Lemmaless Induction in Trace Logic

Ahmed Bhayat[1], Pamina Georgiou[2], Clemens Eisenhofer[2], Laura Kovács[2], and Giles Reger[1]

[1] University of Manchester, Manchester, UK
[2] TU Wien, Vienna, AT

Abstract. We present a novel approach to automate the verification of first-order inductive program properties capturing the partial correctness of imperative program loops with branching, integers and arrays. We rely on trace logic, an instance of first-order logic with theories, to express first-order program semantics by quantifying over program execution timepoints. Program verification in trace logic is translated into a first-order theorem proving problem where, to date, effective reasoning has required the introduction of so-called trace lemmas to establish inductive properties. In this work, we extend trace logic with generic induction schemata over timepoints and loop counters, reducing reliance on trace lemmas. Inferring and proving loop invariants becomes an inductive inference step within superposition-based first-order theorem proving. We implemented our approach in the RAPID framework, using the first-order theorem prover VAMPIRE. Our extensive experimental analysis show that automating inductive verification in trace logic improves over existing approaches.

1 Introduction

Automating the verification of programs containing loops and recursive data structures is an ongoing research effort of growing importance. While different techniques for proving the correctness of such programs are in place [5, 6, 8, 11], most existing tools in this realm are heavily based on *satifiability modulo theories* (SMT) backends [4, 7] that come with strong theory reasoning but have limitations in quantified reasoning. In contrast, first-order theorem provers enable quantified reasoning modulo theories [16, 21, 22], such as linear integer arithmetic and arrays. First-order reasoning can thus complement the aforementioned verification efforts when it comes to proving program properties with complex quantification, as evidenced in our original work on the RAPID framework [9] which utilised the VAMPIRE theorem prover [2, 17].

At a high level, the RAPID framework [9] works by translating a program into *trace logic*, adding a number of ad hoc trace lemmas, asserting a desired property, and then running an automated theorem prover on the result. The effectiveness of this approach depends on the underlying trace lemmas. This paper focuses on building support into the VAMPIRE theorem prover to reduce reliance on these lemmas.

To understand the role of these trace lemmas (and therefore, what support must be added to the theorem prover) we briefly overview trace logic and the RAPID framework in a little more detail. Trace logic is an instance of first-order logic with theories, such that the program semantics of imperative programs with loops, branching, integers, and

arrays can be directly encoded in trace logic. A key feature of this encoding is tracking program executions by quantifying over execution *timepoints* (rather than only over single states), which may themselves be parameterised by *loop iterations*. In principle, we can check whether a translated program entails the desired property in trace logic using an automated theorem prover. In our case, we make use of the saturation-based theorem prover VAMPIRE implementing the superposition calculus [3, 19]. However, a straightforward use of theorem proving often fails in establishing validity of program properties in trace logic, as the proof requires some specific induction, in general not supported by superposition-based reasoning.

In our previous work [9], we overcame this challenge by introducing so-called *trace lemmas* capturing common patterns of inductive loop properties over arrays and integers. Inductive loop reasoning in trace logic is then achieved by generating and adding trace lemma instances to the translated program. However, there are two significant limitations to using trace lemmas:

1. Trace lemmas capture inductive patterns/templates that need to be manually identified, as induction is not expressible in first-order logic. As such, they cannot be inferred by a first-order reasoner, implying that the effectiveness of trace logic reasoning depends on the expressiveness of manually supplied trace lemmas.
2. When instantiating trace lemmas with appropriate inductive program variables, a large number of inductive properties are generated, causing saturation-based proof search to diverge and fail to find program correctness proofs in reasonable time.

In this paper we address these limitations by reducing the need for trace lemmas. We achieve this by introducing a couple of novel induction inferences. Firstly, *multi-clause goal induction* which applies induction in a goal oriented fashion as many safety program assertions are structurally close to useful loop invariants. Secondly, *array mapping induction* which covers certain cases where the required loop invariant does not stem from the goal. Specifically, we make the following contributions:

Contribution 1. We introduce two new inference rules, *multi-clause goal* and *array mapping* induction, for *lemmaless induction* over loop iterations (Sections 4–5). The inference rules are compatible with any saturation-based inference system used for first-order theorem proving and work by carrying out induction on terms corresponding to final loop iterations.

Contribution 2. We implemented our approach in the first-order theorem prover VAMPIRE [17]. Further, we extended the RAPID framework [9] to support inductive reasoning in the automated backend (Section 6). We carry out an extensive evaluation of the new method (Section 7) comparing against state-of-the-art approaches SEAHORN [10, 11] and VAJRA/DIFFY [5, 6].

2 Motivating Example

We motivate our work with the example program in Figure 1. The program iterates over two arrays `a` and `b` of arbitrary, but fixed length `length` and copies array elements into a new array `c`. Each even position in `c` contains an element of `a`, while each odd position an element of `b`. Our task is to prove the safety assertion at line 13: given that `length` is not negative, every element in `c` is an element from `a` or `b`. This prop-

```
 1    func main() {
 2      const Int[] a;
 3      const Int[] b;
 4      Int[] c;
 5      const Int length;
 6      Int i = 0;
 7
 8      while (i < length) {
 9        c[2*i] = a[i]
10        c[(2*i) + 1] = b[i]
11        i = i + 1;
12      }
13    }
14    assert (∀k_I.∃l_I.((0 ≤ k < (2 × length) ∧ length ≥ 0)
15       → c(main_end, k) = a(l) ∨ c(main_end, k) = b(l)))
```

Fig. 1: Copying elements from arrays `a` and `b` to even/odd positions in array `c`.

erty involves (i) alternation of quantifiers and (ii) is expressed in the first-order theories of linear integer arithmetic and arrays. Note that in the safety assertion, the program variable `length` is modeled as a logical constant of the same name of sort integer, whilst the constant arrays `a` and `b` are modeled as logical functions from integers to integers. The mutable array variable `c` is additionally equipped with a timepoint argument `main_end`, indicating that the assertion is referring to the value of the variable at the end of program execution.

Proving the correctness of this example program remains challenging for most state-of-the-art approaches, such as [5, 6, 8, 10], mainly due to the complex quantified structure of our assertion. Moreover, it cannot be achieved in the current RAPID framework either, as existing trace lemmas do not relate the values of multiple program variables, notably equality over multiple array variables. In fact, to automatically prove the assertion, we need an inductive property/trace lemma formalizing that each element at an even position in `c` is an element of `a` or `b` at each valid loop iteration, thereby also restricting the bounds of the loop counter variable `i`. Naïvely adding such a trace lemma would be highly inefficient as automated generation of verification conditions would introduce many instances that are not required for the proof. In this paper, we provide a remedy to this challenge as follows.

