

Combining Lists with Non-stably Infinite Theories

Pascal Fontaine, Silvio Ranise, and Calogero G. Zarba

LORIA and INRIA-Lorraine

Abstract. In program verification one has often to reason about lists over elements of a given nature. Thus, it becomes important to be able to combine the theory of lists with a generic theory T modeling the elements. This combination can be achieved using the Nelson-Oppen method *only if* T is stably infinite.

The goal of this paper is to relax the stable-infiniteness requirement. More specifically, we provide a new method that is able to combine the theory of lists with any theory T of the elements, regardless of whether T is stably infinite or not. The crux of our combination method is to guess an arrangement over a set of variables that is larger than the one considered by Nelson and Oppen.

Furthermore, our results entail that it is also possible to combine T with the more general theory of lists with a length function.

1 Introduction

In program verification one has often to decide the validity or satisfiability of logical formulae involving lists over elements of a given nature. For instance, these formulae may involve lists of integers or lists of booleans.

One way to reason about lists over elements of a given nature is to use the Nelson-Oppen method [12] in order to modularly combine a decision procedure for a theory modeling lists with a decision procedure for a theory modeling the elements. This solution requires that the theory of the elements be *stably infinite*. Unfortunately, this requirement is not satisfied by many interesting theories such as, for instance, the theory of booleans and the theory of integers modulo n .

In this paper, we show how to relax the stable infiniteness requirement. More specifically, let T_{list} be the two-sorted theory of lists involving a sort `elem` for elements, a sort `list` for flat lists of elements, plus the symbols `nil`, `car`, `cdr`, and `cons`. For instance, a valid formula in T_{list} is

$$x \approx \text{cdr}(\text{cons}(a, \text{nil})) \rightarrow x \not\approx \text{cons}(b, y).$$

We consider the theory T_{len} that extends T_{list} with a sort `int` for the integers, the symbols `0`, `1`, `+`, `-`, `<` for reasoning over the integers, and a function symbol `length` whose sort is `list` \rightarrow `int`. For instance, a valid formula in T_{len} is

$$x \not\approx \text{cdr}(\text{cons}(a, \text{nil})) \rightarrow \text{length}(x) > 0.$$

We then provide a combination method that is able to combine T_{len} with any theory T_{elem} modeling the elements, regardless of whether T_{elem} is stably infinite or not.

The core ideas of our combination method are:

- modifying the Nelson-Oppen method in such a way to guess an arrangement over an extended set of free constants, and not just the shared ones.
- appropriately computing a certain minimal cardinality k_0 , so that we can ensure that the domain of the elements must have at least k_0 elements.

1.1 Related Work

The importance of reasoning about lists is corroborated by the numerous flavors of theories of lists [1,3,4,13,14,18] present in literature, as well as by the increasing number of tools [6,7,11,15,16,19] containing some capabilities for reasoning about lists.

The idea of guessing an arrangement over a larger sets of free constants was already used by Zarba in order to combine the theory of sets [24] and the theory of multisets [22] with any arbitrary theory T of the elements, regardless of whether T is stably infinite or not. This idea was also used by Fontaine and Gribomont [8] in order to combine the theory of arrays with any other non-necessarily stably infinite theory T .

The idea of computing minimal cardinalities was used by Zarba [23] in order to combine the theory of finite sets with a non-necessarily stably infinite theory T of the elements, in the presence of the cardinality operator. This idea was also exploited by Tinelli and Zarba [20], who provided a method for combining any *shiny* theory S with any non-necessarily stably infinite theory T . Examples of shiny theories include the theory of equality, the theories of partial and total orders, and the theories of lattices with maximum and minimum.

2 Many-Sorted Logic

2.1 Syntax

We fix the following infinite sets: a set **sorts** of sorts, a set **con** of constant symbols, a set **fun** of functions symbols, and a set **pred** of predicate symbols. We also fix an infinite set of variable symbols for every sort in **sorts**.

A *signature* Σ is a tuple $\langle S, C, F, P \rangle$ where $S \subseteq \mathbf{sorts}$, $C \subseteq \mathbf{con}$, $F \subseteq \mathbf{fun}$, $P \subseteq \mathbf{pred}$, all the symbols in C have sorts in S , and all the symbols in F, P have sorts constructed using the sorts in S . If $\Sigma = \langle S, C, F, P \rangle$ is a signature, we sometimes write Σ^S for S , Σ^C for C , Σ^F for F , and Σ^P for P .

If $\Sigma_1 = \langle S_1, C_1, F_1, P_1 \rangle$ and $\Sigma_2 = \langle S_2, C_2, F_2, P_2 \rangle$ are signatures, we write $\Sigma_1 \subseteq \Sigma_2$ when $S_1 \subseteq S_2$, $C_1 \subseteq C_2$, $F_1 \subseteq F_2$, and $P_1 \subseteq P_2$. If $\Sigma_1 = \langle S_1, C_1, F_1, P_1 \rangle$ and $\Sigma_2 = \langle S_2, C_2, F_2, P_2 \rangle$ are signatures, their *union* is the signature $\Sigma_1 \cup \Sigma_2 = \langle S_1 \cup S_2, C_1 \cup C_2, F_1 \cup F_2, P_1 \cup P_2 \rangle$.

Given a signature Σ and a set of variables, we assume the standard notions of Σ -term, Σ -atom, Σ -literal, Σ -formula. If φ is either a term or a formula, we denote by $\text{vars}_\sigma(\varphi)$ the set of variables of sort σ occurring in φ .

In the rest of this paper we identify a conjunction of formulae $\varphi_1 \wedge \cdots \wedge \varphi_n$ with the set $\{\varphi_1, \dots, \varphi_n\}$. In addition, we abbreviate literals of the form $\neg(s \approx t)$ with $s \not\approx t$.

2.2 Semantics

Definition 1. If Σ is a signature, a Σ -INTERPRETATION \mathcal{A} over a set of variables V is a map which interprets:¹

- each sort $\sigma \in \Sigma^S$ as a non-empty domain A_σ ;
- each variable $x \in V$ of sort σ as an element $x^{\mathcal{A}} \in A_\sigma$;
- each constant symbol $c \in \Sigma^C$ of sort σ as an element $c^{\mathcal{A}} \in A_\sigma$;
- each function symbol $f \in \Sigma^F$ of sort $\sigma_1 \times \cdots \times \sigma_n \rightarrow \tau$ as a function $f^{\mathcal{A}} : A_{\sigma_1} \times \cdots \times A_{\sigma_n} \rightarrow A_\tau$;
- each predicate symbol $p \in \Sigma^P$ of sort $\sigma_1 \times \cdots \times \sigma_n$ as a subset $P^{\mathcal{A}}$ of $A_{\sigma_1} \times \cdots \times A_{\sigma_n}$. □

A Σ -formula is *satisfiable* if it evaluates to true under some Σ -interpretation.

