Verifying Higher-Order Functional Programs with Pattern-Matching Algebraic Data Types

C.-H. Luke Ong
Oxford University Computing Laboratory
lo@comlab.ox.ac.uk

Steven J. Ramsay
Oxford University Computing Laboratory
ster@comlab.ox.ac.uk

Abstract

Type-based model checking algorithms for higher-order recursion schemes have recently emerged as a promising approach to the verification of functional programs. We introduce *pattern-matching recursion schemes* (PMRS) as an accurate model of computation for functional programs that manipulate algebraic data-types. PMRS are a natural extension of higher-order recursion schemes that incorporate pattern-matching in the defining rules.

This paper is concerned with the following (undecidable) verification problem: given a correctness property φ, a functional program $\mathcal{P}$ (*qua* PMRS) and a regular input set $\mathcal{I}$, does every term that is reachable from $\mathcal{I}$ under rewriting by $\mathcal{P}$ satisfy φ? To solve the PMRS verification problem, we present a sound *semi-algorithm* which is based on model-checking and counterexample guided abstraction refinement. Given a no-instance of the verification problem, the method is guaranteed to terminate.

From an order-n PMRS and an input set generated by a regular tree grammar, our method constructs an order-n *weak* PMRS which over-approximates *only* the first-order pattern-matching behaviour, whilst remaining completely faithful to the higher-order control flow. Using a variation of Kobayashi's type-based approach, we show that the (trivial automaton) model-checking problem for weak PMRS is decidable. When a violation of the property is detected in the abstraction which does not correspond to a violation in the model, the abstraction is automatically refined by 'unfolding' the pattern-matching rules in the program to give successively more and more accurate weak PMRS models.

Categories and Subject Descriptors D.2.4 [*Software Engineering*]: Software/Program Verification; F.3.1 [*Logics and Meanings of Programs*]: Specifying and Verifying and Reasoning about Programs

General Terms Languages, Verification

1. Introduction

In the past decade, huge strides have been made in the development of finite-state and pushdown model checking for software verification. Though highly effective when applied to first-order, imperative programs such as C, these techniques are much less useful for higher-order, functional programs. In contrast, the two standard approaches to the verification of higher-order programs are *type-based program analysis* on the one hand, and *theorem-proving* and *dependent types* on the other. The former is sound, but often imprecise; the latter typically requires human intervention.

Recently, a model-checking approach, based on *higher-order recursion schemes* (HORS), has emerged as a verification methodology that promises to combine accurate analysis and push-button automation. HORS are a form of simply-typed lambda-calculus with recursion and uninterpreted function symbols that is presented as a grammar and used as a generator of (possibly infinite) trees. Ong showed that the trees generated by HORS have a decidable modal mu-calculus theory [14] and Kobayashi introduced a novel approach to the verification of higher-order functional programs by reduction to their model-checking problems [6].

This method has been applied successfully to the Resource Usage Verification Problem [3] (and, through it, to such problems as reachability and control-flow analysis) for a simply typed functional language with finite data-types and dynamic resource creation and resource access primitives. The method relies on the existence of certain transformations which, given a functional program and a resource usage specification, reduce the corresponding verification problem to the question of whether the computation tree of the program, generated by a HORS, satisfies a resource-wise specification encoded by an automaton on infinite trees. Despite the high worst-case time complexity of the modal mu-calculus model-checking problem for recursion schemes, which is n-EXPTIME complete for order-n schemes, an implementation of this approach, TRecS, performs remarkably well on realistic inputs [7].

From a verification perspective, a serious weakness of the HORS approach is its inability to naturally model functional programs with infinite data structures, such as integers and algebraic data-types. This severely limits the potential impact of this programme as functions defined by cases on algebraic data types are ubiquitous in functional programming.

A model of functional programs. Our first contribution is the introduction of *pattern-matching recursion schemes*, which are HORS extended with a notion of pattern matching. A PMRS is a kind of restricted term-rewriting system. We believe that PMRS have a very natural syntax into which large classes of functional programs can readily be translated. A typical rule, which is required to be well typed, has the shape:

$$F\, x_1 \cdots x_m\, p(y_1, \cdots, y_k) \;\longrightarrow\; t$$

where the variables $x_1, \cdots, x_m$ are (possibly higher-order) formal parameters of the non-terminal (or *defined operator*) F. The expression $p(y_1, \cdots, y_k)$, which takes the place of the final parameter, is a *pattern* constructed from terminal (or *constructor*) symbols and variables $y_1, \cdots, y_k$.

POPL'11, January 26–28, 2011, Austin, Texas, USA.

Example 1. The following PMRS defines a function $Merge$: $\mathbf{ablist} \to \mathbf{ablist} \to \mathbf{ablist}$ that merges two lists of `a` and `b` by recursively destructing them.

$$\begin{array}{lcl} Merge\ x\ \texttt{nil} & \longrightarrow & x \\ Merge\ x\ (\texttt{cons a}\ y) & \longrightarrow & \texttt{cons a}\ (Merge\ y\ x) \\ Merge\ x\ (\texttt{cons b}\ y) & \longrightarrow & \texttt{cons b}\ (Merge\ y\ x) \end{array}$$

The patterns in the second argument position are used both to decompose compound data structures (so as to select the required components), and to determine control flow. Selected components are communicated to the right-hand side of the chosen rule by means of binding to the variables in the pattern.

Remark 1. Our work is not the first to propose a pattern-matching extension to HORS. A recent paper by Kobayashi, Tabuchi and Unno [9] introduces an extension of HORS called *higher-order multi-parameter tree transducers* (HMTT). HMTT model functions that may employ pattern matching but, in return, must satisfy a rigid type constraint. An HMTT function takes tree arguments of input sort `i` (which are trees that can only be destructed) and returns a tree of sort `o` (which are trees that can only be constructed). Pattern matching is only allowed on trees of sort `i`. Consequently HMTT functions are not compositional in the natural way. We believe our PMRS model to be both simpler and more natural.

A verification problem. This paper is concerned with the following verification problem. Given a correctness property φ, a functional program $\mathcal{P}$ (*qua* deterministic PMRS) and a regular set $\mathcal{I}$ of input (constructor) terms, does every term that is reachable from $\mathcal{I}$ under rewriting by $\mathcal{P}$ satisfy φ? It is straightforward to see that the problem is undecidable.

Example 2. Consider the PMRS $\mathcal{P}$ which, when started from $Main$ takes as input a list of natural numbers and returns the same list with all occurrences of the number zero removed. The defining rules of $\mathcal{P}$ are given by:

$$Main\ m \longrightarrow Filter\ Nz\ m$$

$$\begin{array}{l} If\ a\ b\ \texttt{true} \longrightarrow a \\ If\ a\ b\ \texttt{false} \longrightarrow b \end{array}$$

$$\begin{array}{l} Nz\ \texttt{z} \longrightarrow \texttt{false} \\ Nz\ (\texttt{s}\ n) \longrightarrow \texttt{true} \end{array}$$

$$\begin{array}{l} Filter\ p\ \texttt{nil} \longrightarrow \texttt{nil} \\ Filter\ p\ (\texttt{cons}\ x\ xs) \longrightarrow \\ \quad If\ (\texttt{cons}\ x\ (Filter\ p\ xs))\ (Filter\ p\ xs)\ (p\ x) \end{array}$$

The input set $\mathcal{I}$ is given by a regular tree grammar $\mathcal{G}$ (equivalently order-0 recursion scheme). The defining rules of $\mathcal{G}$ are:

$$S \longrightarrow ListN$$

$$\begin{array}{l} N \longrightarrow \texttt{z} \\ N \longrightarrow \texttt{s}\ N \end{array}$$

$$\begin{array}{l} ListN \longrightarrow \texttt{nil} \\ ListN \longrightarrow \texttt{cons}\ N\ ListN \end{array}$$

As usual, the start symbol of $\mathcal{G}$ is taken to be S. The correctness property φ is: "any outcome of the program is a list containing no zeros". This is easily expressible as a trivial automaton $\mathcal{A}$, whose definition is omitted.

An algorithmic solution. Our second contribution is a sound but incomplete semi-algorithm for solving the problem, which is based on a counterexample-guided abstraction refinement loop [2, 11]. The input to the algorithm consists of a PMRS $\mathcal{P}$ representing the program, a regular tree grammar $\mathcal{G}$ (equivalently an order-0 recursion scheme) representing the set $\mathcal{I}$ of possible inputs to the program and a trivial tree automaton $\mathcal{A}$ (which is to say, an automaton on infinite trees with a trivial acceptance condition) representing a specification φ of good behaviour. The algorithm proceeds according to the diagram in Figure 1.

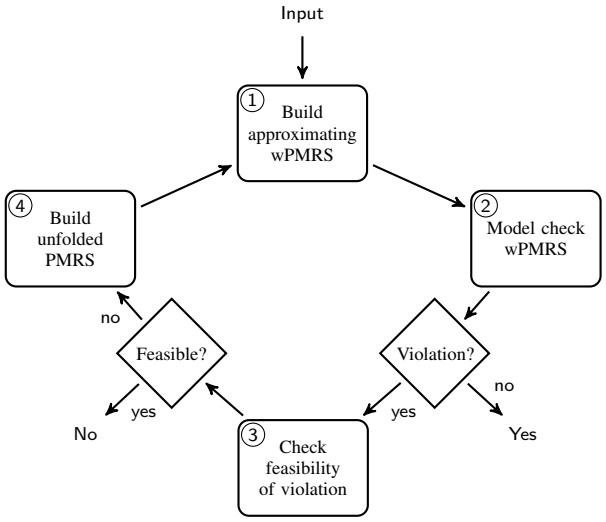


Figure 1. Counterexample-guided abstraction-refinement loop.

In step (1) we compute a sound abstraction of the behaviour of $\mathcal{P}$ when started from terms in $\mathcal{I}$. From an order-n PMRS $\mathcal{P}$ and an order-0 recursion scheme $\mathcal{G}$, we build an order-n *weak pattern-matching recursion scheme* (wPMRS) which over-approximates the set of terms that are reachable from $\mathcal{I}$ under rewriting by $\mathcal{P}$. A wPMRS is similar to a PMRS, except that its pattern matching mechanism is only able to determine control flow; it is unable to decompose data structure.

Our method is a kind of flow analysis. The first – and key – stage of the algorithm is a *binding analysis* which is inspired by Jones and Andersen [5]. It performs a fixpoint construction of a *finite* set Ξ of variable-term bindings such that, for every variable x (formal parameter of rewrite rule), every term that is ever bound to x during the computation is derivable from Ξ. In the second stage, we use the fixpoint set Ξ to build rules of the over-approximating wPMRS. These rules model the bindings of all non-pattern-matching (including all higher-order) variables *precisely*; they only approximate the binding behaviours of the pattern-matching variables. This is in contrast to Jones and Andersen's algorithm, which builds a regular tree grammar that over-approximates the binding set of *every* variable. For an order-n PMRS, our algorithm produces an order-n wPMRS $\widetilde{\mathcal{P}_\mathcal{G}}$ as an abstraction, which is a tighter approximation of the order-n PMRS being analysed than regular tree grammars (which are equivalent to order-0 wPMRS). To our knowledge, our algorithm gives the most accurate reachability / flow analysis of its kind.

The weakened pattern-matching mechanism of wPMRS makes it possible to decide a model checking problem for it, which is the content of step (2). Given a wPMRS $\mathcal{W}$, a closed term t and a Büchi automaton with a trivial acceptance condition $\mathcal{A}$, we decide if every (possibly infinite) tree generated by $\mathcal{W}$ on input t is accepted by $\mathcal{A}$. The proof uses a variation of Kobayashi's type-based approach.

If the model-checker fails to find any violation of the property then, since the approximating wPMRS $\widetilde{\mathcal{P}_\mathcal{G}}$ defines a superset of the terms reachable under $\mathcal{P}$ from $\mathcal{I}$, the loop in Figure 1 will terminate because $\mathcal{P}$ satisfies $\mathcal{A}$ on $\mathcal{I}$. However, if the model-checker reports a counterexample, then it may be that $\mathcal{P}$ also violates the property

(for some term in $\mathcal{I}$), but it may also be that the counterexample is an artifact of an inaccuracy in the abstraction. To determine which of these possibilities is the case, in step (3) we analyse the non-determinism introduced in the abstraction to see whether, in this particular counterexample, it behaves well or behaves badly.

In step (4) the abstraction process is refined. Due to the fact that the abstractions only ever approximate the (first-order) pattern matching variables, whilst remaining faithful to all the others, there is a simple notion of automatic abstraction-refinement, whereby patterns are "unfolded" to a certain depth in the PMRS $\mathcal{P}$, forming a new PMRS $\mathcal{P}'$. In the abstraction $\widetilde{\mathcal{P}'_{\mathcal{G}}}$ of $\mathcal{P}'$, the rules that define the approximation will be more accurate and, in particular, the spurious counterexample will no longer be present. Since any rule in a wPMRS abstraction $\widetilde{\mathcal{P}'_{\mathcal{G}}}$ is *perfectly* accurate whenever the pattern parameter contains no free variables, this method of unfolding gives rise to a semi-completeness property. Given any no-instance of the PMRS verification problem, the loop in Figure 1 will eventually terminate with the answer "No".

Returning to Example 2, whilst performing step (1) we obtain an over-approximation of the binding behaviour of the variables in the program Ξ. This fixpoint set contains, amongst others, the bindings: $x \mapsto N$ and $xs \mapsto ListN$. From this set, we construct an approximating wPMRS $\widetilde{\mathcal{P}_{\mathcal{G}}}$, whose rule-set contains the following:

$$\begin{array}{l} \mathit{Filter}\ p\ \mathtt{nil} \longrightarrow\ \ \mathit{Nil} \\ \mathit{Filter}\ p\ (\mathtt{cons}\ x\ xs) \longrightarrow \\ \quad \mathit{If}\ (\mathit{Cons}\ X\ (\mathit{Filter}\ p\ XS))\ (\mathit{Filter}\ p\ XS)\ (p\ X) \end{array}$$

$$\begin{array}{l} X \longrightarrow N \\ XS \longrightarrow \mathit{ListN} \end{array}$$

together with, amongst others, all the $\mathcal{P}$ and $\mathcal{G}$ rules in Example 2 except those for Filter. Unfortunately the wPMRS is too coarse to be useful: there are trees (representing lists) that are obtained by rewriting from '$\mathit{Main}\ S$' that are not accepted by the trivial automaton $\mathcal{A}$. However, these are spurious counterexamples. For an illustration, consider the error trace in the wPMRS:

$$\begin{array}{l} \mathit{Main}\ S \\ \rightarrow^* \mathit{Main}\ (\mathtt{cons}\ (\mathtt{s}\ \mathtt{z})\ \mathtt{nil}) \\ \rightarrow^* \mathit{Filter}\ \mathit{Nz}\ (\mathtt{cons}\ (\mathtt{s}\ \mathtt{z})\ \mathtt{nil}) \\ \rightarrow\ \ \mathit{If}\ (\mathit{Cons}\ X\ (\mathit{Filter}\ \mathit{Nz}\ \mathit{XS}))\ (\mathit{Filter}\ \mathit{Nz}\ \mathit{XS})\ (\mathit{Nz}\ X) \\ \rightarrow^* \mathit{If}\ (\mathtt{cons}\ \mathtt{z}\ (\mathit{Filter}\ \mathit{Nz}\ \mathtt{nil}))\ (\mathit{Filter}\ \mathit{Nz}\ \mathit{XS})\ (\mathit{Nz}\ (\mathtt{s}\ \mathtt{z})) \\ \rightarrow^* \mathtt{cons}\ \mathtt{z}\ (\mathit{Filter}\ \mathit{Nz}\ \mathtt{nil}) \\ \rightarrow^* \mathtt{cons}\ \mathtt{z}\ \mathtt{nil} \end{array}$$

The problem can be traced to the second clause of Filter in the wPMRS: when replacing the *variable* x by the *non-terminal* X, the connection between the two occurrences of x in the RHS is lost, as the reduction of one occurrence of X is independent of that of the other.

The refinement algorithm produces a new, unfolded PMRS $\mathcal{P}'$ that replaces the two defining rules of Filter by five new rules. The two rules that cover the case when the list is a singleton are shown below:

$$\begin{array}{l} \mathit{Filter}\ p\ (\mathtt{cons}\ \mathtt{z}\ \mathtt{nil}) \longrightarrow \\ \quad \mathit{If}\ (\mathtt{cons}\ \mathtt{z}\ (\mathit{Filter}\ p\ \mathtt{nil}))\ (\mathit{Filter}\ p\ \mathtt{nil})\ (p\ \mathtt{z}) \\ \mathit{Filter}\ p\ (\mathtt{cons}\ (\mathtt{s}\ v_2)\ \mathtt{nil}) \longrightarrow \\ \quad \mathit{If}\ (\mathtt{cons}\ (\mathtt{s}\ v_2)\ (\mathit{Filter}\ p\ \mathtt{nil}))\ (\mathit{Filter}\ p\ \mathtt{nil})\ (p\ (\mathtt{s}\ v_2)) \end{array}$$

Applying the approximation algorithm to PMRS $\mathcal{P}'$ (and input grammar $\mathcal{G}$), we obtain a wPMRS $\widetilde{\mathcal{P}'_{\mathcal{G}}}$ that does accurately capture the set of reachable terms.

$$\frac{}{\Gamma,\ x:\sigma\ \vdash\ x:\sigma}\ (\textsc{Var}) \qquad \frac{\xi:\sigma \in \Sigma \cup \mathcal{N}}{\Gamma\ \vdash\ \xi:\sigma}\ (\textsc{Const})$$

$$\frac{\Gamma\ \vdash\ t_0:\sigma\to\tau \qquad \Gamma\ \vdash\ t_1:\sigma}{\Gamma\ \vdash\ t_0\ t_1:\tau}\ (\textsc{App})$$

Figure 2. A simple type system for applicative terms.