RAPID works by expressing the program semantics in trace logic and then attempting to prove the safety assertion from the semantics using a first-order theorem prover. Let S stand for the trace logic semantics of the program in Figure 1 and P the safety property. RAPID attempts to show the validity of the implication $S \rightarrow P$. To do this, it is obvious that we require a loop invariant for the loop at line 8 of the program. In this paper, we hypothesise that such invariants are often structurally similar to the assertion to be established and use goal-oriented multi-clause induction to find induction hypotheses whose conclusions are potential loop invariants. Given some induction hypothesis,

```
    program ::= function
   function ::= func main() { subprogram }
subprogram ::= statement | context
    context ::= statement; ... ; statement
  statement ::= atomicStatement
              | if ( condition ) { context } else { context }
              | while ( condition ) { context }
```

Fig. 2: Grammar of $\mathcal{W}$.

proving the base case as well as the step case of the induction hypothesis provides the prover with the conclusion necessary to prove the safety assertion.

3 Preliminaries

Many-Sorted First-Order Logic. We consider standard many-sorted first-order logic with built-in equality, denoted by $\simeq$. By $s = F[u]$ we indicate that the term u is a subterm of s surrounded by (a possibly empty) context F.

We use x, y to denote variables, l, r, s, t for terms and sk for Skolem symbols. A *literal* is an atom A or its negation $\neg A$. A *clause* is a disjunction of literals $L_1 \vee ... \vee L_n$, for $n \geq 0$. Given a formula F, we denote by $CNF(F)$ the clausal normal form of F.

For a logical variable x of sort S we write x_S. A *first-order theory* denotes the set of all valid formulas on a class of first-order structures. Any symbol in the signature of a theory is considered *interpreted*. All other symbols are *uninterpreted*. In particular, we use the theory of linear integer arithmetic denoted by $\mathbb{I}$ and the boolean sort $\mathbb{B}$. We consider natural numbers as the term algebra $\mathbb{N}$ with four symbols in the signature: the constructors `0` and successor `suc`, as well as `pred` and $<$ respectively interpreted as the predecessor function and less-than relation. Note that we do not define any arithmetic on naturals. In the rest of this paper, we assume familiarity with the basics of saturation theorem proving. In the next two subsections, we recall our programming model $\mathcal{W}$ and give a brief overview of trace logic $\mathcal{L}$. For more details, we refer to [9].

3.1 Programming Model $\mathcal{W}$

We consider programs written in a WHILE-like programming language $\mathcal{W}$, as given in the (partial) language grammar of Figure 2. Programs in $\mathcal{W}$ contain mutable and immutable integer as well as integer-array program variables and consist of a single top-level function `main` comprising arbitrary nestings of while-loops and if-then-else branching. We consider expressions over booleans and integers without side effects.

Locations and Timepoints. We consider programs as sets of locations over time: given a program statement `s`, we denote its location by l_s of type $\mathbb{L}$, the location/timepoint sort, corresponding to the line of the program where the statement appears. When `s` is a while-loop the corresponding location is revisited at multiple timepoints of the

execution. Thus, we model such locations as functions over *loop iterations* $l_s : \mathbb{N} \mapsto \mathbb{L}$, where $\mathbb{N}$ intuitively corresponds to the natural number sort used for loop iterations. Further, for each loop statement $\mathtt{s}$ we model the last loop iteration nl_s of target sort $\mathbb{N}$. Let S_{Tp} be the set of all location symbols l_s and S_n denote the set of all function symbols nl_s. Let $\mathtt{p}$ be a program statement or context. We use $start_{\mathtt{p}}$ to denote the location at which the execution of $\mathtt{p}$ has started and $end_{\mathtt{p}}$ to denote the location that occurs just after the execution of $\mathtt{p}$. We use $main\_end$ to denote the location at the end of the main function.

Example 1. Consider line 5 of our running example in Figure 1. Term l_5 corresponds to the timepoint of the first assignment of 0 to program variables $\mathtt{i}$ while $l_7(\mathtt{0})$ and $l_7(nl_7)$ denote the timepoints of the loop at the first and last loop iteration respectively. Further, we can quantify over all executions of the loops by quantifying over all iterations smaller than the last: $\forall it_{\mathbb{N}}.it < nl_7 \rightarrow F[l_7(it)]$ where $F[l_7(it)]$ is some first-order formula.

Program Variables. Program variable values are expressed as functions over timepoints of target sort $\mathbb{I}$: we express integer variables $\mathtt{v}$ as functions $v : \mathbb{L} \mapsto \mathbb{I}$, where $v(tp)$ denotes the value of $\mathtt{v}$ at timepoint tp. Additionally we model numeric array variables $\mathtt{v}$ with an additional argument of sort $\mathbb{I}$ to denote the position of an array access. We obtain $v : \mathbb{L} \times \mathbb{I} \mapsto \mathbb{I}$. Immutable variables are modelled as per their mutable counterparts, but without the timepoint argument. Let S_V be a set of function symbols corresponding to program variables.

Example 2. To denote program variable $\mathtt{i}$ at the location of the assignment in line 5, we write $i(l_5) = 0$. For the first assignment of $\mathtt{c}$ within the loop, we write $c(l_8(it), 2 \times i(l_8(it))) = a(i(l_8(it)))$ for some iteration it. As $\mathtt{a}$ is a constant array, we omit the timepoint argument and only keep the integer denoting the position of the array access.

Program Expressions. In this section we only consider program expressions that contain integer variables and not those containing array variables. To see how these are dealt with, please refer to [9]. Let $\mathtt{e}$ be an arbitrary program expression. We write $[\![\mathtt{e}]\!](tp)$ to denote the value of the evaluation of $\mathtt{e}$ at timepoint tp. For a program expression $\mathtt{e}$ that contains no mutable program variables, we abuse notation and use $[\![\mathtt{e}]\!]$ to denote the translation of $\mathtt{e}$ in first-order logic.

Let $\mathtt{e}, \mathtt{e}_1, \mathtt{e}_2$ be program expressions, tp_1, tp_2 be two timepoints and $v \in S_V$ denote the functional representation of a program variable. The trace logic formula $v(tp_1) \simeq v(tp_2)$ asserts that the variable v has the same value at timepoints tp_1 and tp_2. We introduce a definition for the formula that expresses that the value of a variable v changes between timepoints tp_1 and tp_2 whilst the values of all other variables remain the same.

$$Update(v, e, tp_1, tp_2) \quad := \quad v(tp_2) \simeq [\![\mathtt{e}]\!](tp_1) \wedge \textstyle\bigwedge_{v' \in S_V \setminus \{v\}} v'(tp_1) \simeq v'(tp_2),$$

3.2 Trace Logic $\mathcal{L}$

Trace logic, denoted as $\mathcal{L}$, is an instance of many-sorted first-order logic with theories. Its signature is

$$\Sigma(\mathcal{L}) := S_{\mathbb{N}} \cup S_{\mathbb{I}} \cup S_{Tp} \cup S_V \cup S_n,$$

respectively including the signatures of the theory of natural numbers $\mathbb{N}$ (as a term algebra), the in-built integer theory $\mathbb{I}$, the set of timepoints S_{Tp}, the set of functions representing program variables S_V as well as the one of last iteration symbols S_n as defined in section 3.1.