Let \mathcal{A} be an Σ -interpretation over the set of variables V , and let $\Sigma' \subseteq \Sigma$ and $V' \subseteq V$. We denote by $\mathcal{A}^{\Sigma', V'}$ the interpretation obtained from \mathcal{A} by restricting it to interpret only the symbols in Σ' and variables in V' . For convenience, $\mathcal{A}^{\Sigma'}$ also denotes $\mathcal{A}^{\Sigma', V}$.

A Σ -*structure* is a Σ -interpretation over an empty set of variables.

2.3 Theories

Following Ganzinger [9], we define theories as sets of structures rather than as sets of formulas. More formally:

Definition 2. A Σ -THEORY is a pair $\langle \Sigma, \mathbf{A} \rangle$ where Σ is a signature and \mathbf{A} is a set of Σ -structures. □

Definition 3. Let T be a Σ -theory, and let $\Omega \subseteq \Sigma$. An Ω -interpretation \mathcal{A} is a T -INTERPRETATION if $\mathcal{A}^{\Sigma, \emptyset} \in T$. □

A formula is *T-satisfiable* if it evaluates to true under some T -interpretation.

Given a Σ -theory T , the *ground satisfiability problem* of T is the problem of deciding, for each ground Σ -formula φ , whether or not φ is T -satisfiable.

Definition 4. Let Σ be a signature, let $S \subseteq \Sigma^S$ be a nonempty set of sorts, and let T be a Σ -theory. We say that T is STABLY INFINITE with respect to S if every ground Σ -formula φ is T -satisfiable if and only if there exists a T -interpretation satisfying φ such that A_σ is infinite, for each sort $\sigma \in S$. □

¹ Unless otherwise specified, we use the convention that calligraphic letters denote interpretations, and that the corresponding Roman letters, appropriately subscripted, denote the domains of the interpretations.

Definition 5 (Combination of theories). Let $T_i = \langle \Sigma_i, \mathbf{A}_i \rangle$ be a theory, for $i = 1, 2$. The COMBINATION of T_1 and T_2 is the theory $\text{comb}(T_1, T_2) = \langle \Sigma, \mathbf{A} \rangle$ where $\Sigma = \Sigma_1 \cup \Sigma_2$ and $\mathbf{A} = \{ \mathcal{A} \mid \mathcal{A}^{\Sigma_1} \in \mathbf{A}_1 \text{ and } \mathcal{A}^{\Sigma_2} \in \mathbf{A}_2 \}$. \square

2.4 The Theory of Integers

Let us fix a signature Σ_{int} containing a sort `int` for the integers, plus the constant symbols `0` and `1` of sort `int`, the function symbols `+` and `-` of sort `int` \times `int` \rightarrow `int`, and the predicate symbol `<`, of sort `int` \times `int`.

Definition 6. The STANDARD `int`-STRUCTURE is the Σ_{int} -structure \mathcal{A} specified by letting $A_{\text{int}} = \mathbb{Z}$ and interpreting the symbols `0`, `1`, `+`, `-`, `<` according to their intuitive meaning over \mathbb{Z} . \square

Definition 7. The THEORY OF INTEGERS is the pair $T_{\text{int}} = \langle \Sigma_{\text{int}}, \{ \mathcal{A} \} \rangle$, where \mathcal{A} is the standard `int`-structure. \square

The ground satisfiability problem of T_{int} can be decided by using methods based on integer automata [21], the omega test [2,17], or appropriate extensions of the Fourier-Motzkin method [10].

2.5 Lists

Let A be a non-empty set, and assume that the special object \perp does not belong to A .² A *list* x over A of *length* n is a map $x : \mathbb{N} \rightarrow A \cup \{ \perp \}$ such that $x(i) \in A$, for $i < n$, and $x(i) = \perp$, for $i \geq n$. We write $|x| = n$ to indicate that the length of the list x is n . We denote by A^* the set of lists over A .

We denote by *nil* the empty list, that is, $\text{nil}(i) = \perp$, for each $i \in \mathbb{N}$. We denote by *car* and *cons* the partial functions defined as follows: given a list $x \neq \text{nil}$, we let $\text{car}(x) = x(0)$, whereas $\text{cdr}(x)$ is the unique list y such that $y(n) = x(n+1)$, for each $n \in \mathbb{N}$.

Given an element $e \in A$ and a list x in A^* , we denote by $\text{cons}(e, x)$ the list y such that $y(0) = e$, and $y(n+1) = x(n)$, for each $n \in \mathbb{N}$.

2.6 The Theory of Lists

We fix a signature Σ_{list} containing a sort `elem` for elements and a sort `list` for lists of elements, plus the constant symbol \perp_{elem} of sort `elem`, the constant symbols *nil* and \perp_{list} of sort `list`, the function symbols *car* of sort `list` \rightarrow `elem`, the function symbol *cdr* of sort `list` \rightarrow `list`, and the function symbol *cons* of sort `elem` \times `list` \rightarrow `list`.

Definition 8. A STANDARD `list`-STRUCTURE \mathcal{A} is a Σ_{list} -structure satisfying the following conditions:

² Using this special object \perp to define lists is not fundamental but it is convenient for the following.

- $\perp \notin A_{\text{elem}}$;
- $A_{\text{list}} = (A_{\text{elem}})^*$;
- $\text{nil}^A = \text{nil}$;
- $\text{car}^A(\text{nil}) = (\perp_{\text{elem}})^A$;
- $\text{cdr}^A(\text{nil}) = (\perp_{\text{list}})^A$;
- $\text{car}^A(x) = \text{car}(x)$, for each $x \in A_{\text{list}}$ such that $x \neq \text{nil}$;
- $\text{cdr}^A(x) = \text{cdr}(x)$, for each $x \in A_{\text{list}}$ such that $x \neq \text{nil}$;
- $\text{cons}^A(e, x) = \text{cons}(e, x)$, for each $e \in A_{\text{elem}}$ and $x \in A_{\text{list}}$. □

Note that although car and cdr are partial functions, standard list-structures interpret the symbols car and cdr as total functions. In particular, all standard list-structures ensure that the constants \perp_{elem} and \perp_{list} have the same interpretations of the terms $\text{car}(\text{nil})$ and $\text{cdr}(\text{nil})$, respectively. However \perp_{elem} and \perp_{list} may be interpreted by any element and list in the respective domain. There are thus many standard list-structures.

