Implementing Reasoning Modules in Implicit Induction Theorem Provers
Sorin Stratulat

To cite this version:

Implementing Reasoning Modules in Implicit Induction Theorem Provers

Sorin Stratulat
LITA, Department of Computer Science
Université de Lorraine
Ile du Saulcy, Metz, 57000, FRANCE
Email: sorin.stratulat@univ-lorraine.fr

I. INTRODUCTION

In [1] it has been shown that concrete induction-based theorem provers can be built by implementing abstract rules with reasoning modules. The abstract rules express the declarative part of the reasoning and can be considered as the logical components of the provers since they define what information can be soundly used during proof derivations, for example the induction hypotheses. On the other hand, the reasoning modules represent their operational components that define how new formulas can be derived from existing ones, based on the information provided by the abstract rules.

In this paper, we discuss two integration schemas of reasoning modules in SPIKE [1]–[3], an automated first-order theorem prover based on implicit induction. The first follows ideas from [4]. It allows for a tight cooperation between the prover and the reasoning modules. Compared with the previous schema, the reasoning modules can provide in addition new information in terms of logical consequences of the given input, which can be further used by the prover in subsequent steps of the proof.

The second integration schema follows a ‘black box’ approach such that the conjectures to be proved, together with other information (axioms, induction hypotheses) useful for proving the conjecture, are processed by an external tool, in our case an SMT solver. The answers provided by the tool back to the prover are rather limited; they can inform SPIKE only if the conjecture is true, false or has an unknown truth value.

The rest of the paper has 3 sections. Section II introduces the backgrounds of reasoning by implicit induction and describes the SPIKE prover. Section III details two extensions of SPIKE, following the above integration schemas, with reasoning modules operating over naturals combined with interpreted symbols, then compare the performances of their implementations when tested on a non-trivial application. The last section concludes.

II. THE SPIKE PROVER

In the following, we present the syntax and semantics of SPIKE specifications, as well as the underlying induction principle and the SPIKE’s inference system.

Syntax. SPIKE can reason on many-sorted conditional specifications built from quantifier-free conditional equalities, denoted by axioms, that define function symbols. In addition, a (disjoint) set of constructor function symbols is attached to each sort. We denote by $\mathcal{F}$ the disjoint union between the sets of constructor and defined function symbols, and by $\mathcal{V}$ a denumerable set of variables. Each variable from $\mathcal{V}$ has a sort and the profile of each function $f$ from $\mathcal{F}$ is sorted, of the form $f : s_1 \times \ldots \times s_n \to s$, where $s_1, \ldots, s_n$ are sorts. In this case, we say that $f$ has the sort $s$. The set of terms built from function symbols from $\mathcal{F}$ and variables from $\mathcal{V}$ is denoted by $T(\mathcal{F}, \mathcal{V})$. If $S$ is a denumerable set of sorts, by $T(\mathcal{F}, \mathcal{V})_{s \in S}$ we can recursively define the set of terms of sort $s$, which can be either a variable of sort $s$ or a non-variable term of the form $f(t_1, \ldots, t_n)$, where $f$ has the profile $f : s_1 \times \ldots \times s_n \to s$ and $t_1$ (resp., $t_2, \ldots, t_n$) is in $T(\mathcal{F}, \mathcal{V})_{s \in S}$ (resp., $T(\mathcal{F}, \mathcal{V})_{s_2 \in S}, \ldots, T(\mathcal{F}, \mathcal{V})_{s_n \in S}$). Therefore, $T(\mathcal{F}, \mathcal{V})$ can be defined as $\cup T(\mathcal{F}, \mathcal{V})_{s \in S}$, i.e., the disjoint union of the sets of terms of same sort, by considering all the sorts occurring in $S$.

An (unconditional) equality is a binary relation between two terms $l$ and $r$ of same sort, denoted by $l = r$. A conditional equality is represented under the form of the implication $l_1 = r_1 \land \ldots \land l_n = r_n \Rightarrow l = r$, where the conclusion $l = r$ and each $l_i = r_i$ ($i \in [1..n]$) from the condition part are unconditional equalities. SPIKE accepts specifications built on free constructors, for which there is no equality relation between any two different constructor symbols. A term or equality is ground if it has no variables. The equality $l_1 = r_1 \land \ldots \land l_n = r_n \Rightarrow l = r$ can be oriented into the rewrite rule $l_1 = r_1 \land \ldots \land l_n = r_n \Rightarrow l \rightarrow r$ if $l$ is greater than $r$ (w.r.t. an ordering over terms) as well as any $l$, for any $i \in [1..n]$. In SPIKE, the ordering over terms is the recursive path ordering (rpo) with status [5], denoted in the following by $<_{\text{rpo}}$.

New terms and equalities can be built from replacing (subsets of their) variables with terms of the same sort, by the means of substitutions. A substitution is a mapping $\{x_i \mapsto t_i\}$, where the variable $x_i$ and the term $t_i$ have the same sort. It is ground if each replacing term is ground. If $\sigma$ is a substitution and $t$ a term or equality, then $t\sigma$ is an instance of $t$. 

Abstract—We detail the integration in SPIKE, an implicit induction theorem prover, of two reasoning modules operating over naturals combined with interpreted symbols. The first integration schema is à la Boyer-Moore, based on the combination of a congruence closure procedure with a decision procedure for linear arithmetic over rationals/reals. The second follows a ‘black-box’ approach and is based on external SMT solvers. It is shown that the two extensions significantly increase the power of SPIKE; their performances are compared when proving a non-trivial application.

The second integration schema follows a 'black box' approach and is based on external SMT solvers. It is shown that the two extensions significantly increase the power of SPIKE; their performances are compared when proving a non-trivial application.
Semantics. Let $Ax$ be the set of axioms of a given SPIKE specification and $M$ a set of Herbrand models of $Ax$. An (unconditional or conditional) equality is a $M$-consequence (or just consequence) of a set of equalities $\Phi$, denoted by $\Phi \models_M \phi$, if $\phi$ is valid in the model $m$ whenever $\psi$ is valid in $m$, for any $\psi \in \Phi$ and $m \in M$. An equality $\phi$ is $M$-valid (or just valid), denoted by $\models_M \phi$, iff it is a consequence of $Ax$.