Outline. The rest of the paper is organised as follows. Section 2 introduces PMRS, wPMRS and other technical preliminaries. In Section 3, the abstraction algorithm, which takes a program (PMRS) and an input set (order-0 recursion scheme) and returns a wPMRS, is presented; termination and soundness of the approximation are proved. Section 4 presents a type-inference algorithm for deciding if every tree generated by a given wPMRS is accepted by a trivial automaton. The abstraction refinement algorithm is the topic of Section 5. Finally Section 6 presents related work. Note: a long version of the paper is available [15], which contains the proofs and additional material.

2. Preliminaries

We introduce PMRS, a model for functional programs manipulating algebraic data types; wPMRS, a restriction of PMRS with good algorithmic properties and the PMRS Verification Problem, whose solution is the subject of the remainder of this work.

2.1 Types, terms and substitutions

Fix a finite set $(b,\ o\ \in)\ \mathbb{B}$ of *base types*. The *simple types* $(\sigma,\ \tau \in)\ \mathbb{S}$ are those expressions that can be constructed from the base types using the arrow:

$$\sigma,\ \tau\quad ::=\quad b\ \mid\ \sigma \to \tau.$$

We adopt the usual convention that arrows associate to the right and omit parenthesis accordingly. The *order* of a type τ, denoted $\mathbf{ord}(\tau)$, is a measure of the nestedness of the arrow constructor on the left; it is defined by $\mathbf{ord}(b) = 0$ and $\mathbf{ord}(\sigma \to \tau) = \mathsf{max}\{\mathbf{ord}(\sigma)+1,\ \mathbf{ord}(\tau)\}$.

Applicative terms. Fix a finite, simply-typed alphabet $(f, g, a \in)$ Σ of first-order *terminal symbols* (or *constructors*), a finite, simply-typed alphabet $(F, G, H\ \in)\ \mathcal{N}$ of (arbitrary-order) *non-terminal symbols* (or *defined operators*) and a denumerable set $(x,\ y,\ z\ \in)$ $\mathcal{V}$ of *variables*.

- The *constructor terms* $T(\Sigma)$ are those expressions that can be built from terminals using application.
- The *closed terms* $T(\Sigma,\ \mathcal{N})$ are those expressions that can be built from terminals and non-terminals using application.
- The *patterns* are those expressions $p,\ q$ of base type that can be built from variables of base type and terminals.
- The *applicative terms* $T(\Sigma,\ \mathcal{N},\ \mathcal{V})$ are those expressions that can be built from terminals, non-terminals and variables using application.

We denote the free variables of a term t by $\mathsf{FV}(t)$.

Standardly, applicative terms may be assigned simple types via a formal system of typing judgements, $\Gamma\ \ \vdash\ \ s\ :\ \tau$ (where Γ is a finite set of *type bindings*) defined by the rules in Figure 2. When an applicative term t can be assigned a simple type $\tau_1 \to \cdots \to \tau_m \to b$ we say that it has *arity* m and write $\mathsf{ar}(t) = m$. Henceforth, by *term* we shall mean well-typed, applicative term.

Labelled trees. Given a ranked alphabet Ω, an Ω-*labelled tree* t is a map from $\{1, \cdots, m\}^*$ to Ω, where m is the largest arity of

symbols in Ω, such that $\mathbf{dom}(t)$ is prefix-closed, and if $t(x) = f$ then $\{\, i \mid x\,i \in \mathbf{dom}(t)\,\} = \{\, 1, \cdots, \mathsf{ar}(f)\,\}$. Standardly we identify $T(\Sigma)$ with finite Σ-labelled trees, and write $T^{\infty}(\Sigma)$ for the collection of (possibly infinite) Σ-labelled trees.

Let $\Sigma^{\perp}$ be $\Sigma \cup \{\,\perp\,\}$ with $\mathsf{ar}(\perp) = 0$. Given a closed term t, we write $t^{\perp}$ for the finite, $\Sigma^{\perp}$-labelled tree defined by recursion as follows: for $m \geq 0$

$$(\xi\ s_1\ \cdots\ s_m)^{\perp} := \begin{cases} \perp & \text{if } \xi = F \in \mathcal{N} \\ f\ s_1^{\perp} \cdots s_m^{\perp} & \text{otherwise } \xi = f \in \Sigma \end{cases}$$

E.g. $(f(g(G\,a))b)^{\perp} = f(g\,\perp)b$. $\Sigma^{\perp}$-labelled trees can be endowed with a natural complete partial order $\sqsubseteq$ in which, for all trees t, $\perp \sqsubseteq t$ and $f\ s_1\ \cdots\ s_m \sqsubseteq f\ t_1\ \cdots\ t_m$ iff for all i, $s_i \sqsubseteq t_i$.

Substitutions. A *substitution* is just a partial function θ in $\mathcal{V} \to T(\Sigma, \mathcal{N}, \mathcal{V})$. By convention, we do not distinguish between a substitution and its homomorphic extension to the free algebra $T(\Sigma, \mathcal{N}, \mathcal{V})$ and we will write the application of both using prefix juxtaposition. A term t is said to *match* a term u precisely when there exists a substitution θ such that $t = \theta u$. We shall say that a substitution θ is *closed* whenever every term in its image is closed.

2.2 Pattern-matching recursion scheme (PMRS)

A *pattern-matching recursion scheme* (PMRS) is a quadruple $\mathcal{P} = \langle \Sigma, \mathcal{N}, \mathcal{R}, \mathit{Main} \rangle$ with Σ and $\mathcal{N}$ as above. $\mathcal{R}$ is a finite set of rewrite rules, each of which is one of the following shapes ($m \geq 0$):

$$\begin{array}{rl} (\textit{pure}) & F\ x_1\ \cdots\ x_m \longrightarrow t \\ (\textit{pattern-matching}) & F\ x_1\ \cdots\ x_m\ p \longrightarrow t \end{array}$$

where p is a pattern (which may be trivial). $\mathit{Main} : b \to o$ is a distinguished non-terminal symbol whose defining rules are always pattern-matching rules. In this paper we will assume that the variables appearing as formal parameters to defining rules in a PMRS will always be distinct.

A pure rule $F\ x_1\ \cdots\ x_m \longrightarrow t$ is well-typed when $F : \tau_1 \to \cdots \to \tau_m \to o \in \mathcal{N}$ and the judgement:

$$x_1 : \tau_1, \ldots, x_m : \tau_m \ \vdash\ t : o$$

is provable. A pattern-matching rule $F\ x_1\ \cdots\ x_m\ p \longrightarrow t$ is well-typed when $F : \tau_1 \to \cdots \to \tau_m \to b \to o \in \mathcal{N}$ and there exist base-types $b_1, \ldots, b_k$ such that the judgements:

$$y_1 : b_1, \ldots, y_k : b_k \ \vdash\ p : b \qquad \text{and}$$
$$x_1 : \tau_1, \ldots, x_m : \tau_m, y_1 : b_1, \ldots, y_k : b_k \ \vdash\ t : o$$

are provable. We say that a PMRS is *well-typed* just when each of its rules is well-typed. We will only consider well-typed PMRS in the following.

We define the *order* of a PMRS to be the maximum order of (the type of) any of the non-terminal symbols in $\mathcal{N}$. Since a pure rule can be simulated by a pattern-matching rule with a trivial pattern (e.g. a nullary terminal of a distinguished base type), we shall sometimes find it convenient to treat all PMRS rules as pattern-matching rules.

Reduction. We associate with each PMRS a notion of reduction as follows. A *redex* is a term of the form $F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ whenever θ is a closed substitution and $F\ x_1\ \cdots\ x_m\ p \longrightarrow t$ is a rule in $\mathcal{P}$. The *contractum* of the redex is $\theta\, t$. We define the one-step reduction relation, $\Rightarrow\ \subseteq T(\Sigma, \mathcal{N}) \times T(\Sigma, \mathcal{N})$, by $C[s] \Rightarrow C[t]$ whenever s is a redex, t is its contractum and C is a one-hole context.

We say that a PMRS is *deterministic* just if, given some redex $F\ s_1\ \cdots\ s_n$ there is exactly one rule $l \longrightarrow r \in \mathcal{R}$ such that $F\ s_1\ \cdots\ s_n = \theta l$ for some θ.

Given a PMRS $\mathcal{P} = \langle \Sigma, \mathcal{N}, \mathcal{R}, \mathit{Main} \rangle$, let $s \in T(\Sigma, \mathcal{N})$ be a closed term of base type. We write $\mathcal{L}(\mathcal{P}, s)$ to mean the language of $\Sigma^{\perp}$-labelled trees obtained by infinitary rewriting of the term s. More precisely, define $\mathcal{L}(\mathcal{P}, s)$ as the collection of $\Sigma^{\perp}$-labelled trees t such that there are $\langle t_i \rangle_{i \in \omega}$ with $s \Rightarrow t_1 \Rightarrow t_2 \Rightarrow t_3 \cdots$ a *fair* reduction sequence (in the sense that for each i, every outermost redex in t_i is eventually contracted) and $t = \bigsqcup \{\, t_i^{\perp} \mid i \in \omega \,\}$. In case $\mathcal{P}$ is a deterministic PMRS, $\mathcal{L}(\mathcal{P}, s)$ is a singleton set; we write the unique $\Sigma^{\perp}$-labelled tree as $[\![s]\!]_{\mathcal{P}}$.

Example 3. Let $\Sigma = \{\mathtt{zero} : \mathbf{nat}, \mathtt{succ} : \mathbf{nat} \to \mathbf{nat}, \mathtt{nil} : \mathbf{natlist}, \mathtt{cons} : \mathbf{nat} \to \mathbf{natlist} \to \mathbf{natlist}\}$ and $\mathcal{N} = \{\mathit{Rev} : \mathbf{natlist} \to \mathbf{natlist}, \mathit{RevA} : \mathbf{natlist} \to \mathbf{natlist} \to \mathbf{natlist}\}$. The following deterministic, order-1 PMRS contains rewrite rules that implement list reversal with an accumulating parameter:

$$\begin{array}{rcl} \mathit{Main}\ zs & \longrightarrow & \mathit{RevA}\ \mathtt{nil}\ zs \\ \\ \mathit{RevA}\ xs\ \mathtt{nil} & \longrightarrow & xs \\ \mathit{RevA}\ xs\ (\mathtt{cons}\ y\ ys) & \longrightarrow & \mathit{RevA}\ (\mathtt{cons}\ y\ xs)\ ys \end{array}$$

When started from the term $t = \mathtt{cons\ z\ nil}$, the only possible reduction sequence is:

$$\mathit{Main}\ t \Rightarrow \mathit{RevA}\ \mathtt{nil}\ t \Rightarrow \mathit{RevA}\ t\ \mathtt{nil} \Rightarrow t$$

and hence $[\![\mathit{Main}\ t]\!]_{\mathcal{P}} = t$, as expected.

2.3 Weak pattern matching recursion schemes (wPMRS)

A *weak pattern-matching recursion scheme* (wPMRS) is a quadruple $\mathcal{W} = \langle \Sigma, \mathcal{N}, \mathcal{R}, \mathit{Main} \rangle$ with $\Sigma, \mathcal{N}$ and Main as for PMRS. The (finite) set $\mathcal{R}$ consists of rewrite rules of the shape ($m \geq 0$):

$$\begin{array}{rl} (\textit{pure}) & F\ x_1\ \cdots\ x_m \longrightarrow t \\ (\textit{weak-matching}) & F\ x_1\ \cdots\ x_m\ p \longrightarrow t \end{array}$$

in which $\mathsf{FV}(p) \cap \mathsf{FV}(t) = \emptyset$. A pure rule is well typed according to the same criteria as for pure PMRS rules. A weak-matching rule $F\ x_1\ \cdots\ x_m\ p \longrightarrow t$ is well-typed just when $F : \tau_1 \to \cdots \to \tau_m \to b \to o \in \mathcal{N}$ and there exist base-types $b_1, \ldots, b_k$ such that the judgements:

$$y_1 : b_1, \ldots, y_k : b_k \ \vdash\ p : b$$
$$\text{and} \quad x_1 : \tau_1, \ldots, x_m : \tau_m \ \vdash\ t : o$$

are provable (note that none of the pattern-matching variables y_j occurs in t). Henceforth we will only consider wPMRS with well-typed rules.

wPMRS have exactly the same notion of reduction as PMRS: a *redex* is a term of the form $F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ whenever θ is a substitution and $F\ x_1\ \cdots\ x_m\ p \longrightarrow t$ is a rule in $\mathcal{P}$. The *contractum* of the redex is $\theta\, t = t[\theta\, x_1/x_1] \cdots [\theta\, x_m/x_m]$ (as the pattern-matching variables do not occur in t). The one-step reduction relation, $\rightarrow$, is defined as for PMRS.

We define the order, determinism and language of a wPMRS analogously with PMRS.

2.4 A verification problem

We are interested in solving the following verification problem. Given a program in the form of a PMRS $\mathcal{P}$, a regular set $\mathcal{I}$ of "input" terms, and a correctness property φ, does the output $[\![\mathit{Main}\ t]\!]$ of the program '$\mathit{Main}\ t$' satisfy φ, for every input $t \in \mathcal{I}$? To propose a solution, we require two further stipulations, both of which concern the representation of the entities involved.

Higher-order recursion schemes. A *higher-order recursion scheme* (HORS) is a quadruple $\mathcal{G} = \langle \Sigma, \mathcal{N}, \mathcal{R}, S \rangle$ with Σ and $\mathcal{N}$ as before and $\mathcal{R}$ is a finite set of well-typed, pure wPMRS rewrite rules. The component S is a distinguished non-terminal called the

"start symbol". The reduction relation for HORS, $\rightarrow$, is just that of wPMRS, noting that all redexes will necessarily be of the form $F\ \theta x_1\ \cdots\ \theta x_m$ since there are no pattern-matching arguments. We can associate with a recursion scheme $\mathcal{G}$ its language $\mathcal{L}(\mathcal{G})$ of terms in $T(\Sigma)$ that can be derived from the start symbol S by rewriting away all occurrences of non-terminals. More precisely, we make the following definition:

$$\mathcal{L}(\mathcal{G}) := \{\, t \mid S \rightarrow^* t,\ t \in T(\Sigma) \,\}$$

We define the *order* of a recursion scheme analogously with PMRS and wPMRS. Note that (as generators of finite ranked trees) order-0 recursion schemes are equivalent to regular tree grammars.

Trivial automata. Let Σ be as before. A *Büchi tree automaton with a trivial acceptance condition* (or simply, *trivial automaton*) is a quadruple $\mathcal{A} = \langle \Sigma, Q, \Delta, q_0 \rangle$ where Σ is as before, Q is a finite set of states, $q_0 \in Q$ is the initial state, and Δ, the transition relation, is a subset of $Q \times \Sigma \times Q^*$ such that if $(q, f, q_1 \cdots q_n) \in \Delta$ then $n = \mathsf{ar}(f)$. A Σ-labelled tree t is *accepted* by $\mathcal{A}$ if there is a Q-labelled tree r such that

(i) $\mathbf{dom}(t) = \mathbf{dom}(r)$,

(ii) for every $x \in \mathbf{dom}(r)$, $(r(x), t(x), r(x\,1) \cdots r(x\,m)) \in \Delta$ where $m = \mathsf{ar}(t(x))$.

The tree r is called a *run-tree* of $\mathcal{A}$ over t. We write $\mathcal{L}(\mathcal{A})$ for the set of Σ-labelled trees accepted by $\mathcal{A}$.

The PMRS Verification Problem. Given a deterministic PMRS $\mathcal{P} = \langle \Sigma, \mathcal{N}_\mathcal{P}, \mathcal{R}_\mathcal{P}, \mathit{Main} \rangle$, a (non-deterministic) order-0 recursion scheme $\mathcal{G} = \langle \Sigma, \mathcal{N}_\mathcal{G}, \mathcal{R}_\mathcal{G}, S \rangle$, and a Büchi tree automaton with a trivial acceptance condition $\mathcal{A} = \langle \Sigma, Q, \Delta, q_0 \rangle$, we write:

$$\models (\mathcal{P}, \mathcal{G}, \mathcal{A}) \quad \text{iff} \quad \forall t \in \mathcal{L}(\mathcal{G}) \cdot [\![\mathit{Main}\ t]\!]_\mathcal{P} \in \mathcal{L}(\mathcal{A})$$

The *PMRS Verification Problem* is to decide the truth of $\models(\mathcal{P}, \mathcal{G}, \mathcal{A})$.

3. Constructing an abstraction

In this section we will present an algorithm which, given an order-n deterministic PMRS $\mathcal{P}$ and an order-0 recursion scheme $\mathcal{G}$, constructs an order-n wPMRS $\widetilde{\mathcal{P}_\mathcal{G}}$ whose language of Σ-labelled trees is an over-approximation of the set of Σ-labelled trees reachable from $\mathcal{L}(\mathcal{G})$ under rewriting by $\mathcal{P}$.