Definition 9. The THEORY OF LISTS is the pair $T_{\text{list}} = \langle \Sigma_{\text{list}}, \mathbf{A} \rangle$, where \mathbf{A} is the set of all standard list-structures. □

As a by product of the results of this paper, we will see that the ground satisfiability problem of T_{list} can be decided by appropriately adapting Oppen’s decision procedure for a one-sorted theory of lists without nil [14].

2.7 The Theory of Lists with a Length Function

We fix a signature Σ_{len} containing all the symbols in Σ_{int} and Σ_{list} , plus the function symbol length of sort $\text{list} \rightarrow \text{int}$.

Definition 10. A STANDARD len-STRUCTURE \mathcal{A} is a Σ_{len} -structure satisfying the following conditions:

- $\mathcal{A}^{\Sigma_{\text{int}}}$ is the standard int-structure;
- $\mathcal{A}^{\Sigma_{\text{list}}}$ is a standard list-structure;
- $\text{length}^{\mathcal{A}}(x) = |x|$, for each $x \in A_{\text{list}}$. □

Definition 11. The THEORY OF LISTS WITH A LENGTH FUNCTION is the pair $T_{\text{len}} = \langle \Sigma_{\text{len}}, \mathbf{A} \rangle$, where \mathbf{A} is the set of all standard len-structures. □

The ground satisfiability problem of T_{len} can be decided by appropriately adapting a decision procedure for a two-sorted theory of recursively defined data structures with integer constraints [25].

3 The Combination Method

Let Σ_{elem} be a signature such that $\Sigma^{\text{S}} = \{\text{elem}\}$, and let T_{elem} be any Σ_{elem} -theory, not necessarily stably infinite with respect to the sort elem . Assume that the ground satisfiability problem of T_{elem} is decidable. We now describe a

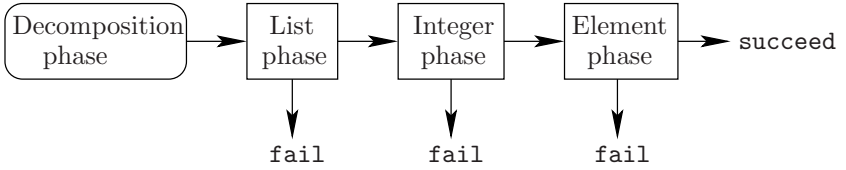


Fig. 1. The phases of our combination method

combination-based decision procedure for the ground satisfiability problem of $T = \text{comb}(T_{\text{elem}}, T_{\text{len}})$.

In our combination method we use as black boxes a decision procedure for the ground satisfiability problem of T_{elem} and a decision procedure for the ground satisfiability problem of T_{int} . We also use—albeit not strictly as a black box—Oppen’s decision procedure for recursively defined data structures.

Without loss of generality, we restrict ourselves to conjunctions Γ of literals in *separate* form: $\Gamma = \Gamma_{\text{elem}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{length}}$ where:

- (a) Γ_{elem} contains only Σ_{elem} -literals;
- (b) Γ_{int} contains only Σ_{int} -literals;
- (c) Γ_{list} contains only flat Σ_{list} -literals of the form

$$\begin{array}{lll}
 x \approx y, & x \not\approx y, & x \approx \text{nil}, \\
 e \approx \perp_{\text{elem}}, & x \approx \perp_{\text{list}}, & x \approx \text{cons}(e, y),
 \end{array}$$

where e_1, e_2, e are elem-variables and x, y are list-variables;

- (d) Γ_{length} contains only literals of the form $u \approx \text{length}(x)$ where u is an int-variable and x is a list-variable;
- (e) for each list-variable $x \in \text{vars}_{\text{list}}(\Gamma)$, either $x \approx \text{nil}$ or $x \not\approx \text{nil}$ is in Γ_{list} .

Notice that, given a set of literals in T , it is easy to build an equisatisfiable separation verifying (a),(b),(d) the usual way [12] by introducing fresh variables. However to furthermore ensure (c) and (e), and in particular to eliminate all occurrences of `car` and `cdr`, it is necessary to include disjunctions to the set of literals. For efficiency concerns, this transformation is done at the formula level; it is described in Section 5.

Our combination method consists of the four phases depicted in Figure 1, and described below.

3.1 Decomposition Phase

Let $\Gamma = \Gamma_{\text{elem}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{length}}$ be a conjunction of literals in separate form. Also let $V_{\text{elem}} = \text{vars}_{\text{elem}}(\Gamma_{\text{list}}) \cup \{\perp_{\text{elem}}\}$ and $V_{\text{list}} = \text{vars}_{\text{list}}(\Gamma)$. In the decomposition phase we non-deterministically guess an equivalence relation \sim_{elem} of V_{elem} , and we construct the following set of literals:

$$\alpha_{\text{elem}} = \{e_1 \approx e_2 \mid e_1 \sim_{\text{elem}} e_2\} \cup \{e_1 \not\approx e_2 \mid e_1, \not\sim_{\text{elem}} e_2\}.$$

Note that our decomposition phase differs from the one of Nelson-Oppen method. In fact, in the Nelson-Oppen method one guesses an equivalence relation over the smaller set of variables $vars_{\text{elem}}(\Gamma_{\text{elem}}) \cap vars_{\text{elem}}(\Gamma_{\text{list}})$. We need to use the larger set V_{elem} because we do not have any stable infiniteness assumption over the theory T_{elem} of the elements.

3.2 List Phase

In the list phase we essentially employ Oppen's decision procedure for recursively defined data structures. By not using Oppen's procedure just as a black box, we will later be able to use the information constructed in this phase in the later phases of our method. (Cf. Section 5.)