In the following, we consider $M$ as being the singleton built from the initial model of $Ax$. The consequence relation is referred to as inductive. An equality is a counterexample if there is a ground instance of it which is not valid. We say that an equality has a counterexample if there is a ground instance of it which is a counterexample. A set of equalities has a counterexample if there is an equality from the set that has a counterexample.

The induction principle. SPIKE implements an instance of the Noetherian induction principle. Given $(\mathcal{E}, \prec)$ a non-empty well-founded poset of equalities to prove, the formula-based instance of the Noetherian induction principle [6] states that if, for any equality $\delta \in \mathcal{E}, \{\gamma \in \mathcal{E} \mid \prec \gamma \} \models_M \delta$ then $\forall \rho \in \mathcal{E}, \models_M \rho$. The principle allows that, during the proof of an equality conjecture, smaller formulas can be used in terms of induction hypotheses (IHs), which makes it a natural choice for reasoning on proofs requiring lazy and mutual induction steps. The proof method used by SPIKE is an application of this principle, called implicit induction, as shown in [7] and formally presented in [8]. In SPIKE, $\prec$ is a multiset extension of $\leq_c$.

The inference system. The inference system is a set of inference rules representing transitions between pairs of sets of equalities of the form $(E, H) \vdash (E', H')$, where $E, E'$ are two sets of conjectures and $H, H'$ are two sets of premises. The application of such a rule replaces a (current) conjecture $\phi$ with a potentially empty set of new conjectures $\Phi$. This can be summarized as $(E \cup \{\phi\}, H) \vdash (E \cup \Phi, H')$. Moreover, $\phi$ may be added as a premise in order to participate to further inference steps. Therefore, $H'$ can be either $H$ or $H \cup \{\phi\}$. An implicit induction derivation for a set of equalities $E_0$ is any sequence of the form $(E_0, H_0) \vdash \cdots \vdash (E_n, H_n) \vdots$ resulted from successive applications of inference rules starting from $(E_0, H_0)$. A proof is a finite derivation $(E_0, H_0) \vdash \cdots \vdash (E_n, H_n)$ for which $H_0$ and $E_n$ are empty sets.

The induction principle can be applied to validate a proof $(E_0, \emptyset) \vdash \cdots \vdash (\emptyset, H_n)$ by considering $\mathcal{E}$ as the set $\{\phi \mid \phi \in \bigcup_i E_i\}$ and the inference system as reductive, i.e., for any ground instance $\phi\tau$ of the current conjecture $\phi$ from any step $(E \cup \{\phi\}, H) \vdash (E \cup \Phi, H')$ of the proof, either i) $\phi\tau$ is valid, or ii) $\phi\tau$ is a counterexample and there is a smaller or equal counterexample in $E \cup \Phi$. In addition, we assume that any premise does not have minimal counterexamples in $\mathcal{E}$. It can be easily noticed that any reductive system is sound, i.e., $\models_M E_0$. Otherwise, if $E_0$ is not valid, $\mathcal{E}$ should have a counterexample. Since $\prec$ is well-founded, there is a minimal counterexample $\tau$ in $\mathcal{E}$. The proof ends with an empty set of conjectures, therefore there is a last proof step where $\phi$ is the current conjecture, of the form $(E \cup \{\phi\}, H) \vdash (E \cup \Phi, H')$, and $\phi$ has a counterexample equal to $\tau$. On the other hand, the inference system is reductive, so there is a counterexample smaller than $\tau$ in $E \cup \Phi$, as $E \cup \Phi \cup H$ does not have counterexamples equal to $\tau$, hence in $\mathcal{E}$. Contradiction.

Concrete reductive inference systems, for which the premises do not have minimal counterexamples, can be built using a methodology based on contextual cover sets (CCSs) [1]. We denote by the context $C$ the pair of sets of equalities $(C^1, C^2)$. We say that the set of equalities $\Psi$ contextually covers the set of equalities $\Phi$ in $C$, denoted by $\Phi \sqsubseteq_C \Psi$, iff $C^1 \leq_{\phi\sigma} \cup C^2_{\phi\sigma} \cup \Psi_{\phi\sigma} \models_M \phi\sigma$, for any $\phi \in \Phi$ and ground instance $\phi\sigma$. When $\Phi = \{\phi\}$, we say that $\Psi$ is a CCS of $\phi$. When $\Phi$ is empty, the CCS is empty, and when $\Psi_{\phi\sigma}^1 \leq_{\phi\sigma}$ is replaced in the definition by $\Psi_{\phi\sigma}^1 <_{\phi\sigma}$, then the CCS is strict and denoted by $\Phi \sqsubseteq_C \Psi$.

The core of the methodology is an abstract inference system made of two rules: ADDPREMISE and SIMPLIFY, as shown in Fig. 1.

1-ADD-PREMISE: $(E \cup \{\phi\}, H) \vdash (E \cup \Phi, H \cup \{\phi\})$
if $\{\phi\} \sqsubseteq_{(H, E)} \Phi$

1-SIMPLIFY: $(E \cup \{\phi\}, H) \vdash (E \cup \Phi, H)$
if $\{\phi\} \sqsubseteq_{(E \cup H, \emptyset)} \Phi$

Fig. 1: Abstract inference rules.

Each rule defines the set of new conjectures $\Phi$ as a CCS of the current conjecture $\phi$, by explicitly stating the content of the context. It can be noticed that the current conjecture of ADDPREMISE does not have minimal counterexamples, so the rule adds it as a premise. SIMPLIFY does not satisfy this property but allows bigger contexts. A very useful instance of SIMPLIFY is when $\Phi$ is empty, also referred to as the DELETE rule. It can be noticed that more complex abstract inference rules can be built. In [1], it was shown that abstract inference rules can build in two steps the CCS of the current conjecture, thanks to the compositional properties of CCSs and that they define the biggest contexts compared to similar abstract rules proposed in the literature.