At the heart of the algorithm is an analysis of the composite PMRS $\mathcal{P}_\mathcal{G} := \langle \Sigma, \mathcal{N}_\mathcal{G} \cup \mathcal{N}_\mathcal{P}, \mathcal{R}_\mathcal{G} \cup \mathcal{R}_\mathcal{P}, \mathit{Main} \rangle$. Since every term s reachable from $\mathcal{L}(\mathcal{G})$ under rewriting by $\mathcal{P}$ (i.e. $\mathit{Main}\ t \Rightarrow^*_\mathcal{P} s$, for some $t \in \mathcal{L}(\mathcal{G})$) is certainly reachable from S under rewriting by $\mathcal{P}_\mathcal{G}$ (i.e. $\mathit{Main}\ S \Rightarrow^*_{\mathcal{P}_\mathcal{G}} \mathit{Main}\ t \Rightarrow^*_{\mathcal{P}_\mathcal{G}} s$), it suffices to look only at the behaviours of $\mathcal{P}_\mathcal{G}$ in order to construct a safe abstraction of those of $\mathcal{P}$. We detail the nature of this analysis and its properties separately before showing how it underlies the construction of the approximating wPMRS $\widetilde{\mathcal{P}_\mathcal{G}}$.

Some nomenclature. A *simple term* is a subterm of the RHS of a $\mathcal{P}_\mathcal{G}$-rule or is the "starting term" $\mathit{Main}\ S$. A *compound term* has the shape $\xi\, t_1 \cdots t_m$ with $m \geq 0$, where the head symbol ξ is either a variable, or a terminal, or a non-terminal, and each t_i is simple. It follows from the definition that a simple term is compound, but the converse is not true.

3.1 Binding analysis

In a PMRS, the pattern matching rules use pattern matching both to determine control flow (by selecting which of a number of defining rules is used to reduce a redex) as well as to decompose compound data structure (by binding components to variables in the pattern that then occur on the RHS of the rule). However, the weak pattern matching mechanism in a wPMRS exhibits only the former capability: although patterns are matched, since there are no pattern-matching variables on the RHS of defining rules, data structures cannot be decomposed. Therefore, to build an effective abstraction of a PMRS requires some knowledge of the substitutions that can occur in redex/contractum pairs during PMRS reduction.

To this end, we define a *binding analysis*, which determines a (finitary) over-approximation Ξ to the set of variable-term bindings $\bigcup\{\, \theta \mid \mathit{Main}\ S \Rightarrow^* C[F\ \theta x_1\ \cdots\ \theta x_m\ \theta p] \Rightarrow C[\theta t] \,\}$ which occur in redex/contractum substitutions θ arising in $\mathcal{P}_\mathcal{G}$-reductions from '$\mathit{Main}\ S$'. The analysis is based on the observation that every such redex is either '$\mathit{Main}\ S$', or arises as an instance of a simple term. It proceeds by an iterative process in which bindings, by which instances of simple terms can be derived, give rise to redexes which in turn give rise, via contraction, to more bindings, until the desired set Ξ is reached in the limit.

Before we give the details of the analysis, let us make precise what it means for a set of bindings $\mathcal{S}$ to give rise to an instance of a term. Given such a set $\mathcal{S}$, we define the relation $s \preccurlyeq_\mathcal{S} t$, which is a subset of $T(\Sigma, \mathcal{N}, \mathcal{V}) \times T(\Sigma, \mathcal{N})$, inductively, by the system RS:

(R) $t \preccurlyeq_\mathcal{S} t$

(S) If $x \mapsto s \in \mathcal{S}$ and $C[s] \preccurlyeq_\mathcal{S} t$, then $C[x] \preccurlyeq_\mathcal{S} t$

where C ranges over one-hole contexts. We say that an instance of rule (S) is a *head-instance* just if the hole in $C[\,]$ occurs in head position.

Example 4. Let $\mathcal{S}_1 = \{x \mapsto y\ \mathtt{b},\ x \mapsto N,\ y \mapsto \mathtt{f}\ z,\ z \mapsto \mathtt{a}\}$. Then, using the system RS, it is possible to derive:

$$F\ x\ z \preccurlyeq_{\mathcal{S}_1} F\ (\mathtt{f\ a\ b})\ \mathtt{a} \qquad \text{and} \qquad F\ x\ z \preccurlyeq_{\mathcal{S}_1} F\ N\ \mathtt{a}$$

Note that the form of rule (S) does not constrain bindings to be used consistently within non-linear terms. Let $\mathcal{S}_2 = \{x \mapsto \mathtt{f}\ y\ z,\ y \mapsto z,\ y \mapsto \mathtt{a},\ z \mapsto \mathtt{b}\}$. Then we have, for example:

$$F\ x\ (G\ x) \preccurlyeq_{\mathcal{S}_2} F\ (\mathtt{f\ a\ b})\ (G\ (\mathtt{f\ b\ b}))$$

in which the binding $y \mapsto \mathtt{a}$ has been used in the derivation of the first argument of F whereas $y \mapsto z$ has been used in the derivation of the second argument.

To ensure that the analysis is computable, we cannot afford to work with instances of simple terms directly. We instead work with terms in which bindings have been applied only where strictly necessary in order to uncover new redexes. The construction of such terms is the purpose of the function head.

The head function. Given a set $\mathcal{S}$ of bindings, we define the *head function*, $\mathsf{head}_\mathcal{S} : T(\Sigma, \mathcal{N}, \mathcal{V}) \longrightarrow 2^{T(\Sigma, \mathcal{N}, \mathcal{V})}$ given by:

$$\mathsf{head}_\mathcal{S}(\xi\ t_1\ \cdots\ t_m) = \{\, \delta\ t_1\ \cdots\ t_m \mid \delta \in \mathsf{hs}_\mathcal{S}(\xi, \emptyset) \,\}$$

where $\mathsf{hs}_\mathcal{S}$ is an auxiliary function defined by the following:

$\mathsf{hs}_\mathcal{S}(k, X) = \{k\}$ (whenever $k \in \Sigma \cup \mathcal{N}$)

$\mathsf{hs}_\mathcal{S}(x, X) =$ if $x \in X$ then $\emptyset$ else

$$\{\, \delta\ t_1\ \cdots\ t_m \mid x \mapsto \zeta\ t_1\ \cdots\ t_m \in \mathcal{S},\ \delta \in \mathsf{hs}_\mathcal{S}(\zeta, X \cup \{x\}) \,\}$$

Thus $\mathsf{head}_\mathcal{S}(u)$ is the set of terms that are obtainable from u by iteratively replacing the head symbol—provided it is a variable—by a term bound to it in $\mathcal{S}$. The second argument of $\mathsf{hs}_\mathcal{S}$ disregards any cyclic chain of bindings. For example, let $\mathcal{S} = \{\, x \mapsto y, y \mapsto x \,\}$, then: $\mathsf{head}_\mathcal{S}(x) = \mathsf{hs}_\mathcal{S}(x, \emptyset) = \mathsf{hs}_\mathcal{S}(y, \{\, x \,\}) = \mathsf{hs}_\mathcal{S}(x, \{\, x, y \,\}) = \emptyset$

Example 5. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be as in Example 4. Then:

$$\mathsf{head}_{\mathcal{S}_1}(x\ \mathtt{c}) = \{\, N\ \mathtt{c},\ \mathtt{f}\ z\ \mathtt{b}\ \mathtt{c} \,\}$$

$$\mathsf{head}_{\mathcal{S}_2}(x\ \mathtt{c}) = \{\mathtt{f}\ y\ z\ \mathtt{c}\} \quad \mathsf{head}_{\mathcal{S}_2}(F\ x\ (G\ x)) = \{F\ x\ (G\ x)\}$$

Notice that since head performs variable-substitutions according to bindings from $\mathcal{S}$, its behaviour is consistent with a strategy

for constructing initial prefixes of derivations in the system RS. Each use of the recursive clause of $\mathsf{hs}_{\mathcal{S}}$ corresponds to a head-instance of rule (S). A consequence of this relationship is made precise by the following lemma.

Lemma 1. *If $u \preccurlyeq_{\mathcal{S}} \xi\ v_1\ \cdots\ v_m$ then there is a compound term $\xi\ u_1\ \cdots\ u_m \in \mathsf{head}_{\mathcal{S}}(u)$ and, for all $1 \leq i \leq m$, $u_i \preccurlyeq_{\mathcal{S}} v_i$.*

One final property to note about head is that, whenever its argument is compound and all the variables in $\mathcal{S}$ are bound to simple terms, the terms in (sets in) its image are all compound. This is due to the fact that, in this case, the action of the head function is to construct new, compound terms by prepending old, simple terms into head position. This limited behaviour of the head-function will contribute towards guaranteeing the termination of the analysis.

Lemma 2. *We say that a set of bindings $\mathcal{S}$ is image-simple just if every term in the image of $\mathcal{S}$ is simple. Suppose $\mathcal{S}$ is image-simple. If u is compound, then every term in $\mathsf{head}_{\mathcal{S}}(u)$ is compound.*

The goal of the analysis is to discover the possible redexes $F\theta x_1\ \cdots\ \theta x_m\ \theta p$ that occur during reduction sequences of $\mathcal{P}_{\mathcal{G}}$ starting from $Main\ \ S$. The head function $\mathsf{head}_{\mathcal{S}}(u)$ is able to determine, in a way that is computable, when an F-redex is an instance (according to $\mathcal{S}$) of a simple term u. In this case, according to Lemma 1, a term of the shape $F\ t_1\ \cdots\ t_m\ s$ is an element of $\mathsf{head}_{\mathcal{S}}(u)$. However, to know which defining rule of F is triggered, it is necessary to find out which patterns are matched by residuals of instances of s.

The approximate reduction. To this end, we introduce a new notion of reduction $\triangleright_{\mathcal{S}} \subseteq T(\Sigma,\mathcal{N},\mathcal{V}) \times T(\Sigma,\mathcal{N},\mathcal{V})$ parametrised by a set of bindings $\mathcal{S}$. This reduction approximates the usual PMRS reduction by performing redex/contractum substitutions only where absolutely necessary and only when the relevant bindings are contained in $\mathcal{S}$. A *$\triangleright_{\mathcal{S}}$-redex* is a term of the form $F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ whenever there is a $\mathcal{P}_{\mathcal{G}_0}$-rule of the form $F\ x_1\ \cdots\ x_m\ p \longrightarrow t$ and θ is a substitution (not necessarily closed). The contractum of the redex is t, no substitution is performed upon contraction.

We define the one step reduction $\triangleright_{\mathcal{S}}$ by the following rules. Let C range over one-hole contexts.

$$\frac{(s,t)\text{ a }\triangleright\text{-redex/contractum pair}}{C[s] \triangleright_{\mathcal{S}} C[t]} \qquad \frac{t \in \mathsf{head}_{\mathcal{S}}(x\,t_1\ \cdots\ t_m)}{C[x\ t_1\ \cdots\ t_m] \triangleright_{\mathcal{S}} C[t]}$$

As is standard, we write $\triangleright_{\mathcal{S}}^*$ to mean the reflexive, transitive closure of $\triangleright_{\mathcal{S}}$, and $\triangleright_{\mathcal{S}}^n$ to mean a n-long chain of $\triangleright_{\mathcal{S}}$.

Example 6. Consider the composite PMRS $\mathcal{P}_{\mathcal{G}}$ constructed from the PMRS and grammar given in Example 2 and let $\mathcal{S}$ contain the bindings $p \mapsto Nz$ and $x \mapsto N$. Then the following:

$$p\ x \triangleright_{\mathcal{S}} Nz\ x \triangleright_{\mathcal{S}} Nz\ N\ \triangleright_{\mathcal{S}} Nz\ (\mathtt{s}\ N) \triangleright_{\mathcal{S}} \mathtt{true}$$

is a $\triangleright_{\mathcal{S}}$-reduction. Observe how, as demonstrated by the third step, approximate reduction is accurate for order-0 $\mathcal{G}$-rules.

Given a substitution θ and a pattern p, we say that a $\triangleright_{\mathcal{S}}$-reduction $s \triangleright_{\mathcal{S}}^i \theta p$ is *minimal* just if it is *not* the case that there exist $j < i$ and substitution θ' such that $s \triangleright_{\mathcal{S}}^j \theta'\, p$. Consider the two rules defining $\triangleright_{\mathcal{S}}$-reduction. In the RHS of the conclusion of each rule is the term t. In both cases, assuming $\mathcal{S}$ is image simple, t is a compound term. Since there are only finitely many such terms t and since there are only finitely many patterns (drawn from the PMRS) p, the problem of finding such *minimal* reductions is computable.

Lemma 3. *Assume $\mathcal{S}$ is image-simple. Given a compound term s and a pattern p drawn from the defining rules of $\mathcal{P}_{\mathcal{G}}$, the problem of finding a substitution θ and a minimal reduction $s \triangleright_{\mathcal{S}}^* \theta p$ is computable.*

The fixpoint construction. Let $\mathcal{S}$ be a set of bindings. We define $\mathcal{F}(\mathcal{S})$ as the least set X of bindings that contains $\mathcal{S}$ and is closed under ***Rule* C**: if

(i) u is simple term of base type,

(ii) $F\ t_1\ \cdots\ t_m\ s \in \mathsf{head}_{\mathcal{S}}(u)$,

(iii) $F\ x_1\ \cdots\ x_m\ p \longrightarrow t$ is a $\mathcal{P}_{\mathcal{G}_0}$-rule,

(iv) there is a minimal reduction $s \triangleright_{\mathcal{S}}^* \theta\, p$

then $\theta \cup \{\, x_i \mapsto t_i \mid 1 \leq i \leq m \,\} \subseteq X$.

Thus $\mathcal{F} : 2^{\mathcal{V}\times T(\Sigma,\mathcal{N},\mathcal{V})} \longrightarrow 2^{\mathcal{V}\times T(\Sigma,\mathcal{N},\mathcal{V})}$ is, by construction, a monotone (endo)function on the complete lattice $2^{\mathcal{V}\times T(\Sigma,\mathcal{N},\mathcal{V})}$ ordered by subset-inclusion. By the Tarski-Knaster Fixpoint Theorem, the least fixpoint of $\mathcal{F}$, which we shall denote Ξ, exists, and is constructable as the supremum of the chain

$$\emptyset \subseteq \mathcal{F}\,\emptyset \subseteq \mathcal{F}(\mathcal{F}\,\emptyset) \subseteq \mathcal{F}(\mathcal{F}(\mathcal{F}\,\emptyset)) \subseteq \cdots$$

Example 7. Consider again the composite PMRS $\mathcal{P}_{\mathcal{G}}$ composed from the PMRS $\mathcal{P}$ and tree grammar $\mathcal{G}$ given in Example 2. We shall apply the fixpoint construction to this structure.

Initially, the only fruitful choice of simple term is the "starting term" $Main\ \ S$ which otherwise trivially satisfies the premises of Rule **C** and yields the single binding $m \mapsto S$. Subsequently, taking $u = Filter\ Nz\ m$ matches both the defining rules for $Filter$ after approximate-reductions of:

$$m \triangleright_{\{m\mapsto S\}}^* \mathtt{nil} \qquad \text{and} \qquad m \triangleright_{\{m\mapsto S\}}^* \mathtt{cons}\ N\ ListN$$

respectively. This choice adds the bindings $p \mapsto Nz$, $x \mapsto N$ and $xs \mapsto ListN$. Examining the term $p\ \ x$ in the RHS of the second defining rule for $Filter$ then gives $n \mapsto N$. Finally, taking u as the entire RHS of the second defining rule for $Filter$ and approximate-reducing $p\ \ x$ as in Example 6 gives bindings $a \mapsto \mathtt{cons}\ x\ (Filter\ p\ xs)$ and $b \mapsto Filter\ p\ xs$. In this case, no other choices of simple term yield any new bindings, so the fixpoint Ξ is obtained as:

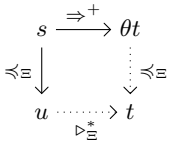
$$m \mapsto S, \quad p \mapsto Nz, \quad x \mapsto N, \quad xs \mapsto ListN$$
$$n \mapsto N, \quad a \mapsto \mathtt{cons}\ x\ (Filter\ p\ xs), \quad b \mapsto Filter\ p\ xs$$

Though the complete lattice $2^{\mathcal{V}\times T(\Sigma,\mathcal{N},\mathcal{V})}$ is infinite, the least fixpoint Ξ is finitely constructable (i.e. the closure ordinal of $\mathcal{F}$ is finite); it is in fact a finite set. Observe that, in Example 7, the form of every binding in the fixpoint is $v \mapsto t$ in which t is a simple term. This is the key to showing the convergence of the analysis. Since every term $F\ t_1\ \cdots\ t_m\ s \in \mathsf{head}_{\mathcal{S}}(u)$ is compound (whenever $\mathcal{S}$ is image-simple and u is compound) so every binding $x_i \rightarrow t_i$ is image-simple. Since, whenever $\mathcal{S}$ is image-simple, every $\triangleright_{\mathcal{S}}$-contractum is compound, so the bindings due to θp are image-simple. Since there are only finitely many simple terms, termination follows.

Theorem 1 (Termination)**.** *The least fixpoint of $\mathcal{F}$, Ξ, is a finite set.*

To see that this finite set of bindings Ξ is sufficient to describe all the all the substitutions that occur during redex contractions in reduction sequences of $\mathcal{P}_{\mathcal{G}}$ starting from $Main\ \ S$, one should first notice that the approximate reduction, when instantiated with the fixpoint, acts on simple terms in a way which is consistent with the way PMRS reduction acts on their instances in a trivial context.