More in detail, in the list phase we construct the least equivalence relation \sim_{list} of V_{list} satisfying the following conditions:

- (a) if $x \approx y$ is in Γ_{list} then $x \sim_{\text{list}} y$;
- (b) if $x_1 \approx \text{cons}(e_1, y_1)$ and $x_2 \approx \text{cons}(e_2, y_2)$ are in Γ_{list} , and $e_1 \sim_{\text{elem}} e_2$ and $y_1 \sim_{\text{list}} y_2$ then $x_1 \sim_{\text{list}} x_2$;
- (c) if $x_1 \approx \text{cons}(e_1, y_1)$ and $x_2 \approx \text{cons}(e_2, y_2)$ are in Γ_{list} , and $x_1 \sim_{\text{list}} x_2$ then $e_1 \sim_{\text{elem}} e_2$ and $y_1 \sim_{\text{list}} y_2$.

Furthermore, we construct the relation \prec_{list} of V_{list} defined by letting $x \prec_{\text{list}} y$ if and only if there are list-variables $x', y' \in V_{\text{list}}$ and an elem-variable $e \in V_{\text{elem}}$ such that $x \sim_{\text{list}} x'$, $y \sim_{\text{list}} y'$, and the literal $y' \approx \text{cons}(e, x')$ is in Γ_{list} .

We end our method by outputting `fail` if at least one of the following conditions does not hold:

- (C1) If $x \sim_{\text{list}} y$ then the literal $x \not\approx y$ is not in Γ_{list} ;
- (C2) There are no two literals $x \approx \text{nil}$ and $y \approx \text{cons}(e, z)$ in Γ_{list} for which $x \sim_{\text{list}} y$;
- (C3) The relation \prec_{list} is well-founded.

If instead all conditions (C1)–(C3) hold, we proceed to the next phase.

3.3 Integer Phase

In this phase we extract integer constraints from the conjunctions Γ_{list} and Γ_{length} , as well as from the equivalence relation \sim_{list} constructed in the list phase.

More in detail, we generate a fresh int-variable u_x , for each list-variable x in V_{list} , and we construct the following set of literals

$$\begin{aligned} \alpha_{\text{int}} = & \{u_x \approx 0 \mid x \approx \text{nil} \text{ is in } \Gamma_{\text{list}}\} \cup \\ & \{u_x > 0 \mid x \not\approx \text{nil} \text{ is in } \Gamma_{\text{list}}\} \cup \\ & \{u_x = u_y + 1 \mid x \approx \text{cons}(e, y) \text{ is in } \Gamma_{\text{list}}\} \cup \\ & \{u \approx u_x \mid u \approx \text{length}(x) \text{ is in } \Gamma_{\text{length}}\} \cup \\ & \{u_x \approx u_y \mid x \sim_{\text{list}} y\}. \end{aligned}$$

Then, we check whether $\Gamma_{\text{int}} \cup \alpha_{\text{int}}$ is T_{int} -satisfiable. If this is not the case, we end our method by outputting `fail`; otherwise we proceed to the next phase.

3.4 Element Phase

We will prove later that when we reach this point we can already conclude that $\alpha_{\text{elem}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{length}}$ is T_{len} -satisfiable.³ Therefore, we can effectively compute the minimal integer k_0 for which there exists a T_{len} -interpretation \mathcal{A} satisfying $\alpha_{\text{elem}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{length}}$ such that $k_0 = |A_{\text{elem}}|$.⁴

Let $\{|\text{elem}| \geq k_0\}$ denotes the set of disequalities $\{e_i \not\approx e_j \mid 1 \leq i < j \leq k_0\}$, where the e_i are fresh **elem**-variables. The last step of the element phase consists of checking whether $\Gamma_{\text{elem}} \cup \alpha_{\text{elem}} \cup \{|\text{elem}| \geq k_0\}$ is T_{elem} -satisfiable. If this is not the case, we end the method by outputting **fail**; otherwise we happily output **succeed**.

4 Correctness

In this section we prove that our combination method is correct. Clearly, our method is terminating. The following proposition shows that our method is also partially correct.

Proposition 12. *Let T_{elem} be a Σ_{elem} -theory such that $\Sigma^S = \{\text{elem}\}$, let $T = \text{comb}(T_{\text{elem}}, T_{\text{len}})$, and let $\Gamma = \Gamma_{\text{elem}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{length}}$ be a conjunction of literals in separate form. Then the following are equivalent:*

1. Γ is T -satisfiable.
2. There exists an equivalence relation \sim_{elem} of $\text{vars}_{\text{elem}}(\Gamma_{\text{list}}) \cup \{\perp_{\text{elem}}\}$ for which our method outputs **succeed**. □

PROOF. Remember that $V_{\text{elem}} = \text{vars}_{\text{elem}}(\Gamma_{\text{list}}) \cup \{\perp_{\text{elem}}\}$ and $V_{\text{list}} = \text{vars}_{\text{list}}(\Gamma)$. ($1 \Rightarrow 2$). Let \mathcal{M} be a T -interpretation satisfying Γ . We define an equivalence relation \sim_{elem} over V_{elem} by letting

$$e_1 \sim_{\text{elem}} e_2 \iff e_1^{\mathcal{M}} = e_2^{\mathcal{M}}, \quad \text{for each } e_1, e_2 \in V_{\text{elem}}.$$

We claim that if we guess \sim_{elem} as defined above, then our method outputs **succeed**. To see this, let \sim_{list} be the equivalence relation constructed in the list phase, and let \equiv_{list} be the equivalence relation of V_{list} defined as follows:

$$x \equiv_{\text{list}} y \iff x^{\mathcal{M}} = y^{\mathcal{M}}, \quad \text{for each } x, y \in V_{\text{list}}.$$

By construction \equiv_{list} satisfies conditions (a)–(c) in the list phase. Therefore, we have $\sim_{\text{list}} \subseteq \equiv_{\text{list}}$, that is:

$$x \sim_{\text{list}} y \implies x \equiv_{\text{list}} y, \quad \text{for each } x, y \in V_{\text{list}}.$$

³ A T_{len} -interpretation satisfying $\alpha_{\text{elem}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{length}}$ is denoted by \mathcal{C} in the second part of the proof of Proposition 12.

⁴ One way of computing k_0 is to use [25] to check, for increasing k , whether there exists a T_{len} -interpretation \mathcal{A} satisfying $\alpha_{\text{elem}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{length}}$ such that $|A_{\text{elem}}| = k$.

By using the fact that $\sim_{\text{list}} \subseteq \equiv_{\text{list}}$, one can verify that \sim_{list} satisfies all conditions (C1)–(C3) of the list phase. Therefore, our method does not output **fail** when executing the list phase.

