let $P_0, P_1, \ldots, P_n$ be a transformation sequence where, for $i = 0, \ldots, n$, P_i is a set of definite clauses. Let us assume that every definition in $Defs_n$ is unfolded during the construction of this sequence (that is, for every definition $D \in Defs_n$, there exists i, with $0 \leqslant i \leqslant n-1$, such that P_{i+1} is derived from P_i by unfolding D). Then, for every predicate p and $(a_1, \ldots, a_m) \in U^m$,

$$p(a_1, \ldots, a_m) \in lm(P_0 \cup Defs_n) \text{ if and only if } p(a_1, \ldots, a_m) \in lm(P_n).$$

From Theorem 1 it follows that, as we now show, the transformation rules R1–R4 derive sets of clauses that are equivalent with respect to $\mathbb{D}$-satisfiability.

Theorem 2 (*Equivalence with respect to $\mathbb{D}$-Satisfiability*)
Let $P_0, P_1, \ldots, P_n$ be a transformation sequence such that every definition in $Defs_n$ is unfolded during the construction of this sequence. Then, P_0 is $\mathbb{D}$-satisfiable if and only if P_n is $\mathbb{D}$-satisfiable.

Proof
First, we observe that P_0 is $\mathbb{D}$-satisfiable iff $P_0 \cup Defs_n$ is $\mathbb{D}$-satisfiable. Indeed: (1) if M is a $\mathbb{D}$-model of P_0, then the $\mathbb{D}$-interpretation $M \cup \{newp(a_1, \ldots, a_k) \mid newp$ is a head predicate in $Defs_n$ and $(a_1, \ldots, a_k) \in U^k\}$ is a $\mathbb{D}$-model of $P_0 \cup Defs_n$, and (2) if M is a $\mathbb{D}$-model of $P_0 \cup Defs_n$, then by the definition of $\mathbb{D}$-model, M is a $\mathbb{D}$-model of P_0.

Now let us consider a new sequence $P_0', P_1', \ldots, P_n'$ obtained from the transformation sequence $P_0, P_1, \ldots, P_n$ by replacing each occurrence of *false* in the head of a clause by a new predicate symbol f. The sequence $P_0', P_1', \ldots, P_n'$ satisfies the hypothesis of Theorem 1, and hence $f \in lm(P_0' \cup Defs_n)$ iff $f \in lm(P_n')$.

We have that: $P_0 \cup Defs_n$ is $\mathbb{D}$-satisfiable iff $P_0' \cup Defs_n \cup \{\neg f\}$ is $\mathbb{D}$-satisfiable iff $f \notin lm(P_0' \cup Defs_n)$ iff {by Theorem 1} $f \notin lm(P_n')$ iff $P_n' \cup \{\neg f\}$ is $\mathbb{D}$-satisfiable iff P_n is $\mathbb{D}$-satisfiable. □

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Theorem 2 is *not* sufficient to ensure that a transformation sequence preserves the existence of an $\mathscr{A}$-definable model. Indeed, as shown by Example 2 in Section 2, for some set $\mathscr{A}$ of constraints, $\mathbb{D}$-satisfiability does not imply the existence of an $\mathscr{A}$-definable model.

Now, in order to study the preservation of $\mathscr{A}$-models during the construction of a transformation sequence over $\mathscr{C}$, for $\mathscr{A} \subseteq \mathscr{C}$, we introduce the notions of *$\mathscr{A}$-soundness* and *$\mathscr{A}$-completeness*.

Definition 2 (*$\mathscr{A}$-Soundness, $\mathscr{A}$-Completeness*)
Let $P_0, P_1, \ldots, P_n$ be a transformation sequence. (1) If P_0 has an $\mathscr{A}$-definable model implies that P_n has an $\mathscr{A}$-definable model, we say that the sequence is *$\mathscr{A}$-sound*. (2) If P_n has an $\mathscr{A}$-definable model implies that P_0 has an $\mathscr{A}$-definable model, we say that the sequence is *$\mathscr{A}$-complete*.

In order to prove the $\mathscr{A}$-soundness of a transformation sequence (see Theorem 4 below), we need the following definition and theorem.

Definition 3
Let $\mathscr{S}$ be a CHC set. A symbolic interpretation Σ is said to be *tight on* $\mathscr{S}$ if for all clauses $A \leftarrow c, G$ in $\mathscr{S}$, $\mathbb{D} \models \forall(\Sigma(A) \leftrightarrow \exists X_1 \ldots \exists X_k(c \wedge \Sigma(G)))$, where $\{X_1, \ldots, X_k\} = Fvars(c \wedge \Sigma(G)) \backslash Fvars(\Sigma(A))$.

For instance, given the singleton set of clauses $\mathscr{S} = \{p(X) \leftarrow q(X)\}$, the symbolic interpretation Σ_1 that maps both $p(X)$ and $q(X)$ to $X = 0$ is tight on $\mathscr{S}$, while the symbolic interpretation Σ_2 that maps $p(X)$ to *true* and $q(X)$ to $X = 0$ is not tight on $\mathscr{S}$. Both Σ_1 and Σ_2 are models.

Theorem 3
Let $P_0, P_1, \ldots, P_n$ be a transformation sequence. For $i = 0, \ldots, n-1$, if P_i has an $\mathscr{A}$-definable model that is tight on $Defs_i$, then P_{i+1} has an $\mathscr{A}$-definable model that is tight on $Defs_{i+1}$.

Proof See Appendix. □

The hypothesis that P_i has an $\mathscr{A}$-definable model that is tight on $Defs_i$ is needed to guarantee that the folding rule replaces a conjunction consisting of constraints and atoms by a single atom that is *equivalent* in the given model.

From Theorem 3 and the fact that, if P_0 has an $\mathscr{A}$-definable model, then P_0 has an $\mathscr{A}$-definable model that is tight on $Defs_0$, which is the empty set, we get the following result.

Theorem 4 (*$\mathscr{A}$-Soundness*)
Every transformation sequence is $\mathscr{A}$-sound.

Now we prove that, if some suitable hypotheses hold, a transformation sequence is also $\mathscr{A}$-complete. First, we need the following two definitions.

Definition 4
An application of the unfolding rule R2 to a clause $C\colon H \leftarrow c, G_1, p(X_1, \ldots, X_k), G_2$ in P_i is said to be a *self-unfolding* if the predicate of H is p.

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Definition 5
Let $P_0, P_1, \ldots, P_n$ be a transformation sequence. An application of the folding rule R3 to a clause C in P_i using a definition D in $Defs_i$ is said to be a *reversible folding* if D belongs to P_i and is different from C.

Theorem 5 ($\mathscr{A}$*-Completeness*)
Let $P_0, P_1, \ldots, P_n$ be a transformation sequence over a set $\mathscr{C}$ of constraints. Let $\mathscr{A}$ be equal to $\mathscr{C}$. Suppose that: (1) no application of the unfolding rule is a self-unfolding, and (2) every application of the folding rule is a reversible folding. Then, $P_0, P_1, \ldots, P_n$ is $\mathscr{A}$-complete.

Proof See Appendix. □

In Section 4, we will present the PP transformation strategy and we will show that it generates transformation sequences that are $\mathscr{A}$-sound, but not necessarily $\mathscr{A}$-complete (see Theorem 6).

Normally, $\mathscr{A}$-soundness is a desirable property of a transformation sequence P_0, $P_1, \ldots, P_n$.Indeed, suppose we have a CHC solver, call it *SOLVE*, that finds an $\mathscr{A}$-definable model of a set of clauses, whenever it exists. As already mentioned, such an ideal solver exists, as long as the validity problem for the formulas in $\mathscr{A}$ is decidable. Then, $\mathscr{A}$-soundness guarantees that if the satisfiability of P_0 can be proved by using *SOLVE*, then also the satisfiability of P_n can be proved by using *SOLVE*. In other terms, the effectiveness of the solver is not worsened by the transformation.

In contrast, $\mathscr{A}$-completeness might not always be a desirable property. Indeed, in many cases we may want to transform clauses for which *SOLVE* cannot find an $\mathscr{A}$-definable model, because such a model does not exist, and derive equisatisfiable clauses with an $\mathscr{A}$-definable model, which can be constructed by using *SOLVE*.

In practice, the existing solvers do not guarantee that they find an $\mathscr{A}$-definable model of a set of clauses whenever it exists. Thus, the theoretical properties of $\mathscr{A}$-soundness and $\mathscr{A}$-completeness might not hold in some cases. We will show through the experiments reported in Section 6, that these unfortunate cases are rare.

We conclude this section by showing that there are $\mathbb{D}$-safisfiable CHC sets that have no $\mathscr{A}$-definable models, and yet can be transformed, by applying rules R1–R4, into CHC sets that have $\mathscr{A}$-definable models.

Example 3
Let us continue Example 2, where $\mathbb{D}$ is $\mathbb{Z}$, $\mathscr{C}$ is LIA, and $\mathscr{A}$ is 2VAR. Let P_0 be the set consisting of clauses 2–6.

Starting from P_0, we construct a transformation sequence P_0, P_1, P_2, P_3, P_4, as we now indicate. First, by applying the definition rule, we introduce the following new predicate:

(7) $su_sq(M, R0, Sum, N, S0, Sqr) \leftarrow M = Y,\ su(M, R0, Sum),\ sq(N, Y, S0, Sqr)$

We derive the clause set $P_1 = P_0 \cup \{7\}$ and $Defs_1 = \{7\}$. Now, by unfolding the atoms *su* and *sq* of clause 7, and then performing some more unfoldings of the derived atoms, we get:

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

(8) $su_sq(M, R0, Sum, N, S0, Sqr) \leftarrow M = Y,\ M \leqslant 0,\ Sum = R0,\ Sqr = S0$
(9) $su_sq(M, R0, Sum, N, S0, Sqr) \leftarrow M = Y,\ M \leqslant 0,\ Sum = R0,\ Y > 0,$
$Y1 = Y - 1,\ S1 = S0 + N,\ sq(N, Y1, S1, Sqr)$
(10) $su_sq(M, R0, Sum, N, S0, Sqr) \leftarrow M = Y,\ M > 0,\ M1 = M - 1,\ R1 = R0 + M,$
$Y \leqslant 0,\ Sqr = S0,\ su(M1, R1, Sum)$
(11) $su_sq(M, R0, Sum, N, S0, Sqr) \leftarrow M = Y,\ M > 0,\ M1 = M - 1,\ R1 = R0 + M,$
$Y > 0,\ Y1 = Y - 1,\ S1 = S0 + N,\ su(M1, R1, Sum),\ sq(N, Y1, S1, Sqr)$

We get $P_2 = P_0 \cup \{8, 9, 10, 11\}$ and $Defs_2 = \{7\}$. (Here and in the rest of the example, for reasons of conciseness, we feel free to avoid to list some intermediate CHC sets in the transformation sequence.) By the constraint replacement rule R4 we can remove clauses 9 and 10, whose bodies have unsatisfiable constraints, and replace clauses 8 and 11 by

(12) $su_sq(M, R0, Sum, N, S0, Sqr) \leftarrow M \leqslant 0,\ Sum = R0,\ Sqr = S0$
(13) $su_sq(M, R0, Sum, N, S0, Sqr) \leftarrow M > 0,\ M1 = M - 1,\ R1 = R0 + M,$
$S1 = S0 + N,\ M1 = Y1,\ su(M1, R1, Sum),\ sq(N, Y1, S1, Sqr)$

We get $P_3 = P_0 \cup \{12, 13\}$ and $Defs_3 = \{7\}$. Then, by the folding rule R3, we fold clause 4 (in P_0) using clause 7 and we derive the following clause:

(14) $false \leftarrow Sum > Sqr, M \geqslant 0, M = N, R0 = 0, S0 = 0, su_sq(M, R0, Sum, N, S0, Sqr)$

Finally, we fold clause 13 using clause 7 and we derive the following clause:

(15) $su_sq(M, R0, Sum, N, S0, Sqr) \leftarrow M > 0,\ M1 = M - 1,\ R1 = R0 + M,$
$S1 = S0 + N,\ su_sq(M1, R1, Sum, N, S1, Sqr)$

We get the final set of clauses $P_4 = (P_0 \setminus \{4\}) \cup \{12, 14, 15\}$. Now, it is easy to check that the symbolic interpretation that maps the atom $su_sq(M, R0, Sum, N, S0, Sqr)$ to the 2VAR constraint $Sum \leqslant Sqr \vee R0 > S0 \vee M > N$, and the su and sq atoms to $true$, is a 2VAR-definable model of P_4. This check can be done by replacing the atoms in P_4 by the corresponding symbolic interpretations, and then verifying the validity of the formulas obtained in that way by using an SMT solver for linear arithmetic, such as the popular Z3 solver (de Moura and Bjørner 2008).

Let us now make some remarks on the derivation above.