Axiomatic Semantics of $\mathcal{W}$ in $\mathcal{L}$. The semantics of a program in $\mathcal{W}$ is given by the conjunction of the respective axiomatic semantics of each program statement of $\mathcal{W}$ occurring in the program. In general, we define reachability of program statements over timepoints rather than program states. We briefly recall the axiomatic semantics of assignments and while-loops respectively, again ignoring the array variable case.

Assignments. Let `s` be an assignment `v = e`, where `v` is an integer-valued program variable and `e` is an expression. The evaluation of `s` is performed in one step such that, after the evaluation, the variable `v` has the same value as `e` before the evaluation while all other variables remain unchanged. We obtain

$$[\![\mathtt{s}]\!] := \mathit{Update}(v, e, \mathit{end}_{\mathtt{s}}, \mathit{start}_{\mathtt{s}}) \tag{1}$$

While-Loops. Let `s` be the while-statement `while(Cond){c}` where `Cond` is the *loop condition*. The semantics of `s` is given by the conjunction of the following properties: (2a) the iteration nl_s is the first iteration where `Cond` does not hold anymore, (2b) jumping into the loop body does not change the values of the variables, (2c) the values of the variables at the end of evaluating the loop `s` are equal to the values at the loop condition location in iteration nl_s. As such, we have

$$\begin{aligned}
[\![\mathtt{s}]\!] := \quad & \forall it^s_{\mathbb{N}}.\ (it^s < nl_{\mathtt{s}} \rightarrow [\![\mathtt{Cond}]\!](tp_{\mathtt{s}}(it^s))) \\
\wedge \quad & \neg[\![\mathtt{Cond}]\!](tp(nl_{\mathtt{s}})) && \text{(2a)} \\
\wedge \quad & \forall it^s_{\mathbb{N}}.\ (it^s < nl_{\mathtt{s}} \rightarrow \mathit{EqAll}(\mathit{start}_{\mathtt{c}}, tp_{\mathtt{s}}(it^s))) && \text{(2b)} \\
\wedge \quad & \mathit{EqAll}(\mathit{end}_{\mathtt{s}}, tp_s(nl_{\mathtt{s}})) && \text{(2c)}
\end{aligned}$$

3.3 Trace Lemma Reasoning

Trace logic $\mathcal{L}$ allows to naturally express common program behavior over timepoints. Specifically, it allows us to reason about (i) all iterations of a loop, and the (ii) the existence of specific timepoints. In [9], we leveraged such reasoning with the use of so-called *trace lemmas*, capturing common inductive properties of program loops. An example of a trace lemma would formalise that a certain program variable value remains unchanged from a specific timepoint to the end of program execution. In our current work, we show how induction for trace logic can be directly supported in the first-order prover making the majority of trace lemmas redundant.

Nonetheless, our semantics is not always strong enough for the prover to resolve the more complex step case on its own. Specifically, we need to nudge the prover to deduce that a loop counter expression will at the end of loop execution have the value of the expression it is compared against in the loop condition.

(A) Equal Lengths Trace Lemma We define a common property of loop counter expressions. We call a program expression $\mathtt{e}$ *dense* at loop $\mathtt{w}$ if:

$$Dense_{w,e} := \forall it_{\mathbb{N}}.\Big(it < nl_{\mathtt{w}} \rightarrow \begin{pmatrix} [\![\mathtt{e}]\!](tp_{\mathtt{w}}(\mathtt{suc}(it))) = [\![\mathtt{e}]\!](tp_{\mathtt{w}}(it)) \vee \\ [\![\mathtt{e}]\!](tp_{\mathtt{w}}(\mathtt{suc}(it))) = [\![\mathtt{e}]\!](tp_{\mathtt{w}}(it)) + 1 \end{pmatrix}\Big).$$

Let $\mathtt{w}$ be a while-statement, $C_{\mathtt{w}} := \mathtt{e} < \mathtt{e'}$ be the loop condition where $\mathtt{e'}$ is a program expression that remains constant during iterations of $\mathtt{w}$. The *equal lengths trace lemma of* $\mathtt{w}$*,* $\mathtt{e}$ *and* $\mathtt{e'}$ is defined as

$$\begin{aligned}\big(Dense_{w,e} \wedge [\![\mathtt{e}]\!](tp_{\mathtt{w}}(\mathtt{0})) \leq [\![\mathtt{e'}]\!](tp_{\mathtt{w}}(\mathtt{0}))\big) \rightarrow \\ [\![\mathtt{e}]\!](tp_{\mathtt{w}}(nl_{\mathtt{w}})) = [\![\mathtt{e'}]\!](tp_{\mathtt{w}}(nl_{\mathtt{w}})).\end{aligned} \tag{A}$$

Trace lemma A states that a dense expression $\mathtt{e}$ smaller than or equal to some expression $\mathtt{e'}$ that does not change in the loop, will eventually, specifically in the last iteration, reach the same value as $\mathtt{e'}$. This follows from the fact that we assume termination of a loop, hence we assume the existence of a timepoint $nl_{\mathtt{w}}$ where the loop condition does not hold anymore. As a consequence, given that the loop condition held at the beginning of the execution, we can derive that the loop counter value immediately after the loop execution $[\![\mathtt{e}]\!](tp_{\mathtt{w}}(nl_{\mathtt{w}}))$ will necessarily equate to $[\![\mathtt{e'}]\!](tp_{\mathtt{w}}(\mathtt{0})) = [\![\mathtt{e'}]\!](tp_{\mathtt{w}}(nl_{\mathtt{w}}))$. In the special case where $\mathtt{e'}$ contains no mutable variables, the conclusion of the lemma can be simplified to $[\![\mathtt{e}]\!](tp_{\mathtt{w}}(nl_{\mathtt{w}})) = [\![\mathtt{e'}]\!]$. Note that a similar lemma can just as easily be added for dense but decreasing loop counters.