Next, we claim that our method also does not output **fail** when executing the integer phase. To justify the claim, we need to show that $\Gamma_{\text{int}} \cup \alpha_{\text{int}}$ is T_{int} -satisfiable. Indeed, by again using the fact that $\sim_{\text{list}} \subseteq \equiv_{\text{list}}$, it is possible to verify that a T_{int} -interpretation satisfying $\Gamma_{\text{int}} \cup \alpha_{\text{int}}$ can be obtained by extending \mathcal{M} to the variables u_x by letting

$$u_x^{\mathcal{M}} = |x^{\mathcal{M}}|, \quad \text{for each list-variable } x \in V_{\text{list}}.$$

It remains to show that our method outputs **succeed** when executing the element phase. To see this, let k_0 be the minimal integer computed in the element phase. By construction, \mathcal{M} satisfies $\Gamma_{\text{elem}} \cup \alpha_{\text{elem}}$. Moreover, since \mathcal{M} satisfies $\alpha_{\text{elem}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{length}}$, it must have at least k_0 elements. It follows that \mathcal{M} is a T_{elem} -interpretation satisfying $\Gamma_{\text{elem}} \cup \alpha_{\text{elem}} \cup \{|\text{elem}| \geq k_0\}$.

(2 \Rightarrow 1). Let \sim_{elem} be an equivalence relation of V_{elem} for which our method outputs **succeed**. Denote by \sim_{list} and \prec_{list} the relations of V_{list} constructed in the list phase, and denote by k_0 the minimal integer computed in the element phase. Next, note that there exists an interpretation \mathcal{A} satisfying $\Gamma_{\text{elem}} \cup \alpha_{\text{elem}} \cup \{|\text{elem}| \geq k_0\}$ and a T_{int} -interpretation \mathcal{B} satisfying $\Gamma_{\text{int}} \cup \alpha_{\text{int}}$.

Using \mathcal{A} and \mathcal{B} , we define a T_{len} -interpretation \mathcal{C} satisfying $\alpha_{\text{elem}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{length}}$ by first letting $C_{\text{elem}} = A_{\text{elem}} \cup X$, where X is any infinite set disjoint from A_{elem} . We also let:

$$\begin{aligned} e^{\mathcal{C}} &= e^{\mathcal{A}}, & \text{for all } e \in \text{vars}_{\text{elem}}(\Gamma), \\ u^{\mathcal{C}} &= u^{\mathcal{B}}, & \text{for all } u \in \text{vars}_{\text{int}}(\Gamma). \end{aligned}$$

In order to define \mathcal{C} over the list-variables in V_{list} , we fix an injective function $h : (V_{\text{list}} / \sim_{\text{list}}) \rightarrow X$. Note that h exists because V_{list} is finite and X is infinite.

Next, we proceed by induction on the well-founded relation \prec_{list} . Thus, let $x \in V_{\text{list}}$. Then:

- In the **base case**, we let $x^{\mathcal{C}}$ be the unique list of length $u_x^{\mathcal{B}}$ containing only the element $h([x]_{\sim_{\text{list}}})$. In other words, $x^{\mathcal{C}}(i) = h([x]_{\sim_{\text{list}}})$ for $i < u_x^{\mathcal{B}}$, and $x^{\mathcal{C}}(i) = \perp$ for $i \geq u_x^{\mathcal{B}}$.
- In the **inductive case**, fix a list-variable y such that $x \prec_{\text{list}} y$. Then there exists variables x', y', e such that $x \sim_{\text{list}} x'$, $y \sim_{\text{list}} y'$, and the literal $x' \approx \text{cons}(e, y')$ is in Γ_{list} . We let $x^{\mathcal{C}} = \text{cons}(e^{\mathcal{C}}, (y')^{\mathcal{C}})$.

Note that \mathcal{C} is well-defined over the list-variables. Furthermore, by construction \mathcal{C} is a T_{len} -interpretation satisfying $\alpha_{\text{elem}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{length}}$.

It follows that there exists a T_{len} -interpretation \mathcal{D} satisfying $\alpha_{\text{elem}} \cup \Gamma_{\text{int}} \cup \Gamma_{\text{list}} \cup \Gamma_{\text{length}}$ and such that $|D_{\text{elem}}| = k_0$. But then, we can use \mathcal{D} and \mathcal{A} to obtain

```

1:  $\varphi := preprocess(\varphi)$ 
2:  $\varphi^a \leftarrow abs(\varphi)$ 
3: while  $\varphi^a \neq false$  do
4:    $\Gamma^a \leftarrow pick\_assign(\varphi^a)$ 
5:    $\Gamma \leftarrow prop2fol(\Gamma^a)$ 
6:    $(\rho, \pi) \leftarrow check\_sat(\Gamma)$ 
7:   if  $\rho = fail$  then
8:      $\varphi^a \leftarrow \varphi^a \wedge \neg fol2prop(\pi)$ 
9:   else
10:    return succeed
11:   end if
12: end while

```

Fig. 2. **haRVey**'s main loop

a T -interpretation \mathcal{M} satisfying Γ by letting $M_{\text{elem}} = A_{\text{elem}}$ and

$$\begin{aligned}
e^{\mathcal{M}} &= e^{\mathcal{A}}, & \text{for all } e \in \Sigma_{\text{elem}}^{\text{C}} \cup vars_{\text{elem}}(\Gamma), \\
f^{\mathcal{M}} &= f^{\mathcal{A}}, & \text{for all } f \in \Sigma_{\text{elem}}^{\text{F}}, \\
p^{\mathcal{M}} &= p^{\mathcal{A}}, & \text{for all } p \in \Sigma_{\text{elem}}^{\text{P}}, \\
u^{\mathcal{M}} &= u^{\mathcal{D}}, & \text{for all } u \in vars_{\text{int}}(\Gamma).
\end{aligned}$$

In order to define \mathcal{M} over the list-variables, fix an injective function $g : D_{\text{elem}} \rightarrow A_{\text{elem}}$. For convenience, also let $g(\perp) = \perp$. Note that g exists because $|D_{\text{elem}}| = k_0 \leq |A_{\text{elem}}|$. We let:

$$x^{\mathcal{M}}(i) = g(x^{\mathcal{D}}(i)), \quad \text{for all } x \in vars_{\text{list}}(\Gamma) \text{ and } i \in \mathbb{N}.$$

By construction, \mathcal{M} is a T -interpretation satisfying Γ . ■

From Proposition 12 and the fact that our combination method is terminating, we obtain the following decidability result.