(1) The applications of the transformation rules satisfy the hypothesis of Theorem 2 and hence P_0 is $\mathbb{Z}$-satisfiable if and only if P_4 is $\mathbb{Z}$-satisfiable (indeed, they are both $\mathbb{Z}$-satisfiable).
(2) The fact that P_4 has a 2VAR-definable model, while P_0 has no such model, is due to the fact that the applications of the folding rule R3 are not reversible foldings, and hence the transformation sequence $P_0, \ldots, P_4$ does not satisfy the hypothesis of Theorem 5. Indeed, clause 7 occurs in $Defs_3$, but not in P_3. More in general, the derivation of a CHC set with an $\mathscr{A}$-definable model from a CHC set without an $\mathscr{A}$-definable model is due to the introduction of new predicates and also to the derivation (via non-reversible foldings) of clauses that constitute recursive definitions of these new predicates.

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

(3) Finally, note that at every step during the transformation from P_0 to P_4 we have handled linear constraints only. However, the introduction of the new predicate $su_sq(M, R0, Sum, N, S0, Sqr)$, defined in terms of the conjunction of $su(M, R0, Sum)$ and $sq(N, Y, S0, Sqr)$, allows us to discover linear relations between Sum and Sqr without having to deal with non-linear constraints.

4 Predicate Pairing

In Section 3, we have seen that by the applying unfold/fold rules, in some cases one may transform a given $\mathbb{D}$-satisfiable set of clauses which does *not* admit an $\mathscr{A}$-definable model, into an equisatisfiable set of clauses that admits an $\mathscr{A}$-definable model. Then, the $\mathbb{D}$-satisfiability of the set of clauses can be proved by a CHC solver that constructs an $\mathscr{A}$-definable model. Example 3 of Section 3 suggests that the crucial step in that transformation is a predicate pairing step, that is, the introduction of a new predicate, say t, whose defining clause has in its body the conjunction of two atoms, one with predicate, say q, and the other with predicate, say r, whose definitions are provided by the original set of clauses. Indeed, the predicate pairing allows us to derive suitable relations between arguments of the conjunction of q and r, which cannot be expressed by using constraints on the arguments of q and r separately.

In general, in order to do the transformation and perform some required folding steps, it may be necessary to introduce, by predicate pairing, more than one definition. The introduction of these new definitions is a major issue in the case where the predicates to be paired should be chosen from the various predicates occurring in the conjunction of several atoms in the body of a clause. In particular, this issue arises when predicates are defined by *non-linear* clauses, and hence their repeated unfolding may generate unbounded conjunctions of atoms.

In this section, we will present a strategy, called *Predicate Pairing*, for making the choice of the predicates to be paired. This strategy, which is realized by Algorithm 1, takes as input a set of clauses and derives a new, equisatisfiable set of clauses, which by Theorem 4 is guaranteed to admit an $\mathscr{A}$-definable model, whenever the original clauses had one. Actually, as we will show, in many interesting cases the PP strategy constructs new sets of clauses for which a CHC solver is able to construct one such model, while the same solver is unable to do so for the original set of clauses. We present the PP strategy with the help of an example. Suppose that we are given the following specification of (a variant of) the Ackermann function:

S1. $ackermann(m,n) = n+1$ *if* $m \leqslant 0$

S2. $ackermann(m,n) = ackermann(m-1, 1)$ *if* $m > 0 \wedge n = 0$

S3. $ackermann(m,n) = ackermann(m-1, ackermann(m, n-1))$ *if* $m > 0 \wedge n > 0$

and the following two programs each of which implements that specification:

```
int ackermann1(int m, int n) {
    if (m =< 0) { return n+1; }
    else if (m > 0 && n = 0) { ackermann1(m-1,1); }
    else if (m > 0 && n > 0) { ackermann1(m-1,ackermann1(m,n-1)); }
  }
```

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Algorithm 1: The *Predicate Pairing* strategy

Input: (1) a clause C_{init} of the form: $false \leftarrow c_{init}, q(X), r(Y)$, and
(2) two disjoint sets Q and R of clauses such that: q occurs in Q, r occurs in R, and Q and R have no predicates in common.
Output: a set *TransfCls* of clauses such that there is no occurrence of an atom with predicate in Q and an atom with predicate in R in the same clause body.

Notation:
For all constraints d, for all atoms A and B,
let $Eq(d,A,B)$ be $\{X{=}Y \mid X{\in}vars(A),\ Y{\in}vars(B),\ \mathbb{D} \models \forall(d{\rightarrow}(X{=}Y))\}$.

$InCls := \{C_{init}\}$; $Defs := \emptyset$; $TransfCls := Q \cup R$;
while in *InCls* there is a clause C of the form: $L \leftarrow c, A_Q, B_R$, where:
(1) A_Q occurs in Q, and (2) B_R occurs in R **do**

- UNFOLDING: Unfold once the atom A_Q and once the atom B_R in clause C using $Q \cup R$, thereby deriving the set *UnfoldedCls* of clauses ;
- DEFINITION & FOLDING: *FoldedCls* := *UnfoldedCls* ;
  while in *FoldedCls* there is clause E of the form: $H \leftarrow d, A, B, G$, where:
  (1) A and B are atoms whose predicates occur in Q and R, respectively, and
  (2) G is a conjunction of atoms, such that:
  for all atoms M and N in (A,B,G), whose predicates occur in Q and R, respectively, we have that $|Eq(d,A,B)| \geqslant |Eq(d,M,N)|$ **do**
  if in *Defs* there is a clause D' of the form: $H' \leftarrow d', A', B'$ such that, for some substitution ϑ, we can fold (A, B) in clause E using D',
  then $FoldedCls := (FoldedCls \backslash \{E\}) \cup \{H \leftarrow d, H'\vartheta, G\}$;
  else
  let D be the clause $newp(Z) \leftarrow e, A, B$, where:
  (1) *newp* is a predicate symbol not occurring elsewhere,
  (2) $Z = vars(A) \cup vars(B)$, and
  (3) e is the conjunction of the equalities in $Eq(d,A,B)$;
  $FoldedCls := (FoldedCls \backslash \{E\}) \cup \{H \leftarrow d, newp(Z), G\}$;
  $Defs := Defs \cup \{D\}$; $InCls := InCls \cup \{D\}$;
  $InCls := InCls \backslash \{C\}$; $TransfCls := TransfCls \cup FoldedCls$;

```
int ackermann2(int m, int n) {
  while (m > 0) {
    if (n == 0) { m = m-1;   n = 1; }
    else { n = ackermann2(m,n-1);   m = m-1; }
  }
  return n+1;
}
```

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

We want to prove the equivalence of these two implementations, in the sense that, for all non-negative integers `m`$\geqslant$`0` and `n`$\geqslant$`0`, `ackermann1(m,n)` returns the same integer returned by `ackermann2(m,n)`.

Given the programs `ackermann1(m,n)` and `ackermann2(m,n)`, we first generate the following two sets Q and R of clauses that encode the operational semantics of the programs. These sets of clauses can be derived by specializing the interpreter of the imperative language with respect to the programs (De Angelis *et al.* 2017).

Q : (1) $ackermann1(M1, N1, A1) \leftarrow ack1(M1, N1, A1)$
(2) $ack1(M1, N1, A1) \leftarrow M1 \leqslant 0, A1 = N1+1$
(3) $ack1(M1, N1, A1) \leftarrow M1 > 0, N1 = 0, X1 = M1-1, Y1 = 1,\ ack1(X1, Y1, A1)$
(4) $ack1(M1, N1, A1) \leftarrow M1 > 0, N1 > 0, X1 = M1-1, Y1 = N1-1,$
$ack1(M1, Y1, Z1),\ ack1(X1, Z1, A1)$

R : (5) $ackermann2(M2, N2, A2) \leftarrow A3+1 = A2, ack2(M2, N2, A3)$
(6) $ack2(M2, N2, A2) \leftarrow M2 \leqslant 0, A2 = N2$
(7) $ack2(M2, N2, A2) \leftarrow M2 > 0, N2 = 0, M2 = X2+1, Y2 = 1,\ ack2(X2, Y2, A2)$
(8) $ack2(M2, N2, A2) \leftarrow M2 > 0, N2 \neq 0, X2 = M2-1, Y2 = N2-1, Z2 = Z3-1,$
$ack2(M2, Y2, Z2),\ ack2(X2, Z3, A2)$

The equivalence of the functions computed by the programs for `ackermann1` and `ackermann2` is expressed in terms of the predicates defined by clauses 1–8 as follows: for all integers $M1$, $M2$, $N1$, $N2$, we have that if $M1 \geqslant 0$, $M1 = M2$, $N1 \geqslant 0$, $N1 = N2$, and $ackermann1(M1, N1, A1)$ and $ackermann2(M2, N2, A2)$ both hold, then $A1 = A2$ holds. Thus, given the clause:

(9) $false \leftarrow A1 \neq A2,\ M1 \geqslant 0,\ M1 = M2,\ N1 \geqslant 0,\ N1 = N2,$
$ackermann1(M1, N1, A1),\ ackermann2(M2, N2, A2)$

the proof that `ackermann1` and `ackermann2` are equivalent is reduced to the construction of a model for clauses 1–9 (note that the constraint $A1 \neq A2$ in clause 9 states that the values returned by the two programs are different).

Now, we have that no CHC solver that constructs LIA-definable models, can prove the satisfiability of clauses 1–9. Indeed, in order to make that proof, the solver should discover that the atoms $ackermann1(M1, N1, A1)$ and $ackermann2(M2, N2, A2)$ imply the two equalities $A1 = ackermann(M1, N1)$ and $A2 = ackermann(M2, N2)$, respectively, where *ackermann* is the function specified by equations $S1$–$S3$, and these equalities cannot be expressed as linear integer constraints.

Thus, in order to allow a CHC solver to construct a LIA-definable model for clauses 1–9, one should avoid reasoning on the two predicates $ackermann1$ and $ackermann2$ in a separate way, and instead, one should reason on the *conjunction* of those predicates. Indeed, in what follows we will derive for that conjunction a new, equisatisfiable set of clauses that has a LIA-definable model. In this new set of clauses, we will discover suitable LIA constraints relating the arguments of the predicates $ackermann1(M1, N1, A1)$ and $ackermann2(M2, N2, A2)$. Using these constraints the CHC solver Z3 can show the existence of a LIA-definable model for

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

clauses 1–9, thereby proving the desired equivalence between programs `ackermann1` and `ackermann2`.

This new set of clauses will be derived from clause 9 by applying a sequence of transformation rules according to the PP strategy, as indicated in Algorithm 1. The algorithm takes as input a set of clauses $\{C_{init}\} \cup Q \cup R$, that is, $\{9\} \cup \{1,2,3,4\} \cup \{5,6,7,8\}$ in our case, and produces as output a new set of clauses by applying the unfolding, definition, and folding rules. During the application of that strategy, we silently apply the constraint replacement rule to remove clauses which have unsatisfiable constraints in their body.

• *First iteration of the body of the while-loop of Predicate Pairing.*

Since $InCls = \{9\}$, $ackermann1$ occurs in Q, and $ackermann2$ occurs in R, we start off by unfolding $ackermann1(M1, N1, A1)$ and $ackermann2(M2, N2, A2)$ in clause 9. These unfoldings correspond to a symbolic evaluation step of each of the two atoms. We get

(10) $false \leftarrow A1 \neq A2,\ M1 \geqslant 0,\ M1 = M2,\ N1 \geqslant 0,\ N1 = N2,\ A2 = A3+1,$
$ack1(M1, N1, A1), ack2(M2, N2, A3)$

Then in order to fold clause 10, we introduce the following definition clause 11 that pairs together the atoms with predicate $ack1$ occurring in Q and predicate $ack2$ occurring in R. In the body of this definition, we have the equality constraints $M1 = M2$ and $N1 = N2$ between the arguments of $ack1$ and $ack2$.

(11) $new1(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, N1 = N2,$
$ack1(M1, N1, A1), ack2(M2, N2, A2)$

The definition of $new1$ is then used for replacing, by folding, the conjunction of the atoms with predicates $ack1$ and $ack2$ in the body of clause 10. Thus, from clause 10, by folding, we derive

(12) $false \leftarrow A1 \neq A2,\ M1 \geqslant 0,\ M1 = M2,\ N1 \geqslant 0,\ N1 = N2,\ A2 = A3+1,$
$new1(M1, N1, A1, M2, N2, A3)$

• *Second iteration of the body of the while-loop of Predicate Pairing.*

Since $InCls = \{11\}$, $ack1$ occurs in Q, and $ack2$ occurs in R, we have to perform a second iteration of the body of the while-loop of the PP strategy.