According to the methodology, any concrete inference rule is built as an instance of one of the abstract rules by implementing its elementary CCSs, i.e. the CCSs that are not built with composition operations, by the means of reasoning modules. A reasoning module can produce a CCS with a particular reasoning technique using as IHs formulas from the context defined by the instantiated abstract rule. A reasoning module $M$ can be characterised by a function $f_M$ defined as $f_M(\phi, C) = \Phi$, where $\Phi$ is a CCS of $\phi$ in the context $C$. The main reasoning techniques used by SPIKE are based on rewriting, case analysis and variable instantiations. Since any proof ends with an empty set of conjectures, a class of interest is represented by the reasoning modules that build empty CCSs.

III. IMPLEMENTING REASONING MODULES FOR NATURALS COMBINED WITH INTERPRETED SYMBOLS

We will detail the implementation of two reasoning modules for naturals combined with interpreted symbols using: i) a tight integration schema à la Boyer-Moore [4] with homemade components, and ii) a ‘black-box’ approach based on an
external SMT solver. Both reasoning modules take as input an equality, as well as information from the context of the CCS defined by implemented abstract rule, negate the equality and validate it if an inconsistency is detected. We further point out the major role they played during the validation of the conformity algorithm for a telecommunications protocol.

A. Integrating the reasoning module à la Boyer-Moore

The reasoning module is built from the cooperation between a decision procedure for the quantifier-free theory of equality, $T_{ge}$, and a decision procedure for linear arithmetic over rationals/reals, $T_{la}$. It served as a case study to validate an approach integrating decision procedures in a proof environment mixing induction and rewriting techniques [9]. The cooperation schema is defined by the function symbols and variables from the above definition. New rewrite rules and linear inequalities issued from the negation of the equality given as input.

Definition 1 ($\mathcal{C}$-structure): Let $C$ be a conditional equality. A $\mathcal{C}$-structure is a data structure consisting of the quadruple $\langle \mathcal{C} \cup \mathcal{G} \cup \mathcal{L} \rangle$, where

- $\mathcal{C}$ is built from unconditional rewrite rules;
- $\mathcal{A}$ consists of atomic formulas issued from the negation of $C$;
- $\mathcal{G}$ contains unconditional equalities and disequalities;
- $\mathcal{L}$ is the pair $(\mathcal{P} \cdot \mathcal{IE})$, where $\mathcal{P}$ has linear inequalities and $\mathcal{IE}$ stores the (implicit) equalities issued from $\mathcal{P}$ after applying the decision procedure for $T_{la}$.

The components. The cooperation schema is defined by transitions between states, called $\mathcal{C}$-structures. A $\mathcal{C}$-structure consists of constraint stores containing equalities, rewrite rules and linear inequalities issued from the negation of the equality given as input.

Example 1: Let $C$ be the conjecture from Example 5.2 of [10] (also presented in [11]):

$$(p(x) = \text{True} \land z \leq f(\text{max}(x, y)) = \text{True} \land 0 < \text{min}(x, y) = \text{True} \land x \leq \text{max}(x, y) = \text{True} \land \text{max}(x, y) \leq x = \text{True}) \Rightarrow z < g(x) + y = \text{True}.$$

The $\mathcal{C}$-structure is initialized by

$$\langle \mathcal{C} | \mathcal{A} | \mathcal{G} | \mathcal{L} \rangle$$

The cooperation schema allows to normalize monomials from $\mathcal{P}$ and equalities from $\mathcal{IE}$ with rewrite rules from $\mathcal{CR}$. We denote by linearize the procedure to transform an atomic formula into a conjunction of linear inequalities. It returns the empty set if no transformation has been operated on the input; in this case, we say that the input is non-linearizable, otherwise it is linearizable. The decision procedure for $T_{la}$ will be applied on the linear inequalities from $\mathcal{P}$ based on the Fourier-Motzkin variable elimination method [12] over rationals and implemented by the function arith. It returns inconsistent if an unsatisfiable linear inequality (for example, $1 \leq 0$) is generated. A set of implicit equalities is returned if the inequality $0 \leq 0$ is produced by the method. The decision procedure is incomplete because a consistency reported with an interpretation over rationals can be an inconsistency with an interpretation over naturals. In spite of this limitation and its doubly exponential complexity, it has been chosen for its simple implementation. The disjunction operations, mainly resulted when processing disequalities, can be avoided as explained in the following. It can be noticed that the negation of the equality $l_1 = r_1 \land \ldots \land l_n = r_n \Rightarrow l = r$ can be represented as the conjunction $l_1 = r_1 \land \ldots \land l_n = r_n \land l \neq r$. For the case when the linearization of $l \neq r$ yields the disjunction $l - r + 1 \leq 0 \lor r - l + 1 \leq 0$, the disequality $l \neq r$ is not processed (but can be processed by the decision procedure for $T_{ge}$). This is another source of incompleteness but guarantees that there is always only one set of conjunctions of monomials on which the Fourier-Motzkin method operates. An example where the disjunctive representation can be avoided is when one of the sides of the disequality is a boolean value (True or False), for example $f(x) \leq g(x) \neq \text{False}$ can be linearized to $f(x) - g(x) \leq 0$.

The monomials from a linear inequality can be ordered by a total ordering, similar to [13] and defined as follows.

Definition 2 (total ordering over terms): Let $t_1$ and $t_2$ be two terms.

We write $t_1 <_{gt} t_2$ iff

- the number of variable positions in $t_1$ is smaller than that for $t_2$, or
- the number of variable positions in both terms is the same, but the number of strict positions in $t_1$ is smaller than that for $t_2$, or
- the number of variable and strict positions in both terms is the same, but $t_1$ is smaller than $t_2$, according to some lexicographic ordering [5, p. 19].