Theorem 13 (Decidability). *Let T_{elem} be a Σ_{elem} -theory such that the ground satisfiability problem is decidable. Then the ground satisfiability problem of the theory $comb(T_{\text{elem}}, T_{\text{len}})$ is decidable.* □

5 Using the Combination Method

In this Section, we describe how to lift the proposed combination method to efficiently (at least in practice) handle arbitrary Boolean combinations of ground literals. The method is a refinement of the main loop of **haRVey** [6] (cf. Figure 2), a prover based on a combination of Boolean solving and satisfiability checking modulo theories. The idea is to obtain a propositional abstraction φ^a of a formula φ (cf. *abs*) and to enumerate all the propositional assignments (cf. *pick_assign*).

If an assignment, refined to a conjunction of first-order literals (cf. *prop2fol*), is found satisfiable modulo the background theory (cf. *check_sat* returns with $\rho = \text{fail}$), then we are entitled to conclude the satisfiability of φ . Otherwise, a new assignment is considered. For efficiency, it is crucial to reduce the number of invocations to *check_sat*. To this end, it is required that *check_sat* returns a conflict set π (which is a subset of the input set of literals) so that all the propositional assignments sharing that set can be eliminated in one shot.⁵

We now give some details of the implementation of the functionalities in Figure 2 which are peculiar to using the combination method in Section 3. In particular, we describe how to satisfy the requirements necessary for the method to work correctly (see beginning of Section 3) and, most importantly, we explain how to compute the \sim_{list} and \prec_{list} of Section 3.2.

Function preprocess. A *flat atom* is an atom of the form $p(c_1, \dots, c_n)$, $c \approx f(c_1, \dots, c_m)$, $c_1 \approx c_2$ or $c_1 \approx d$, where p is n -ary predicate symbol ($n \geq 0$), f is an m -ary function symbol ($m > 0$), c_i is an element of **par**, and d is a constant. A *flat literal* is either a flat atom or the negation of a flat atom of one of the two forms $\neg p(c_1, \dots, c_n)$ or $c_1 \not\approx c_2$. A formula is said to be flattened if all its literals are flat. It is easy to get an equisatisfiable flattened formula from any ground formula by introducing fresh variables to name subterms.

The preprocessing step also removes all occurrences of **car** and **cdr** in the formula using the following equivalences

$$\begin{aligned} e \approx \text{car}(x) &\equiv (x \approx \text{nil} \wedge e \approx \perp_{\text{elem}}) \vee (x \not\approx \text{nil} \wedge (\exists_{\text{list}} y)(x \approx \text{cons}(e, y))) \\ x \approx \text{cdr}(y) &\equiv (y \approx \text{nil} \wedge x \approx \perp_{\text{list}}) \vee (y \not\approx \text{nil} \wedge (\exists_{\text{elem}} e)(y \approx \text{cons}(e, x))) \end{aligned}$$

For instance, $\varphi[a \approx \text{car}(x)]$ is equisatisfiable to $\varphi[a \approx e] \wedge e \approx \text{car}(x)$. In this last formula, the atom $e \approx \text{car}(x)$ has always positive polarity. In a later step, it can be replaced by $(x \approx \text{nil} \wedge e \approx \perp_{\text{elem}}) \vee (x \not\approx \text{nil} \wedge (\exists_{\text{list}} y)(x \approx \text{cons}(e, y)))$ and since the polarity is positive, the existential quantifier can be Skolemized by simply introducing a fresh variable. Exhaustively applying this transformation gives a new ground formula, without **car** and **cdr**.

Finally, and still by introducing fresh variables, functions **cons** and **length** are made to appear only in unit clauses of the form $\text{cons}(e, x) \approx y$ or $\text{length}(x) \approx u$. For instance formula $\varphi[\text{cons}(e, x) \not\approx y]$ is replaced by $\varphi[y' \not\approx y] \wedge y' \approx \text{cons}(e, x)$.

Function pick_assign. The function *pick_assign* is implemented by the Boolean solver and returns a propositional assignment satisfying φ^a . It is easy to tune the solver to make *pick_assign* return a propositional assignment Γ^a such that *prop2fol*(Γ^a) contains the literals representing the fact that each list variable is equal to **nil** or **not**.

⁵ Best results are obtained in practice when this set is chosen to be minimal: an unsatisfiable set such that each subset is satisfiable.

Function check_sat. First of all, we notice that, thanks to *preprocess*, the function *pick_assign* returns a set Γ of literals which can be put in separate form satisfying conditions (a)–(e) at the beginning of Section 3 by simply partitioning the literals.

Our combination method uses decision procedures for the quantifier-free fragment of arithmetic and for the theory of acyclic lists. While we use a decision procedure for the first theory as a black box, we require the decision procedure for the theory of acyclic lists to be able to return \sim_{list} and \prec_{list} . For this reason, we detail below how to do this.

Reasoning About Acyclic Lists

We introduce a graph structure encapsulating all constraints on the T_{list} -models of a set of equalities of the form $x \approx y$, $e \approx e'$, $x \approx \text{cons}(e, y)$, where x, y are list-variables, and e, e' are elem-variables. In fact, this structure is implicitly computed by the algorithm described in [14]. We here make it explicit, and explain how to extract relations \sim_{list} and \prec_{list} from it. The structure may also be used in order to guide the guessing in Section 3.1.

From now on, if not otherwise specified, nil is treated as any other variable. An equality $x \approx \text{nil}$ can thus be seen as an equality between two different list variables. Given finite sets of list and element variables, a list-graph is a tuple $\langle V_{\text{list}}, V_{\text{elem}}, s_{\text{list}}, s_{\text{elem}} \rangle$ with

- V_{list} (V_{elem}) is a partition of list (resp. element) variables. It is the set of list (resp. element) nodes. Variables in a node are *labels* for that node;
- s_{list} (s_{elem}) is a function from V_{list} to subsets of V_{list} (resp. V_{elem}). Given a list node u , $s_{\text{list}}(u)$ ($s_{\text{elem}}(u)$) is the set of list (resp. element) successors of u .

A T_{list} -interpretation \mathcal{A} agrees with a list-graph if the following conditions are met:

- if x and y label the same node then $\mathcal{A} \models x \approx y$, where x and y are both element variables or both list variables;
- if y labels the list successor of x then $\mathcal{A} \models \exists e x \approx \text{cons}(e, y)$;
- if e labels the element successor of x then $\mathcal{A} \models \exists y x \approx \text{cons}(e, y)$.