We unfold the atoms with predicate $ack1$ and $ack2$ in the premise of clause 11, and we get the following three clauses:

(13) $new1(M1, N1, A1, M2, N2, N2) \leftarrow M1 \leqslant 0, M1 = M2, N1 = N2, A1 = N1+1$
(14) $new1(M1, N1, A1, M2, N2, A2) \leftarrow M1 > 0, M1 = M2, N1 = 0, N1 = N2,$
$X1 = M1-1, Y1 = 1, X2 = M2-1, Y2 = 1, ack1(X1, Y1, A1),\ ack2(X2, Y2, A2)$
(15) $new1(M1, N1, A1, M2, N2, A2) \leftarrow M1 > 0, M1 = M2, N1 > 0, N1 = N2, N2 \neq 0,$
$Y1 = N1-1, X1 = M1-1, Y2 = N2-1, X2 = M2-1, Z3 = Z2+1,$
$ack1(M1, Y1, Z1),\ ack1(X1, Z1, A1),\ ack2(M2, Y2, Z2),\ ack2(X2, Z3, A2)$

Clause 13 need not be folded because it has no atoms in its body. Clause 14 can be folded using clause 11 (the conditions for folding given in Section 3 are indeed satisfied), and we get

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

(16) $new1(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, M1 > 0, N1 = N2, N1 = 0,$
$X1 = M1 - 1, Y1 = 1, X2 = M2 - 1, Y2 = 1, new1(X1, Y1, A1, X2, Y2, A2)$

Clause 15 should be folded, but first we need to choose the atoms to be paired together. According to our goal of proving a relation between the arguments of *ackermann*1 and *ackermann*2, we should pair together an *ack*1 atom with an *ack*2 atom. However, the choice of the atoms to be paired can be made in different ways because in clause 15 there are two *ack*1 atoms and two *ack*2 atoms. The strategy we propose looks at the arguments of the atoms and selects the two atoms that share a maximal number of equality constraints holding between an argument of *ack*1 and an argument of *ack*2.

According to this strategy, we have that $ack1(M1, Y1, Z1)$ should be paired with $ack2(M2, Y2, Z2)$ because these two atoms share the two equalities $M1 = M2$ and $Y1 = Y2$ (this last equality follows from $N1 = N2$), while $ack1(M1, Y1, Z1)$ shares no equalities with $ack2(X2, Z3, A2)$. Moreover, $ack1(X1, Z1, A1)$ shares one equality only, namely $X1 = X2$ (this equality follows from $M1 = M2$) with $ack2(X2, Z3, A2)$. Thus, we pair $ack1(M1, Y1, Z1)$ with $ack2(M2, Y2, Z2)$ and then we take clause 11 for folding these two atoms.

In order to fold the other two atoms occurring in clause 15, that is, $ack1(X1,Z1,A1)$ and $ack2(X2, Z3, A2)$, we introduce the following clause:

(17) $new2(M1,N1,A1,M2,N2,A2) \leftarrow M1 = M2, ack1(M1,N1,A1), ack2(M2,N2,A2)$

Thus, by folding clause 15 using clauses 11 and 17, we get

(18) $new1(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, M1 > 0, N1 = N2, N1 > 0, N2 \neq 0,$
$Y1 = N1 - 1, X1 = M1 - 1, Y2 = N2 - 1, X2 = M2 - 1, Z3 = Z2 + 1,$
$new1(M1, Y1, Z1, M2, Y2, Z2),\ new2(X1, Z1, A1, X2, Z3, A2)$

The basic idea of our pairing strategy is that the atoms that are paired together, having some of their arguments equal, have a somewhat synchronized behaviour and this synchronization may determine, for the other arguments, the existence of simple relations that are easy to express in the theory of constraints one considers.

At this point of the application of the PP strategy, we have that $TransfCls = Q \cup R \cup \{12, 13, 16, 18\}$, $Defs = \{11, 17\}$, and $InCls = \{17\}$.

• *Third iteration of the body of the while-loop of Predicate Pairing.*

Since $InCls = \{17\}$ and clause 17 has the atom *ack*1 that occurs in Q and the atom *ack*2 that occurs in R, we have to perform a new iteration of the body of the while-loop of the PP strategy.

In clause 17, we unfold once the atom with predicate *ack*1 and the atom with predicate *ack*2, and we get

(19) $new2(M1, N1, A1, M2, N2, N2) \leftarrow M1 = M2, M1 \leqslant 0, A1 = N2 + 1$

(20) $new2(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, M1 > 0, N1 = 0, N2 = 0,$
$X1 = M1 - 1, Y1 = 1, X2 = M2 - 1, Y2 = 1, ack1(X1, Y1, A1), ack2(X2, Y2, A2)$

(21) $new2(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, M1 > 0, N1 = 0, N2 \neq 0,$
$X1 = M1 - 1, Y1 = 1, X2 = M2 - 1, Y2 = N2 - 1, Z3 = Z2 + 1,$
$ack1(X1, Y1, A1),\ ack2(M2, Y2, Z2),\ ack2(X2, Z3, A2)$

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

(22) $new2(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, M1 > 0, N1 > 0, N2 = 0,$
$X1 = M1 - 1, Y1 = N1 - 1, X2 = M2 - 1, Y2 = 1,$
$ack1(M1, Y1, Z1), ack1(X1, Z1, A1), ack2(X2, Y2, A2)$

(23) $new2(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, M1 > 0, N1 > 0, N2 \neq 0,$
$X1 = M1 - 1, Y1 = N2 - 1, X2 = M2 - 1, Y2 = N2 - 1, Z2 + 1 = Z3,$
$ack1(M1, Y1, Z1), ack1(X1, Z1, A1), ack2(M2, Y2, Z2), ack2(X2, Z3, A2)$

Clause 19 need not be folded. Clause 20 can be folded using definition 17 and we get

(24) $new2(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, M1 > 0, N1 = 0, N2 = 0,$
$X1 = M1 - 1, Y1 = 1, X2 = M2 - 1, Y2 = 1, new2(X1, Y1, A1, X2, Y2, A2)$

In order to fold clause 21, first we select the two atoms with the predicates to be paired. We have that $ack1(X1, Y1, A1)$ shares one equality with $ack2(X2, Z3, A2)$, that is, $X1 = X2$ (this equality follows from $M1 = M2$), and shares no equalities with $ack2(M2, Y2, Z2)$. Hence, we select the atoms $ack1(X1, Y1, A1)$ and $ack2(X2, Z3, A2)$ in the body of clause 21, and we fold that clause by using clause 17, thereby deriving the following clause:

(25) $new2(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, M1 > 0, N1 = 0, N2 \neq 0,$
$X1 = M1 - 1, Y1 = 1, X2 = M2 - 1, Y2 = N2 - 1, Z3 = Z2 + 1,$
$new2(X1, Y1, A1, X2, Z3, A2), ack2(M2, Y2, Z2)$

By processing clauses 22 and 23 in a similar manner, we get

(26) $new2(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, M1 > 0, N1 > 0, N2 = 0,$
$X1 = M1 - 1, Y1 = N1 - 1, X2 = M2 - 1, Y2 = 1,$
$new2(X1, Z1, A1, X2, Y2, A2), ack1(M1, Y1, Z1)$

(27) $new2(M1, N1, A1, M2, N2, A2) \leftarrow M1 = M2, M1 > 0, N1 > 0, N2 \neq 0,$
$X1 = M1 - 1, Y1 = N1 - 1, X2 = M2 - 1, Y2 = N2 - 1, Z3 = Z2 + 1,$
$new2(M1, Y1, Z1, M2, Y2, Z2), new2(X1, Z1, A1, X2, Z3, A2)$

Since $InCls = \emptyset$ no new iteration of the body of the while-loop of the PP strategy is required. Thus, the application of that strategy terminates. The resulting set of clauses is $TransfCls = Q \cup R \cup \{12, 13, 16, 18, 19, 24, 25, 26, 27\}$.

Now, the CHC solver Z3, when given as input the set *TransfCls* of clauses, constructs a LIA-definable model of *TransfCls*. In particular, it constructs a LIA-definable model of clause 12 by inferring that $new1(M1, N1, A1, M2, N2, A3)$ implies $A1 = A3 + 1$, which together with $A2 = A3 + 1$, implies $A1 = A2$, and hence the body of the clause is shown to be false.

By Theorem 2, the existence of a LIA-definable model for *TransfCls* entails that clauses 1–9 have a $\mathbb{Z}$-model, and this concludes the proof that programs `ackermann1` and `ackermann2` are equivalent.

We end this section by stating some results about the PP strategy. First, we have that the PP strategy always terminates because the number of new predicate definitions that can be introduced is bounded by the number k of different conjunctions of the form (e, A, B), where A and B are atoms whose predicates

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

occur in $Q \cup R$, and e is a conjunction of equalities between variables in (A, B). Hence, the number of executions of the body of the while-loop of PP is at most k.

It is easy to see that the sequence of applications of the transformation rules performed by the PP strategy constructs a transformation sequence where every definition in *Defs* is unfolded at least once. Thus, from Theorem 2, which states the equivalence with respect to $\mathbb{D}$-satisfiability, and Theorem 4, which states the preservation of $\mathscr{A}$-definable models, we get the following result.

Theorem 6 (*Termination and soundness of the Predicate Pairing strategy*)
Let the sets $\{C_{init}\}$, Q, and R of clauses be the input of the PP strategy. Then the strategy terminates and returns a set *TransfCls* of clauses such that

(1) $\{C_{init}\} \cup Q \cup R$ is $\mathbb{D}$-satisfiable iff *TransfCls* is $\mathbb{D}$-satisfiable, and
(2) if $\{C_{init}\} \cup Q \cup R$ has an $\mathscr{A}$-definable model, then *TransfCls* has an $\mathscr{A}$-definable model.

Finally, note that the application of the PP strategy may be iterated, and hence, at the end of the transformation of a set of clauses, more than two predicates may turn out to be tupled together.

5 Case studies: Relational program properties

In this section, we illustrate the application of the PP strategy to some relevant classes of relational program properties. In particular, we have considered the following classes of properties: (1) the equivalence of programs implementing non-linear and/or nested recursive functions, (2.1) the injectivity of programs, (2.2) the monotonicity of programs, (2.3) the functional dependence of programs, (3) non-interference of programs, (4) equivalence of loop-optimized versions of programs with respect to the corresponding non-optimized versions, and (5) the equivalence of programs that manipulate integer arrays. We will consider these classes of properties in separate subsections.

Now, we briefly show how to encode relational properties between the executions of two programs P and Q (De Angelis *et al.* 2016). We assume that the operational semantics of programs P and Q is represented by predicates $p(A, B)$ and $q(X, Y)$, where A and X represent (tuples of) input values, and B and Y represent (tuples of) output values, respectively. As already mentioned in the previous sections, the clauses defining p and q can be derived by specializing the interpreter of the imperative language with respect to the programs (De Angelis *et al.* 2017).

Let us now consider the relational property stating that, if the constraint $pre(A,B)$ holds before the execution of P and Q and the execution of these programs terminates, then the constraint $post(A, B, X, Y)$ holds after the execution. This property can be verified by testing the satisfiability of the CHC set consisting of the clauses defining predicates p and q, together with the following clause:

$$RP: \ false \leftarrow pre(A, B), \ notpost(A, B, X, Y), \ p(A, B), \ q(X, Y)$$

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

where $notpost(A, B, X, Y)$ is a constraint which is equivalent to the negation of $post(A, B, X, Y)$.[1]

The application of our method based on the use of the PP strategy, is often crucial for solving satisfiability problems that encode relational program properties. In Section 6, we will discuss the results we have obtained in an extensive experimental evaluation that we have conducted.

5.1 Functions with non-linear and/or nested recursion

Similarly to what has been considered in Section 4, where we have presented two imperative programs implementing the Ackermann function specification and then proved their equivalence, in this section, we consider various equivalence problems for pairs of imperative programs implementing some functional specifications. In each pair one imperative program uses recursion only and the other one uses recursion and iteration.

The operational semantics of the two imperative programs is encoded using two distinct sets of CHCs, each defining a predicate for each program. The recursive structure of these predicate definitions mirrors the control flow of two imperative programs.

Let us consider two predicates $p1$ and $p2$ that encode the operational semantics of the two imperative programs, say $P1$ and $P2$, implementing a given function specification. The *equivalence* between $P1$ and $P2$ holds if, under some precondition on the input values, $p1$ and $p2$ define the same input/output relation. This property holds if the following clause, together with the set of clauses defining the predicates $p1$ and $p2$, is satisfiable:

$$EQ:\ \mathit{false} \leftarrow c(X1),\ X1 = X2,\ Y1 \neq Y2,\ p1(X1, Y1),\ p2(X2, Y2)$$

where: (1) $X1$, $X2$ represent tuples of input values, (2) $Y1$, $Y2$ represent tuples of output values, and (3) $c(X1)$ is a precondition on the input values. The reader may note that clause EQ is an instance of clause RP defining the general relational property. Note also that clause 9 in Section 4, encoding the equivalence relation between the two implementations of the Ackermann function, is an instance of EQ.

We have considered equivalence problems for imperative programs implementing *non-linear* recursive functional specifications, that is, functional specifications with two or more recursive calls that depend on the same call (as in the case of the Fibonacci function). Also, several of these specifications have *nested recursions*, that is, they have recursive calls that are arguments of other recursive calls (as in the case of the Ackermann function), thus making the verification problem more challenging.

In particular, in our experiments we have considered the following specifications of variants of the Ackermann function.[2] (Here and in the other function definitions we assume that x, y, and z are non-negative integers.)