4 Multi-Clause Goal Induction for Lemmaless Induction

We now focus on fully automating the reasoning about inductive program properties, using only the trace logic progam semantics without extra lemmas. Such inductive program properties express inductive loop invariants, are defined by *multiple clauses*, and typically have well-defined *bounds* on induction variables (e.g. loop counters). Reasoning about such properties therefore allows us to leverage recent theorem proving efforts using *bounded (integer) induction* [12,13]. However, as illustrated in the following, these recent efforts cannot be directly used in trace logic reasoning: (i) we need to adjust bounded induction for the setting of natural numbers case, and (ii) generalise to multi-clause induction. We discuss these steps using Figure 1. Verifying the safety assertion of Figure 1 requires proving the trace logic formula:

$$\begin{aligned}\forall pos_{\mathbb{I}}.\, \exists j_{\mathbb{I}}.\, (0 \leq pos < (2 \times length) \\ \rightarrow (c(main\_end, pos) = a(j) \vee c(main\_end, pos) = b(j))\end{aligned} \tag{3}$$

For proving (3), it suffices to prove that the following, slightly modified statement is a loop invariant of Figure 1:

$$\begin{aligned}\forall it_{\mathbb{N}}.\, it < nl_{\mathtt{w}} \rightarrow \forall pos_{\mathbb{I}}.\, \exists j_{\mathbb{I}}.\, (0 \leq pos < (2 \times i(tp_{\mathtt{w}}(it)))) \\ \rightarrow (c(tp_{\mathtt{w}}(it), pos) = a(j) \vee c(tp_{\mathtt{w}}(it), pos) = b(j))\end{aligned} \tag{4}$$

where $\mathtt{w}$ refers to the loop statement in Figure 1. As part of the program semantics in trace logic, we have formula (5) which links the value of c at the end of the loop to its value at the end of the program. Moreover, using the trace lemma A, we also derive formula (6) in trace logic:

$$\forall x_{\mathbb{I}}.c(tp_{\mathtt{w}}(nl_{\mathtt{w}}), x) \simeq c(main\_end, x) \tag{5}$$

$$i(tp_{\mathtt{w}}(nl_{\mathtt{w}}) \simeq length \tag{6}$$

It is tempting to think that in the presence of these clauses (5)–(6), a saturation-based prover would rewrite the negated conjecture (3) to

$$\begin{aligned}\neg(\forall pos_{\mathbb{I}}.\, \exists j_{\mathbb{I}}.\, (0 \leq pos < (2 \times i(tp_{\mathtt{w}}(nl_{\mathtt{w}})))) \\ \rightarrow (c(tp_{\mathtt{w}}(nl_{\mathtt{w}}), pos) = a(j) \vee c(tp_{\mathtt{w}}(nl_{\mathtt{w}}), pos) = b(j)))\end{aligned}$$

from which a bounded natural number induction inference (similar to the $\mathtt{IntInd}_<$ rule of [13]) would quickly introduce an induction hypothesis with (4) as the conclusion, by induction over $nl_{\mathtt{w}}$. However, this is not the case, as most saturation provers work by first *clausifying* their input. The negated conjecture (3) would not remain a single formula, but be split into the following clauses where sk is a Skolem symbol:

$$\begin{array}{cc} a(x) \not\simeq c(main\_end, sk) & b(x) \not\simeq c(main\_end, sk) \\ \neg(sk \leq 0) & sk \leq 2 \times length \end{array}$$

These clauses can be rewritten using (5)–(6). For example, the first clause can be rewritten to $a(x) \not\simeq c(tp_{\mathtt{w}}(nl_{\mathtt{w}}, sk)$. However, attempting to prove the negation of any of the rewritten clauses individually via induction would merely result in the addition of useless induction formulas to the search space. For example, attempting to prove $\forall it_{\mathbb{N}}.\, it < nl_{\mathtt{w}} \rightarrow (\exists x_{\mathbb{I}}.\, a(x) \simeq c(tp_{\mathtt{w}}(it), sk)$, is pointless as it is clearly false. *The solution we propose in this work is to use multi-clause induction*, whereby we attempt to prove the negation of the conjunction of multiple clauses via a single induction inference. For our running example Figure 1, we can use the following rewritten versions of clauses from the negated conjecture $a(x) \not\simeq c(tp_{\mathtt{w}}(nl_{\mathtt{w}}, sk)$, $b(x) \not\simeq c(tp_{\mathtt{w}}(nl_{\mathtt{w}}, sk)$, and $sk \leq 2 \times i(tp_{\mathtt{w}}(nl_{\mathtt{w}}))$, with induction term $nl_{\mathtt{w}}$, to obtain the induction formula:

$$\begin{array}{l} \neg\Big(\quad \forall x_{\mathbb{I}}.\, a(x) \not\simeq c(i(tp_{\mathtt{w}}(0)), sk) \\ \quad \wedge\ \forall x_{\mathbb{I}}.\, b(x) \not\simeq c(i(tp_{\mathtt{w}}(0), sk)) \\ \quad \wedge\ sk \leq 2 \times i(tp_{\mathtt{w}}(0))\Big) \\ \wedge\ StepCase \end{array} \rightarrow \begin{array}{l} \forall it_{\mathbb{N}}.\, it < nl_{\mathtt{w}} \rightarrow \\ \quad \neg\Big(\quad \forall x_{\mathbb{I}}.\, a(x) \not\simeq c(i(tp_{\mathtt{w}}(it), sk) \\ \quad\ \wedge \forall x_{\mathbb{I}}.\, b(x) \not\simeq c(i(tp_{\mathtt{w}}(it), sk) \\ \quad\ \wedge\ sk \leq 2 \times i(tp_{\mathtt{w}}(it))\Big) \end{array} \tag{7}$$

where $StepCase$ is the formula:

$$\begin{array}{l} \forall it_{\mathbb{N}}.\, it < nl_{\mathtt{w}} \wedge \\ \neg\Big(\quad \forall x_{\mathbb{I}}.\, a(x) \not\simeq c(i(tp_{\mathtt{w}}(it)), sk) \\ \quad \wedge\ \forall x_{\mathbb{I}}.\, b(x) \not\simeq c(i(tp_{\mathtt{w}}(it)), sk) \\ \quad \wedge\ sk \leq i(tp_{\mathtt{w}}(y))\Big) \end{array} \rightarrow \begin{array}{l} \neg\Big(\quad \forall x_{\mathbb{I}}.\, a(x) \not\simeq c(i(tp_{\mathtt{w}}(\mathtt{suc}(it)), sk) \\ \quad \wedge \forall x_{\mathbb{I}}.\, b(x) \not\simeq c(i(tp_{\mathtt{w}}(\mathtt{suc}(it)), sk) \\ \quad \wedge\ sk \leq 2 \times i(tp_{\mathtt{w}}(\mathtt{suc}(it)))\Big) \end{array}$$

Using the induction formula (7), a contradiction can then easily be derived, establishing validity of (3). In what follows, we formalize the multi-clause induction principle we used above. To this end, we introduce a generic multi-clause induction inference rule in trace logic, called *multi-clause goal induction* and denoted as `MCGLoopInd`, as given below:

$$\frac{C_1[nl_{\mathrm{w}}] \qquad C_2[nl_{\mathrm{w}}] \quad \ldots \quad C_n[nl_{\mathrm{w}}]}{\mathrm{CNF}\left(\left(\begin{array}{c} \neg(C_1[0] \wedge C_2[0] \wedge \ldots \wedge C_n[0]) \wedge \\ \forall it_{\mathbb{N}}.\ \left(\begin{array}{c}((it < nl_{\mathrm{w}}) \wedge \neg(C_1[it] \wedge C_2[it] \wedge \ldots \wedge C_n[it])) \rightarrow \\ \neg(C_1[\mathtt{suc}(it)] \wedge C_2[\mathtt{suc}(it)] \wedge \ldots \wedge C_n[\mathtt{suc}(it)]))\end{array}\right) \\ \rightarrow (\forall it_{\mathbb{N}}.\, (it < nl_{\mathrm{w}}) \rightarrow \neg(C_1[it] \wedge C_2[it] \wedge \ldots \wedge C_n[it]))\end{array}\right)\right)}$$

For performance reasons, we mandate that the premises $C_1 \ldots C_n$ be derived from trace logic formulas expressing safety assertions and not from formulas encoding the program semantics. The `MCGLoopInd` rule is formalised only as an induction inference over last loop iteration symbols. While restricting to nl_{w} terms is of purely heuristic nature, our experiments justify the necessity and usefulness of this condition (Section 7).