Assume L is a T_{list} -satisfiable set of equalities of the form $x \approx y$, $e \approx e'$, $x \approx \text{cons}(e, y)$. Then there is a list-graph G such that, for every T_{list} -interpretation \mathcal{A} , \mathcal{A} agrees with G if and only if \mathcal{A} is a model of L . Indeed, the following graph verifies this property:

- x and y label the same node if and only if $L \models_{\text{list}} x \approx y$,⁶ where x and y are both element variables or both list variables;
- y labels the list successor of x if and only if $L \models_{\text{list}} \exists e x \approx \text{cons}(e, y)$;
- e labels the element successor of x if and only if $L \models_{\text{list}} \exists y x \approx \text{cons}(e, y)$.

⁶ \models_{list} denotes logical consequence in the theory of lists. That is $L \models_{\text{list}} x \approx y$ if every T_{list} -model of L is a model of $x \approx y$.

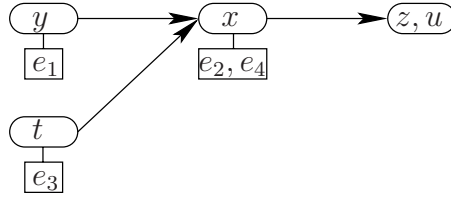


Fig. 3. example of canonical list-graph

This graph is unique. It is such that, for each $v \in V_{\text{list}}$, $s_{\text{list}}(v)$ and $s_{\text{elem}}(v)$ are either a singleton or the empty set. In other words, every list node has at most one list successor, and one element successor. In fact, it can be showed that every node has two or zero successor, since the `cdr` and `car` functions are not explicitly used in the set of equalities. If `nil` labels a list-node, then this node has no list successors. It is acyclic in the sense that s_{list} is acyclic. Finally, for each $u, v \in V_{\text{list}}$, if $s_{\text{list}}(u) = s_{\text{list}}(v)$, $s_{\text{list}}(u) \neq \emptyset$, $s_{\text{elem}}(u) = s_{\text{elem}}(v)$, and $s_{\text{elem}}(u) \neq \emptyset$, then $u = v$. In other words, two different list nodes must not have the same list and element successors.

This graph will thus be called the *canonical list-graph* for a set of equalities. For instance, the canonical list-graph for the set of equalities

$$y \approx \text{cons}(e_1, x), x \approx \text{cons}(e_2, z), x \approx \text{cons}(e_4, u), t \approx \text{cons}(e_3, x)$$

is given in Figure 3.

Given the canonical list-graph for a set of equalities, we have that $x \sim_{\text{list}} y$ is true if and only if x and y both label the same list node and \prec_{list} is the transitive closure of the list successor relation.

Computing Canonical list-Graphs

To compute the canonical graph for a set of equalities, three transformations on list-graphs are necessary:

- a congruence step replaces two lists nodes u and v such that $s_{\text{list}}(u) = s_{\text{list}}(v)$ and $s_{\text{elem}}(u) = s_{\text{elem}}(v)$ by a unique node $u \cup v$.⁷ The new node inherits all successors of the nodes it replaces. All list nodes which had u or v as list successor are made to have $u \cup v$ as list successor.
- a list unification step (Unify-`cdr`) replaces two list successors u and v of one node t by a unique node $u \cup v$. The new node inherits all successors of the nodes it replaces. All list nodes which had u or v as list successor are made to have $u \cup v$ as list successor.
- an element unification step (Unify-`car`) replaces two element successors u and v of one node t by a unique node $u \cup v$. All list nodes which had u or v as element successor are made to have $u \cup v$ as list successor.

⁷ Remember u and v are disjoint sets of list variables.

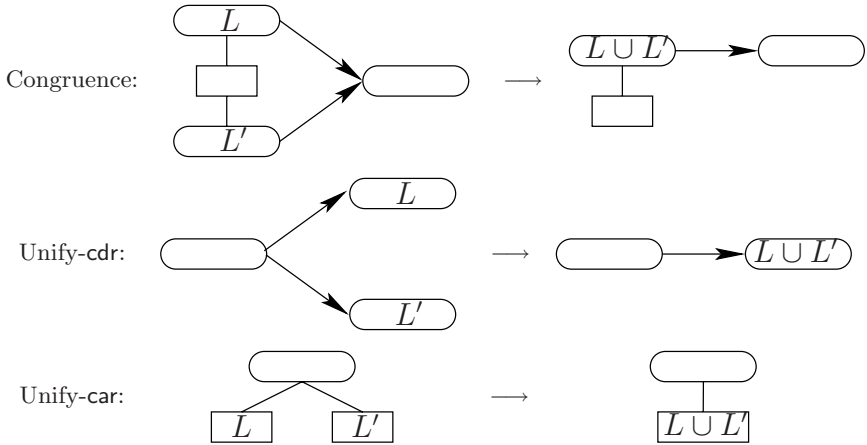


Fig. 4. Transformation steps

These transformations are depicted in Figure 4.

Let L be a set of equalities of the form $x \approx y$, $e \approx e'$, $x \approx \text{cons}(e, y)$. To build the canonical graph for this set, the first operation is to compute the reflexive, symmetric and transitive closure of all equalities between variables in the set L . Second, for every equality $\text{cons}(e, x) \approx y$, the nodes labeled by x and e are made list and element successors of the node labeled by y . Third, the graph is unified, beginning with nodes without parent, finishing with those without successor, using unification steps (beginning with all element unification steps). Last, the congruence rule is applied, from the nodes without successors, to the nodes without parents. In presence of nil , a postprocessing ensures that the node it labels has no successor.

If the graph happens to be cyclic, or if nil happens to have a successor, the procedure fails. In that case the initial set of equalities is unsatisfiable. A careful implementation of this procedure is linear in time [14].

The obtained graph (after a finite number of transformation steps) is indeed the canonical graph: every T_{list} -interpretation \mathcal{A} agreeing with a graph G also agrees with the graph obtained from G by a transformation step. That ensures that every model of L agrees with the final graph. To show that every T_{list} -interpretation agreeing with the graph is also a model for L , it suffices to show that every equality of L is trivially satisfied by any interpretation agreeing with the graph.

There is a T_{list} -interpretation agreeing with a canonical list-graph, such that every node is assigned to a different element or list. As a consequence, satisfiability checking of a set of literals in T_{list} can be simply implemented by building the canonical list-graph for all equalities in the set, and check afterward if no inequality has both members labeling the same node.