[1] If the constraint language has no negation symbol, but the negation of a constraint is equivalent to a disjunction of constraints, as in the case of LIA, then the relational property can be encoded by a set of clauses.

[2] `http://mrob.com/pub/math/ln-2deep.html`

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

(1) Original version by W. Ackermann:
$A(0,y,z) = y+z, \quad A(1,y,0) = 0, \quad A(2,y,0) = 1,$
$A(x+3,y,0) = y, \quad A(x+1,y,z+1) = A(x,y,A(x+1,y,z))$

(2) Variant by H. Edelsbrunner:
$E(0,y,z) = y+z, \quad E(x+1,y,0) = 0, \quad E(x+1,y,1) = y,$
$E(x+1,y,z+2) = E(x,y,E(x+1,y,z+1))$

(3) Variant by R. Robinson (we have used this variant in Section 4):
$R(0,y) = y+1, \quad R(x+1,0) = R(x,1), \quad R(x+1,y+1) = R(x,R(x+1,y))$

(4) Variant by R. Péter:
$P(0,y) = 2y+1, \quad P(x+1,0) = P(x,1), \quad P(x+1,y+1) = P(x,P(x+1,y))$

Note that these variants of the specification of the Ackermann function actually correspond to pairwise different functions. Indeed, we have that $A(2,y,0) \neq E(2,y,0)$ for all $y \geqslant 0$, and $R(0,y) \neq P(0,y)$ for all $y > 0$.

Additionally, we have considered some other equivalence problems for pairs of imperative programs encoding the following functional specifications:

(5) a variant of the Sudan function:
$S(0,y,z) = y+z, \quad S(x+1,y,0) = S(x,y+1,0),$
$S(x+1,y,z+1) = S(x,S(x+1,y,z),y+S(x+1,y,z))$

(6) the B function:
$B(0,y) = y+1, \quad B(x+1,y) = B(x,B(x+1,y-1))$

(7) the G function:
$G(1,y) = y+3, \quad G(x+2,0) = G(x+1,1), \quad G(x+2,y+1) = G(x+1,G(x+2,y))$

(8) the McCarthy 91 function: $M(x) =$ if $x > 100$ then $x-10$ else $M(M(x+11))$

(9) the Dijkstra *fusc* function:
$fusc(0) = 0, \quad fusc(1) = 1, \quad fusc(2x) = fusc(x),$
$fusc(2x+1) = fusc(x+1) + fusc(x), \quad$ and

(10) a function that computes the minimum number of moves needed for the solution of the towers of Hanoi problem.

Our strategy turns out to be very effective in increasing the ability of the CHC solver to prove the satisfiability, or the unsatisfiability, of the clauses encoding the considered problems. As already mentioned, the main reason for this effectiveness is due to the fact that, by pairing together two atoms, the PP strategy often enables the discovery of relations between some of their arguments.

5.2 *Monotonicity, injectivity, and functional dependence*

Some interesting classes of relational properties we have considered are those of monotonicity, injectivity, and functional dependency. These notions relate two different terminating executions of the same program on two distinct input values, say x and y, computing the output values, say m and n, respectively. The definition of these properties are derived in a straightforward manner from those of the mathematical functions.

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

In particular, monotonicity properties state that the application of the program on ordered input values produces ordered output values. For example, a typical monotonicity property is the following: if $x \leqslant y$, then $m \leqslant n$.

Injectivity properties state that any two executions of the same program on different inputs produce different outputs, that is, if $x \neq y$, then $m \neq n$.

Functional dependence properties state that the output of a program is a function of (a possibly proper subset of) its input values: for instance, if $x = y$, then $m = n$.

In particular, let us consider the following CHCs encoding the operational semantics of a given imperative recursive program `Fib` that computes the Fibonacci numbers:

$fib(X,Y) \leftarrow X = 0, Y = 0$
$fib(X,Y) \leftarrow X = 1, Y = 1$
$fib(X,Y) \leftarrow X \geqslant 2, X1 = X-1, X2 = X-2, Y = Y1+Y2,\ fib(X1,Y1),\ fib(X2,Y2)$

where the first argument of *fib* encodes the input and the second argument of *fib* encodes the output. Then, the above mentioned properties of monotonicity, injectivity, and functional dependence of the program `Fib` can be checked by testing the satisfiability of the following clauses:

$false \leftarrow Y2 \geqslant Y1+1,\ X1 \geqslant X2,\ fib(X1,Y1),\ fib(X2,Y2)$ (Monotonicity)
$false \leftarrow Y1 = Y2,\ X1 \neq X2,\ fib(X1,Y1),\ fib(X2,Y2)$ (Injectivity)
$false \leftarrow Y1 \neq Y2,\ X1 = X2,\ fib(X1,Y1),\ fib(X2,Y2)$ (Functional Dependence)

Note that the above clauses are all instances of clause *RP* encoding the general relational property.

Based on this example, the reader will not find it difficult to express monotonicity, injectivity, and functional dependence for other given imperative programs. We have successfully verified these properties for programs computing: (1) the sum of two numbers (by iterated increment), (2) the product of two numbers (by iterated addition), and (3) the square and the cube of a number (by iterated addition). We have also considered some more programs containing simple, sequential, or nested while-loops, possibly combined with conditionals.

5.3 Non-interference

Non-interference is a property that guarantees information-flow security. It can be viewed as a variant of the functional dependence property as we now indicate.

Let us consider an imperative program P whose variables are partitioned into a set of public variables (or low security variables) and a set of private variables (or high security variables). We say that P satisfies the non-interference property if any two terminating executions of P, starting with the same initial values of the public variables, but possibly with different values of the private variables, compute the same values of the public variables. Thus, if a program satisfies the non-interference property, an attacker cannot acquire information about the private variables by observing the input/output relation between the public variables, which are functionally dependent on the public input variables only.

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

To clarify the ideas, let us consider the following simple imperative program HL:

```
while (high >= 1) { high = high-1; low = high; }
```

where low is a public variable and high is a private variable. Program HL violates the non-interference property because there exist two different executions starting with identical values of the variable low and terminating in configurations having different values for the variable low. Indeed, if initially we have that high is at least 1, then the body of while-loop is executed and the final value of low will be 0, otherwise the value of low is left unchanged.

The non-interference property for program P can be verified by checking the satisfiability of the following set of clauses:

$$false \leftarrow OutL \neq OutL1,\ L = L1,\ p(L, H, OutL),\ p(L1, H1, OutL1)$$
$$p(L, H, OutL) \leftarrow H < 1,\ OutL = L$$
$$p(L, H, OutL) \leftarrow H \geqslant 1,\ H1 = H - 1,\ L1 = H1,\ p(L1, H1, OutL)$$

where: (1) the predicate $p(L, H, OutL)$ encodes the input/output relation among the variables of program P, (2) the variables L and H encode the values of the variables low and high at the beginning of the while-loop, and (3) $OutL$ encodes the value of the variable low at the end of the while-loop. Note that the set of clauses shown above is *unsatisfiable* because program P *violates* the non-interference property.

The reader may note that the first clause, encoding the non-interference property for program P, is an instance of clause RP defining the general relational property. The encoding of the non-interference property for other programs can be done in a similar way.

The following program HL1 is representative of a class of programs for which we have successfully verified that the non-interference property holds:

```
low1 = low2;   low1 = low1 + f(high);   low1 = low1 - g(high,low1);
```

where: (1) low1 and low2 are public variables, high is a private variable, and (2) f and g are two functions defined as follows:

```
int f(int m) {
  int i = 0, s = 0;
  while (i <= m) { s += i+m; i++; }
  return s;  }

int g(int m, int n) {
  int i = 0, s = 0;
  if (n <= m) { while ( i<= n) { s += i+m; i++; } } ;
  while (i <= m) { s += i+m; i++; }
  return s; }
```

Note that, in the program HL1 the functions f and g compute the same value. This program HL1 *does satisfy* the non-interference property, and thus the corresponding set of clauses is *satisfiable*. Indeed, in the program HL1 the public variable low1 is first incremented and then decremented by the same value, which, however, is

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

computed by the distinct, yet equivalent functions `f` and `g`, which take the private variable `high` as input.

5.4 Loop optimizations

Modern compilers often perform a series of optimizations for producing a new program that is semantically equivalent to an old program, but whose execution is faster, or requires less memory, or has lower energy consumption.

By applying our method, we have successfully verified equivalence properties between some imperative programs and their optimized versions (Lopes and Monteiro 2016). The CHC encoding of program equivalence is the one defined by clause *EQ* in Section 5.1. For instance, we have proved the equivalence of the following program:

```
while (i < n) {
   if (n > 5) { a = a+n;   i = i+1; }
      else    { a = a+1;   i = i+1; }
}
```

and the one derived from it by the *loop unswitching* optimization:

```
if (n > 5) { while (i < n) { a = a+n;   i = i+1; } }
   else    { while (i < n) { a = a+1;   i = i+1; } }
```

where the conditional statement occurring in the while-loop is moved outside the loop, so that the evaluation of the conditional expression is performed only once, instead of being performed at each loop iteration.

We have also considered some specific instances of other equivalence problems relating original, non-optimized programs to new programs obtained by applying the following loop optimizations:

(1) *loop fission*, that splits the commands occurring in a loop in two blocks that are then executed by two consecutive, independent loops;
(2) *loop fusion*, that merges the commands occurring in consecutive loops and executes them in a single loop;
(3) *loop reversal*, that executes the commands occurring in a loop, in a new loop where the iteration proceeds in reversed order with respect to the order of the original loop;
(4) *strength reduction*, that replaces iterated expensive computations in a loop by cheaper ones (for instance, replacing multiplication by a loop index with addition); and
(5) *code sinking*, that moves code occurring immediately before or after a loop inside the loop itself, possibly using conditionals for keeping the semantics of the program unaltered.

We have also considered other loop optimizations, including *loop tiling*, *loop aligning*, *loop pipelining* as well as other optimizations for removing redundant assignments, expression evaluations, and conditionals.

We are confident that our method of proving equivalence of programs can be extended for proving correctness of code optimizations at a schematic level (Leroy

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

2009; Lopes and Monteiro 2016), and not for some specific instances only. We leave this study for future research.

5.5 *Array-manipulating programs*

We have applied our verification method to relational properties of imperative programs manipulating integer variables and integer arrays.

Let us first introduce some preliminary notions. An *integer array a* (or an *array*, for short) is a finite sequence of integers whose length, called the dimension of the array, is denoted $dim(a)$. An *atomic array constraint* is either $read(a, i, v)$, denoting that the i-th element of the array a is the integer v, or $write(a, i, v, b)$, denoting that, for $k = 1, \ldots, dim(a)$, if $k \neq i$, then the k-th element of a is equal to the k-th element of b, and if $k = i$, then the k-th element of b is the integer v.

The *read* and *write* constraints satisfy the following implicative axioms (Bradley *et al.* 2006), whose variables are assumed to be universally quantified at the front:

$$\begin{array}{lll} (I = J,\ read(A,I,U),\ read(A,J,V)) \rightarrow U = V & & (\textit{array congruence}) \\ (I = J,\ write(A,I,U,B),\ read(B,J,V)) \rightarrow U = V & & (\textit{read-over-write}) \\ (I \neq J,\ write(A,I,U,B),\ read(B,J,V)) \rightarrow read(A,J,V) & & (\textit{read-over-write}) \end{array}$$

For example, the operational semantics of the following imperative program which acts on the array `a`:

```
i = 1;   a[0] = 3;
while ( i < n ) { a[i] = a[i-1]+2;  i++; }
```

can be represented by the following set of CHCs:

$$\begin{array}{l} prog(N, A1, A3) \leftarrow I = 1,\ K = 0,\ U = 3,\ write(A1, K, U, A2),\ loop(N, A2, I, A3) \\ loop(N, A1, I, A3) \leftarrow I + 1 \leqslant N,\ J = I - 1,\ read(A1, J, U), \\ \qquad\qquad V = U + 2,\ write(A1, I, V, A2),\ I1 = I + 1,\ loop(N, A2, I1, A3) \\ loop(N, A, I, A) \leftarrow I \geqslant N \end{array}$$

In these clauses: (1) the predicate *loop(N, A_in, I, A_out)* encodes the while-loop, (2) its arguments N, *A_in*, and I encode the values of variables `n`, `a`, and `i`, respectively, at loop entry, and (3) *A_out* encodes the value of `a` at loop exit.

Now, we show an example of an equivalence property between array manipulating programs that has been proved by using our method based on the PP strategy. Let us consider the programs `P1` and `P2` shown in Table 1, where program `P2` is obtained from program `P1` by applying the loop-pipelining optimization, a commonly used technique for enabling instruction-level parallelism at the hardware level.