Lemma 4. *Assume θt is a contractum and u is a simple term. If $s \Rightarrow^+ \theta t$ and $u \preccurlyeq_{\Xi} s$, then $u \triangleright_{\Xi}^* t$ and $t \preccurlyeq_{\Xi} \theta t$.*



To lift this fact to the level of arbitrary reduction sequences starting from $Main\ S$, it is enough to observe that any redex in such a sequence (apart from the first), can be seen either to be itself a simple term or to arise as a subterm of some previous contractum, regardless of the context in which the redex occurs. As a consequence of Lemma 4, the variable-term bindings necessary to derive the redex as an instance of the corresponding simple term will already be contained in the fixpoint. Hence, if the reduction sequence reaches any contractum, the fixpoint will contain the bindings necessary to reconstruct the substitution associated with the contraction.

Lemma 5. *Assume θt is a contractum. If $Main\ S \Rightarrow^+ C[\theta t]$ is a $\mathcal{P}_\mathcal{G}$-reduction sequence then $t \preccurlyeq_\Xi \theta t$.*

3.2 Construction of the over-approximating wPMRS

We are now ready to define the wPMRS which is an abstraction of the composite PMRS $\mathcal{P}_\mathcal{G} = \langle \Sigma, \mathcal{N}, \mathcal{R}, Main \rangle$. Let Ξ be the fixpoint set of bindings and let $\mathcal{N}_\mathcal{V} = \{ V_x \mid x \in \mathcal{V} \}$ and $\mathcal{N}_\Sigma = \{ K_\mathtt{a} \mid \mathtt{a} \in \Sigma \}$ be two sets of fresh non-terminal symbols which we call *pattern-symbols* and *accounting-symbols* respectively. We define the approximating wPMRS:

$$\widetilde{\mathcal{P}_\mathcal{G}} := \langle \Sigma,\ \mathcal{N} \cup \mathcal{N}_\mathcal{V} \cup \mathcal{N}_\Sigma,\ \mathcal{R}',\ Main \rangle$$

where $\mathcal{R}'$ consists of the following three kinds of rules:

I. *Weak pattern-matching rules*. For each (pure or pattern-matching) $\mathcal{P}_\mathcal{G}$-rule $F\, x_1 \cdots x_m\, p \longrightarrow t$, $\mathcal{R}'$ contains the following rule:

$$F\, x_1 \cdots x_m\, p \longrightarrow t^\dagger$$

II. *Instantiation rules*. For each binding $x \mapsto t$ in Ξ where $\mathsf{FV}(t^\dagger) = \{ x_1, \cdots, x_l \}$, $\mathcal{R}'$ contains the following rule:

$$V_x\, z_1 \cdots z_{\mathsf{ar}(x)} \longrightarrow (t^\dagger[V_{x_1}/x_1] \cdots [V_{x_l}/x_l])\, z_1 \cdots z_{\mathsf{ar}(x)}$$

where each z_i is a fresh variable of the appropriate types.

III. *Accounting rules*. For each terminal symbol $\mathtt{a} : b_1 \to \cdots \to b_n \to o$ in Σ, $\mathcal{R}'$ contains the following rule:

$$K_\mathtt{a}\ z_1\ \cdots\ z_n \longrightarrow \mathtt{a}\ z_1\ \cdots\ z_n$$

where each z_i is a fresh variable of type b_i.

where we have written $t^\dagger$ to denote the term t in which every occurrence of a pattern matching variable $y \in \mathsf{FV}(p)$ has been replaced by the corresponding pattern-symbol V_y and every occurrence of a terminal symbol $\mathtt{a}$ has been replaced by the corresponding accounting-symbol $K_\mathtt{a}$.

Example 8. Consider the following order-2 PMRS, whose defining rules are given by:

$$\begin{array}{l} Main\ m \longrightarrow\ Map2\ KZero\ KOne\ m \\ \\ Map2\ \varphi\ \psi\ \mathtt{nil} \longrightarrow\ \mathtt{nil} \\ Map2\ \varphi\ \psi\ (\mathtt{cons}\ x\ xs) \longrightarrow\ \mathtt{cons}\ (\varphi\ x)\ (Map2\ \psi\ \varphi\ xs) \\ \\ KZero\ x_1 \longrightarrow\ \mathtt{0} \\ KOne\ x_2 \longrightarrow\ \mathtt{1} \end{array}$$

and input grammar $\mathcal{G}$ consisting of two rules:

$$S \longrightarrow\ \mathtt{nil} \mid \mathtt{cons}\ \mathtt{0}\ S$$

The function $Map2$ behaves like the standard Map function, except that it swaps the first two function arguments as it filters through the successive elements of the list argument. The reachable constructor terms are finite lists that are prefixes of $\mathtt{[0\ 1\ 0\ 1\ 0\ 1\ \cdots]}$.

After applying the fixpoint construction to this example, the set of bindings Ξ consists of the following:

$$\begin{array}{cccccccccc} m & \mapsto & S & \quad & \varphi & \mapsto & KZero & \quad & \varphi & \mapsto & \psi \\ x & \mapsto & \mathtt{0} & \quad & \psi & \mapsto & KOne & \quad & \psi & \mapsto & \varphi \\ x_1 & \mapsto & x & \quad & x_2 & \mapsto & x & \quad & xs & \mapsto & S \end{array}$$

and hence the approximating wPMRS $\widetilde{\mathcal{P}_\mathcal{G}}$ is as follows.

$$\begin{array}{l} Main\ m \longrightarrow\ Map2\ KZero\ KOne\ M \\ \\ Map2\ \varphi\ \psi\ \mathtt{nil} \longrightarrow\ Nil \\ Map2\ \varphi\ \psi\ (\mathtt{cons}\ x\ xs) \longrightarrow\ Cons\ (\varphi\ X)\ (Map2\ \psi\ \varphi\ XS) \\ \\ KZero\ x_1 \longrightarrow\ Zero \\ KOne\ x_2 \longrightarrow\ One \\ \\ M \longrightarrow\ S \\ X \longrightarrow\ Zero \\ XS \longrightarrow\ S \\ S \longrightarrow\ Nil \mid Cons\ Zero\ S \\ \\ Zero \longrightarrow\ \mathtt{0} \\ One \longrightarrow\ \mathtt{1} \\ Nil \longrightarrow\ \mathtt{nil} \\ \\ Cons\ v_1\ v_2 \longrightarrow\ \mathtt{cons}\ v_1\ v_2 \end{array}$$

Since φ, ψ, x_1 and x_2 are not pattern-matched variables, the rules for V_φ, V_ψ, V_{x_1} and V_{x_2} are, in this case, never used and so play no part in the approximation process: they have been omitted. It is easy to see that the constructor terms in $\mathcal{L}(\widetilde{\mathcal{P}_\mathcal{G}}, Main\ S)$ are exactly the finite prefixes of $\mathtt{[0\ 1\ 0\ 1\ \cdots]}$ i.e. the approximation is exact in this case.

Given any $\mathcal{P}_\mathcal{G}$-reduction $Main\ S \Rightarrow^+ t$, the reduction can be faithfully simulated in the abstraction $\widetilde{\mathcal{P}_\mathcal{G}}$ using the weak pattern-matching rules and the instantiation rules. Whenever the $\mathcal{P}_\mathcal{G}$-reduction contracts a $\mathcal{P}$-rule which binds data θy to a pattern matching variable y, the simulation can contract the corresponding redex using a weak pattern-matching rule and, by Lemma 5, can then reconstruct the bound data θy from V_y using the instantiation rules.

Theorem 2 (Soundness)**.** *Let the composite PMRS $\mathcal{P}_\mathcal{G}$ and the approximating wPMRS $\widetilde{\mathcal{P}_\mathcal{G}}$ be as before. Then $\mathcal{L}(\mathcal{P}_\mathcal{G}, Main\ S) \subseteq \mathcal{L}(\widetilde{\mathcal{P}_\mathcal{G}}, Main\ S)$.*

The third class of rules is not essential to the achieving soundness. The purpose of the accounting rules is to enforce a strict correspondence between the length of a $\widetilde{\mathcal{P}_\mathcal{G}}$ reduction sequence and the maximum size of any constructor term created within it. This eases the justification of the semi-completeness property of refinement in Section 5.

4. Model checking by type inference

In this section, we exhibit an algorithm to decide the *wPMRS Model Checking Problem*: given a non-deterministic wPMRS $\mathcal{W} = \langle \Sigma, \mathcal{N}, \mathcal{R}, Main \rangle$ in which $Main : b \to o$, a closed term $t : b$ and a trivial automaton $\mathcal{A}$, is $\mathcal{L}(\mathcal{W}, Main\ t) \subseteq \mathcal{L}(\mathcal{A})$? Following work by Kobayashi [6] and Kobayashi and Ong [8], we characterise the model checking problem as a type inference problem in a particular, finitary intersection type system induced by the automaton.

Eliminating non-determinism. The first step we take is to simplify the problem at hand by eliminating the non-determinism in $\mathcal{W}$. To this end we construct a new wPMRS $\mathcal{W}^{\#}$ in which multiple defining rules for a given non-terminal are collapsed using a family $\mathcal{B} := \{ \mathtt{br}_b \mid b \in \mathbb{B} \}$ of "non-deterministic choice" terminal symbols $\mathtt{br}_b$ of type $b \to b \to b$. We define:

$$\mathcal{W}^{\#} := \langle \Sigma \cup \mathcal{B}, \mathcal{N}, \{ l \longrightarrow \mathtt{BR}(l) \mid \exists r \cdot l \longrightarrow r \in \mathcal{R} \}, \mathit{Main} \rangle$$

in which, by way of a short-hand, we define:

$$\mathtt{BR}(F\ t_1\ \cdots\ t_n) := \mathtt{br}_b\ r_1\ (\mathtt{br}_b\ r_2\ (\cdots(\mathtt{br}_b\ r_{m-1}\ r_m)\cdots))$$

where $\{r_1, \ldots, r_m\} = \{ r \mid F\ t_1\ \cdots\ t_n \longrightarrow r \in \mathcal{R} \}$ and the type of F is of the form $\tau_1 \to\ \cdots\ \to \tau_n \to b$. We must modify the automaton $\mathcal{A}$ accordingly, so we define:

$$\mathcal{A}^{\#} := \langle \Sigma \cup \mathcal{B}, Q, \Delta \cup \{ (q, \mathtt{br_b}, q\,q \mid q \in Q, b \in \mathbb{B} \}, q_0 \rangle$$

Lemma 6. *For all terms t of base-type:*

$$\mathcal{L}(\mathcal{W}, \mathit{Main}\ t) \subseteq \mathcal{L}(\mathcal{A}) \qquad \textit{iff} \quad [\![\mathit{Main}\ t]\!]_{\mathcal{W}^{\#}} \in \mathcal{L}(\mathcal{A}^{\#})$$

Model checking as type inference. We first introduce recursion schemes with weak definition-by-cases, which is a term rewriting system similar to (in fact, equi-expressive with) wPMRS; the difference is that (weak) matching is explicitly provided by a case construct. Assume for each base type b, an exhaustive and non-overlapping family of patterns $P_b = \{ p_1, \cdots, p_k \}$. A *recursion scheme with weak definition-by-cases* (wRSC) is a quadruple $\mathcal{G} = \langle \Sigma, \mathcal{N}, \mathcal{R}, S \rangle$ where $\Sigma, \mathcal{N}$, and S are as usual, and $\mathcal{R}$ is a set of (pure) rules of the form

$$F\, x_1 \cdots x_m \longrightarrow t$$

We write $\mathbf{rhs}(F) = \lambda x_1 \cdots x_m.t$. The set of applicative terms is defined as before, except that it is augmented by a definition-by-cases construct $\mathsf{case}_b(t; t_1, \cdots, t_k)$ with typing rule:

$$\frac{\Gamma \vdash t : b \qquad \Gamma \vdash t_i, o \text{ (for } 1 \le i \le k)}{\Gamma \vdash \mathsf{case}_b(t; t_1, \cdots, t_k) : o}$$

We say that $\mathcal{G}$ is deterministic just if there is one rule for each $F \in \mathcal{N}$. There are two kinds of redexes:

(i) $F\, s_1 \cdots s_m$ which contracts to $t[s_1/x_1] \cdots [s_m/x_m]$ for each rule $F\, x_1 \cdots x_m \longrightarrow t$ in $\mathcal{R}$

(ii) $\mathsf{case}_b(t; t_1, \cdots, t_k)$ which contracts to t_i, provided t of base type b matches pattern $p_i \in P_b = \{ p_1, \cdots, p_k \}$.

We define evaluation contexts E as follows

$$E \ ::= \ [\,] \mid f\, t_1 \cdots t_{i-1}\, E\, t_{i+1} \cdots t_{\mathsf{ar}(f)}$$

and write $\to$ for the one-step reduction relation $E[\Delta] \to E[\Delta]$ where (Δ, Δ) ranges over redex/contractum pairs and E over evaluation contexts. Assuming $\mathcal{G}$ is deterministic, we define the $\Sigma^{\perp}$-labelled tree generated by $\mathcal{G}$ by infinitary rewriting from S as $[\![\mathcal{G}]\!] := \{ t^{\perp} \mid S \to^* t \}$.

Lemma 7. *Deterministic wPMRS and deterministic wRSC are equi-expressive as generators of Σ-labelled trees.*

We present an intersection type system for characterising the model checking problem. The *intersection types* of the system are given by the grammar:

$$\sigma, \tau \quad ::= \quad q \ \mid\ p \ \mid\ \bigwedge_{i=1}^{m} \tau_i \to \tau$$

where $q \in Q$ and p is one of the finitely many patterns associated with a definition by cases in the scheme $\mathcal{G}$. Judgements of the type system are sequents of the form $\Gamma \vdash t : \tau$, in which Γ is simply a set of type bindings $\xi : \sigma$ where $\xi \in \mathcal{N} \cup \mathcal{V}$. The defining rules of the system are as follows:

$$\frac{}{\Gamma, x : \tau \vdash x : \tau} \qquad \text{(VAR)}$$

$$\frac{(q,\ f,\ q_1 \cdots q_n) \in \Delta_{\mathcal{A}^{\#}}}{\Gamma \vdash f : q_1 \to \cdots \to q_n \to q} \qquad \text{(TERM)}$$

$$\frac{\exists \theta \cdot s\ p_1\ \cdots\ p_n = \theta\, p}{\Gamma \vdash s : p_1 \to \cdots \to p_n \to p} \qquad \text{(MATCH)}$$

$$\frac{\Gamma \vdash t : p_i \qquad \Gamma \vdash t_i : \tau}{\Gamma \vdash \mathsf{case}_b(t;\ t_1, \cdots, t_i, \cdots, t_n) : \tau} \qquad \text{(CASE)}$$

$$\frac{\Gamma \vdash s : \bigwedge_{i=1}^{n} \tau_i \to \tau \qquad \Gamma \vdash t : \tau_i \quad (\text{for each } 1 \le i \le n)}{\Gamma \vdash s\ t : \tau} \qquad \text{(APP)}$$

$$\frac{\Gamma, x : \tau_1, \cdots, x : \tau_n \vdash t : \tau}{\Gamma \vdash \lambda x.t : \bigwedge_{i=1}^{n} \tau_i \to \tau} \qquad \text{(ABS)}$$

Note that we have the following derived rule from (Match): if a term s (of the appropriate type) matches the pattern p, then $\Gamma \vdash s : p$.

We write $\vdash_{\mathcal{A}} \mathcal{G} : \Gamma$ if $\Gamma \vdash \mathbf{rhs}(F) : \tau$ is provable for every $F : \tau \in \Gamma$. A wRSC is *well-typed*, written $\vdash_{\mathcal{A}} \mathcal{G}$, just if there exists Γ such that (i) $\vdash_{\mathcal{A}} \mathcal{G} : \Gamma$, (ii) $S : q_0 \in \Gamma$, (iii) for each $F : \tau \in \Gamma$, $\tau :: \kappa$, where $F : \kappa \in \mathcal{N}$, meaning that τ is an intersection type *compatible* with type κ (as assigned to F by the wRSC), which is defined by: (i) $q :: o$, (ii) $p :: b$ for each $p \in P_b$, (iii) $\bigwedge_{i=1}^{k} \tau_i \to \tau :: \kappa' \to \kappa$ if $\tau :: \kappa$ and for each $1 \le i \le k$, $\tau_i :: \kappa'$.

Theorem 3. *Let $\mathcal{A}$ be a trivial automaton, and $\mathcal{G}$ be wRSC. Then $\vdash_{\mathcal{A}} \mathcal{G}$ if and only if $[\![\mathcal{G}]\!] \in \mathcal{L}(\mathcal{A})$.*

The proof is omitted as it is very similar to the proof of the soundness and completeness theorems in [6].

Corollary 1. *The wPMRS model checking problem is decidable.*

Proof. This follows from Lemma 6, Lemma 7 and Theorem 3, and the decidability of typability $\vdash_{\mathcal{A}} \mathcal{G}$. The latter follows from the fact that for each non-terminal, there are only finitely many candidate intersection types compatible with a given type. □

5. Abstraction refinement

When the model checking stage reports a counterexample in the form of an error trace, the trace may be "feasible", that is, it corresponds to a concrete reduction sequence in the original PMRS $\mathcal{P}$, or it may be "spurious": an artifact of the abstraction process. In the case the counterexample is spurious, we will want to ignore it and perform the process again, but in a new setting in which we are guaranteed never again to encounter this unwanted trace. To achieve this we restart the cycle from a modified PMRS $\mathcal{P}'$, which has had some of its defining rules unfolded, so as to reduce the amount of non-determinism in the corresponding wPMRS abstraction.