5 Array Mapping Induction for Lemmaless Induction

```
 1      func main(){
 2        const Int alength;
 3        Int[] a;
 4        Int i = 0;
 5        const Int n;
 6
 7        while(i < alength){
 8          a[i] = a[i] + n;
 9          i = i + 1;
10        }
11
12        Int j = 0;
13        while(j < alength){
14          a[j] = a[j] - n;
15          j = j + 1;
16        }
17      }
18    assert (∀k_I.((0 ≤ k <length ∧ length ≥ 0)
                        → a(main_end, k) = a(begin, k)))
```

Fig. 3: Adding and subtracting `n` to every element of array `a`.

Multi-clause goal induction neatly captures goal-oriented application of induction. Nevertheless, there are verification challenges where `MCGLoopInd` fails to prove inductive loop properties. This is particularly the case for benchmarks containing multiple

loops, such as in Figure 3. We first discuss the limitations of `MCGLoopInd` using Figure 3, after which we present our solution, the *array mapping induction* inference.

Let $\mathtt{w}_1$ be the first loop statement of Figure 3 and $\mathtt{w}_2$ be the second loop. Using `MCGLoopInd`, we would attempt to prove

$$\begin{array}{l}\forall it_{\mathbb{N}}.\, it \leq nl_{\mathtt{w}_2} \rightarrow \\ \quad \forall pos_{\mathbb{I}}.\, (0 \leq pos < j(tp_{\mathtt{w}_2}(it))) \rightarrow (a(tp_{\mathtt{w}_2}(it), pos) = a(begin, pos)\end{array} \tag{8}$$

However, formula (8) is not a useful invariant for proving the assertion. Rather, for $\mathtt{w}_2$ we need a loop invariant similar to

$$\begin{array}{l}\forall it_{\mathbb{N}}.\, it \leq nl_{\mathtt{w}_2} \rightarrow \forall pos_{\mathbb{I}}.\, (0 \leq pos < j(tp_{\mathtt{w}_2}(it))) \\ \qquad\qquad\qquad\qquad \rightarrow (a(tp_{\mathtt{w}_2}(it), pos) = a(tp_{\mathtt{w}_2}(0), pos) - n\end{array} \tag{9}$$

and a similar loop invariant for loop $\mathtt{w}_1$. The loop invariant (9) is however not linked to the safety assertion of Figure 3, and thus multi-clause goal induction is unable to infer and prove with it. To aid with the verification of benchmarks such as Figure 3, we introduce an *array mapping induction* inference form of natural number induction, where we trigger induction not on clauses and terms coming from the goal, but on clauses and terms appearing in the program semantics.

Our *array mapping induction* inference rule, denoted as `AMLoopInd` is given below. Essentially, `AMLoopInd` involves analysing a clause set to heuristically devise a suitable loop invariant. Obviously, guessing a candidate loop invariant is a difficult problem. The `AMLoopInd` inference is triggered if clauses of the shapes of C_1 and C_2 defined below are present in the clause set. Intuitively, C_2 can be read as saying that on each round of some loop `w`, some array a at position i is set to some function F of its previous value at that position. Clause C_1 states that i increases by one in each round of the loop. Together the two clauses suggest that the loop is mapping the function F to the array and this is precisely what induction formula attempts to prove. In summary, our `AMLoopInd` rule is

$$\frac{\begin{array}{c}C_1 = i(tp_{\mathtt{w}}(\mathtt{suc}(x))) \simeq i(tp_{\mathtt{w}}(x)) + 1 \ \vee \ \neg(x < nl_{\mathtt{w}}) \\ C_2 = a(tp_{\mathtt{w}}(\mathtt{suc}(x)), i(tp_{\mathtt{w}}(x)) \simeq F[a(tp_{\mathtt{w}}(x), i(tp_{\mathtt{w}}(x))] \ \vee \ \neg(x < nl_{\mathtt{w}})\end{array}}{\begin{array}{c}\text{CNF}(BaseCase_1 \ \wedge \ StepCase_1 \rightarrow Conclusion_1) \\ \text{CNF}(BaseCase_2 \ \wedge \ StepCase_2 \rightarrow Conclusion_2)\end{array}}$$

where $\mathtt{w}$ is some loop, F an arbitrary non-empty context and:

$$\begin{array}{ll}BaseCase_1: & \forall x_{\mathbb{I}}.\, x < i(tp_{\mathtt{w}}(0)) \wedge x \geq i(tp_{\mathtt{w}}(0)) \rightarrow a(tp_{\mathtt{w}}(0), x) = F[a(tp_{\mathtt{w}}(0), x)] \\ StepCase_1: & \forall it_{\mathbb{N}}.\, (\forall y_{\mathbb{I}}.\, it < nl_{\mathtt{w}} \wedge y < i(tp_{\mathtt{w}}(it)) \wedge y \geq i(tp_{\mathtt{w}}(0)) \\ & \qquad \rightarrow a(tp_{\mathtt{w}}(it), y) = F[a(tp_{\mathtt{w}}(0), y)]) \rightarrow \\ & (\forall y_{\mathbb{I}}.\, y < i(tp_{\mathtt{w}}(\mathtt{suc}(it))) \wedge y \geq i(tp_{\mathtt{w}}(0)) \\ & \qquad \rightarrow a(tp_{\mathtt{w}}(\mathtt{suc}(it)), y) = F[a(tp_{\mathtt{w}}(0), y)]) \\ Conclusion_1: & \forall x_{\mathbb{I}}.\, x < i(tp_{\mathtt{w}}(nl_{\mathtt{w}})) \wedge x \geq i(tp_{\mathtt{w}}(0)) \\ & \qquad \rightarrow a(tp_{\mathtt{w}}(nl_{\mathtt{w}}), x) = F[a(tp_{\mathtt{w}}(0), x)] \\ BaseCase_2: & \forall x_{\mathbb{I}}.\, x \geq i(tp_{\mathtt{w}}(0)) \rightarrow a(tp_{\mathtt{w}}(0), x) = a(tp_{\mathtt{w}}(0), x)\end{array}$$

$$
\begin{aligned}
\mathit{StepCase}_2 : \quad & \forall it_{\mathbb{N}}.\, (\forall x_{\mathbb{I}}.\, x \geq i(tp_{\mathrm{w}}(it)) \wedge (it < nl_{\mathrm{w}}) \\
& \qquad \rightarrow a(tp_{\mathrm{w}}(0), x) = a(tp_{\mathrm{w}}(it), x)) \rightarrow \\
& (\forall x_{\mathbb{I}}.\, x \geq i(tp_{\mathrm{w}}(\mathtt{suc}(it))) \\
& \qquad \rightarrow a(tp_{\mathrm{w}}(0), x) = a(tp_{\mathrm{w}}(\mathtt{suc}(it)), x)) \\
\mathit{Conclusion}_2 : \; & \forall x_{\mathbb{I}}\, it_{\mathbb{N}}.\, x \geq i(tp_{\mathrm{w}}(it)) \wedge (it < nl_{\mathrm{w}}) \rightarrow a(tp_{\mathrm{w}}(0), x) = a(tp_{\mathrm{w}}(it), x)
\end{aligned}
$$