The function symbols and variables from the above definition are uniquely interpreted as naturals by using their internal representation into SPIKE. On the other hand, $<_{gt}$ does not need to be well-founded, hence it may be different from $<_t$.

The input for the decision procedure for $T_{ge}$, denoted by cong, is a finite set of equalities and disequalities. The equalities are oriented into rewrite rules using the $<_{gt}$ ordering. New equalities are produced by a rewrite-based ground completion algorithm, similar to [14], and the new rewrite system helps to normalize the members of the disequalities. The output of cong is either i) inconsistent, if a trivial disequality of the form $s \neq s$ is derived, or ii) the new set of equalities and disequalities.

The cooperation schema. We show in the dotted box from Fig. 2 the transitions between the components of a $\mathcal{C}$-structure,
and from the proof environment to the reasoning module. The
definition of each transition is given in Fig. 3.

The atomic formulas from A will be dispatched between
L and G, using the first two transitions, A2L and A2G, such
that any atomic formula that cannot be transformed into a
conjunction (seen as a set) of linear inequalities will be sent
to G, otherwise to P. The transitions initialize respectively
the P and G components with linear inequalities and equal-
ities/disequalities resulted from the atomic formulas from A.
The transition A2L applies in addition arith on the new set
of linear inequalities. Any new rewrite rule, resulted from the
application of congr on G, are transferred to CR by G2CR. The
L2G rule transfers to G all equations from IE. The monomials
of linear equalities from P and the equations from IE are
normalized by rules from CR, using the normalize procedure.

The proof environment interacts with the reasoning module
by means of the augmentation operation, inspired from [4].
Defined in Fig. 4 and denoted by Oracle_A, it is based on
the R(G) set of formulas which may contain (instances of)
conditional equations issued from axioms, lemmas, or the
context of the CCS given as input to the reasoning module.
Oracle_A firstly assumes that R(G) has an instance of a condi-
tional equality \( \phi \) whose conclusion is linearizable and, after
the linearization process, i) the set of inequalities has a maxi-
mal monomial of an inequality from P but with a coefficient with
opposite sign, and ii) no new monomials are introduced in
P. If the conditions of \( \phi \) are consequences of R(G) and the
conditions of C, denoted by \( \text{cond}(C) \), the operation Oracle_A
returns the set of equalities from the condition part of \( \phi \).
The last rule from Fig. 3, augment_L, augments the set of
linear inequalities by using rewrite rules from R(C); it applies
afterwards arith on the new set of inequalities.

The strategy for applying the above transitions can be
implemented with a list of transition labels, initialized by
[A2G, A2L]. The first element from the list determines the
transition to be executed. When the current transition finished

Fig. 2: The cooperation schema.

The cooperation schema succeeds if inconsistent is returned by
arith or congr procedures. The cooperation schema failed if
the list becomes empty.

Properties. The cooperation schema of the reasoning module
is sound and well-founded.

Theorem 1 (soundness): Let \( \phi \) be a conditional equality
and C the context given as input to the reasoning module.
Then, the empty set is an empty CCS of \( \phi \) in C whenever
arith or congr return inconsistent during the execution of
the cooperation schema.\(^1\)

Theorem 2 (termination): The cooperation schema is well-

Proof: Let us assume that there is a chain of \( \overline{C} \)-structure
transformations \( S_1 \rightarrow S_2 \rightarrow \ldots \rightarrow S_n \rightarrow \ldots \). It can be
noticed that the execution graph of the rules from Fig. 2
has two minimal cycles: \( \text{(A2L, L2G, G2CR, normalize)} \) and
\( \text{(A2L, augment_L)} \). Since i) the set of monomials from P
can be modified only by rewriting with rewrite rules from
CR, and ii) \( <_t \) is well-founded, there is a step \( j \) in the
chain such that normalize is no longer applied in further
steps to modify the set of monomials from P. This breaks
the cycle (A2L, L2G, G2CR, normalize). If no new monomials
are produced, Oracle_A will return an empty set of equalities,

A detailed example. Let \( C \) be the conjecture from Example 1.
We will show how the reasoning module can help producing an
inconsistency from its negation, assuming that R(G) consists of
the set of two equalities \{ \( \max(x', y') = x' \Rightarrow \min(x', y') = y' \), \( p(u) = True \Rightarrow f(u) \leq g(u) = True \) \}, and that there is
no additional knowledge on \( f \) and \( g \). The initial \( \overline{C} \)-structure
is (the empty fields are not listed in the following):

\( ^1\)For lack of space, the detailed proof can be found in [15].
Fig. 3: The transition rules for the cooperation schema.

\[\langle CR \mid A \cup E \mid G \mid (P \cdot \text{IE})\rangle \xrightarrow{A2L} \langle CR \mid A \mid G \mid (P' \cdot \text{IE} \cup \text{IE}')\rangle\]

where
1) \(E\) is a maximal set such that \(\forall e \in E, e\) is linearizable, and
2) \((\text{IE}', P') = \text{arith}(\{\cup_{e' \in E} \text{linearize}(e')\} \cup P)\)

Next:
if \(E = \emptyset\) then \([]\) else (if \(IE' = \emptyset\) then \([\text{augment}_L]\) else \([\text{L2G}]\))

\[\langle CR \mid A \cup E \mid G \mid L\rangle \xrightarrow{A2G} \langle CR \mid A \mid G \cup E \mid L\rangle\]

where \(E\) is a maximal set such that \(a = b\) is non-linearizable, \(\forall a \rightarrow b \in E\)

Next: \([\text{G2CR}]\)

\[\langle CR \mid A \mid G \mid L\rangle \xrightarrow{G2CR} \langle CR \cup E \mid A \mid \text{congr}(G) \mid L\rangle\]

where \(\forall a \rightarrow b \in E, a = b \in \text{congr}(G)\) and \(b <_t a\)