5.1 Counterexamples and feasibility.

When the model-checker reports a violation of the property, a counterexample error trace is returned. This error trace is a reduction sequence in the abstract wPMRS $\widetilde{\mathcal{P}_\mathcal{G}}$. Since $\widetilde{\mathcal{P}_\mathcal{G}}$ is not completely faithful to the PMRS $\mathcal{P}$, it is necessary to determine whether such a counterexample trace corresponds to a reduction sequence in $\mathcal{P}$ which itself witnesses the violation or whether it is an artifact of the abstraction.

Anatomy of a counterexample. It is useful to highlight two important features of any given counterexample trace, namely, (i) the shape of the last term in the reduction sequence and (ii) the "type" of each constituent reduction .

Any counterexample trace must end in a term t which witnesses the violation of the property φ. Since the property is a collection of (possibly infinite) Σ-labelled trees, the witnessing term can be seen to be of the form θq where q is a pattern which does not match any prefix of a tree $t \in \varphi$. We say that the pattern q which witnesses the violation of the property is the *error witness*.

In any $\widetilde{\mathcal{P}_\mathcal{G}}$ reduction sequence, each reduction $u \to v$ can be classified into one of two kinds based on the head symbol occurring in the redex. In case we want to emphasise that the head symbol is a non-terminal belonging to $\mathcal{P}$ we say the contraction of this redex is an *abstract $\mathcal{P}$-reduction* and write $u \to_\mathcal{P} v$. Otherwise the head symbol is either a pattern-symbol, an accounting-symbol or it belongs to $\mathcal{G}$. In this case we say that the head symbol in question is a *live-symbol* and that the contraction of this redex is an *abstract $\widetilde{\mathcal{P}_\mathcal{G}}$-reduction*; we write $u \to_{\widetilde{\mathcal{P}_\mathcal{G}}} v$.

Example 9. Consider the following abstract error trace which is derived from the abstraction $\widetilde{\mathcal{P}_\mathcal{G}}$ of the PMRS $\mathcal{P}$ and grammar $\mathcal{G}$ given in Example 2:

$Main\ S$
$\to\ \ Filter\ Nz\ M$
$\to^*\ Filter\ Nz\ (\texttt{cons}\ N\ ListN)$
$\to\ \ If\ (Cons\ X\ (Filter\ Nz\ XS))\ (Filter\ Nz\ XS)\ (Nz\ X)$
$\to^*\ If\ (Cons\ X\ (Filter\ Nz\ XS))\ (Filter\ Nz\ XS)\ (Nz\ (\mathtt{s}\ N))$
$\to\ \ If\ (Cons\ X\ (Filter\ Nz\ XS))\ (Filter\ Nz\ XS)\ True$
$\to^*\ \ Cons\ X\ (Filter\ Nz\ XS)$
$\to\ \ \texttt{cons}\ X\ (Filter\ Nz\ XS)$
$\to^*\ \texttt{cons}\ \mathtt{z}\ (Filter\ Nz\ XS)$

which violates the property since it is the start of a list that contains a zero. The *error-witness* for this trace is $\texttt{cons}\ \mathtt{z}\ v$ (for some variable v). The first reduction is an *abstract $\mathcal{P}$-reduction*, as is the reduction written over lines 3 and 4 and that of lines 5 and 6. All the other reductions in the sequence are *abstract $\widetilde{\mathcal{P}_\mathcal{G}}$-reductions*.

The trace in the above example is spurious since there are no reduction sequences of the PMRS $\mathcal{P}$ (starting from terms in $\mathcal{L}(\mathcal{G})$) from Example 2 which result in a list headed by a zero. Intuitively, we can see that this trace is infeasible because the non-determinism introduced by the abstraction has been resolved in an inconsistent way during the sequence. The data bound by the pattern match for $Filter$, which is given as N (i.e. some number) has been resolved on the one hand (line 5) to a non-zero number and on the other hand (line 9) to zero.

In the following, we define a process of labelling of the counterexample trace that will reveal information about the resolution of non-determinism that has been introduced as a consequence of the abstraction. The information that is exposed will allow us to see whether or not this abstract trace in $\widetilde{\mathcal{P}_\mathcal{G}}$ has any corresponding trace in $\mathcal{P}$ starting form $\mathcal{I}$, that is, whether the trace is feasible.

Labelling. The labelling procedure, $\mathsf{labelSeq}$, keeps track of how non-determinism is resolved in an abstract reduction sequence by annotating each live-symbol X with a set of (possibly open) terms, which represent all the closed terms to which it reduces. When the terms are given by the set l, we write the annotated term X^l and we identify an unlabelled live-symbol X with $X^\emptyset$. Given a term t which may include labelled subterms, we define the *resolution* of t, which is a set of terms $\bar{t}$, defined as follows:

$$\bar{t} = \begin{cases} \{\mathtt{a}\} & \text{when } t = \mathtt{a} \in \Sigma \\ \{F\} & \text{when } t = F \text{ is not a live-symbol} \\ \hat{l} & \text{when } t = F^l \text{ is a live-symbol} \\ \{\, u\ v \mid u \in \overline{t_0},\, v \in \overline{t_1}\,\} & \text{when } t = t_0\ t_1 \end{cases}$$

where $\hat{l}$ denotes the set l when l is non-empty and $\{z\}$ for some fresh variable z otherwise. If any pattern-symbol reduces to two incompatible terms or to a term which is inconsistent with the term that it represents in the matching, then the procedure will detect a conflict and record it in the set $\mathsf{Failures}$.

$\mathsf{labelSeq}(Main\ S)$

If S is labelled by l and there is a term $t \in \mathcal{L}(\mathcal{G})$ which is an instance of $\mathsf{mgci}(l)$ then do nothing else add $(Main, l)$ to $\mathsf{Failures}$.

$\mathsf{labelSeq}(Main\ S \to^* u \to v)$

1. Analyse the reduction $u \to v$:
   $C[F\ \theta x_1\ \cdots\ \theta x_m] \to_{\widetilde{\mathcal{P}_\mathcal{G}}} C[t\ \theta x_1\ \cdots\ \theta x_m]$:
   Label the head symbol F by $\bar{t}$.
   $C[F\ \theta x_1\ \cdots\ \theta x_m\ \theta p] \to_\mathcal{P} C[\theta t^\dagger]$:
   For each $y \in \mathsf{FV}(p)$, let $\{V_y^{l_1}, \ldots, V_y^{l_k}\}$ be the labelled pattern-symbols in v created by the contraction. Perform $\mathsf{labelTm}(\theta y)(\bigcup\{l_1, \ldots, l_k\})$ on the corresponding occurrence of θy in θp. If $\mathsf{labelTm}$ fails, then add $(F, \bigcup\{l_1, \ldots, l_k\})$ to $\mathsf{Failures}$.
2. For each occurrence of an unlabelled live-symbol N in u, let $\{N_1^{l_1}, \ldots, N_k^{l_k}\}$ be the set of labelled descendants in v. Label this occurrence of N with $\bigcup\{l_1, \ldots, l_k\}$.
3. Perform $\mathsf{labelSeq}(Main\ S \to^* u)$.

where the procedure $\mathsf{labelTm}$, which is designed to resolve the data bound in a pattern match and the data created by the abstraction, is given by:

$\mathsf{labelTm}(t)(\{s_1, \ldots, s_k\})$

Analyse the form of t:

$t = \mathtt{a}$**:** If $\forall i$ $\mathtt{a}$ matches s_i then do nothing else fail.

$t = F$**:** If F is not a live-symbol and $\forall i$ F matches s_i then do nothing else, if F is a live-symbol and $w = \mathsf{mgci}(\{s_1, \ldots, s_k\})$ exists then label F by $\{w\}$ else fail.

$t = t_0\ t_1$**:** If $\forall i$ either s_i is a variable or $s_i = s_{i_0}\ s_{i_1}$ then (let $s_{j_0} = z_0$ and $s_{j_1} = z_1$ for fresh z_1, z_2 whenever s_j is a variable) and perform $\mathsf{labelTm}(t_0)(\{s_{1_0}, \ldots, s_{k_0}\})$ and perform $\mathsf{labelTm}(t_1)(\{s_{1_1}, \ldots, s_{k_1}\})$ else fail.

where $\mathsf{mgci}(l)$ denotes the most general common instance (MGCI) of the set of terms l (regarding a single fresh variable as the MGCI of the empty set)[1]. We call a counterexample trace that has been labelled by $\mathsf{labelSeq}$ a *labelled trace*.

[1] For the purposes of calculating MGCIs, terms are considered as first order entities constructed from atomic constants and a single (silent) application operator.

Example 10. Consider again the abstract reduction sequence in Example 9, after performing labelSeq the following labelled trace is produced (the set bracket notation has been elided since all labels are singleton sets):

$Main\ S^{\mathtt{cons}\ v_1\ v_2}$
$\rightarrow\ \ Filter\ Nz\ M^{\mathtt{cons}\ v_1\ v_2}$
$\rightarrow^*\ Filter\ Nz\ (\mathtt{cons}\ N\ ListN^{v_2})$
$\rightarrow\ \ If\ (Cons^{\mathtt{cons}}\ X^{\mathtt{z}}\ (Filter\ Nz\ XS))\ (Filter\ Nz\ XS)\ (Nz\ X^{\mathtt{s}\ v_0})$
$\rightarrow^*\ If\ (Cons^{\mathtt{cons}}\ X^{\mathtt{z}}\ (Filter\ Nz\ XS))\ (Filter\ Nz\ XS)\ (Nz\ (\mathtt{s}\ N^{v_0}))$
$\rightarrow\ \ If\ (Cons^{\mathtt{cons}}\ X^{\mathtt{z}}\ (Filter\ Nz\ XS))\ (Filter\ Nz\ XS)\ True^{\mathtt{true}}$
$\rightarrow^*\ Cons^{\mathtt{cons}}\ X^{\mathtt{z}}\ (Filter\ Nz\ XS)$
$\rightarrow\ \ \mathtt{cons}\ X^{\mathtt{z}}\ (Filter\ Nz\ XS)$
$\rightarrow^*\ \mathtt{cons}\ \mathtt{z}\ (Filter\ Nz\ XS)$

After labelling, we have $\mathsf{Failures} = \{(Filter, \{\mathtt{z}, \mathtt{s}\ v_0\})\}$. Observe that there are no labels on N in line 2, due to the fact that labelTm failed.

Feasibility. For a trace α in $\widetilde{\mathcal{P}_\mathcal{G}}$ to be feasible two properties are required. First, the non-determinism introduced by the abstraction should be well behaved and second, there should be a term in the input that is able to trigger the trace, i.e. when given as an argument to $Main$, the rest of the trace follows. The first of these conditions is the subject of the step case in labelSeq, the second is the subject of the base case. Hence, if after performing $\mathsf{labelSeq}(\alpha)$ it is the case that $\mathsf{Failures} = \emptyset$, then we say α is *feasible*. The justification is the following lemma.

Lemma 8. *Let α be a feasible reduction sequence in $\widetilde{\mathcal{P}_\mathcal{G}}$ with error-witness q. Then there exists a term $t \in \mathcal{L}(\mathcal{G})$ and a finite reduction sequence $Main\ t \Rightarrow \cdots$ in $\mathcal{P}$ with error witness q.*

The witness to soundness, which appears in the proof of Theorem 2, will always be feasible. We say that any trace that is not feasible is *spurious*.

5.2 Refinement

When a reduction sequence in $\widetilde{\mathcal{P}_\mathcal{G}}$ is shown to be spurious, the problem can always be traced back to an occurrence of pattern-matching (notice that, by definition, the single parameter of the defining rule for $Main$ is always a pattern). Since the only loss of accuracy in the abstraction is in the way that data bound in pattern matches is handled during reduction, our remedy for infeasibility is to increase precision in the pattern matching rules of $\widetilde{\mathcal{P}_\mathcal{G}}$.

Our strategy is based on the observation that, due to the particular way in which the abstract wPMRS is constructed from the composite PMRS, the terminal symbol-labelled parts of each pattern are accurately preserved in the RHS of the defining rules of the abstraction. Based on the depth of pattern matches in the counterexample trace, we unfold patterns in the defining rules of $\mathcal{P}$ in a way that preserves the the set of possible reduction sequences.

Pattern-matching depth. To determine how much to unfold we define a measure $\mathsf{depth} : T(\Sigma, \mathcal{N}, \mathcal{V}) \rightarrow \mathbb{N}$, which quantifies the extent to which a term can be matched, as follows:

$$\mathsf{depth}(x) = 0$$
$$\mathsf{depth}(F\ t_1\ \cdots\ t_m) = 1$$
$$\mathsf{depth}(a\ t_1\ \cdots\ t_m) = 1 + \bigsqcup_{\mathbb{N}} \{\, \mathsf{depth}(t_i) \mid 1 \leq i \leq m \,\}$$

Given a set of non-terminals $\mathcal{N}$, a *depth profile* for $\mathcal{N}$ is a map $\mathcal{N} \rightarrow \mathbb{N}$. We assign a depth profile to a set of rules to quantify, for each non-terminal F, how accurately the defining rules for F model pattern-matching. Given a set of rules $\mathcal{R}$ defining non-terminals from $\mathcal{N}$, let the depth profile of $\mathcal{R}$, denoted $\mathsf{dp}(\mathcal{R})$, be the function:

$$\mathsf{dp}(\mathcal{R})(F) = \bigsqcup_{\mathbb{N}} \{\, \mathsf{depth}(p) \mid F\ x_1\ \cdots\ x_m\ p \longrightarrow t \in \mathcal{R} \,\}$$

Depth profiles can be naturally ordered pointwise, so that if d and d' are depth profiles over the same domain $\mathcal{N}$, then $d \leq d'$ iff $d(F) \leq d'(F)$ for all $F \in \mathcal{N}$.

Unfolding. To capture the result of unfolding we first introduce two auxiliary definitions. To aid readability, in each of them we will annotate fresh variables with their implied types by a superscript. The set of *atomic patterns* of type b, $\mathbb{A}_b$ is the set $\{\, a\ z_1^{b_1}\ \cdots\ z_{\mathsf{ar}(a)}^{b_{\mathsf{ar}(a)}} \mid a : b_1 \rightarrow \cdots \rightarrow b_{\mathsf{ar}(a)} \rightarrow b \in \Sigma \,\}$. For each $n \in \mathbb{N}$ we define the non-overlapping, exhaustive set of patterns of type b and depth n, $\mathsf{pats}_b(n)$:

$$\mathsf{pats}_b(0) = \{z^b\}$$
$$\mathsf{pats}_b(n+1) = \{\, p[q_1, \ldots, q_m / x_1^{b_1}, \cdots, x_m^{b_m}] \mid \varphi \,\}$$

where φ stands for the conjunction:

$$p \in \mathsf{pats}_b(n)\ \ \&\ \ \mathsf{FV}(p) = \{x_1^{b_1}, \ldots, x_m^{b_m}\}\ \ \&\ \ \forall i\, q_i \in \mathbb{A}_{b_i}$$

Hence, the depth 2 family of patterns of type $\mathbf{natlist}$ are given by (n, x and xs arbitrary variables):

$$\mathtt{nil},\quad \mathtt{cons}\ \mathtt{z}\ \mathtt{nil},\quad \mathtt{cons}\ \mathtt{z}\ (\mathtt{cons}\ x\ xs),$$
$$\mathtt{cons}\ (\mathtt{s}\ n)\ \mathtt{nil},\quad \mathtt{cons}\ (\mathtt{s}\ n)\ (\mathtt{cons}\ x\ xs)$$

To unfold the rules of a PMRS $\mathcal{P}$ according to a depth profile d, one constructs a new PMRS $\mathcal{P}'$ whose rule-set is enlarged so that, for a given non-terminal F of type $\tau_1 \rightarrow \cdots \rightarrow \tau_m \rightarrow b \rightarrow o$, there is a number of defining rules which is equal to the number of patterns of type b and depth $d(F)$. For each of these rules the corresponding right-hand side is constructed by using the existing $\mathcal{P}$ rules as a template.

Let $\mathcal{P} = \langle \Sigma, \mathcal{N}, \mathcal{R}, Main \rangle$ be a PMRS and let d be a depth profile with domain $\mathcal{N}$ such that $\mathsf{dp}(\mathcal{R}) \leq d$. The d-unfolding of $\mathcal{P}$ is the PMRS $\langle \Sigma, \mathcal{N}, \mathcal{R}', Main \rangle$, where $\mathcal{R}'$ is the set such that, for all substitutions σ, $F\ x_1\ \cdots\ x_m\ \sigma p \longrightarrow \sigma t \in \mathcal{R}'$ iff:

(i) $F\ x_1\ \cdots\ x_m\ p \longrightarrow t \in \mathcal{R}$
(ii) and p is of type b
(iii) and $q \in \mathsf{pats}_b(d(F))$
(iv) and $q = \sigma p$

Example 11. Let $\mathcal{P}$ be as in Example 2 and let d be the depth profile given by the following rule:

$$d(F) = \begin{cases} 2 & \text{when } F = Filter \\ \mathsf{dp}(\mathcal{R})(F) & \text{otherwise} \end{cases}$$

Then the d-unfolding of $\mathcal{P}$ is the PMRS $\mathcal{P}'$, whose rules are the same as $\mathcal{P}$ except that the two rules for $Filter$ have been replaced by the five rules in Figure 3.