To prove $\mathit{StepCase}_1$, it is necessary to be able to reason that positions in the array a remain unchanged until visited by the indexing variable. To this end, we add the second induction formula to the conclusion of the inference. The `AMLoopInd` inference is thus sufficient to prove the assertion of Figure 3. While `AMLoopInd` is a limited approach for guessing inductive loop invariants, we believe it can be extended towards further, more generic methods to guess invariants, as discussed in Section 9. We conclude this section by noting that our induction rules are sound, based on trace logic semantics. Since both rules merely add instances of the bounded induction schema for natural numbers to the search space, we simply state the following theorem without proof.

Theorem 1 (Soundness of Lemmaless Induction). *The inference rules* `MCGLoopInd` *and* `AMLoopInd` *are sound.*

6 Implementation

Our approach is implemented as an extension of the RAPID framework, using the first-order theorem prover VAMPIRE.

***Extensions to* RAPID.** RAPID takes as an input a $\mathcal{W}$ program along with a property expressed in $\mathcal{L}$. It outputs the semantics of the program expressed in $\mathcal{L}$ using SMT-LIB syntax along with the property to be proven. For our "lemmaless induction" framework, we have extended RAPID as follows. Firstly, we prevent the output of all trace lemmas other than trace lemma A (Section 3.3). We added custom extensions to the SMT-LIB language to identify trace logic symbols, such as loop iteration symbols, program variables, within the RAPID encodings. This way, trace logic symbols to be used for induction inferences are easily identified and can also be used for various proving heuristics. We refer to this version (available online[3]) as RAPID^{l-}.

***Extensions to* VAMPIRE.** We implemented our induction inference rules `MCGLoopInd` and `AMLoopInd` in a new branch of VAMPIRE[4]. The main issue with the induction inferences `MCGLoopInd` and `AMLoopInd` is their explosiveness which can cause proof search to diverge. We have, therefore, introduced various heuristics in the implementation to try and control them. For `MCGLoopInd` we not only necessitate that the premises are derived from the conjecture, but that their derivation length from the

[3] See commit `285e54b7e` of `https://github.com/vprover/rapid/tree/ahmed-induction-support`.

[4] See commit `4a0f319f` of `https://github.com/vprover/vampire/tree/ahmed-rapid`

conjecture is below a certain distance controlled by an option. The premises must be unit clauses unless another option `multi_literal_ clauses` is toggled on. The option `induct_all_loop_counts` allows `MCGLoop Ind` induction to take place on all loop counter terms, not just final loop iterators. In order for the `MCGLoopInd` and `AMLoopInd` inferences to be applicable, we need to rewrite terms not containing final loop counters to terms that do. However, rewriting in VAMPIRE is based on superposition, which is parameterised by a term order preventing smaller terms to be rewritten into larger ones. In this case, the term order may work against us and prevent such rewrites from happening. We implemented a number of heuristics to handle this problem. One such heuristic is to give terms representing constant program variables a large weight in the ordering. Then, equations such as $alength \simeq i(tp_{\mathrm{w}}(nl_{\mathrm{w}}))$ will be oriented left to right as desired. We combined these options with others to form a portfolio of strategies[5] that contains 13 strategies each of which runs in under 10s.

7 Experimental Results

Benchmarks. For our experiments, we use a total of 111 input problems of different logical complexity, mainly coming from the SV-COMP repository.[6] Most of these problems were selected by Gurfinkel et al. [11] from the array verification benchmarks of the SV-COMP 2018 repository [1]. We selected a subset of these benchmarks based on the fragment of C constructs supported by RAPID, specifically we omitted any examples containing pointers or memory management. We added a number of hand-crafted problems containing existential and alternating quantification that we intend to submit to SV-COMP. All examples are adapted to our input format, as for example arrays in trace logic are treated as unbounded data structures. Additionally, we added new safety assertions in trace logic to the programs to showcase proving existentially and alternatingly quantified conjectures on top of the universal and/or quantifier-free conjectures of SV-COMP. Our benchmarks are divided into four groups, as indicated in Table 1: the first 13 problems have quantifier-free conjectures; the majority of benchmarks contain universally quantified safety assertions; 7 problems are only existentially quantified; and 23 assertions contain alternation of quantifiers.

Experimental Setting. We used two versions of RAPID. First, **(1)** RAPID^{l+} uses trace lemmas for inductive reasoning, as described in [9]. Further, **(2)** RAPID^{l-} denotes our RAPID approach, using lemmaless induction `MCGLoopInd` and `AMLoopInd` in VAMPIRE. We also compared RAPID^{l-} with other verification tools. In particular, we considered **(3)** SEAHORN and **(4)** VAJRA (and its extension DIFFY that produced for us exactly the same results as VAJRA). SEAHORN converts the program into a constrained horn clause (CHC) problem and uses the SMT solver Z3 for solving. VAJRA and DIFFY implement inductive reasoning and recurrence solving over loop counters; in the background, they also use Z3.

[5] `--mode portfolio --schedule rapid_induction`.

[6] Artifact evaluation: in order to reproduce the results reported in this section, please follow the instructions at `https://github.com/vprover/vampire_publications/tree/master/experimental_data/NFM-2022-RAPID-INDUCTION`

Table 1: Experimental results.