Next: if \(E\) is empty then \([]\) else \([\text{normalize}]\)

\[\langle CR \mid A \mid G \mid (P \cdot \text{IE})\rangle \xrightarrow{L2G} \langle CR \mid A \mid G \cup \text{IE} \mid (P \cdot \emptyset)\rangle\]

Next: \([\text{G2CR}]\)

\[\langle CR \mid A \mid G \mid (P \cdot \text{IE})\rangle \xrightarrow{\text{normalize}} \langle CR \mid A \mid G' \mid (P' \cdot \text{IE}')\rangle\]

where
1) \(P'\) is \(P\) normalized with rules from \(\text{CR}\), and
2) \(\text{IE}'\) is \(\text{IE}\) normalized with rules from \(\text{CR}\)

Next: \([\text{A2L}]\)

\[\langle CR \mid A \mid G \mid (P \cdot \text{IE})\rangle \xrightarrow{\text{augment}_L} \langle CR \mid A \mid G \mid (P' \cdot (\text{IE} \cup \text{IE}'))\rangle\]

where
1) \(E = \text{Oracle}_A((CR \mid A \mid G \mid (P \cdot \text{IE})))\), and
2) \((P', \text{IE}') = \text{arith}(\{\cup_{e \in E} \text{linearize}(e)\} \cup P)\)

Next: if \(E = \emptyset\) then \([]\) else (if \(IE' = \emptyset\) then \([\text{A2L}]\) else \([\text{L2G}]\))

Fig. 4: The \(\text{Oracle}_A\) augmentation operation.

\(\langle A : \{p(x) = \text{True}, z \leq f(max(x, y)) = \text{True}, 0 < \min(x, y) = \text{True}, x \leq \max(x, y) = \text{True}, \max(x, y) \leq x = \text{True}, z < g(x) + y = \text{False}\} \rangle\)

The transition list is initiated by \([\text{A2G, A2L}]\). Firstly, the non-linearizable and non-orientable equality \(p(x) = \text{True}\) is transferred to \(G\), by \(\text{A2G}\). \(\text{A2G}\) is replaced by \(\text{G2CR}\) in the transition list, but its application on the non-orientable equality does not create new rewrite rules. Next, the rest of the (linearizable) equalities from \(A\) are dispatched to \(L\), by \(\text{A2L}\). The equalities from \(L\) are linearized and the new state becomes:

\(\langle G : \{p(x) = \text{True}\} \rangle\)

\(L : \{\{z - f(max(x, y)) \leq 0, 1 - \min(x, y) \leq 0, x - \max(x, y) \leq 0, \max(x, y) - x \leq 0, g(x) + y - z \leq 0\} \cdot \emptyset\}\}

By adding the inequalities \(x - \max(x, y) \leq 0\) and \(\max(x, y) - x \leq 0\), we get the inequality \(0 \leq 0\). Therefore, the decision procedure for \(T_{la}\) can derive the equality \(\max(x, y) = x\). On the other hand, the inequality \(g(x) + y - f(max(x, y)) \leq 0\) is the result of the addition between \(g(x) + y - z \leq 0\) and \(z - f(max(x, y)) \leq 0\). The new state after applying \(\text{A2L}\) is:

\(\langle G : \{p(x) = \text{True}\} \rangle\)

\(L : \{\{z - f(max(x, y)) \leq 0, 1 - \min(x, y) \leq 0, x - \max(x, y) \leq 0, \max(x, y) - x \leq 0, g(x) + y - z \leq 0, 0 \leq 0, g(x) + y - f(max(x, y)) \leq 0\} \cdot \{\max(x, y) = x\}\}

The next transition is \(\text{L2G}\), so \(\max(x, y) = x\) of \(\text{IE}\) is transferred to \(G\), to give:

\(\langle G : \{p(x) = \text{True}, max(x, y) = x\} \rangle\)

\(L : \{\{z - f(max(x, y)) \leq 0, 1 - \min(x, y) \leq 0, x - \max(x, y) \leq 0, \max(x, y) - x \leq 0, g(x) + y - z \leq 0, 0 \leq 0, g(x) + y - f(max(x, y)) \leq 0\} \cdot \{\emptyset\}\}\}

The execution of \(\text{congr}\) on \(\text{CR}\) within the next \(\text{G2CR}\) transition produces no changes. However, since \(\max(x, y) >_t x\), a copy
of the equality \(\max(x, y) = x\) is transformed into the rewrite rule \(\max(x, y) \rightarrow x\) and stored to CR:

\[
\langle CR \rangle : \{\max(x, y) \rightarrow x\}
\]

\[G : \{p(x) = True, \max(x, y) = x\}\]

\[L : \{(z - f(x) \leq 0, \min(x, y) \leq 0, g(x) + y - f(x) \leq 0) \Rightarrow \emptyset\}\}

Next, we apply normalize to convert \(\max(x, y)\) to \(x\) in all inequalities from \(P\):

\[
\langle CR \rangle : \{\max(x, y) \rightarrow x\}
\]

\[G : \{p(x) = True, \max(x, y) = x\}\]

\[L : \{(z - f(x) \leq 0, \min(x, y) \leq 0, g(x) + y - z \leq 0, 0 \leq 0, g(x) + y - f(x) \leq 0) \Rightarrow \emptyset\}\}

By applying A2L, arith is again called, to give the new state:

\[
\langle CR \rangle : \{\max(x, y) \rightarrow x\}
\]

\[G : \{p(x) = True, \max(x, y) = x\}\]

\[L : \{(z - f(x) \leq 0, \min(x, y) \leq 0, g(x) + y - z \leq 0, 0 \leq 0, g(x) + y - f(x) \leq 0) \Rightarrow \emptyset\}\}