Consider an abstraction $\widetilde{\mathcal{P}'_\mathcal{G}}$ of the PMRS $\mathcal{P}'$ in Example 11 (with $\mathcal{G}$ as given in Example 2). The only non-determinism that is introduced in constructing the abstraction is in replacing the pattern-matching variables in the right-hand sides of the defining rules by pattern-symbols. Due to the unfolding of the $Filter$ rules in $\mathcal{P}'$ (and hence in $\widetilde{\mathcal{P}'_\mathcal{G}}$), there is no longer a possibility to make the problematic reduction:

$Filter\ (Nz\ (\mathtt{cons}\ N\ ListN)$
$\rightarrow\ \ If\ (Cons\ X\ (Filter\ Nz\ XS))\ (Filter\ Nz\ XS)\ (Nz\ X)$

since the unfolded rules require more of the non-determinism (in the non-terminal symbols N and $ListN$) to be resolved earlier.

$$
\begin{aligned}
\mathit{Filter}\ p\ \mathtt{nil} &\longrightarrow \mathtt{nil}\\
\mathit{Filter}\ p\ (\mathtt{cons}\ \mathtt{z}\ \mathtt{nil}) &\longrightarrow \mathit{If}\ (\mathtt{cons}\ \mathtt{z}\ (\mathit{Filter}\ p\ \mathtt{nil}))\ (\mathit{Filter}\ p\ \mathtt{nil})\ (p\ \mathtt{z})\\
\mathit{Filter}\ p\ (\mathtt{cons}\ \mathtt{z}\ (\mathtt{cons}\ v_0\ v_1)) &\longrightarrow \mathit{If}\ (\mathtt{cons}\ \mathtt{z}\ (\mathit{Filter}\ p\ (\mathtt{cons}\ v_0\ v_1)))\ (\mathit{Filter}\ p\ (\mathtt{cons}\ v_0\ v_1))\ (p\ \mathtt{z})\\
\mathit{Filter}\ p\ (\mathtt{cons}\ (\mathtt{s}\ v_2)\ \mathtt{nil}) &\longrightarrow \mathit{If}\ (\mathtt{cons}\ (\mathtt{s}\ v_2)\ (\mathit{Filter}\ p\ \mathtt{nil}))\ (\mathit{Filter}\ p\ \mathtt{nil})\ (p\ (\mathtt{s}\ v_2))\\
\mathit{Filter}\ p\ (\mathtt{cons}\ (\mathtt{s}\ v_3)\ (\mathtt{cons}\ v_4\ v_5)) &\longrightarrow \mathit{If}\ (\mathtt{cons}\ (\mathtt{s}\ v_3)\ (\mathit{Filter}\ p\ (\mathtt{cons}\ v_4\ v_5)))\ (\mathit{Filter}\ p\ (\mathtt{cons}\ v_4\ v_5))\ (p\ (\mathtt{s}\ v_3))
\end{aligned}
$$

Figure 3. Depth-2 unfolding of the defining rules for Filter.

Refinement. Given a PMRS $\mathcal{P}$ and an infeasible error trace α in the abstraction of $\mathcal{P}$, we can obtain refined abstractions by unfolding the rules of $\mathcal{P}$ according to the depths of terms in the Failures set, then using the unfolded PMRS as the input to the next cycle of the abstraction-refinement loop.

Lemma 9. *Let $\mathcal{P} = \langle \Sigma, \mathcal{N}, \mathcal{R}, \mathit{Main}\rangle$ be a PMRS and $\widetilde{\mathcal{P}_\mathcal{G}}$ be the abstraction of $\mathcal{P}$ (starting from terms in $\mathcal{L}(\mathcal{G})$). Let α be a counterexample trace of $\widetilde{\mathcal{P}_\mathcal{G}}$ which is spurious with* Failures *set S. Let d be the depth profile with domain $\mathcal{N}$ defined by:*

$$d(F) = \mathsf{dp}(\mathcal{R})(F) + \bigsqcup_{\mathbb{N}} \{\, \mathsf{depth}(t) \mid (F,\, P) \in S,\, t \in P \,\}$$

and let $\mathcal{P}'$ be the d-unfolding of $\mathcal{P}$. Then α is not a reduction sequence in the abstraction $\widetilde{\mathcal{P}'_\mathcal{G}}$ of $\mathcal{P}'$.

Although it is clear that, given any spurious trace in some abstraction $\widetilde{\mathcal{P}_\mathcal{G}}$, it is possible to construct a refinement that eliminates it from any future abstraction $\widetilde{\mathcal{P}'_\mathcal{G}}$, the set of traces of $\widetilde{\mathcal{P}_\mathcal{G}}$ and the set of traces of $\widetilde{\mathcal{P}'_\mathcal{G}}$ are incomparable since, in general, there are new pattern-variables introduced in the refinement and hence new pattern-symbols into $\widetilde{\mathcal{P}'_\mathcal{G}}$. However, there is a very close relationship between the depth of a PMRS and the feasibility of reduction sequences in its abstraction.

Lemma 10. *Fix $n \in \mathbb{N}$. Then given any PMRS $\mathcal{P}$ and input grammar $\mathcal{G}$, there is a depth-profile d such that, if $\widetilde{\mathcal{P}'_\mathcal{G}}$ is the abstraction of the d-unfolding of $\mathcal{P}$, then all length-$m \leq n$ reduction sequences in $\widetilde{\mathcal{P}'_\mathcal{G}}$ are feasible.*

A consequence of this close relationship between depth and feasibility is that, under the assumption that the model-checker always reports the shortest counterexample trace, if the PMRS $\mathcal{P}$ (when run from a term in $\mathcal{I}$) does violate the property then eventually the abstraction-refinement cycle will produce a feasible counterexample trace demonstrating the fact.

Theorem 4 (Semi-completeness)**.** *Assume the model checker always produces shortest counterexamples and let $(\mathcal{P}, \mathcal{I}, \mathcal{A})$ be a no-instance of the verification problem. Then the algorithm terminates with a feasible counterexample trace.*

6. Related work

We compare and contrast our work with a number of topics in the literature broadly related to flow analysis and verification of functional programs.

Higher-order multi-parameter tree transducer. As discussed in the Introduction, Kobayashi [6] introduced a type-based verification method for temporal properties of higher-order functional programs generated from finite base types. In a follow-up paper [9], Kobayashi et al. introduced a kind of tree transducer, called HMTT, that uses pattern-matching, taking trees as input and returning an output tree. They studied the problem of whether the tree generated by a given HMTT meets the output specification, assuming that the input trees meet the input specification, where both input and output specifications are regular tree languages. A sound but incomplete algorithm has been proposed for the HMTT verification problem by reduction to a model checking problem for recursion schemes with finite data domain (which can then be solved by a variation of Kobayashi's type-based algorithm). Though our algorithm in the present paper solves a similar kind of verification problem, it is not straightforward to compare it with the HMTT work [9]. It would appear that PMRS is a more general (and natural) formalism than HMTT. What is clear is that our approach to the over-approximation is very different: we use binding analysis to obtain a wPMRS which generates an over-approximation of the reachable term-trees, whereas Kobayashi et al. use automaton states to approximate input trees.

Approximating collecting semantics and flow analysis. In a seminal paper [5], Jones and Andersen studied the (data) flow analysis of functional programs by safely approximating the behaviour of a certain class of untyped, first-order, term rewriting systems with pattern matching. Their algorithm takes a regular set $\mathcal{I}$ of input terms, a program $\mathcal{P}$ and returns a regular tree grammar which is a "safe" description of the set of all reachable (constructor) terms of the computation of $\mathcal{P}$ with inputs form $\mathcal{I}$. Precisely, the algorithm computes a safe approximation of the *collecting semantics* of $\mathcal{P}$ relative to $\mathcal{I}$, which assigns to each rewrite rule a set of pairs (θ, g_θ) such that θ is a substitution (*realisable* in the course of such a computation) of actual parameters to the formal parameters of the rule, and g_θ is a term reachable from the RHS of the rule with the substitution θ. The collecting semantics is undecidable in general. Jones and Andersen was able to obtain, for each rewrite rule, a regular over-approximation of the set of realisable bindings $\{\, x \mapsto \theta\, x \mid \text{ realisable } \theta \,\}$ for each formal parameter x of the rule, and the set of reachable terms $\{\, g_\theta \mid \text{ realisable } \theta \,\}$, by decoupling the pair (θ, g_θ).

There are two directions in which Jones and Andersen's algorithm may be refined. Consider the setting of simply-typed functional programs with pattern-matching algebraic data types. Recent advances in the model checking of higher-order recursion schemes (notably the decidability of MSO theories of trees generated by higher-order recursion schemes [14]) indicate that the bindings of *non* pattern-matching variables, whether higher-order or not, can be precisely analysed algorithmically (though with extremely high asymptotic complexity). Jones and Andersen's algorithm builds a regular approximation of the binding set of every variable. A natural question is whether one can improve it by approximating *only* the bindings of pattern-matching variables, while analysing other variables (including all higher-order variables) precisely using the method in [14]. The work presented here offers a positive answer to the question. Another direction worth investigating is to seek to preserve, for each rewrite rule, as much of the connection between realisable substitutions θ and reachable terms g_θ as one can get away with. In an recent dissertation [10], Kochems has presented just such an algorithm using a kind of linear indexed tree grammars

(which are equivalent to context-free tree grammars) whereby the indices are the realisable substitutions.

To compare our algorithm with Jones and Andersen's, it is instructive to apply their algorithm to our Example 8. Their framework can be extended to simply-typed and higher-order programs. It is an old idea in functional programming that an higher-order expression, such as an "incompletely applied" function $(\cdots(f\,e_1)\cdots)e_m$ where the type of f has arity greater than m, may be viewed as a closure. (Indeed, closures are a standard implementation technique.) From this viewpoint, a higher-order non-terminal is regarded, not as a defined operator, but as a constructor, and closures are formed using a *binary closure-forming operator* @. Thus, the second clause of $Map2$ is written in their system as

$$@\ (@\ (@\ Map2\ \varphi)\ \psi)\ (\texttt{cons}\ x\ xs) \longrightarrow \texttt{cons}\ (@\ \varphi\ x)\ (@\ (@\ (@\ Map2\ \psi)\ \varphi)\ xs)$$

Observe that in this setting, $Map2$ is a constructor (i.e. terminal) symbol, and the expression $(@\ (@\ Map2\ \varphi)\ \psi)$ a pattern. Call the *binding set* of a variable the set of terms that may be bound to it at some point in the course of a computation. The approximating grammar produced by Jones and Andersen's algorithm is always regular (equivalently an order-0 recursion scheme). This is achieved by over-approximating the binding set of *every* variable (including higher-order ones, such as φ). The resultant grammar generates all finite lists of 0's and 1's, which is less precise than our algorithm.

Control flow analysis. Established in the 80's by Jones [4], Shivers [17] and others, *Control Flow Analysis (CFA)* of functional programs has remained an active research topic ever since (see e.g. Midtgaard's survey [12] and the book by Nielson et al. [13]). The aim of CFA is to approximate the flow of control within a program phrase in the course of a computation.

In a functional computation, control flow is determined by a sequence of function calls (possibly unknown at compile time); thus CFA amounts to approximating the values that may be substituted for bound variables during the computation. Since these values are (denoted by) pieces of syntax, CFA reduces to an algorithm that assigns *closures* (subterms of the examined term paired with substitutions for free variables) to bound variables. Reachability analysis and CFA are clearly related: for example, the former can aid the latter because unreachable parts of the term can be safely excluded from the range of closure assignment. There are however important differences: on one hand, CFA algorithms are *approximation algorithms* designed to address a more general problem; on the other, because CFA considers terms in isolation of its possible (program) contexts, the corresponding notion of reachability essentially amounts to reachability in the reduction graph.

Functional reachability. Based on the fully abstract game semantics, *traversals* [1, 14] are a (particularly accurate) model of the flow of control within a term; they can therefore be viewed as a CFA method. Using traversals, a new notion of reachability of higher-order functional computation (in the setting of PCF) is studied in [16], called *Contextual Reachability*: given a PCF term M of type A and a subterm N^α with occurrence α, is there a program context $C[-]$ such that $C[M]$ is a closed term of ground type and the evaluation of $C[M]$ causes control to flow to N^α?

7. Conclusion

Recursion schemes with pattern matching (PMRS) are an accurate and natural model of computation for functional programs have pattern-matching algebraic data types. We have given an algorithm that, given a PMRS $\mathcal{P}$ and a regular set $\mathcal{I}$ of input terms, constructs a *recursion scheme with weak pattern-matching* (wPMRS) that (i) over-approximates the set of terms reachable from under rewriting from $\mathcal{P}$ (ii) has a decidable model checking problem (relative to trivial automata). Finally, because of the precise analysis at higher-orders, we show that there is a simple notion of automatic abstraction-refinement, which gives rise to a semi-completeness property.

For future work, we plan to build an implementation of the verification algorithm for a real functional programming language. We shall be especially interested in investigating the scalability of our approach.

Acknowledgements. We would like to thank the anonymous reviewers for many useful comments.

References

[1] W. Blum and C.-H. L. Ong. Path-correspondence theorems and their applications. Preprint, 2009.

[2] E. M. Clarke, O. Grumberg, S. Jha, Y. Lu, and H. Veith. Counterexample-guided abstraction refinement. In *CAV '00: Proceedings of the 12th International Conference on Computer Aided Verification*, pages 154–169, London, UK, 2000. Springer-Verlag.

[3] A. Igarashi and N. Kobayashi. Resource usage analysis. *ACM Trans. Program. Lang. Syst.*, 27(2):264–313, 2005.

[4] N. D. Jones. Flow analysis of lambda expressions (preliminary version). In *Proceedings of the 8th Colloquium on Automata, Languages and Programming*, pages 114–128. Springer-Verlag, 1981. ISBN 3-540-10843-2.

[5] N. D. Jones and N. Andersen. Flow analysis of lazy higher-order functional programs. *Theoretical Computer Science*, 375:120–136, 2007.

[6] N. Kobayashi. Types and higher-order recursion schemes for verification of higher-order programs. In *Proceedings of POPL 2009*, pages 416–428. ACM Press, 2009.

[7] N. Kobayashi. Model-checking higher-order functions. In *PPDP*, pages 25–36, 2009.

[8] N. Kobayashi and C.-H. L. Ong. A type theory equivalent to the modal mu-calculus model checking ofhigher-order recursion schemes. In *Proceedings of LICS 2009*. IEEE Computer Society, 2009.

[9] N. Kobayashi, N. Tabuchi, and H. Unno. Higher-order multi-parameter tree transducers and recursion schemes for program verification. In *POPL*, pages 495–508, 2010.

[10] J. Kochems. Approximating reachable terms of functional programs. University of Oxford MMathsCompSc thesis, 2010.

[11] R. P. Kurshan. *Computer Aided Verification of Coordinating Processes*. Princeton University Press, 1994.

[12] J. Midtgaard. Control-flow analysis of functional programs. Technical Report BRICS RS-07-18, DAIMI, Department of Computer Science, University of Aarhus, Aarhus, Denmark, Dec 2007. URL `http://www.brics.dk/RS/07/18/BRICS-RS-07-18.pdf`.

[13] F. Nielson, H. R. Nielson, and C. Hankin. *Principles of Program Analysis*. Springer-Verlag New York, 1999.

[14] C.-H. L. Ong. On model-checking trees generated by higher-order recursion schemes. In *Proceedings 21st Annual IEEE Symposium on Logic in Computer Science, Seattle*, pages 81–90. Computer Society Press, 2006. Long version (55 pp.) downloadable at `users.comlab.ox.ac.uk/luke.ong/`.

[15] C.-H. L. Ong and S. J. Ramsay. Verifying higher-order functional programs with pattern-matching algebraic data types. Long version, available from: `https://mjolnir.comlab.ox.ac.uk/papers/pmrs.pdf`.

[16] C.-H. L. Ong and N. Tzevelekos. Functional reachability. In *LICS*, pages 286–295, 2009.

[17] O. Shivers. *Control-flow analysis of higher-order languages*. PhD thesis, Carnegie-Mellon University, 1991.

A. Soundness and termination of the approximation.

We state some simple properties of the system RS without proof.

Proposition 1. *Four easy consequences of the definition.*

(i) $\xi' u_k \cdots u_m \preccurlyeq_{\mathcal{S}} \xi\ v_1\ \cdots v_k\ \cdots\ v_m$ *iff* $\xi' \preccurlyeq_{\mathcal{S}} \xi\ u_1\ \cdots\ u_m$ *and, for all* $k \leq i \leq m$, $u_i \preccurlyeq_{\mathcal{S}} v_i$.

(ii) If there is a derivation of $u \preccurlyeq v$, *then there is such a derivation in which no non-head instance occurs at a greater depth than any head-instance.*

(iii) If there is a derivation of $u \preccurlyeq v$, *then there is such a derivation in which, for all variables* z, *there is at most one head-instance with* $x = z$.

(iv) If s *is a simple term and* $s \preccurlyeq_{\mathcal{S}} C[t]$ *then there is some simple* u *such that* $u \preccurlyeq t$

Proposition 2. $\xi'\ u_{k+1}\ \cdots\ u_m \preccurlyeq_{\mathcal{S}} \xi\ v_1\ \cdots\ v_m$ *iff* $\xi' \preccurlyeq_{\mathcal{S}} \xi\ v_1\ \cdots\ v_k$ *and, for all* $k+1 \leq i \leq m$, $u_i \preccurlyeq_{\mathcal{S}} v_i$.

A.1 Proof of Lemma 1

The statement of Lemma 1 is obtained as a corollary of the following.

Lemma 11. *Suppose* ζ *is an atomic term. If there is a derivation* δ *of* $\zeta \preccurlyeq_{\mathcal{S}} \xi\ v_1\ \cdots\ v_m$ *and, for all* $z \in X$, δ *does not contain any head-instance with* $x = z$, *then there exists a term* $\xi\ u_1\ \cdots\ u_m \in \mathsf{hs}_{\mathcal{S}}(\zeta, X)$ *such that, for all* i, $u_i \preccurlyeq_{\mathcal{S}} v_i$.