Benchmark	(1)	(2)	(3)	(4)
atleast_one_iteration_0	✓	✓	✓	✓
atleast_one_iteration_1	✓	✓	✓	✓
count_down	✓	-	-	-
eq	✓	-	✓	-
find_sentinel	✓	✓	-	-
find1_0	✓	✓	✓	-
find1_1	✓	✓	✓	-
find2_0	✓	✓	✓	-
find2_1	✓	✓	✓	-
indexn_is_arraylength_0	✓	✓	✓	-
indexn_is_arraylength_1	✓	✓	✓	-
set_to_one	✓	✓	✓	✓
str_cpy_3	✓	✓	✓	-
add_and_subtract	✓	-	-	✓
both_or_none	✓	✓	-	✓
check_equal_set_flag_1	✓	✓	-	✓
collect_indices_eq_val_0	✓	✓	-	✓
collect_indices_eq_val_1	✓	✓	-	-
copy	✓	✓	-	✓
copy_absolute_0	✓	✓	-	✓
copy_absolute_1	✓	✓	-	✓
copy_and_add	✓	-	-	✓
copy_nonzero_0	✓	✓	-	✓
copy_partial	✓	✓	-	✓
copy_positive_0	✓	✓	-	✓
copy_two_indices	✓	✓	-	-
find_max_0	✓	✓	-	✓
find_max_2	✓	✓	-	✓
find_max_from_second_0	✓	-	-	✓
find_max_local_2	-	-	-	-
find_max_up_to_0	-	-	-	-
find_max_up_to_2	-	-	-	-
find_min_0	✓	✓	-	✓
find_min_2	✓	✓	-	-
find_min_local_2	-	-	-	-
find_min_up_to_0	-	-	-	-
find_min_up_to_2	-	-	-	-
find1_4	-	✓	-	-
find2_4	✓	✓	-	-
in_place_max	✓	✓	-	✓
inc_by_one_0	✓	✓	-	✓
inc_by_one_1	✓	✓	-	✓
inc_by_one_harder_0	✓	✓	-	✓
inc_by_one_harder_1	✓	✓	-	✓
init	✓	✓	-	-
init_conditionally_0	✓	✓	-	-
init_conditionally_1	✓	✓	-	✓
init_non_constant_0	✓	✓	-	-
init_non_constant_1	✓	✓	-	✓
init_non_constant_2	✓	✓	-	✓
init_non_constant_3	✓	✓	-	✓
init_non_constant_easy_0	✓	✓	-	-
init_non_constant_easy_1	✓	✓	-	✓
init_non_constant_easy_2	✓	✓	-	✓
init_non_constant_easy_3	✓	✓	-	✓
init_partial	✓	✓	-	✓
init_prev_plus_one_0	✓	✓	-	-
init_prev_plus_one_1	✓	✓	-	-
init_prev_plus_one_alt_0	✓	✓	-	-
init_prev_plus_one_alt_1	✓	✓	-	-
insertion_sort	-	-	-	-
max_prop_0	✓	✓	-	✓
max_prop_1	✓	✓	-	✓
merge_interleave_0	✓	-	-	✓
merge_interleave_1	✓	-	-	✓
min_prop_0	✓	✓	-	✓
min_prop_1	✓	✓	-	✓
partition_0	✓	✓	-	✓
partition_1	✓	✓	-	✓
push_back	✓	✓	-	✓
reverse	✓	✓	-	-
rewnifrev	✓	-	-	✓
rewrev	✓	-	-	✓
skipped	✓	-	-	✓
str_cpy_0	✓	✓	-	-
str_cpy_1	✓	✓	-	-
str_cpy_2	✓	✓	-	-
swap_0	-	✓	✓	✓
swap_1	-	✓	✓	✓
vector_addition	✓	✓	-	✓
vector_subtraction	✓	✓	-	✓
check_equal_set_flag_0	✓	✓	-	-
find_max_1	-	-	-	-
find_max_from_second_1	✓	-	-	-
find1_2	✓	✓	-	-
find1_3	✓	✓	-	-
find2_2	✓	✓	-	-
find2_3	✓	✓	-	-
collect_indices_eq_val_2	-	✓	-	-
collect_indices_eq_val_3	✓	-	-	-
copy_nonzero_1	✓	✓	-	-
copy_positive_1	✓	✓	-	-
find_max_local_0	-	-	-	-
find_max_local_1	✓	-	-	-
find_max_up_to_1	-	-	-	-
find_min_1	-	-	-	-
find_min_local_0	-	-	-	-
find_min_local_1	✓	-	-	-
find_min_up_to_1	-	-	-	-
merge_interleave_2	✓	-	-	-
partition_2	✓	✓	-	-
partition_3	✓	✓	-	-
partition_4	-	-	-	-
partition_5	-	✓	-	-
partition_6	-	-	-	-
partition-harder_0	✓	✓	-	-
partition-harder_1	✓	✓	-	-
partition-harder_2	✓	-	-	-
partition-harder_3	✓	-	-	-
partition-harder_4	✓	-	-	-
str_len	✓	✓	-	-
Total solved	**93**	**78**	**13**	**47**

RAPID *Experiments.* Table 1 shows that RAPID^{l-} is superior to RAPID^{l+}, as it solves a total of 93 problems, while RAPID^{l+} only proved 78 conjectures correct. Particularly, RAPID^{l-} can solve benchmark `merge_interleave_2` corresponding to our motivating example 1, and other challenging problems such as `find_max_local_1` also containing quantifier alternations.

While RAPID^{l-} can solve a total of ten problems more than RAPID^{l+}, it is interesting to look into which problems can now be solved. Many of the newly solved problems are structurally very close to the loop invariants needed to prove them. This is where multi-clause goal-oriented induction `MCGoalInd` makes the biggest impact. For instance, this allows RAPID^{l-} to prove the partial correctness of `find_max_from_second_0` and `find_max_from_second_1`.

On the other hand, RAPID^{l-} also lost two challenging benchmarks that were previously solved by RAPID^{l+}, namely `swap_0` and `partition_5`. This could be for two reasons: (1) the strategies in the induction schedule of RAPID^{l-} are too restrictive for such benchmarks, or (2) the step case of the induction axiom introduced by our two rules are too difficult for VAMPIRE to prove. Strengthening lemmaless induction with additional trace lemmas from RAPID^{l+} is an interesting line of further work.

Comparing with other tools. Both, SEAHORN and VAJRA/DIFFY require C code as input, whereas RAPID uses its own syntax. We translated our benchmarks to C code expressing the same problem. However, a direct comparison of RAPID, and in particular RAPID^{l-}, with most other verifiers requiring standard C code as an input is not possible as we consider slightly different semantics. In contrast to SEAHORN and VAJRA/DIFFY, we assume that integers and arrays are unbounded and that all array positions are initialized by arbitrary data. Further, we can read/write at any array position without allocating the accessed memory beforehand.

Apart from semantic differences, RAPID can directly express assertions and assumptions containing quantifiers and put variable contents from different points in time into relation. In order to deal with the latter, we introduced history variables in the code handed over to SEAHORN and VAJRA/DIFFY. The quantification in the RAPID benchmarks was simulated by non-deterministically assigned variables and by loops.

As a result, SEAHORN verified 13 examples, whereas VAJRA/DIFFY 47 of our benchmarks. As VAJRA/DIFFY restrict their input programs to contain only loops having very specific loop-conditions, several of our benchmarks failed. For example, $i < length$ is permitted, whereas $a[i] \neq 0$ is not. VAJRA/DIFFY could prove correctness for nearly all the programs satisfying these restrictions. SEAHORN, on the other hand, has problems with the complexity introduced by the arrays. It could solve especially those benchmarks whose correctness do not depend on the arrays' content.