The equivalence between programs `P1` and `P2` with respect to the output array `c`, is expressed by the following clause F (which is an instance of clause EQ of Section 5.1):

$$\begin{array}{l} F:\ false \leftarrow X \neq Y,\ N \geqslant 1,\ J \geqslant 0,\ J \leqslant N - 1,\ read(C1, J, X), read(C2, J, Y), \\ \qquad\qquad new11(N, A, C1),\ new21(N, A, C2) \end{array}$$

where $new11(N, A, C1)$ represents the input/output relation of the source program `P1`, and in particular, N is the value of the integer variable `n`, A is the

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Table 1. *The source program* P1 *(left) and the optimized program* P2 *obtained by applying the loop pipelining transformation (right)*

Program P1	Program P2

`i=0;`
`while (i < n) {`
`  a[i]++;`
`  b[i] += a[i];`
`  c[i] += b[i];`
`  i++;`
`}`
	`i = 0;`
	`a[0]++;`
	`b[0] += a[0];`
	`a[1]++;`
	`while ( i < n-2 ) {`
	`  a[i+2]++;`
	`  b[i+1] += a[i+1];`
	`  c[i] += b[i];`
	`  i++;`
	`}`
	`c[i] += b[i];`
	`b[i+1] += a[i+1];`
	`c[i+1] += b[i+1];`

value of the array a at the beginning of program execution, and $C1$ is the value of the array c at the end of program execution. Similarly, $new21$ represents the input/output relation of the optimized program P2.

Thus, by proving the satisfiability of the set of clauses consisting of F together with the clauses defining $new11$ and $new21$, we have been able to prove that programs P1 and P2 produce identical values for the array c as output, when provided with identical values for the array a as input.

6 Experimental evaluation

In this section, we present the experimental evaluation we have performed for assessing the effectiveness of the PP strategy presented in Section 4.

Implementation. We have implemented Algorithm 1 by using the VeriMAP system (De Angelis *et al.* 2014b), which is a tool for software model checking based on transformation techniques for CHCs. Then, we have used the SMT solver Z3 (de Moura and Bjørner 2008) for checking the satisfiability of the clauses generated by VeriMAP. In particular, we have used Z3 version 4.5.0 with the Duality fixed-point engine (McMillan and Rybalchenko 2013), which provides support for constraints defined on linear integers and integer arrays.

Our prototype implementation consists of two components: (1) a module that realizes Algorithm 1, and (2) a module that translates the generated CHCs into the SMT-LIB format, which is the format accepted by Z3 (see Fig. 3). The VeriMAP system also provides a front-end module (T) that takes a pair of C programs, together with a relational property to be verified, and translates them into the CHCs that encode the verification problem (De Angelis *et al.* 2016).

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

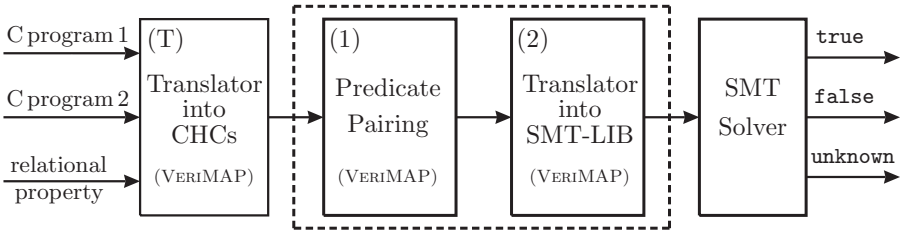


Fig. 3. The verification System: (1) the Predicate Pairing module, implementing Algorithm 1, and (2) the Translator to SMT-LIB module.

Verification problems. We have considered a benchmark[3] of 163 sets of CHCs (153 of which are satisfiable and the other 10 are unsatisfiable), representing verification problems (all acting on integers and 14 of them acting on integer arrays), which refer to relational properties of small, yet non-trivial, imperative programs mostly taken from the literature (Benton 2004; Barthe *et al.* 2011; Felsing *et al.* 2014; De Angelis *et al.* 2015a; De Angelis *et al.* 2016; Lopes and Monteiro 2016).

We have considered the following categories of relational verification problems. (1) The category NLIN, which refers to equivalence properties between programs implementing functions with non-linear and/or nested recursion. For instance, we have verified the equivalence of two programs for computing the Ackermann function (see the running example presented in Section 4 and the examples of Section 5.1). (2)–(6) The categories MON, INJ, FUN, NINT, and LOPT, which refer to monotonicity, injectivity, functional dependence, non-interference, and loop optimizations problems, respectively (see Section 5). (7) The category ITE, which refers to equivalence properties and inequality properties relating two iterative programs acting on integers (by an inequality property we mean that the values computed by the programs are related by $\leqslant$). (8) The category ARR, which refers to equivalence and inequality properties between iterative programs acting on integer arrays. (9) The category REC, which refers to equivalence properties between recursive programs. (10) The category I-R, which refers to equivalence properties between an iterative program and a (non-tail) recursive program. For example, we have verified the equivalence of the iterative and recursive versions of programs for computing: (10.1) the greatest common divisor of two integers, and (10.2) the m-th triangular number, that is, $\sum_{x=1}^{m} x$. (11) The category COMP, which refers to equivalence and inequality properties between programs that contain compositions of different numbers of loops acting on integers (3 problems) and integer arrays (7 problems). (12) The category PCOR, which refers to partial correctness properties of an iterative program with respect to a recursive functional postcondition (De Angelis *et al.* 2015a).

Note that in our benchmark set, 31 problems (belonging to the categories NLIN, NINT, and COMP) are encoded by sets of clauses that include non-linear clauses,

[3] The benchmark suite can be found at `http://map.uniroma2.it/predicate-pairing`. The suite includes the CHCs that encode the relational verification problems (both in Prolog and SMT-LIB format), the transformation logs, the transformed CHCs, the statistics on the transformations and on the Z3 satisfiability checks. For the verification problems generated from C programs, the suite also contains the C sources.

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Table 2. *Results of the experiments.*

Problems			Z3 before PP		PP	Z3 after PP	
Category		P	S_1	T_1	T_{PP}	S_2	T_2
(1)	NLIN	13	4	16.11	25.80	13	13.12
(2)	MON	18	1	1.04	2.27	12	3.72
(3)	INJ	11	0	–	1.36	8	1.39
(4)	FUN	7	4	1.39	1.24	7	1.48
(5)	NINT	18	3	0.27	55.80	17	41.33
(6)	LOPT	20	2	4.83	2.98	15	10.71
(7)	ITE	22	5	26.67	4.53	18	17.01
(8)	ARR	6	1	7.45	2.04	5	3.25
(9)	REC	15	6	2.89	1.50	13	4.28
(10)	I-R	4	0	–	0.65	3	1.02
(11)	COMP	10	0	–	16.35	7	6.46
(12)	PCOR	19	5	83.93	17.84	17	17.65
Total number		163	31	144.58	132.36	135	121.42
Average time				4.66	0.81		0.90

The first two columns report the names of the problem categories and the number P of problems in each category, respectively.
Columns S_1 and S_2 report the number of verification problems solved by Z3 before and after the application of the PP strategy, respectively. Columns T_1 and T_2 report the time taken by Z3 to solve the problems reported in columns S_1 and S_2, respectively. Column T_{PP} reports the time taken by VeriMAP to apply the PP strategy on the P problems. Times are in seconds. The timeout is of 300 seconds. No timeout occurred during the application of the PP strategy.

besides the ones that encode the relational properties, which are always non-linear (see Sections 4 and 5 for some examples).

Technical resources. The experimental evaluation has been performed on a single core of an Intel Core Duo E7300 2.66 GHz processor with 4 GB of memory running Ubuntu. For all problems, we have set the timeout limit of 300 seconds.

Experimental processes. We have considered the following two experimental processes. (E1) The first experimental process consists in running Z3 for checking the satisfiability of the original sets of CHCs that encode the verification problems. (E2) The second experimental process consists in running Algorithm 1 on the original sets of CHCs for each verification problem, and then running Z3 for checking the satisfiability of the derived CHCs. In this second process, the PP strategy has been iterated (see the end of Section 4), if more than one pair of atoms was present in the bodies of the original sets of clauses. In particular, PP has been iterated for 24 sets of CHCs, in total, belonging to categories NINT, ARR, COMP, and PCOR.

Results. The results of the experimental evaluation are summarized in Table 2. The times reported are the CPU seconds spent in user mode and kernel mode by (1) VeriMAP for transforming the clauses, and (2) Z3 for checking their satisfiability. The times required for translating the CHCs into the SMT-LIB format (this translation is necessary for running Z3) are: 6.39 seconds for the first experimental

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

process E1, and 22.59 seconds for the second experimental process E2. Those translation times are not very high with respect to the solving times which are 144.58 seconds and 121.42 seconds, respectively.

As expected, the use of the PP strategy significantly increases the number of problems that are solved by Z3. In particular, the number of solved problems increases from 31 (Total of Column S_1) to 135 (Total of Column S_2). Note, however, the PP strategy can derive a set of clauses, which is larger than the original set. In our benchmark, we have observed that size increases of about 2.16 times on average.

Note also that the PP strategy increases the efficiency of the satisfiability check. Indeed, the average time taken to run the PP strategy and then Z3 (1.88 seconds) is lower than the average time taken to run Z3 on the original set of clauses (4.66 seconds). Although we have proved that the application of the PP strategy preserves all LIA-definable models, in our experiments we have found three problems which Z3 was able to solve before the application of the PP strategy (i.e., in the first experimental process), and it was no longer able to solve, within the given time limit, after the application of the PP strategy (i.e., in the second experimental process). This phenomenon may be due to the fact that the termination of the algorithms implemented by the Z3 solver is sensitive to the syntactic form of the input clauses, and the PP strategy modifies that form. Further work is needed to improve the termination behaviour of the solver.

Finally, we would like to point out that the application of the PP strategy does not decrease the efficiency of the whole verification process. If we consider only the 28 problems for which Z3 is able to solve both before and after the application of the PP strategy, the average time taken to run the PP strategy and then Z3 (1.58 seconds) is slightly lower than the time taken by Z3 alone (1.77 seconds).

7 Related work and conclusions

The basic idea behind the PP transformation strategy for CHCs is that, by finding a recursive definition of a predicate denoting the conjunction of two atoms, it is often possible to infer relations among the variables occurring in the two atoms, which would have been impossible to discover by considering each atom separately.

Techniques for transforming logic programs by deriving new predicates defined in terms of conjunctions of atoms have been largely studied. Let us recall, for instance, the well-known *tupling* transformation strategy (Pettorossi and Proietti 1994) and the *conjunctive partial deduction* technique for logic program specialization (De Schreye *et al.* 1999). The main objective of tupling and conjunctive partial deduction is the derivation of more efficient logic programs by avoiding multiple traversals of data structures and repeated evaluations of predicate calls, and by producing specialized program versions.

Thus, the PP strategy shares with tupling and conjunctive partial deduction the idea of promoting a conjunction of atoms to a new predicate. However, in this paper, we have shown that the application of this idea to CHCs can also play a key role in improving the effectiveness of CHC solvers for proving properties of imperative

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

programs, besides the optimization of the execution of logic programs, which is the objective of tupling and conjunctive partial deduction. This is the case especially when the CHC solvers are required to test the satisfiability of clauses that encode relational program properties, that is, properties that relate two programs, or two executions of the same program. Indeed, as shown by many examples considered in this paper, state-of-the-art solvers often fail to prove the satisfiability of sets of clauses encoding relational properties because they can only infer relations among the variables of individual atoms.

We have considered CHC solvers that prove satisfiability by using predicate abstraction, that is, by looking for models that are definable in a specific class $\mathscr{A}$ of constraints (Bjørner *et al.* 2015). We have shown that, in principle, the PP strategy cannot worsen the effectiveness of the CHC solver. Indeed, we have proved a very general result concerning the unfold/fold transformation rules used by the strategy: if a set of clauses is transformed by applying the unfold/fold rules, and the original set of clauses has an $\mathscr{A}$-definable model, then also the transformed set of clauses has an $\mathscr{A}$-definable model. Thus, if the CHC solver is able to find an $\mathscr{A}$-definable model whenever it exists (and this is indeed possible if the validity problem for the constraints in $\mathscr{A}$ is decidable), then every set of clauses that can be proved satisfiable by the solver before the transformation, will also be proved satisfiable by the solver after the transformation. We have shown that, in practice, for the Z3 solver this property is guaranteed with very few exceptions.

We have also given some restrictions on the use of the rules that guarantee that the converse of the above property holds, that is: if a set of clauses is transformed by applying the unfold/fold rules, and the transformed set of clauses has an $\mathscr{A}$-definable model, then also the original set of clauses has an $\mathscr{A}$-definable model. However, this property is not always desirable. Indeed, the fact that in some cases the PP strategy is able to transform (satisfiable) clauses that do not have an $\mathscr{A}$-definable model into clauses that have an $\mathscr{A}$-definable model, may be a great advantage. Indeed, this means that while a CHC solver that looks for $\mathscr{A}$-definable models is not able to prove the satisfiability of the original clauses, the same solver may be able to prove the satisfiability of the transformed clauses.