Proof. Assume the hypotheses and further that, without loss of generality, δ has properties (2) and (3) of Proposition 1. The proof continues by induction on the depth of δ $(= n)$.

When $n = 1$, $\zeta = \xi$, $m = 0$ and the result follows trivially.

When $n > 0$, δ is necessarily rooted at some instance of rule (S) and ζ is a variable $z \notin X$. In this case, it follows that there is a binding $z \mapsto \xi'\ u_k\ \cdots\ u_m \in \mathcal{S}$ and a sub-derivation δ' of $\xi'\ u_k\ \cdots\ u_m \preccurlyeq_{\mathcal{S}} \xi\ v_1\ \cdots\ v_m$. By Proposition 1, for each $k \leq i \leq m$, $u_i \preccurlyeq_{\mathcal{S}} v_i$. Since we assume properties (2) and (3) of δ, they hold also of δ'. In particular, since (2) holds of δ', we can construct a derivation ϵ of $\xi'\ \preccurlyeq_{\mathcal{S}} \xi\ v_1\ \cdots\ v_{k-1}$ by omitting the last $m - k$ terms in the conclusion of every sub-derivation of δ' and, by construction, properties (2) and (3) hold of ϵ. Since property (3) holds of ϵ, it follows that ϵ contains no head-instance with $x = z$, for all $z \in X$. Hence, by the induction hypothesis, $\xi\ u_1\ \cdots\ u_{k-1} \in \mathsf{hs}_{\mathcal{S}}(\xi', X \cup \{z\})$ and, for each $1 \leq i \leq k-1$, $u_i \preccurlyeq_{\mathcal{S}} v_i$. Thus, by unfolding the definition, we obtain $\xi\ u_1\ \cdots\ u_m \in \mathsf{hs}_{\mathcal{S}}(z, X)$ with properties as required. □

A.2 Proof of Lemma 2

Notice that, by definition, every term in $\mathsf{head}(u)$ is of the form $k\ t_1\ \cdots t_m$ in which $k \in \Sigma \cup \mathcal{N}$ and every t_i is either an argument of a simple term in the image of $\mathcal{S}$ or an argument of the compound term u.

A.3 Proof of Theorem 1

The theorem is a corollary of the following lemma.

Lemma 12. *If* $\mathcal{S}$ *is image simple, so is* $\mathcal{F}(\mathcal{S})$.

Proof. We prove by induction on Rule **C**, and use the notation therein. Assume that every term in the image of $\mathcal{S}$ is simple. Since u is simple (hence compound), by Lemma 2, each t_i is simple. It remains to prove that each term in the image of θ is simple. Let y be a variable that occurs in p. It follows that $\theta\, p = g[\theta\, y]$ for some context $g[\,]$. Since $\theta\, y$ is a ground-type subterm of a term obtained from s after a finite number of $\triangleright$-reductions

- either $\theta\, y$ is a *proper* subterm of a type-1 or type-2 contractum
- or $\theta\, y$ must in fact contain such a contractum

(We omit the trivial case where $\theta\, y$ is a subterm of s, which is a simple term, thanks to Lemma 2.)

We first show that the latter is impossible. Suppose, for a contradiction, $\theta\, y$ contains the contractum Δ of Δ. In other words, $\theta\, y = g'[\Delta]$ for some context $g'[\,]$, and the reduction $s \triangleright^* \theta\, p$ could be seen as $s \triangleright^{n-1} g[g'[\Delta]] \triangleright g[g'[\Delta]]$. Since y only occurs once in p, we could define θ' as the substitution that is obtained from θ by mapping y to $g'[\Delta]$ instead, so that $s \triangleright^{n-1} \theta' p$, contradicting the minimality assumption.

Suppose the former. We show that, whenever u is compound and $u \triangleright^*_{\mathcal{S}} \theta p$ minimally, then every contractum of this sequence is compound. Since every ground type subterm of a compound term is simple, it follows that any such $\theta\, y$ is simple. Let m be the length of the reduction sequence and n the size of u. The proof is by lexicographic induction on $(m,\, n)$. When $m = 0$, the result follows trivially. When $m > 0$, since the reduction is minimal it is necessary that p is of the form $a\ p_1\ \cdots\ p_m$. We analyse u:

When $u = a\ t_1\ \cdots\ t_m$, we have, for each $1 \leq i \leq m$, $t_i \triangleright^*_{\mathcal{S}} \theta p$ minimally. Since each t_i is simple, it follows from the induction hypothesis that every contractum in these sub-sequences – and hence in $u \triangleright^*_{\mathcal{S}} \theta\, p$ – is compound.

When $u = F\ t_1\ \cdots\ t_m$, it is necessary that the reduction sequence is of the form $u = C[v] \triangleright^*_{\mathcal{S}} C[\theta'\, q] \triangleright_{\mathcal{S}} t \triangleright_{\mathcal{S}} \theta p$, for some rule $F\ x_1\ \cdots\ x_m\ q \longrightarrow t$ in $\mathcal{P}_{\mathcal{G}_0}$. Since both v and t are simple, it follows that every contractum in each of the (minimal) sub-reductions – and hence in $u \triangleright^*_{\mathcal{S}} \theta\, p$ – is compound.

When $u = x\ t_1\ \cdots\ t_m$, necessarily the reduction sequence is of the form $u \triangleright_{\mathcal{S}} t \triangleright^*_{\mathcal{S}} \theta p$. Since u is compound, by Lemma 2, so is t whence it follows from the induction hypothesis that every contractum contained in the latter (minimal) sub-reduction – and hence in $u \triangleright^*_{\mathcal{S}} \theta p$ – is compound.

□

A.4 Proof of Lemma 4

The next lemma is a very simple property of the approximation.

Lemma 13. *If* s *is a simple term and* $s \preccurlyeq_{\mathcal{S}} C[t]$ *then there is some simple term* u *such that* $u \preccurlyeq_{\mathcal{S}} t$.

Proof. By induction on the shape of $C[\,]$.

$C[\,] = [\,]$**:** Then, in fact, $s = u = t$.

$C[\,] = \xi\ t_1\ \cdots\ t_{i-1}\ D[t]\ t_{i+1}\ \cdots\ t_m$**:** Then, by Lemma 1, $\xi\ u_1\ \cdots\ u_m \in \mathsf{head}(s)$, $u_i \preccurlyeq_{\mathcal{S}} D[t]$ and u_i simple. Hence, it follows from the inductive hypothesis that there is some term u such that $u \preccurlyeq_{\mathcal{S}} t$, as required.

□

The following is a short lemma characterising the behaviour of the function $\mathcal{F}$ wrt. the fixpoint set.

Lemma 14. *If*

(i) $F\ \theta' x_1\ \cdots\ \theta' x_m\ v \preccurlyeq_{\Xi} F\ \theta x_1\ \cdots\ \theta x_m\ s'$

(ii) $F\ x_1\ \cdots\ x_m\ p \longrightarrow t \in \mathcal{R}_{\mathcal{P}}$

(iii) $v \triangleright_{\Xi} \theta' p$

(iv) $\theta' p \preccurlyeq_{\Xi} \theta p$

(v) for all i, $\theta' x_i \preccurlyeq_{\Xi} \theta x_i$

then $F\ \theta' x_1\ \cdots\ \theta' x_m\ v \triangleright_{\Xi} t$ *and* $t \preccurlyeq_{\Xi} \theta t$.

Proof. By (iii) $F\ \theta' x_1\ \cdots\ \theta' x_m\ v \triangleright_{\Xi} t$. By (iv), (v), Proposition 2 and Lemma 13, $\theta' t \preccurlyeq_{\Xi} \theta t$. Assumptions (i) to (iii) constitute the premisses to the fixpoint construction, so we are assured that $\theta' \subseteq \mathcal{S}$ hence, by a second invocation of Proposition 2, $t \preccurlyeq_{\Xi} \theta t$. □

Proof. The proof of the main lemma is by induction on the length of the reduction ($= n$).

$n = 1$**:** When θt is a contractum, the reduction $s \Rightarrow \theta t$ is necessarily of the form:

$$F\ \theta x_1\ \cdots\ \theta x_m\ \theta p \Rightarrow \theta t$$

with $F\ x_1\ \cdots\ x_m\ p \longrightarrow t \in \mathcal{R}_\mathcal{P}$. Since $u \preccurlyeq_\Xi s$, by Corollary 1, $F\theta' x_1\ \cdots\ \theta' x_m\ v \in \mathsf{head}(u)$ and by Lemma 2 v is simple, $v \preccurlyeq_\Xi \theta p$ and, for all i, $\theta' x_i \preccurlyeq_\Xi \theta x_i$.

We proceed to show that, for all patterns q, subtitutions σ and simple terms w, if $w \preccurlyeq_\Xi \sigma q$, then $w \rhd_\Xi^* \sigma' q$ for some σ' and $\sigma' q \preccurlyeq_\Xi \sigma q$. The proof is by induction on q.

$q = x$**:** Then $w = \sigma' q$ for σ' mapping x to w, $w \rhd_\Xi^* w$ and $w \preccurlyeq_\Xi \sigma q$.

$q = a\ q_1\ \cdots\ q_k$**:** Since $w \preccurlyeq_\Xi a\ \sigma q_1\ \cdots\ \sigma q_k$, it follows from Corollary 1 that $a\ w_1\ \cdots\ w_k \in \mathsf{head}(w)$ with $w_i \preccurlyeq_\Xi \sigma q_i$ for all i. By the induction hypothesis, for all i there exists a substitution σ_i' such that $w_i \rhd_\Xi^* \sigma_i' q_i$ and $\sigma_i' q_i \preccurlyeq_\Xi \sigma q_i$. Hence $w \rhd_\Xi^* (\bigcup_{i=1}^k \sigma_i')q$ and $(\bigcup_{i=1}^k \sigma_i')q \preccurlyeq_\Xi \sigma q$.

It follows from this result that $v \rhd_\Xi^* \theta' p$ for some θ' such that $\theta' p \preccurlyeq_\Xi \theta p$. An appeal to Lemma 14 gives $F\ \theta x_1'\ \cdots\ \theta' x_m\ v \rhd_\Xi t$ whence $u \rhd_\Xi^* t$ and $t \preccurlyeq_\Xi \theta t$.

$n > 1$**:** Then the reduction can be written in the form:

$$s \Rightarrow^+ F\ \theta x_1\ \cdots\ \theta x_m\ s \Rightarrow^* F\ \theta x_1\ \cdots\ \theta x_m\ \theta p \Rightarrow \theta t$$

with $F\ \theta x_1 \cdots\ \theta x_m\ s$ identified with some contractum θr. It follows from the induction hypothesis that $r \preccurlyeq_\Xi F\ \theta x_1\ \cdots\ \theta x_m\ s$ and a second application of the induction hypothesis gives $t \preccurlyeq_\Xi \theta t$.

□

A.5 Proof of Lemma 5

Proof. By induction on the length of $S \Rightarrow^n C[\theta t]$.

$n = 1$**:** Then $S \Rightarrow \mathbf{rhs}(S)$ and $\mathbf{rhs}(S) \preccurlyeq_\mathcal{S} \mathbf{rhs}(S)$.

$n > 1$**:** Then the reduction can be written as:

$$S \Rightarrow^+ C[F\ \theta x_1\ \cdots\ \theta x_m\ s] \Rightarrow^* C[F\ \theta x_1\ \cdots\ \theta x_m\ \theta p] \Rightarrow C[\theta t]$$

with $F\ \theta x_1\ \cdots\ \theta x_m\ s$ a subterm of some contractum $\sigma t''$. It follows from the induction hypothesis that $t''\ \preccurlyeq_\Xi\ \sigma t''$. Hence, by Proposition 1, there is some simple term u such that $u \preccurlyeq_\Xi F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ and, by Lemma 4, $t \preccurlyeq_\Xi \theta t$.

□

A.6 Proof of Soundness

Intuitively, the soundness of the process is a consequence of Lemma 5. Given a $\mathcal{P}_\mathcal{G}$-reduction $\alpha = \mathit{Main}\ S \Rightarrow^*\ t$, one can reconstruct the reduction β in the approximating wPMRS as follows: whenever there is a pure reduction in α (arising from either from a pure $\mathcal{P}$-rule or a $\mathcal{G}$-rule) perform the same reduction in β. Whenever there is a pattern-matching reduction $C[F\ \theta x_1\ \cdots\ \theta x_m\ \theta p] \Rightarrow C[\theta t]$ in α, perform the weak-pattern matching reduction $C[F\ \theta x_1\ \cdots\ \theta x_m] \rightarrow C[\theta t^\dagger]$ in β *and then* rewrite away all the occurrences of pattern-symbols and live-symbols by using the Instantiation and Accounting rules derived from the fixpoint Ξ.

Let us refer to $\widetilde{\mathcal{P}_\mathcal{G}}$-reductions that use the Instantiation and Accounting rules as "administrative-reductions" and write them $s \rightarrow_\Xi t$. Unfortunately, things are not so straight forward as the above strategy might suggest, as there is one techinal detail to overcome. Given a $\mathcal{P}_\mathcal{G}$-reduction:

$$C[F\ \theta x_1\ \cdots\ \theta x_m\ \theta p] \Rightarrow C[\theta t]$$

it is not always possible to perform $\widetilde{\mathcal{P}_\mathcal{G}}$-reductions:

$$C[F\ \theta x_1\ \cdots\ \theta x_m\ \theta p] \rightarrow C[\theta t^\dagger] \rightarrow_\Xi^* C[\theta t]$$

even though all the necessary bindings are in present in Ξ (i.e. $y \preccurlyeq \theta y$ for all $y \in \mathsf{FV}(p)$). The reason is that reductions $V_y \rightarrow_\Xi^* \theta y$ are more restrictive than proofs $y \preccurlyeq \theta y$. In the former Ξ-bindings may be used to rewrite completely arbitrary subterms, whereas in the latter the rewriting is "weak" in the sense of combinatory reduction: rewriting may only occur at occurrences of completely applied terms. By way of an example of this phenomenon, assume y is a variable of type $o \rightarrow o$ and that there is a binding $y \mapsto G \in \Xi$; then $F\ V_y\ \mathtt{a} \preccurlyeq_\Xi F\ G\ \mathtt{a}$ but not $F\ V_y\ \mathtt{a} \rightarrow_\Xi^* F\ G\ \mathtt{a}$. Of course, in the end, a constructor term is produced and this necessitates the fact that all non-terminals are eventually erased or completely applied.

Simulation relation. To reflect this we define a relation $\lesssim_\tau\ \subseteq T(\Sigma, \mathcal{N} \cup \mathcal{N}_\mathcal{V}) \times T(\Sigma, \mathcal{N} \cup \mathcal{N}_\mathcal{V})$ which is parameterised by a simple type τ. Let $\xi : \alpha_1 \rightarrow \cdots \rightarrow \alpha_k \rightarrow \cdots \rightarrow \alpha_m \rightarrow o$ and $\tau = \alpha_{k+1} \rightarrow \cdots \rightarrow \alpha_m \rightarrow o$, then $u \lesssim_\tau \xi\ v_1 \cdots v_k$ just if:

(i) $\forall k+1 \leq i \leq m, \forall s_i : \alpha_i \cdot$
$u\ s_{k+1} \cdots\ s_m \rightarrow_\Xi^* \xi\ u_1 \cdots u_k\ s_{k+1} \cdots s_m$ and

(ii) $\forall 1 \leq j \leq k\ \cdot\ u_i \lesssim_{\alpha_i} v_i$

The relation is clearly reflexive, in fact it is a preorder.

Lemma 15. $\forall\, v,\ u,\ w, \forall \tau\ \cdot\ u \lesssim_\tau v \wedge v \lesssim_\tau w \Rightarrow u \lesssim_\tau w.$

Proof. By induction on the long term-formation rule for v i.e.:

$$\frac{\xi \in (\Sigma \cup \mathcal{N} \cup \mathcal{N}_\mathcal{V}) \qquad \mathsf{ar}(\xi) = k \geq 0 \qquad \text{for all } 1 \leq i \leq k \cdot\ u_i \in T(\Sigma, \mathcal{N}, \mathcal{N}_\mathcal{V})}{\xi\ u_1\ \cdots\ u_k \in T(\Sigma, \mathcal{N}, \mathcal{N}_\mathcal{V})}$$

Let $u = \xi\ u_{n-l+1} \cdots u_n$, $v = \zeta\ v_{n-m+1} \cdots v_n$ and $w = \chi\ w_1 \cdots w_n$ for $0 \leq l \leq m \leq n$. Then assume the hypotheses, i.e, for all i from 1 to n and for all terms s_i:

(H1) $\xi\ u_{n-l+1} \cdots u_n\ s_1 \cdots s_k \rightarrow_\Xi^* \zeta\ u_{n-m+1} \cdots u_n\ s_1 \cdots s_k$ and, for all i from $n-m+1$ to n, $u_i \lesssim_{\alpha_i} v_i$.

(H2) $\zeta\ v_{n-m+1} \cdots v_n\ s_1 \cdots s_k \rightarrow_\Xi^* \zeta\ v_1 \cdots v_n\ s_1 \cdots s_k$ and, for all i from 1 to n, $v_i \lesssim_{\alpha_i} w_i$.