No new equalities are produced this time, so augment_L is called to add new inequalities to \(P\). By instantiating the first equality of \(R(C)\) with the substitution \(\{x' \rightarrow x; y' \rightarrow y\}\), we get \(\max(x, y) = p \Rightarrow \min(x, y) = y\) which can eliminate the maximal monomial \(\min(x, y)\) from \(1 - \min(x, y) \leq 0\). By adding the set of inequalities \(\{(x, y - \min(x, y) \leq 0)\}\) resulted from the linearization of \(\min(x, y) = y\), the transition A2L is fired to get the equality \(\min(x, y) = y\). It will be transferred afterwards to \(G\) by L2G, to give:

\[
\langle CR \rangle : \{\max(x, y) \rightarrow x\}
\]

\[G : \{p(x) = True, \max(x, y) = x, \min(x, y) = y\}\]

\[L : \{(z - f(x) \leq 0, \min(x, y) \leq 0, g(x) + y - z \leq 0, 0 \leq 0, g(x) + y - f(x) \leq 0, 1 - y \leq 0, \min(x, y) - y \leq 0, y - \min(x, y) \emptyset)\}\]

As for \(\max(x, y) = x\), the equality \(\min(x, y) = y\) can be oriented from left to right, and a copy of it is transformed into a rewrite rule, then added to CR by applying G2CR:

\[
\langle CR \rangle : \{\max(x, y) \rightarrow x, \min(x, y) \rightarrow y\}
\]

\[G : \{p(x) = True, \max(x, y) = x, \min(x, y) = y\}\]

\[L : \{(z - f(x) \leq 0, \min(x, y) \leq 0, g(x) + y - z \leq 0, 0 \leq 0, g(x) + y - f(x) \leq 0, 1 - y \leq 0, \min(x, y) - y \leq 0, y - \min(x, y) \emptyset)\}\]

The inequalities from \(P\) are normalized with \(\min(x, y) \rightarrow y\) to get:

\[
\langle CR \rangle : \{\max(x, y) \rightarrow x, \min(x, y) \rightarrow y\}
\]

\[G : \{p(x) = True, \max(x, y) = x, \min(x, y) = y\}\]

\[L : \{(z - f(x) \leq 0, 1 - y \leq 0, g(x) + y - z \leq 0, 0 \leq 0, g(x) + y - f(x) \leq 0) \Rightarrow \emptyset\}\}

The transition augment_L is applied a second time, now with the instance \(p(x) = True \Rightarrow f(x) \leq g(x) = True\) of the second equality of \(R(C)\) using the substitution \(\{u \rightarrow x\}\). The condition \(p(x) = True\) is satisfied, so the inequality \(f(x) - g(x) \leq 0\) is added to \(P\):

\[
\langle CR \rangle : \{\max(x, y) \rightarrow x, \min(x, y) \rightarrow y\}
\]

\[G : \{p(x) = True, \max(x, y) = x, \min(x, y) = y\}\]

\[L : \{(f(x) - g(x) \leq 0, z - f(x) \leq 0, 1 - y \leq 0, g(x) + y - z \leq 0, 0 \leq 0, g(x) + y - f(x) \leq 0, 1 \leq 0) \Rightarrow \emptyset\}\}

Finally, arith yields the unsatisfiable inequality \(1 \leq 0\) after adding \(1 - y \leq 0, f(x) - g(x) \leq 0\) and \(g(x) + y - f(x) \leq 0\). Hence, the procedure returns inconsistent.

**Related works.** Boyer and Moore [4] have been the first to define a heuristics using the augmentation operation. They integrated it into a decision procedure for linear arithmetic that can also manipulate disjunctions of linear inequalities and conditional linear inequalities. After conducting many experiments, they concluded that the augmentation operation increases significantly the performance of the decision procedure, hence improving the automatisation degree of their prover [16]. However, their integration schema is informal and lacks of termination proof. Several works have been inspired from Boyer-Moore’s integration schema. [10] presented an enriched schema with a congruence closure procedure, implemented in the Tecton prover [17]. An informal termination proof of their integration schema is given in [11]. To our knowledge, the first formalisation of the Boyer-Moore’s integration schema was given in [18], where it instantiated a more general ‘plug and play’ reasoning framework. [19] abstracts the interactions between rewriting and decision procedures via constraint contextual rewriting (CCR) rules parameterized by decision procedures. In the same line, [20] presents a flexible environment to integrate decision procedures into heuristic theorem provers.

Different cooperation schemas could have been used instead, as those inspired by Nelson-Oppen [21], Shostak [22] or, more recently, based on Delayed Theory Combination [23].

**B. Integrating the external SMT solver**

SPIKE can translate conditional specifications into SMT specifications following the SMT 2.0 format. By using a ‘black-box’ integration approach, the calls to the previous reasoning module have been replaced in SPIKE by calls to the Z3 SMT solver [24], version 4.3.2. For any call, the axioms and the negated conjecture are translated one-to-one into assert constructions. They are saved in a separate .smt2 file which is finally tested for satisfiability by an external Z3 process. If Z3 returns unsat, the conjecture is interpreted by SPIKE as valid and deleted from the current set of conjectures.

Z3 is an efficient SMT solver that combines different first-order theories using the Model Based Theory Combination approach [25], among which the theories for equality reasoning and arithmetics. It integrates a decision procedure for linear
arithmetic over integers that can be activated by interpreting the naturals from the SPIKE specification as non-negative integers. In order to do this, i) the natural sort is translated to the built-in integer sort ‘Int’, and ii) any SPIKE equality including a set of natural variables V will add the constraints x ≥ 0, for any x ∈ V, as conditions in the corresponding assert construction.

The translation process is automatic, excepting for the user-defined sorts which should be manually translated.

C. Applications

We have previously used the reasoning module à la Boyer-Moore (BM) on several non-trivial applications [3], [9], [26]. Here, we detail our experience with the validation proof of the conformity algorithm for a telecommunications protocol, fully developed with the PVS [27] system in [28]. The proof is about showing the equivalence between two functions defined over naturals and lists. It mixes induction reasoning, case analysis and arithmetic reasoning. It also requires non-trivial user interaction, among which 79 user-defined lemmas.