Two final remarks are in order. First, the list-graph may be build before guessing an arrangement of the element variables, and may be used to guide this guessing. Indeed it is not necessary to consider an α_{elem} implying that two variables labeling the same node in the list-graph are different. Second, for the algorithm in Figure 2 to be efficient, it is required also that *check_sat* returns a small (minimal, if possible) conflict set π out of the input set of literals. For instance, the decision procedure for acyclic lists should produce small unsatisfiable subsets of the input set of literals, or be able to give the equations necessary to deduce a given equality from a satisfiable set. We believe this is possible by adapting the method developed for congruence closure in [5].

6 Conclusion

We presented a combination method that is able to combine a many-sorted theory T_{len} modeling lists of elements in the presence of the length operator with a theory T_{elem} modeling the elements.

Our method works regardless of whether the theory of the elements is stably infinite or not. We were able to relax the stable infiniteness requirement by employing the following basic ideas:

- guess an arrangement larger than the one computed by Nelson and Oppen;
- compute a certain minimal cardinality k_0 , so that we can ensure that the domain of the elements must have at least k_0 elements.

Future works include implementing the proposed method in **haRVey**, and in particular, study heuristics to make it more efficient, and investigate extending the procedure for acyclic lists to compute minimal conflict sets. On the theoretical side, it remains to determine the exact complexity of the algorithm, and examine the proposed combination when some sorts (*elem*, *list*, *int*) are equal.

Acknowledgments

We are grateful to Christophe Ringeissen for insightful discussions on the problem of combining non-stably infinite theories. We would also like to thank the reviewers for their comments.

References

1. A. Armando, S. Ranise, and M. Rusinowitch. A rewriting approach to satisfiability procedures. *Information and Computation*, 183(2):140–164, 2003.
2. S. Berezin, V. Ganesh, and D. L. Dill. An Online Proof-Producing Decision Procedure for Mixed-Integer Linear Arithmetic. In *Proceedings of TACAS'03*, volume 2619 of *LNCS*, Warsaw, Poland, April 2003.
3. N. S. Bjørner. *Integrating Decision Procedures for Temporal Verification*. PhD thesis, Stanford University, 1998.

4. R. S. Boyer and J. S. Moore. *A Computational Logic*. ACM Monograph SERIES, 1979.
5. L. de Moura, H. Rueß, and N. Shankar. Justifying equality. In *PDPAR*, 2004.
6. D. Déharbe and S. Ranise. Light-Weight Theorem Proving for Debugging and Verifying Units of Code. In *Proc. of the International Conference on Software Engineering and Formal Methods (SEFM03)*. IEEE Computer Society Press, 2003.
7. J.-C. Filliâtre, S. Owre, H. Rueß, and N. Shankar. ICS: integrated canonizer and solver. In G. Berry, H. Comon, and A. Finkel, editors, *Computer Aided Verification (CAV)*, volume 2102 of *LNCS*, pages 246–249. Springer-Verlag, 2001.
8. P. Fontaine and P. Gribomont. Combining non-stably infinite, non-first order theories. In S. Ranise and C. Tinelli, editors, *Pragmatics of Decision Procedures in Automated Reasoning*, 2004.
9. H. Ganzinger. Shostak light. In A. Voronkov, editor, *Automated Deduction – CADE-18*, volume 2392 of *LNCS*, pages 332–346. Springer, 2002.
10. D. Kapur and X. Nie. Reasoning about Numbers in Tecton. In *Proc. 8th Intl. Symp. Methodologies for Intelligent Systems*, pages 57–70, 1994.
11. T. F. Melham. Automating Recursive Type Definitions in Higher Order Logic. In *Current Trends in Hardware Verification and Theorem Proving*, LNCS, pages 341–386. Springer-Verlag, 1989.
12. G. Nelson and D. C. Oppen. Simplifications by cooperating decision procedures. *ACM Trans. on Programming Languages and Systems*, 1(2):245–257, Oct. 1979.
13. G. Nelson and D. C. Oppen. Fast decision procedures based on congruence closure. *Journal of the Association for Computing Machinery*, 27(2):356–364, 1980.
14. D. C. Oppen. Reasoning about recursively defined data structures. *Journal of the ACM*, 27(3):403–411, 1980.
15. S. Owre and N. Shankar. Abstract Datatypes in PVS. Technical Report CSL-93-9R, SRI International, 1997.
16. L. C. Paulson. A fixedpoint approach to implementing (co)inductive definitions. In A. Bundy, editor, *Automated Deduction – CADE-12*, LNAI 814, pages 148–161. Springer, 1994. 12th international conference.
17. W. Pugh. The omega test: a fast integer programming algorithm for dependence analysis. *Supercomputing*, pages 4–13, 1991.
18. R. E. Shostak. Deciding combination of theories. *Journal of the Association for Computing Machinery*, 31(1):1–12, 1984.
19. A. Stump, C. W. Barrett, and D. L. Dill. CVC: a cooperating validity checker. In E. Brinksma and K. G. Larsen, editors, *Computer Aided Verification (CAV)*, volume 2404 of *LNCS*, pages 500–504. Springer, 2002.
20. C. Tinelli and C. G. Zarba. Combining non-stably infinite theories. *Journal of Automated Reasoning*, 2004. To appear.
21. P. Wolper and B. Boigelot. On the construction of automata from linear arithmetic constraints. In S. Graf and M. I. Schwartzbach, editors, *TACAS*, volume 1785 of *LNCS*, pages 1–19, Berlin, Mar. 2000. Springer-Verlag.
22. C. G. Zarba. Combining multisets with integers. In A. Voronkov, editor, *Automated Deduction – CADE-18*, volume 2392 of *LNCS*, pages 363–376. Springer, 2002.
23. C. G. Zarba. Combining sets with integers. In A. Armando, editor, *Frontiers of Combining Systems*, volume 2309 of *LNCS*, pages 103–116. Springer, 2002.
24. C. G. Zarba. Combining sets with elements. In N. Dershowitz, editor, *Verification: Theory and Practice*, volume 2772 of *LNCS*, pages 762–782. Springer, 2004.
25. T. Zhang, H. B. Sipma, and Z. Manna. Decision procedures for recursive data structures with integer constraints. In D. A. Basin and M. Rusinowitch, editors, *Automated Reasoning*, volume 3097 of *LNCS*, pages 152–167. Springer, 2004.