The study of the properties that relate unfold/fold transformations and the existence of $\mathscr{A}$-definable models is not present in the literature on tupling and conjunctive partial deduction.

Then, we have presented an algorithm that implements the PP strategy. One of the novel points of this algorithm with respect to tupling and conjunctive partial deduction is that it realizes a heuristic to choose the appropriate atoms to be paired together in a new predicate definition, by maximizing the number of equality constraints that relate the variables occurring in a pair of atoms. This heuristic is crucially needed when dealing with non-linear clauses that, by unfolding, may generate clauses with more than two atoms in their body. We have implemented our algorithm on the VeriMAP transformation and verification system (De Angelis *et al.* 2014b), and we have evaluated its effectiveness on a benchmark of over 160 problems encoding relational properties of small, yet nontrivial C-like programs. The properties were of various kinds, including equivalence, injectivity, functional

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

dependence, and non-interference (a property of interest for enforcing software security). The results show that the use of the PP strategy as a preprocessor greatly improves the ability of the Z3 CHC solver (with the Duality fixed-point computation engine (McMillan and Rybalchenko 2013)) to prove satisfiability.

Several transformation-based techniques for CHCs, or constraint logic programs, have been proposed as a means of facilitating program verification. Many of them are non-conjunctive specialization techniques, which work by propagating the constraints occurring in the goal, thereby producing clauses with strengthened constraints in their bodies (Peralta *et al.* 1998; Albert *et al.* 2007; Méndez-Lojo *et al.* 2008; De Angelis *et al.* 2014a; De Angelis *et al.* 2015a; De Angelis *et al.* 2017; Kafle and Gallagher 2017a; Kafle and Gallagher 2017b). Even if specialization techniques have been shown to be very successful, in most cases they cannot achieve the same effect as the PP strategy. Indeed, as already mentioned, the PP strategy works by introducing new predicates corresponding to conjunctions of old predicates, whereas non-conjunctive specialization can only introduce new predicates that correspond to instances of old predicates. We have experimentally checked that most of the problems considered in Section 6 cannot be solved via (non-conjunctive) specialization alone. Due to lack of space we have not reported these results.

The query-answer transformation (and variants thereof) is another pre-processing technique that is sometimes applied before performing satisfiability tests using CHC solvers (De Angelis *et al.* 2014a; Kafle and Gallagher 2017a; Kafle and Gallagher 2017b). The aim of this transformation is to simulate the top-down, goal oriented evaluation of the clauses in a bottom-up framework. The results we have presented here, showing that the PP strategy is able to transform clauses without an $\mathscr{A}$-definable model into clauses with an $\mathscr{A}$-definable model, are independent of the evaluation strategy adopted by the CHC solvers, and hence the query-answer transformation and the PP strategy should be viewed as orthogonal techniques.

The PP strategy is an extension of the *Linearization* transformation, whose objective is to transform a set of linear clauses (that is, clauses with at most one atom in their body) together with a non-linear goal, into a set of linear clauses and linear goals (De Angelis *et al.* 2015b). The PP strategy does not need any linearity assumption, and indeed in Sections 4, 5, and 6 we have shown that this strategy can solve several verification problems encoded by sets of non-linear clauses. It has also been shown that Linearization preserves the existence of LIA-definable models (De Angelis *et al.* 2015b). Here, we have generalized this result by proving that the application of the unfold/fold transformation rules, independently of the strategy, preserves the existence of $\mathscr{A}$-definable models, for any class $\mathscr{A}$ of constraints.

The PP algorithm presented here is an improvement of the one reported in previous work presented at the SAS Symposium (De Angelis *et al.* 2016). Indeed, as already mentioned, here we use an equality-based heuristic to choose the appropriate atoms to be paired together, and this technique has been shown very effective in practice for handling non-linear clauses. Also the case studies and the benchmark set we consider in the present paper are much larger, and include verification problems such as non-interference, correctness of loop optimizations, and equivalence of non-linear recursive programs that have not been considered in our SAS paper (De Angelis

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

et al. 2016). Moreover, in the present paper we include general results concerning the preservation of $\mathscr{A}$-models.

Bjørner *et al.* have shown that unfolding preserves $\mathscr{A}$-definable models provided that the set $\mathscr{A}$ of constraints admits Craig interpolation (Bjørner *et al.* 2015). In the present paper, we have generalized this result by considering also other transformations, and in particular folding, and we dropped the assumption about Craig interpolation. Moreover, we assume that $\mathscr{A}$ is a subset of the set $\mathscr{C}$ of constraints over which the clauses are defined, while Bjørner *et al.* take $\mathscr{A}$ to be equal to $\mathscr{C}$. Our generalization is significant because sometimes CHC solvers that make use of predicate abstraction look for models defined in subsets of the constraints used for the clauses, such as the popular domain of the octagons (Miné 2006).

The problem of verifying relational properties is very relevant in the context of software engineering. Indeed, during software development it is often the case that one modifies the program text, and hence needs a proof that the semantics of the new program version has some specified relation to the semantics of the old version. This kind of proof is sometimes called *regression verification* (Godlin and Strichman 2008).

Several logics and methods have been presented in the literature for reasoning about various relational program properties. A Hoare-like axiomatization of relational reasoning for simple while-programs has been proposed by Benton (Benton 2004), who however does not present any technique for the automation of the proofs.

Program equivalence is one of the relational properties that have been extensively studied in the past, and still receives remarkable attention in recent work (Godlin and Strichman 2008; Zaks and Pnueli 2008; Barthe *et al.* 2011; Verdoolaege *et al.* 2012; Chaki *et al.* 2012; Ciobâcă *et al.* 2014; Felsing *et al.* 2014; Fedyukovich *et al.* 2016; Lopes and Monteiro 2016; Strichman and Veitsman 2016). A fruitful idea for easing the problem of proving program equivalence is to reduce it to a standard verification task by using some *composition* operator between imperative programs (Zaks and Pnueli 2008; Barthe *et al.* 2011; Lahiri *et al.* 2013). The application of these operators requires human ingenuity, and it is still necessary to provide the suitable invariants to be used by the program verifier.

A method for reusing available verification techniques to prove program equivalence is proposed by Ganty *et al.* (2013), who identify a class of recursive programs for which it is possible to precisely compute the so called summaries. This method can be used to reduce the problem of checking the equivalence of two recursive programs to the problem of checking the equivalence of their summaries.

Lopes and Monteiro proposed a different method for proving program equivalence that is based on the computation of precise (that is, not over-approximated) summaries (Lopes and Monteiro 2016). This method considers programs over the integers and is based on a transformation into integer (possibly non-linear) polynomials. The equivalence checking algorithm then works on loop-free programs. This method has been applied to prove the correctness of several loop optimizations. As shown in Section 5, the PP strategy is able to prove similar correctness properties, by avoiding the use of non-linear arithmetic.

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Felsing *et al.* propose a technique for proving relational properties of imperative programs, which is based on a translation of special purpose proof rules into constrained Horn clauses (Felsing *et al.* 2014). The satisfiability of these clauses is then checked by state-of-the-art CHC solvers. The main difference between our approach and the one of Felsing *et al.* is that we generate a translation of the relational properties into sets of CHCs starting from the semantics of the imperative language, and hence we do not need any special purpose proof rule that depends on the programming language and the class of properties under consideration. Instead, we use language independent transformation rules for CHCs. In particular, unlike Felsing *et al.*, by using our approach we are able to verify relations between programs that have different structure, because the transformation rules are independent of the syntax of the source programs.

In conclusion, we would like to stress that our work confirms once again the great advantages offered by the program verification approach based on the use of CHCs. Indeed, by reducing the problem of verifying properties of programs in a given language to the problem of reasoning with CHCs, we are able to use general purpose techniques and very effective tools developed over the last four decades in the fields of logic programming, constraint-based reasoning, and automated theorem proving. In this way, we get verification methods with a very high level of flexibility and parametricity with respect to the language in which programs are written.

Acknowledgements

We warmly thank the anonymous referees for their very helpful comments and criticism. This work has been partially supported by the National Group of Computing Science (GNCS-INDAM). E. De Angelis, F. Fioravanti, and A. Pettorossi are research associates at CNR-IASI, Roma, Italy.

References

Albert, E., Gómez-Zamalloa, M., Hubert, L. and Puebla, G. 2007. Verification of Java bytecode using analysis and transformation of logic programs. In *Proc. of Practical Aspects of Declarative Languages*, M. Hanus, Ed., Lecture Notes in Computer Science, Vol. 4354. Springer, 124–139.

Barthe, G., Crespo, J. M. and Kunz, C. 2011. Relational verification using product programs. In *Proc. of FM 2011: Formal Methods – 17th International Symposium on Formal Methods*, Limerick, Ireland, June 20-24, 2011. Lecture Notes in Computer Science, Vol. 6664. Springer, 200–214.

Benoy, F. and King, A. 1997. Inferring argument size relationships with CLP(R). In *Proc. of the 6th International Workshop on Logic Program Synthesis and Transformation, LOPSTR'96*, Stockholm, Sweden, August 28-30, 1996, J. P. Gallagher, Ed., Lecture Notes in Computer Science, Vol. 1207. Springer, 204–223.

Benton, N. 2004. Simple relational correctness proofs for static analyses and program transformations. In *Proc. of the 31st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2004,* Venice, Italy, January 14-16, 2004. ACM, 14–25.

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Bjørner, N., Gurfinkel, A., McMillan, K. L. and Rybalchenko, A. 2015. Horn clause solvers for program verification. In *Fields of Logic and Computation II – Essays Dedicated to Yuri Gurevich on the Occasion of His 75th Birthday*, L. D. Beklemishev, A. Blass, N. Dershowitz, B. Finkbeiner, and W. Schulte, Eds., Lecture Notes in Computer Science, Vol. 9300. Springer, Switzerland, 24–51.

Bradley, A. R., Manna, Z. and Sipma, H. B. 2006. What's decidable about arrays? In *Proc. of the 7th International Conference on Verification, Model Checking, and Abstract Interpretation. VMCAI '06*. Lecture Notes in Computer Science, Vol. 3855. Springer, 427–442.

Chaki, S., Gurfinkel, A. and Strichman, O. 2012. Regression verification for multi-threaded programs. In *Proc. of Verification, Model Checking, and Abstract Interpretation – 13th International Conference, VMCAI 2012,* Philadelphia, PA, USA, January 22-24, 2012, V. Kuncak and A. Rybalchenko, Eds., Lecture Notes in Computer Science, Vol. 7148. Springer, 119–135.

Ciobâcă, S., Lucanu, D., Rusu, V. and Rosu, G. 2014. A language-independent proof system for mutual program equivalence. In *Proc. of Formal Methods and Software Engineering – 16th International Conference on Formal Engineering Methods, ICFEM 2014,* Luxembourg, Luxembourg, November 3-5, 2014, S. Merz and J. Pang, Eds., Lecture Notes in Computer Science, Vol. 8829. Springer, 75–90.

Cousot, P. and Cousot, R. 1977. Abstract interpretation: A unified lattice model for static analysis of programs by construction of approximation of fixpoints. In *Proc. of the 4th ACM-SIGPLAN Symposium on Principles of Programming Languages, POPL '77*. ACM, 238–252.

De Angelis, E., Fioravanti, F., Pettorossi, A. and Proietti, M. 2014a. Program verification via iterated specialization. *Science of Computer Programming 95, Part 2*, 149–175. Selected and extended papers from Partial Evaluation and Program Manipulation 2013.

De Angelis, E., Fioravanti, F., Pettorossi, A. and Proietti, M. 2014b. VeriMAP: A tool for verifying programs through transformations. In *Proc. of the 20th International Conference on Tools and Algorithms for the Construction and Analysis of Systems, TACAS '14*. Lecture Notes in Computer Science, Vol. 8413. Springer, 568–574. Available at: http://www.map.uniroma2.it/VeriMAP.

De Angelis, E., Fioravanti, F., Pettorossi, A. and Proietti, M. 2015a. Proving correctness of imperative programs by linearizing constrained Horn clauses. *Theory and Practice of Logic Programming 15,* 4–5, 635–650.

De Angelis, E., Fioravanti, F., Pettorossi, A. and Proietti, M. 2015b. A rule-based verification strategy for array manipulating programs. *Fundamenta Informaticae 140,* 3–4, 329–355.

De Angelis, E., Fioravanti, F., Pettorossi, A., and Proietti, M. 2016. Relational verification through Horn clause transformation. In *Proc. of the 23rd International Symposium on Static Analysis, SAS 2016,* Edinburgh, UK, September 8-10, 2016, X. Rival, Ed., Lecture Notes in Computer Science, Vol. 9837. Springer, 147–169.

De Angelis, E., Fioravanti, F., Pettorossi, A. and Proietti, M. 2017. Semantics-based generation of verification conditions via program specialization. *Science of Computer Programming 147,* 3–4, 78–108.

de Moura, L. M. and Bjørner, N. 2008. Z3: An efficient SMT solver. In *Proc. of the 14th International Conference on Tools and Algorithms for the Construction and Analysis of Systems, TACAS '08*. Lecture Notes in Computer Science, Vol. 4963. Springer, 337–340.