Later on [26], a previous version (p.v.) of SPIKE integrating BM has proved completely automatically 48 user defined lemmas. To measure the impact of decision procedures on the automatisation degree of the prover, TABLE I indicates that a number of 69 lemmas was required by the successfully proved lemmas using the current version (c.v.) without the integration of reasoning modules. The current version, this time integrating the reasoning modules, helped to completely automatically prove 46 lemmas by using the same proof strategy; only the last ‘final’ lemma required one additional lemma. It can be noticed that the overall BM-based proofs have been done 20 times faster than the overall Z3-based proofs. This can be explained by the non-negligible time needed to launch the Z3 processes. On the other hand, the incompleteness of BM didn’t penalize its effectiveness on this example, the two reasoning modules being able to prove the same conjectures. Notice that, in general, the two sets of conjectures proved with BM and Z3, respectively, may differ. For example, we can imagine divergent BM-derivations of BM false negatives that cannot be automatically proved by induction, i.e., without providing additional lemmas, but which can be successfully conducted by Z3.

Only one successful proof, done without the help of reasoning modules, was unsuccessful when using reasoning modules (see ‘null_wind2’). In the other direction, there are 7 lemmas proved with reasoning modules but not proved using lemmas; in fact, we have found difficulties to prove the required lemmas leading to a successful proof. Also, it can be noticed that very few conjectures are not proved by SPIKE integrating reasoning modules. This is due to the slight differences between the specifications, mainly in the definition of the induction orderings and the parameterization of the inference rules. The implementation of some inference rules also changed, which explains why conjectures like ‘null_insat’ and ‘null_insat’ are proved by the previous version and not by the current version. For example, the implementation of some augment-like inference rule from [26] unsoundly adds to the

TABLE I: Statistics about the proof with (w/) and without (w/o) reasoning modules (r.m.).

<table>
<thead>
<tr>
<th>#</th>
<th>lemma</th>
<th>p.v. w/ BM</th>
<th>c.v. w/o r.m.</th>
<th>time (s) w/ r.m.</th>
<th>time (s) w/o r.m.</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>firstat_timeat</td>
<td>0</td>
<td>1</td>
<td>0.052</td>
<td>0.749</td>
</tr>
<tr>
<td>02</td>
<td>firstat_progart</td>
<td>0</td>
<td>1</td>
<td>0.049</td>
<td>0.226</td>
</tr>
<tr>
<td>03</td>
<td>sorted_sorted</td>
<td>0</td>
<td>0</td>
<td>0.034</td>
<td>1.265</td>
</tr>
<tr>
<td>04</td>
<td>sorted_insat1</td>
<td>0</td>
<td>4</td>
<td>0.054</td>
<td>0.399</td>
</tr>
<tr>
<td>05</td>
<td>sorted_insat2</td>
<td>0</td>
<td>4</td>
<td>0.082</td>
<td>1.626</td>
</tr>
<tr>
<td>06</td>
<td>sorted_e_two</td>
<td>0</td>
<td>0</td>
<td>0.037</td>
<td>1.258</td>
</tr>
<tr>
<td>07</td>
<td>sorted_e_insat</td>
<td>0</td>
<td>3</td>
<td>0.049</td>
<td>1.633</td>
</tr>
<tr>
<td>08</td>
<td>member_t_insat</td>
<td>0</td>
<td>2</td>
<td>0.060</td>
<td>5.182</td>
</tr>
<tr>
<td>09</td>
<td>member_t_insat</td>
<td>0</td>
<td>2</td>
<td>0.057</td>
<td>4.164</td>
</tr>
<tr>
<td>10</td>
<td>member_firstat</td>
<td>0</td>
<td>2</td>
<td>0.046</td>
<td>6.172</td>
</tr>
<tr>
<td>11</td>
<td>timel_insat_t</td>
<td>0</td>
<td>0</td>
<td>0.037</td>
<td>3.121</td>
</tr>
<tr>
<td>12</td>
<td>erl_insat</td>
<td>0</td>
<td>0</td>
<td>0.035</td>
<td>2.134</td>
</tr>
<tr>
<td>13</td>
<td>erl_insat</td>
<td>0</td>
<td>0</td>
<td>0.036</td>
<td>2.134</td>
</tr>
<tr>
<td>14</td>
<td>erl_prog</td>
<td>0</td>
<td>2</td>
<td>0.058</td>
<td>7.340</td>
</tr>
<tr>
<td>15</td>
<td>time_progar</td>
<td>0</td>
<td>1</td>
<td>0.041</td>
<td>1.085</td>
</tr>
<tr>
<td>16</td>
<td>timeat_tim</td>
<td>0</td>
<td>0</td>
<td>0.042</td>
<td>4.147</td>
</tr>
<tr>
<td>17</td>
<td>timel_timeat_max</td>
<td>0</td>
<td>3</td>
<td>0.075</td>
<td>1.116</td>
</tr>
<tr>
<td>18</td>
<td>null_insat</td>
<td>0</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>19</td>
<td>null_insat</td>
<td>0</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>20</td>
<td>null_listat</td>
<td>0</td>
<td>1</td>
<td>0.035</td>
<td>5.164</td>
</tr>
<tr>
<td>21</td>
<td>listat</td>
<td>0</td>
<td>0</td>
<td>0.031</td>
<td>0.063</td>
</tr>
<tr>
<td>22</td>
<td>cons_insat</td>
<td>0</td>
<td>0</td>
<td>0.031</td>
<td>3.115</td>
</tr>
<tr>
<td>23</td>
<td>cons_listat</td>
<td>0</td>
<td>0</td>
<td>0.028</td>
<td>0.058</td>
</tr>
<tr>
<td>24</td>
<td>progat_two_timeat</td>
<td>0</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>25</td>
<td>progat_timee</td>
<td>0</td>
<td>3</td>
<td>0.068</td>
<td>2.145</td>
</tr>
<tr>
<td>26</td>
<td>progat_insat</td>
<td>0</td>
<td>4</td>
<td>0.252</td>
<td>11.410</td>
</tr>
<tr>
<td>27</td>
<td>progat_insat</td>
<td>0</td>
<td>3</td>
<td>0.116</td>
<td>4.183</td>
</tr>
<tr>
<td>28</td>
<td>progat_insat</td>
<td>0</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>29</td>
<td>progat_insat</td>
<td>0</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>30</td>
<td>listat_listat</td>
<td>0</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>31</td>
<td>progat_inf_t</td>
<td>0</td>
<td>3</td>
<td>0.508</td>
<td>15.511</td>
</tr>
<tr>
<td>32</td>
<td>listat_listat</td>
<td>0</td>
<td>no</td>
<td>0.189</td>
<td>1.775</td>
</tr>
<tr>
<td>33</td>
<td>null_listupt</td>
<td>0</td>
<td>no</td>
<td>0.068</td>
<td>1.147</td>
</tr>
<tr>
<td>34</td>
<td>listupt</td>
<td>0</td>
<td>4</td>
<td>0.061</td>
<td>4.178</td>
</tr>
<tr>
<td>35</td>
<td>sorted_e_listat</td>
<td>0</td>
<td>no</td>
<td>no</td>
<td>0.168</td>
</tr>
<tr>
<td>36</td>
<td>timel_listup</td>
<td>0</td>
<td>0</td>
<td>0.049</td>
<td>1.094</td>
</tr>
<tr>
<td>37</td>
<td>sorted_listupt</td>
<td>0</td>
<td>4</td>
<td>0.054</td>
<td>6.206</td>
</tr>
<tr>
<td>38</td>
<td>progat_listupto</td>
<td>0</td>
<td>no</td>
<td>0.539</td>
<td>13.511</td>
</tr>
<tr>
<td>39</td>
<td>leftmax</td>
<td>0</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>40</td>
<td>leftmax_max</td>
<td>0</td>
<td>4</td>
<td>0.104</td>
<td>1.167</td>
</tr>
<tr>
<td>41</td>
<td>right_windo</td>
<td>0</td>
<td>4</td>
<td>0.064</td>
<td>1.108</td>
</tr>
<tr>
<td>42</td>
<td>time_listat</td>
<td>0</td>
<td>1</td>
<td>0.049</td>
<td>1.119</td>
</tr>
<tr>
<td>43</td>
<td>listupt</td>
<td>0</td>
<td>4</td>
<td>0.218</td>
<td>13.46</td>
</tr>
<tr>
<td>44</td>
<td>sorted_cons_listat</td>
<td>0</td>
<td>no</td>
<td>0.110</td>
<td>11.345</td>
</tr>
<tr>
<td>45</td>
<td>nullwind</td>
<td>0</td>
<td>1</td>
<td>0.048</td>
<td>1.135</td>
</tr>
<tr>
<td>46</td>
<td>nullwind</td>
<td>0</td>
<td>1</td>
<td>0.048</td>
<td>1.135</td>
</tr>
<tr>
<td>47</td>
<td>member_t_timeat</td>
<td>0</td>
<td>no</td>
<td>0.074</td>
<td>8.229</td>
</tr>
<tr>
<td>48</td>
<td>timel_insat</td>
<td>0</td>
<td>1</td>
<td>0.040</td>
<td>1.077</td>
</tr>
<tr>
<td>49</td>
<td>null_listupto</td>
<td>0</td>
<td>0</td>
<td>0.030</td>
<td>2.203</td>
</tr>
<tr>
<td>50</td>
<td>sorted_e_cons</td>
<td>0</td>
<td>0</td>
<td>0.034</td>
<td>1.118</td>
</tr>
<tr>
<td>51</td>
<td>erl_cons</td>
<td>0</td>
<td>0</td>
<td>0.036</td>
<td>1.115</td>
</tr>
<tr>
<td>52</td>
<td>no_time</td>
<td>0</td>
<td>2</td>
<td>0.130</td>
<td>7.245</td>
</tr>
<tr>
<td>53</td>
<td>final</td>
<td>0</td>
<td>3</td>
<td>0.042</td>
<td>5.183</td>
</tr>
</tbody>
</table>