8 Related Work

Most of recent research in verifying indutive properties of array-manipulating programs focuses on quantified invariant generation and/or is mostly restricted to proving universally quantified program properties. The works [8, 11] generate universally quantified inductive invariants by iteratively inferring and strengthening candidate invariants.

These methods use SMT solving and as such restricted to first-order theories with a finite model property. Similar logical restrictions also apply to [20], where linear recurrence solving is used in combination with array-specific proof tactics to prove quantified program properties. A related approach is described in [6], where relational invariants instead of recurrence equations are used to handle universal and quantifier-free inductive properties. Unlike these works, our work is not limited to universal invariants but can both infer and prove inductive program properties with alternations of quantifiers.

With the use of extended expressions and induction schemata, our work shares some similarity with template-based approaches [14, 18, 23]. These works [14, 18, 23] infer and prove universal inductive properties based on Craig interpolation, formula slicing and/or SMT generalizations over quantifier-free formulas. Unlike these works, we do not require any assumptions on the syntactic shape of the first-order invariants. Moreover, our invariants are not restricted to the shape of our induction schemata. Rather, we treat inductive (invariant) inferences as additional rules of first-order theorem provers, maintaining thus the efficient handling of arbitrary first-order quantifiers. Our framework can be used in arbitrary first-order theories, even with theories that have no interpolation property and/or a finite axiomatization, as exemplified by our experimental results using inductive reasoning over arrays and integers.

First-order theorem proving has previously been used to derive invariants with alternations of quantifiers in our previous work [9]. Our current work generalizes the inductive capabilities of [9] by reducing the expert knowledge of [9] in introducing inductive lemmas to guide the process of proving inductive properties.

9 Future Directions and Conclusion

We introduced lemmaless induction to fully automate the verification of inductive properties of program loops with unbounded arrays and integers. We introduced goal-oriented and array mapping induction inferences, triggered by loop counters, in superposition-based theorem proving. Our results show that lemmaless induction in trace logic outperforms other state-of-the-art approaches in the area.

There are various ways to further develop lemmaless induction in trace logic. On larger benchmarks, particularly those containing multiple loops, our approach struggles. For loops where the required invariant is not connected to the conjecture, we introduced array mapping induction. However, the array mapping induction inference is limited in the form of invariants it can generate. We would like to investigate other methods, such as machine learning for synthesising loop invariants that are not too prolific. A completely different line of research that we are currently working on, is updating the trace logic syntax and semantics of $\mathcal{W}$ to deal with memory and memory allocation, aiming to efficiently reason about loop operations over the memory.

As shown in [15], the validity problem for first-order formulas of linear arithmetic extended with non-theory function symbols is Π^1_1-complete. Therefore, we do not expect any completeness result for inductive theorem proving. Proving relative completeness results for our verification framework is an interesting question. In extension to the relative completeness of trace logic [9], we are, for example, interested in identifying further classes of inductive properties strong enough to prove loop properties.

References

1. "sv-comp repository". https://gitlab.com/sosy-lab/benchmarking/sv-benchmarks.
2. Vampire website. https://vprover.github.io/.
3. L. Bachmair and H. Ganzinger. Resolution theorem proving. In A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*, volume I, chapter 2, pages 19–99. Elsevier Science, 2001.
4. Clark Barrett, Christopher L Conway, Morgan Deters, Liana Hadarean, Dejan Jovanović, Tim King, Andrew Reynolds, and Cesare Tinelli. CVC4. In *CAV*, pages 171–177, 2011.
5. Supratik Chakraborty, Ashutosh Gupta, and Divyesh Unadkat. Verifying array manipulating programs with full-program induction. In *TACAS*, pages 22–39, 2020.
6. Supratik Chakraborty, Ashutosh Gupta, and Divyesh Unadkat. Diffy: Inductive Reasoning of Array Programs Using Difference Invariants. In *CAV*, pages 911–935, 2021.
7. Leonardo De Moura and Nikolaj Bjørner. Z3: An Efficient SMT Solver. In *TACAS*, pages 337–340, 2008.
8. Grigory Fedyukovich, Sumanth Prabhu, Kumar Madhukar, and Aarti Gupta. Quantified invariants via syntax-guided synthesis. In *CAV*, pages 259–277, 2019.
9. Pamina Georgiou, Bernhard Gleiss, and Laura Kovács. Trace Logic for Inductive Loop Reasoning. In *FMCAD*, pages 255–263, 2020.
10. Arie Gurfinkel, Temesghen Kahsai, Anvesh Komuravelli, and Jorge A Navas. The seahorn verification framework. In *CAV*, pages 343–361, 2015.
11. Arie Gurfinkel, Sharon Shoham, and Yakir Vizel. Quantifiers on Demand. In *ATVA*, pages 248–266, 2018.
12. Márton Hajdu, Petra Hozzová, Laura Kovács, Johannes Schoisswohl, and Andrei Voronkov. Induction with generalization in superposition reasoning. In *CICM*, pages 123–137, 2020.
13. Petra Hozzová, Laura Kovács, and Andrei Voronkov. Integer induction in saturation. In *CADE*, pages 361–377. Springer, 2021.
14. E. G. Karpenkov and D. Monniaux. Formula Slicing: Inductive Invariants from Preconditions. In *HVC*, pages 169–185, 2016.
15. Konstantin Korovin and Andrei Voronkov. Integrating linear arithmetic into superposition calculus. In *CSL*, pages 223–237, 2007.
16. Laura Kovács, Simon Robillard, and Andrei Voronkov. Coming to Terms with Quantified Reasoning. In *POPL*, pages 260–270, 2017.
17. Laura Kovács and Andrei Voronkov. First-Order Theorem Proving and Vampire. In *CAV*, pages 1–35, 2013.
18. D. Larraz, E. Rodríguez-Carbonell, and A. Rubio. SMT-Based Array Invariant Generation. In *VMCAI*, pages 169–188, 2013.
19. R. Nieuwenhuis and A. Rubio. Paramodulation-based theorem proving. In A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*, volume I, chapter 7, pages 371–443. Elsevier Science, 2001.
20. Pritom Rajkhowa and Fangzhen Lin. Extending viap to handle array programs. In *VSTTE*, pages 38–49, 2018.
21. Giles Reger, Nikolaj Bjorner, Martin Suda, and Andrei Voronkov. AVATAR modulo theories. In Christoph Benzmüller, Geoff Sutcliffe, and Raul Rojas, editors, *GCAI*, volume 41 of *EPiC Series in Computing*, pages 39–52, 2016.
22. Giles Reger, Johannes Schoisswohl, and Andrei Voronkov. Making theory reasoning simpler. In Jan Friso Groote and Kim Guldstrand Larsen, editors, *TACAS*, volume 12652 of *Lecture Notes in Computer Science*, pages 164–180, 2021.
23. S. Srivastava and S. Gulwani. Program Verification using Templates over Predicate Abstraction. In *PLDI*, pages 223–234, 2009.