De Schreye, D., Glück, R., Jørgensen, J., Leuschel, M., Martens, B. and Sørensen, M. H. 1999. Conjunctive partial deduction: Foundations, control, algorithms, and experiments. *Journal of Logic Programming 41,* 2–3, 231–277.

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Debray, S. K. and Ramakrishnan, R. 1994. Abstract interpretation of logic programs using magic transformations. *Journal of Logic Programming 18*, 149–176.

Etalle, S. and Gabbrielli, M. 1996. Transformations of CLP modules. *Theoretical Computer Science 166*, 101–146.

Fedyukovich, G., Gurfinkel, A. and Sharygina, N. 2016. Property directed equivalence via abstract simulation. In *Proc. of Computer Aided Verification: 28th International Conference, CAV 2016, Part II* , Toronto, Canada, July 17-23, 2016, S. Chaudhuri and A. Farzan, Eds., Lecture Notes in Computer Science, Vol. 7792. Springer International Publishing.

Felsing, D., Grebing, S., Klebanov, V., Rümmer, P. and Ulbrich, M. 2014. Automating regression verification. In *Proc. of ACM/IEEE International Conference on Automated Software Engineering, ASE '14,* Vasteras, Sweden - September 15-19, 2014, I. Crnkovic, M. Chechik, and P. Grünbacher, Eds., 349–360.

Ganty, P., Iosif, R. and Konečný, F. 2013. Underapproximation of procedure summaries for integer programs. In *Proc. of Tools and Algorithms for the Construction and Analysis of Systems: 19th International Conference, TACAS 2013, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2013,* Rome, Italy, March 16-24, 2013, N. Piterman, Scott A. Smolka, Eds., Lecture Notes in Computer Science, Vol. 7795. Springer, 245–259.

Godlin, B. and Strichman, O. 2008. Inference rules for proving the equivalence of recursive procedures. *Acta Informatica 45,* 6, 403–439.

Grebenshchikov, S., Lopes, N. P., Popeea, C. and Rybalchenko, A. 2012. Synthesizing software verifiers from proof rules. In *Proc. of the ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI '12*. 405–416.

Gurfinkel, A., Kahsai, T., Komuravelli, A. and Navas, J. 2015. The SeaHorn verification framework. In *Proc. of Computer Aided Verification: 27th International Conference, CAV 2015,* San Francisco, CA, USA, July 18-24, 2015. Lecture Notes in Computer Science 9206. Springer, 343–361.

Hoare, C. 1969. An axiomatic basis for computer programming. *Communications of the ACM 12,* 10 (October), 576–580.

Hoder, K., Bjørner, N. and de Moura, L. M. 2011. $\mu Z-$ An efficient engine for fixed points with constraints. In *Proc. of Computer Aided Verification, 23rd International Conference, CAV '11,* Snowbird, UT, USA, July 14-20, 2011, G. Gopalakrishnan and S. Qadeer, Eds., Lecture Notes in Computer Science, Vol. 6806. Springer, 457–462.

Hojjat, H., Konecný, F., Garnier, F., Iosif, R., Kuncak, V. and Rümmer, P. 2012. A verification toolkit for numerical transition systems. In *Proc. of FM '12: Formal Methods, 18th International Symposium,* Paris, France, August 27-31, 2012, D. Giannakopoulou and D. Méry, Eds., Lecture Notes in Computer Science, Vol. 7436. Springer, 247–251.

Jaffar, J. and Maher, M. 1994. Constraint logic programming: A survey. *Journal of Logic Programming 19/20*, 503–581.

Jaffar, J., Santosa, A. and Voicu, R. 2009. An interpolation method for CLP traversal. In *Principles and Practice of Constraint Programming, CP '09*, I. Gent, Ed., Lecture Notes in Computer Science, Vol. 5732. Springer, 454–469.

Kafle, B. and Gallagher, J. P. 2017a. Constraint specialisation in Horn clause verification. *Science of Computer Programming 137,* 125–140.

Kafle, B. and Gallagher, J. P. 2017b. Horn clause verification with convex polyhedral abstraction and tree automata-based refinement. *Computer Languages, Systems & Structures 47*, 2–18.

Kafle, B., Gallagher, J. P., and Morales, J. F. 2016. RAHFT: A tool for verifying Horn clauses using abstract interpretation and finite tree automata. In *Proc. of Computer Aided*

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Verification – 28th International Conference, CAV 2016, Part I Toronto, ON, Canada, July 17-23, 2016. Lecture Notes in Computer Science, Vol. 9779. Springer, 261–268.

Lahiri, S. K., McMillan, K. L., Sharma, R. and Hawblitzel, C. 2013. Differential assertion checking. In *Proc. of Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering, ESEC/FSE'13,* Saint Petersburg, Russian Federation, August 18-26, 2013, B. Meyer, L. Baresi, and M. Mezini, Eds., ACM, 345–355.

Leroy, X. 2009. Formal verification of a realistic compiler. *Communications of the ACM 52,* 7, 107–115.

Leuschel, M. and Bruynooghe, M. 2002. Logic program specialisation through partial deduction: Control issues. *Theory and Practice of Logic Programming 2,* 4&5, 461–515.

Lloyd, J. W. 1987. *Foundations of Logic Programming*. Springer-Verlag, Berlin. 2nd Edition.

Lopes, N. P. and Monteiro, J. 2016. Automatic equivalence checking of programs with uninterpreted functions and integer arithmetic. *International Journal on Software Tools for Technology Transfer 18,* 4, 359–374.

McMillan, K. L. and Rybalchenko, A. 2013. Solving constrained Horn clauses using interpolation. MSR Technical Report 2013-6, Microsoft Report.

Mendelson, E. 1997. *Introduction to Mathematical Logic*. Chapman & Hall, London, UK. 4th Edition.

Méndez-Lojo, M., Navas, J. A. and Hermenegildo, M. V. 2008. A flexible, (C)LP-based approach to the analysis of object-oriented programs. In *Proc. of 17th International Symposium on Logic-Based Program Synthesis and Transformation, LOPSTR '07,* Kongens Lyngby, Denmark, August 23-24, 2007. Lecture Notes in Computer Science, Vol. 4915. Springer, 154–168.

Miné, A. 2006. The octagon abstract domain. *Higher-Order and Symbolic Computation 19,* 1, 31–100.

Peralta, J. C., Gallagher, J. P. and Saglam, H. 1998. Analysis of imperative programs through analysis of constraint logic programs. In *Proc. of the 5th International Symposium on Static Analysis, SAS '98*, G. Levi, Ed., Lecture Notes in Computer Science, Vol. 1503. Springer, 246–261.

Pettorossi, A. and Proietti, M. 1994. Transformation of logic programs: Foundations and techniques. *Journal of Logic Programming 19,20*, 261–320.

Podelski, A. and Rybalchenko, A. 2007. ARMC: The logical choice for software model checking with abstraction refinement. In *Proc. of Practical Aspects of Declarative Languages, PADL '07*, M. Hanus, Ed., Lecture Notes in Computer Science, Vol. 4354. Springer, 245–259.

Rümmer, P., Hojjat, H. and Kuncak, V. 2013. Disjunctive interpolants for Horn clause verification. In *Proc. of the 25th International Conference on Computer Aided Verification, CAV '13,* Saint Petersburg, Russia, July 13-19, 2013, N. Sharygina and H. Veith, Eds., Lecture Notes in Computer Science, Vol. 8044. Springer, 347–363.

Strichman, O. and Veitsman, M. 2016. Regression verification for unbalanced recursive functions. In *Proc. of FM 2016: Formal Methods – 21st International Symposium,* Limassol, Cyprus, November 9-11, 2016. Lecture Notes in Computer Science, Vol. 9995. Springer International Publishing, 645–658.

Tamaki, H. and Sato, T. 1984. Unfold/fold transformation of logic programs. In *Proc. of the Second International Conference on Logic Programming, ICLP '84*, S.-Å. Tärnlund, Ed., Uppsala University, Uppsala, Sweden, 127–138.

Verdoolaege, S., Janssens, G. and Bruynooghe, M. 2012. Equivalence checking of static affine programs using widening to handle recurrences. *ACM Trans. Program. Lang. Syst. 34,* 3, 11.

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Zaks, A. and Pnueli, A. 2008. CoVaC: Compiler validation by program analysis of the cross-product. In *Proc. of the 15th International Symposium on Formal Methods (FM 2008)*, Turku, Finland, May 26-30, 2008. J. Cuéllar, T. S. E. Maibaum, and K. Sere, Eds., Lecture Notes in Computer Science, Vol. 5014. Springer, 35–51.

Appendix

Proof of Theorem 3

Proof

Let us assume that there exists an $\mathscr{A}$-definable model Σ of P_i that is tight on $Defs_i$. We will construct an $\mathscr{A}$-definable model Σ' of P_{i+1} that is tight on $Defs_{i+1}$. The proof proceeds by cases on the transformation rule applied to derive P_{i+1} from P_i.

(*Case* R1) Suppose that P_{i+1} is derived from P_i by applying the definition rule. Thus, $P_{i+1} = P_i \cup \{D\}$ and $Defs_{i+1} = Defs_i \cup \{D\}$, where D is the clause $newp(X_1, \ldots, X_k) \leftarrow c, G$, and the following conditions hold: (1) *newp* is a new predicate symbol, (2) $c \in \mathscr{A}$, (3) all predicates occurring in G also occur in P_0, and (4) $X_1, \ldots, X_k$ are distinct variables occurring free in (c, G).

Let Σ' be a symbolic interpretation that is equal to Σ for all atoms whose predicate is different from *newp*, and $\Sigma'(newp(X_1, \ldots, X_k)) = \exists Y_1 \ldots \exists Y_m (c \wedge \Sigma(G))$ where $\{Y_1, \ldots, Y_m\} = Fvars(c \wedge \Sigma(G)) \backslash \{X_1, \ldots, X_k\}$.

Now we have that Σ' is an $\mathscr{A}$-definable model of P_{i+1}, as the following two points hold:

Point (1): Σ' is an $\mathscr{A}$-definable model of P_i because Σ' is equal to Σ for all atoms whose predicates occur in P_i, and

Point (2): Σ' is an $\mathscr{A}$-definable model of D, that is,

$$\mathbb{D} \models \forall (c \wedge \Sigma'(G) \rightarrow \Sigma'(newp(X_1, \ldots, X_k))).$$

Point (2) is shown as follows. Since $Y_1, \ldots, Y_m$ do not occur free in the formula $\Sigma'(newp(X_1, \ldots, X_k))$ and *newp* does not occur in G,

$$\mathbb{D} \models \forall (c \wedge \Sigma'(G) \rightarrow \Sigma'(newp(X_1, \ldots, X_k)))$$

$$\text{iff } \mathbb{D} \models \forall (\exists Y_1 \ldots \exists Y_m (c \wedge \Sigma(G)) \rightarrow \Sigma'(newp(X_1, \ldots, X_k)))$$

and the latter implication holds by the definition of Σ'. Moreover, from the definition of Σ' and from the hypothesis that Σ is tight on $Defs_i$, it follows immediately that Σ' is tight on $Defs_{i+1}$.

(*Case* R2) Suppose that P_{i+1} is derived from P_i by applying the unfolding rule. Thus, $P_{i+1} = (P_i \backslash \{C\}) \cup \{H \leftarrow c, c_j, G_1, B_j, G_2 \mid j = 1, \ldots, m\}$, where C is the clause $H \leftarrow c, G_1, p(X_1, \ldots, X_k), G_2$ in P_i and $\{p(X_1, \ldots, X_k) \leftarrow c_j, B_j \mid j = 1, \ldots, m\}$ is the set of clauses in P_i whose head predicate is p.

Now, we show that Σ is an $\mathscr{A}$-definable model of P_{i+1} that is tight on $Defs_{i+1}$. By the hypothesis that Σ is an $\mathscr{A}$-definable model of P_i we have that

$$\mathbb{D} \models \forall (c \wedge \Sigma(G_1) \wedge \Sigma(p(X_1, \ldots, X_k)) \wedge \Sigma(G_2) \rightarrow \Sigma(H))$$

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

and, for $j=1,\ldots,m$,

$$\mathbb{D} \models \forall(c_j \wedge \Sigma(B_j) \rightarrow \Sigma(p(X_1,\ldots,X_k))).$$

Then, for $j=1,\ldots,m$,

$$\mathbb{D} \models \forall(c \wedge c_j \wedge \Sigma(G_1) \wedge \Sigma(B_j) \wedge \Sigma(G_2) \rightarrow \Sigma(H))$$

and hence Σ is an $\mathscr{A}$-definable model of P_{i+1}.

Obviously, Σ is tight on $Defs_{i+1}$, which is equal to $Defs_i$.