The experiments have been performed on a MacBook Pro notebook featuring a 2.6 GHz Intel Core i5 processor and 8 Go RAM.  

IV. CONCLUSIONS AND FUTURE WORK

We detailed the implementation in SPIKE of two reasoning modules integrating components for arithmetic and equality reasoning. The integration schema à la Boyer-Moore was

---

2 This solution was suggested by Pascal Fontaine in a private communication.

3 ‘no’ means that the proof was not successful.
firstly presented in [15] and served later as example for an approach that combined CCR and implicit induction techniques [9]. Compared to [4], it manages neither conditional linear inequalities nor disjunctions of linear inequalities. Moreover, it does not use abstraction variables as in [21], which makes impossible the application of the decision procedure for linear arithmetic on subterms of monomials or non-linearizable equality and disequality sides. On the other hand, it is fast and has been successfully used to automatize the validation proof from [28], representing one of the most challenging case studies ever tested with a reasoning specialist based on the Boyer-Moore’s integration schema. However, the cooperation schema is rather complex and error-prone. Relaunching the proofs using the Z3-based reasoning module can help checking the (implementation) soundness of the cooperation schema. Moreover, the completeness property of the arithmetic component in Z3 is an added-value to SPIKE; we expect to prove more conjectures in a completely automatic way.

In the future, we intend to tweak around Z3 for speeding up its performance when checking for unsatisfiability. It would be interesting to combine induction with reasoning modules for other first-order theories that Z3 integrates (fixed-sized bit-vectors, arrays, etc). We also intend to try other STM solvers compatible with the SMT format (Simplify, CVC3, Yices, etc).

Acknowledgments. The author would like to thank Pascal Fontaine and Christophe Ringeissen for useful discussions. Anisia Maria Magdalena Tudorescu has partially implemented the integration of Z3 in SPIKE during a three-month INRIA internship.

References