(*Case* R3) Suppose that P_{i+1} is derived from P_i by applying the folding rule. Thus, $P_{i+1} = (P_i \backslash \{C\}) \cup \{E\}$, where C is the clause $H \leftarrow c, G_1, Q, G_2$ in P_i and E is the clause $H \leftarrow e, G_1, K\vartheta, G_2$ derived using the clause $D\colon K \leftarrow d, B$ in $Defs_i$ according to rule R3. Moreover, Conditions (1)–(3) listed above when introducing rule R3, do hold.

Now we show that Σ is an $\mathscr{A}$-definable model of P_{i+1} that is tight on $Defs_{i+1}$. By the hypothesis that Σ is an $\mathscr{A}$-definable model of P_i we have that

$$\mathbb{D} \models \forall(c \wedge \Sigma(G_1) \wedge \Sigma(Q) \wedge \Sigma(G_2) \rightarrow \Sigma(H)).$$

By Conditions (2) and (3) of rule R3 and the definition of symbolic interpretation, we get that

$$\mathbb{D} \models \forall(e \wedge \Sigma(G_1) \wedge (\exists Y_1 \ldots \exists Y_m(d \wedge \Sigma(B)))\vartheta \wedge \Sigma(G_2) \rightarrow \Sigma(H))$$

where $\{Y_1,\ldots,Y_m\} = Fvars(d \wedge \Sigma(B)) \backslash Fvars(\Sigma(K))$. Since Σ is a symbolic interpretation that is tight on $Defs_i$, we have that

$$\mathbb{D} \models \forall(e \wedge \Sigma(G_1) \wedge \Sigma(K\vartheta) \wedge \Sigma(G_2) \rightarrow \Sigma(H)).$$

Thus, Σ is an $\mathscr{A}$-definable model of P_{i+1}.

Obviously, Σ is tight on $Defs_{i+1}$, which is equal to $Defs_i$.

(*Case* R4) Suppose that P_{i+1} is derived from P_i by applying the constraint replacement rule. Thus, $P_{i+1} = (P_i \backslash \{(H \leftarrow c_1, G), \ldots, (H \leftarrow c_k, G)\}) \cup \{(H \leftarrow d_1, G), \ldots, (H \leftarrow d_m, G)\}$, where

$$\mathbb{D} \models \forall\, (\exists Y_1 \ldots \exists Y_r\ (c_1 \vee \ldots \vee c_k) \leftrightarrow \exists Z_1 \ldots \exists Z_s\ (d_1 \vee \ldots \vee d_m)),$$

$\{Y_1,\ldots,Y_r\} = Fvars(c_1 \vee \ldots \vee c_k) \backslash vars(\{H, G\})$, and $\{Z_1,\ldots,Z_s\} = Fvars(d_1 \vee \ldots \vee d_m) \backslash vars(\{H, G\})$.

Now, we show that Σ is an $\mathscr{A}$-definable model of P_{i+1} that is tight on $Defs_{i+1}$. By the hypothesis that Σ is an $\mathscr{A}$-definable model of P_i, the fact that $Y_1,\ldots,Y_r$ do not occur in (G, H), and the distributivity law, we have that

$$\mathbb{D} \models \forall(\exists Y_1 \ldots \exists Y_r\ (c_1 \vee \ldots \vee c_k) \wedge \Sigma(G) \rightarrow \Sigma(H))$$

and hence

$$\mathbb{D} \models \forall(\exists Z_1 \ldots \exists Z_s\ (d_1 \vee \ldots \vee d_m) \wedge \Sigma(G) \rightarrow \Sigma(H)).$$

Thus, by using again the distributivity law, and the fact that $Z_1,\ldots,Z_s$ do not occur in (G, H), we get that Σ is an $\mathscr{A}$-definable model of P_{i+1}. Moreover, Σ is tight on $Defs_{i+1}$, which is equal to $Defs_i$. □

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

Proof of Theorem 5

Proof

Let us assume that there exists an $\mathscr{A}$-definable model Σ' of P_{i+1}, for $i=0,\ldots,n-1$. We will construct an $\mathscr{A}$-definable model Σ of P_i. The proof proceeds by cases on the transformation rule applied to derive P_{i+1} from P_i.

(*Case* R1) Suppose that P_{i+1} is derived from P_i by applying the definition rule. Thus, $P_{i+1} = P_i \cup \{D\}$, where D is a new clause. Let Σ be equal to Σ'. Σ is an $\mathscr{A}$-definable model of every subset of P_{i+1}, and hence it is an $\mathscr{A}$-definable model of P_i.

(*Case* R2) Suppose that P_{i+1} is derived from P_i by applying the unfolding rule. Thus, $P_{i+1} = (P_i \backslash \{C\}) \cup \{H \leftarrow c, c_j, G_1, B_j, G_2 \mid j=1,\ldots,m\}$, where C is the clause $H \leftarrow c, G_1, p(X_1,\ldots,X_k), G_2$ in P_i and $\{p(X_1,\ldots,X_k) \leftarrow c_j, B_j \mid j=1,\ldots,m\}$ is the set of clauses in P_i whose head predicate is p. From the hypothesis that the transformation is not a self-unfolding, that is, H is either *false* or its predicate is different from p, it follows that the set $\{p(X_1,\ldots,X_k) \leftarrow c_j, B_j \mid j=1,\ldots,m\}$ is a subset of P_{i+1}. Since Σ' is an $\mathscr{A}$-definable model of P_{i+1}, we have that, for $j=1,\ldots,m$,

$$\mathbb{D} \models \forall(c \wedge c_j \wedge \Sigma'(G_1) \wedge \Sigma'(B_j) \wedge \Sigma'(G_2) \rightarrow \Sigma'(H)) \quad \text{and} \tag{A1}$$

$$\mathbb{D} \models \forall(c_j \wedge \Sigma'(B_j) \rightarrow \Sigma'(p(X_1,\ldots,X_k))) \tag{A2}$$

Now let us define

$$\Sigma(q(X_1,\ldots,X_l)) = \Sigma'(q(X_1,\ldots,X_l)) \qquad \text{for } q \text{ different from } p, \tag{A3}$$

$$\Sigma(p(X_1,\ldots,X_k)) = \exists Y_1 \ldots \exists Y_n \bigvee_{j=1}^{m} (c_j \wedge \Sigma'(B_j)) \tag{A4}$$

where $\{Y_1,\ldots,Y_n\} = Fvars(\bigvee_{j=1}^{m}(c_j \wedge \Sigma'(B_j))) \backslash \{X_1,\ldots,X_k\}$.
From (A2) and (A4) it follows that

$$\mathbb{D} \models \forall(\Sigma(p(X_1,\ldots,X_k)) \rightarrow \Sigma'(p(X_1,\ldots,X_k))) \tag{A5}$$

and hence, for every conjunction of atoms G,

$$\mathbb{D} \models \forall(\Sigma(G) \rightarrow \Sigma'(G)) \tag{A6}$$

Now, we show that Σ is an $\mathscr{A}$-definable model of P_i, that is, Σ is an $\mathscr{A}$-definable model of each clause D in P_i. We consider the following three subcases.

(*Subcase* 1) D is the clause C: $H \leftarrow c, G_1, p(X_1,\ldots,X_k), G_2$ to which the unfolding rule is applied. By definition of Σ, since H is either *false* or its predicate is different from p, by (A3), we get

$$\Sigma(H) = \Sigma'(H).$$

and hence, by (A1), we get

$$\mathbb{D} \models \forall(c \wedge \Sigma'(G_1) \wedge \bigvee_{j=1}^{m} (c_j \wedge \Sigma'(B_j)) \wedge \Sigma'(G_2) \rightarrow \Sigma(H))$$

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

By (A4), we get

$$\mathbb{D} \models \forall(c \wedge \Sigma'(G_1) \wedge \Sigma(p(X_1,\ldots,X_k)) \wedge \Sigma'(G_2) \rightarrow \Sigma(H))$$

and, finally, by (A6),

$$\mathbb{D} \models \forall(c \wedge \Sigma(G_1) \wedge \Sigma(p(X_1,\ldots,X_k)) \wedge \Sigma(G_2) \rightarrow \Sigma(H))$$

(*Subcase* 2) D is one of the clauses $p(X_1,\ldots,X_k) \leftarrow c_j, B_j$ used for unfolding $p(X_1,\ldots,X_k)$ in C. From the definition of $\Sigma(p(X_1,\ldots,X_k))$ given by (A4), it follows that

$$\mathbb{D} \models \forall(c_j \wedge \Sigma'(B_j) \rightarrow \Sigma(p(X_1,\ldots,X_k)))$$

and hence, by (A6),

$$\mathbb{D} \models \forall(c_j \wedge \Sigma(B_j) \rightarrow \Sigma(p(X_1,\ldots,X_k)))$$

(*Subcase* 3) D is a clause in $P_i \backslash (\{C\} \cup \{p(X_1,\ldots,X_k) \leftarrow c_j, B_j \mid j = 1,\ldots,m\})$. Let D be a clause of the form $K \leftarrow e, Q$. Since K is either *false* or its predicate is different from p, by (A3) it follows that $\Sigma(K) = \Sigma'(K)$. Since D is different from C, it also belongs to P_{i+1}, and by the hypothesis that Σ' is an $\mathscr{A}$-definable model of P_{i+1}, we have that

$$\mathbb{D} \models \forall(e \wedge \Sigma'(Q) \rightarrow \Sigma(K))$$

Thus, by (A6),

$$\mathbb{D} \models \forall(e \wedge \Sigma(Q) \rightarrow \Sigma(K)).$$

(*Case* R3) Suppose that P_{i+1} is derived from P_i by applying the folding rule. Thus, $P_{i+1} = (P_i \backslash \{C\}) \cup \{E\}$, where C is the clause $H \leftarrow c, G_1, Q, G_2$ in P_i and E is the clause $H \leftarrow e, G_1, K\vartheta, G_2$ derived using the clause $D\colon K \leftarrow d, B$ in $Defs_i$ according to rule R3. Moreover, Conditions (1)–(3) listed above when introducing rule R3, do hold.

Now we show that Σ' is an $\mathscr{A}$-definable model of P_i, and hence we can take Σ to be equal to Σ'. From the hypothesis that the application of folding is reversible it follows that D belongs to P_{i+1}, and since Σ' is an $\mathscr{A}$-definable model of P_{i+1}, we have that

$$\mathbb{D} \models \forall(d \wedge \Sigma'(B) \rightarrow \Sigma'(K)) \text{ and}$$

$$\mathbb{D} \models \forall(e \wedge \Sigma'(G_1) \wedge \Sigma'(K\vartheta) \wedge \Sigma'(G_2) \rightarrow \Sigma'(H))$$

and hence, by the definition of symbolic interpretation,

$$\mathbb{D} \models \forall(e \wedge d\vartheta \wedge \Sigma'(G_1) \wedge \Sigma'(B\vartheta) \wedge \Sigma'(G_2) \rightarrow \Sigma'(H))$$

By Conditions (1)–(2) of rule R3, we get

$$\mathbb{D} \models \forall(c \wedge \Sigma'(G_1) \wedge \Sigma'(Q) \wedge \Sigma'(G_2) \rightarrow \Sigma'(H))$$

and thus Σ' is an $\mathscr{A}$-definable model of P_i.

(*Case* R4) Suppose that P_{i+1} is derived from P_i by applying the constraint replacement rule. Thus, $P_{i+1} = (P_i \backslash \{(H \leftarrow c_1, G), \ldots, (H \leftarrow c_k, G)\}) \cup \{(H \leftarrow d_1, G),$

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press

$\ldots, (H \leftarrow d_m, G)\}$, where

$$\mathbb{D} \models \forall (\exists Y_1 \ldots \exists Y_r \ (c_1 \vee \ldots \vee c_k) \leftrightarrow \exists Z_1 \ldots \exists Z_s \ (d_1 \vee \ldots \vee d_m)),$$

$\{Y_1, \ldots, Y_r\} = Fvars(c_1 \vee \ldots \vee c_k) \backslash vars(\{H, G\})$, and $\{Z_1, \ldots, Z_s\} = Fvars(d_1 \vee \ldots \vee d_m) \backslash vars(\{H, G\})$.

Now we show that Σ' is an $\mathscr{A}$-definable model of P_i, and hence we can take Σ to be equal to Σ'. By the hypothesis that Σ' is an $\mathscr{A}$-definable model of P_{i+1}, the fact that $Z_1, \ldots, Z_s$ do not occur in (G, H), and the distributivity law, we have that

$$\mathbb{D} \models \forall (\exists Z_1 \ldots \exists Z_s \ (d_1 \vee \ldots \vee d_m) \wedge \Sigma'(G) \rightarrow \Sigma'(H))$$

and hence

$$\mathbb{D} \models \forall (\exists Y_1 \ldots \exists Y_r \ (c_1 \vee \ldots \vee c_k) \wedge \Sigma'(G) \rightarrow \Sigma'(H)).$$

Thus, by using again the distributivity law and the fact that $Y_1, \ldots, Y_r$ do not occur in (G, H), we get that Σ' is an $\mathscr{A}$-definable model of P_i. □

https://doi.org/10.1017/S1471068417000497 Published online by Cambridge University Press