To show the conclusion, let us construct the following reduction sequence:

$$\begin{aligned}\xi\ u_{n-l+1} \cdots u_n\ s_1 \cdots s_k &\rightarrow_\Xi^* \zeta\ u_{n-m+1} \cdots u_n\ s_1 \cdots s_k \\ &\rightarrow_\Xi^* \chi\ u_1 \cdots u_n\ s_1 \cdots s_k\end{aligned}$$

where, for all $i \in [1..(n-m)]$, $u_i = v_i$ — which is possible because every administrative reduction is of the form

$$C[F\ \theta x_1 \cdots \theta x_m] \rightarrow_\Xi C[t\ \theta x_1 \cdots \theta x_m]$$

for some rule $F\ x_1 \cdots x_m \longrightarrow t\ x_1 \cdots x_m$. To see that, for each $i \in [1..n]$, $u_i \lesssim_{\alpha_i} w_i$ we consider two cases:

(i) When $1 \leq i \leq n-m$ then $u_i = v_i$ and, by (H2), $v_i \lesssim_{\alpha_i} w_i$.

(ii) When $n - m < i \leq n$ then, by (H1), $u_i \lesssim_{\alpha_i} v_i$ and by (H2), $v_i \lesssim_{\alpha_i} w_i$. It follows from the induction hypothesis that $u_i \lesssim_{\alpha_i} w_i$.

□

Furthermore, the simulation relation is compatible with one-hole contexts.

Lemma 16. *For all one-hole contexts $C[]$ of type β (in which the hole is of type α), for all terms u and v: if $u \lesssim_\alpha v$ then $C[u] \lesssim_\beta C[v]$.*

Proof. By induction on the shape of $C[]$:

- When $C[] = []$ then the result is trivial.
- When $C[] = \xi\ s_1 \cdots D[] \cdots s_k$ (and $\xi : \tau_1 \to \cdots \to \tau_k \to \alpha$, to see that

$$\xi\ s_1 \cdots s_i = D[u] \cdots s_k \lesssim_\beta \xi\ s_1 \cdots s_i' = D[v] \cdots s_k$$

notice that it follows from the induction hypothesis that $D[u] \lesssim_{\tau_i} D[v]$.

□

Corollary 2. *If $s \lesssim_\alpha t$ and, for all terms $u : \alpha$, $C[u] \lesssim_\beta D[u]$, then $C[s] \lesssim_\beta D[t]$.*

Proof. By compatibility (Lemma 16), $C[s] \lesssim_\beta C[t]$. By the assumption, $C[t] \lesssim_\beta D[t]$ and the result follows by transitivity of the simulation relation (Lemma 15). □

Terms are related by a change in accounting non-terminal symbols. That is, when there are no free variables, the dagger operation only introduces accounting non-terminal symbols in place of terminal symbols, which preserves the simulation.

Lemma 17. *For all terms $t : \tau$, if $\mathsf{FV}(t) = \emptyset$ then $t^\dagger \lesssim_\tau t$.*

Proof. By induction on the shape of t:

- When $t = a : b_1 \to \cdots \to b_k \to o$, then $t^\dagger = K_a$. To see that $K_a \lesssim_{b_1 \to \cdots \to b_k \to o} a$ let $s_i : b_i$ for all $i \in [1..k]$, then $K_a\ s_1 \cdots s_k \to_\Xi^* a\ s_1 \cdots s_k$ as required.
- When $t = F$, also $t^\dagger = F$ and the result follows trivially.
- When $t = u\ v$, then $t^\dagger = u^\dagger\ v^\dagger$. Since the simulation relation is compatible with one-hole contexts (Lemma 16) and transitive (Lemma 15) it follows from the induction hypothesis (twice) that:

$$u^\dagger\ v^\dagger \lesssim u\ v^\dagger \lesssim u\ v$$

□

Whenever t can arise as an instance of s by substitutions in Ξ, then the double-dagger transform of s simulates t.

Lemma 18. *For all terms s and t of type τ, if $s \preccurlyeq_\Xi t$ then $s^\# \lesssim_\tau t$.*

Proof. By induction on the proof of $s \preccurlyeq_\Xi t$ in the system RS.

(R) Then $s = t$ and $\mathsf{FV}(s) = \mathsf{FV}(t) = \emptyset$. By Lemma 17 $s^\# = s^\dagger \lesssim_\tau t$.

(S) Then $s = C[x] \preccurlyeq_\Xi t$ for some one-hole context $C[]$ and necessarily there is some term $u : \alpha_1 \to \cdots \to \alpha_k \to \beta$ such that $xmapstou \in \Xi$ and $C[u] \preccurlyeq_\Xi t$. It follows from the induction hypothesis that $C[u]^\# \lesssim_\tau t$. By transitivity of the simulation relation (Lemma 15) it remains to show that $C[x]^\# \lesssim_\tau C[u]^\#$, i.e., by compatibility (Lemma 16) that $x^\# \lesssim_{\alpha_1 \to \cdots \alpha_k \to \beta} u^\#$. So let $s_i : \alpha_i$ for each $i \in [1..k]$, then, by definition, $V_x\ s_1 \cdots s_k \to_\Xi u^\#\ s_1 \cdots s_k$, as required.

□

The simulation relation reflects pattern matching.

Lemma 19. *For all terms $s : o$, patterns p and substitutions θ, if $s \lesssim_o \theta p$ then there exists a substitution σ such that $s = \sigma p$.*

Proof. By induction on the shape of the pattern p.

- When $p = x$, then let $\sigma(x) = s$.
- When $p = f\ p_1 \cdots p_k$ (with, for all i, $p_i : b_i$), then $s \to_\Xi^* f\ s_1 \cdots s_k$ with, for all i, $s_i \lesssim_{b_i} \theta p_i$. It follows from the induction hypothesis that, for each i, there exists a substitution σ_i such that $s_i = \sigma_i p_i$. Since patterns are linear, we can form $\sigma = \bigcup \{\, \sigma_i \mid i \in [1..k] \,\}$.

□

The simulation relation relfects redexes in context.

Lemma 20. *For all terms u, one-hole contexts $C[]$ and redexes $F\ \theta x_1 \cdots \theta x_m\ \theta p$, $u \lesssim_o C[F\ \theta x_1 \cdots \theta x_m\ \theta p]$ implies there exists some one-hole context $C'[]$ and some substitution θ' such that $u \to_\Xi^* C'[F\ \theta' x_1 \cdots \theta' x_m\ \theta' p]$ and, for all s, $C'[s] \lesssim_o C[s]$.*

Proof. Let $F : \alpha_1 \to \cdots \to \alpha_m \to o$. The proof proceeds by induction on the shape of $C[]$.

- When $C[] = []$, then we have that $u \lesssim_o F\ \theta x_1 \cdots \theta x_m\ \theta p$, i.e. $u \to_\Xi^* F\ s_1 \cdots s_m\ s'$ with, for all i, $s_i \lesssim_{\alpha_i} \theta x_i$. By Lemma **??** there exists a substitution σ such that $s' = \sigma p$. Since the variables on the LHS of a rule are all distinct, we can form the substitution $\theta' = \sigma \cup \{\, x_i \mapsto s_i \mid i \in [1..m] \,\}$.
- When $C[] = \xi\ u_1 \cdots u_i = D[] \cdots u_k$ (with $u_i : \tau_i$ for all i), then $u \to_\Xi^* \xi\ u_1' \cdots u_i' \cdots u_k'$ and, for all i, $u_i' \lesssim_{\tau_i} u_i$. It follows from the induction hypothesis that there is some $D'[]$ and some substitution θ' such that $u_i' \to_\Xi^* D'[F\ \theta' x_1 \cdots \theta' x_m\ \theta' p]$ and, for all terms s, $D'[s] \lesssim_{\tau_i} D[s]$. Hence:

$$\begin{aligned} u &\to_\Xi^* \xi\ u_1' \cdots u_i' \cdots u_k' \\ &\to_\Xi^* \xi\ u_1' \cdots D'[F\ \theta' x_1 \cdots \theta' x_m\ \theta' p] \cdots u_k \end{aligned}$$

So let $C'[] = \xi\ u_1' \cdots D'[] \cdots u_k'$ and the result is immediate.

□

Before proving the soundness result, it is necessary to show that the "approximate" contracta $\theta' t^\dagger$ simulate the real ones θt.

Lemma 21. *For all terms $t : \tau$ and substitutions θ, θ': if $t \preccurlyeq_\Xi \theta t$ and, for all variables $x : \sigma$, $\theta' x \lesssim_\sigma \theta x$, then $\theta' t^\dagger \lesssim_\tau \theta t$.*

Proof. By induction on t:

- When $t = a$, then $\theta' t^\dagger = t^\dagger$, $\theta t = t$ and the result follows from Lemma 17.
- When $t = F$, then $\theta' t^\dagger = F = \theta t$.
- When $t = x$, there are two cases:
  1. When $x \in \mathbf{dom}(\theta')$, then by assumption $\theta' t^\dagger = \theta' x \lesssim_\sigma \theta x = \theta t$.
  2. When $x \notin \mathbf{dom}(\theta')$, then $\theta' t^\dagger = \theta' x^\#$ and the result follows from Lemma 18.
- When $t = u\ v$ for some terms $u : \tau_1$ and $v : \tau_2$, then $\theta' t^\dagger = \theta' u^\dagger\ \theta' v^\dagger$. It follows from the induction hypothesis that $\theta' u^\dagger \lesssim_{\tau_1} \theta u$ and $\theta' v^\dagger \lesssim_{\tau_2} \theta v$. By transitivity and compatibility (Lemma 15 and Lemma 16) it follows that:

$$\theta' u^\dagger\ \theta' v^\dagger \lesssim_\tau \theta u\ \theta' v^\dagger \lesssim_\tau \theta u\ \theta v$$

□

Finally, we come to the proof of soundness. We aim to show

$$\mathcal{L}(\mathcal{P}_\mathcal{G},\ Main\ S) \subseteq \mathcal{L}(\widetilde{\mathcal{P}_\mathcal{G}},\ Main\ S)$$

Proof. Assume $t \in \mathcal{L}(\mathcal{P}_\mathcal{G},\ Main\ \ S)$, then there is a reduction sequence $Main\ S \Rightarrow t_1 \Rightarrow t_2 \Rightarrow \cdots$ such that $t = \bigsqcup \{\, t_i^\perp \mid i \in \omega \,\}$. We show that, for all $i \in \omega$, there is a term $\widetilde{t_i}$ such that $Main\ S \to^* \widetilde{t_i}$ in $\widetilde{\mathcal{P}_\mathcal{G}}$ and $\widetilde{t_i} \lesssim_o t_i$.

The proof proceeds by induction on $i > 0$:

- When $i = 1$ there are two cases:
  1. $Main\ S \Rightarrow Main\ t$ and $S \longrightarrow t$ is a $\mathcal{G}$-rule. In this case, $Main\ S \rightarrow t^\dagger$ and the result follows by Lemma 17.
  2. $Main\ S \Rightarrow \theta t$ where $\theta(m) = S$ and $Main\ m \longrightarrow t$ is a $\mathcal{P}$-rule. By Lemma 5 we have that $t \preccurlyeq_\Xi \theta t$ and by Lemma 21 (with $\theta' = \emptyset$) we have $Main\ S \rightarrow t^\dagger \lesssim_o \theta t$.
- When $i = k + 1 > 1$ we have a PMRS reduction sequence:

  $$Main\ S \Rightarrow \cdots \Rightarrow t_k \Rightarrow t_{k+1}$$

  and, it follows from the induction hypothesis that there is a corresponding simulating wPMRS sequence:

  $$Main\ S \rightarrow^* \cdots \rightarrow^* \widetilde{t_k}$$

  such that $\widetilde{t_k} \lesssim_o t_k$. We analyse the contraction $t_k \Rightarrow t_{k+1}$. Necessarily, in general t_k is of the form $C[F\ \theta x_1 \cdots \theta x_m\ \theta p]$ and t_{k+1} of the form $C[\theta t]$ whenever $F\ x_1 \cdots x_m\ p \longrightarrow t$ is a rule in $\mathcal{P}_\mathcal{G}$. Since the simulation relation reflects redexes in context (Lemma 20), there is some context $C'[]$ and some substitution θ' such that $\widetilde{t_k} \rightarrow^* C'[F\ \theta' x_1 \cdots \theta' x_m\ \theta' p]$ and, for all terms s, $\widetilde{C'[s]} \lesssim_o C[s]$ and for all variables $x : \sigma$, $\theta' x \lesssim_\sigma \theta x$. Let $\widetilde{t_{k+1}}$ be the contractum of this redex: $C'[\theta' t^\dagger]$. Then $Main\ S \rightarrow^* \widetilde{t_{k+1}}$ and, since by Lemma 21 we have $\theta' t^\dagger \lesssim_o \theta t$, it follows from Corollary 2 that $\widetilde{t_{k+1}} = C'[\theta' t^\dagger] \lesssim_o C[\theta t] = t_{k+1}$.
  To complete the proof, notice that if $s \lesssim_o t$ then $s^\perp = t^\perp$.

□

B. Computability of minimal $u \triangleright^* \theta p$ reductions

Recall that $s \triangleright^i t$ is a *minimal* reduction from s to t whenever there is no reduction $s \triangleright^j t$ such that $j < i$.

We say that a $\triangleright$-reduction sequence $s \triangleright^* t$ *contains* a reduction sequence $s' \triangleright^* t'$ when the former is of the shape $s \triangleright^* C[s'] \triangleright^* C[t'] \triangleright^* t$. We say that a $\triangleright$-reduction sequence α *properly contains* a $\triangleright$-reduction sequence β just if α contains β and $\alpha \neq \beta$.

Proposition 3. *If $s \triangleright^* t$ is a minimal reduction from s to t then it does not properly contain any reduction from s to t.*

We propose the function rmatch, defined in Figure 4 which, given a term s and a pattern p, computes the finite set $\{\, \theta \mid s \triangleright^* \theta\, p \text{ minimally } \}$. We define $\bar{X}$ as the complement of X with respect to the (finite) set of all pairs of compound terms and patterns (for a fixed PMRS $\mathcal{P}$). Notice that, since the size of $\bar{X}$ strictly decreases in every recursive call, rmatch is well-defined.

Lemma 22. *Assume u is a compound term and p a pattern. $\theta \in \mathsf{rmatch}(u,\ p,\ X)$ iff there is a minimal reduction $u \triangleright^* \theta\, p$ which, for all substitutions θ' and $(v,\ q) \in X$, contains no reduction from v to $\theta'\, q$.*

Proof. We prove the forward direction by induction on the size of $\bar{X}$. We assume the antecedent and so note one fact: (*) since $\mathsf{rmatch}(u,\ p,\ X)$ is non-empty it is necessary that $(u,\ p) \notin X$. When $\bar{X}$ is empty, $(u,\ p) \in X$ and the result follows trivially. When $\bar{X}$ is non-empty, analyse the form of $(u,\ p)$:

When $p = x$, $\mathsf{rmatch}(u,\ p,\ X) = \{\{x \mapsto u\}\}$ and $u = \theta u$.

When $p = a\ p_1\ \cdots\ p_m$ and $u = a\ t_1\ \cdots\ t_m$, if $\theta \in \mathsf{rmatch}(u,\ p,\ X)$ then necessarily there exist m substitutions θ_i such that $\theta_i \in \mathsf{rmatch}(t_i,\ p_i,\ X \cup \{(u,\ p)\})$. Since each t_i is simple, it follows from the induction hypothesis that there exists a reduction $t_i \triangleright^* \theta_i\, p_i$ which does not contain any reduction described by a pair in $X \cup \{(u,\ p)\}$ and hence by any pair in X. Since the reductions are disjoint, their concatenation is a reduction with the required property.

When $p \neq x$ and $u = x\ t_1\ \cdots\ t_m$, there must be some $t \in head(u)$ such that $\theta \in \mathsf{rmatch}(t,\ p,\ X \cup \{(u,\ p)\})$. In this case, necessarily $(t,\ p) \notin X$ (**). Since a term constructed by head in this way is compound, it follows from the induction hypothesis that there is a reduction $t \triangleright^* \theta\, p$ which does not contain any reduction described by $X \cup \{(u,\ p)\}$ and hence by X. Consequently, there is a reduction $u \triangleright t \triangleright^* \theta\, p$ with no containment violation since, by (*) and (**), $(u,\ p)$, $(t,\ p) \notin X$.

When $p \neq x$ and $u = F\ t_1\ \cdots\ t_m\ v$, there must be some rule $F\ x_1\ \cdots\ x_m\ q$ in $\mathcal{P}_{\mathcal{G}_0}$ with $\mathsf{rmatch}(v,\ q,\ X \cup \{(u,\ p)\})$ not empty such that $\theta \in \mathsf{rmatch}(t,\ p,\ X \cup \{(u,\ p)\})$. In this case $(t,\ p) \notin X$ (**) and $F\ t_1\ \cdots\ t_m\ v \triangleright t$. Since v is simple and t is simple, it follows from the induction hypothesis that there are reductions $v \triangleright^* \theta' q$ and $t \triangleright^* \theta p$ which do not contain any reduction described by $X \cup \{(u,\ p)\}$ and hence X. Consequently, $u \triangleright^* t \triangleright^* \theta p$ also does not contain any violation since, by (*) and (**), $(u,\ p)$, $(t,\ p) \notin X$.

We prove the backward direction by induction on the length of the reduction ($= n$). We assume the antecedent and so note two facts: (i) since the reduction is minimal, by Proposition 3 it does not contain any reduction from u to $\theta' p$ for any θ' and (ii) since any reduction contains itself it is necessary that $(u,\ p) \notin X$. When $n = 0$, $u = \theta p$ and an easy induction on p gives $\theta \in \mathsf{rmatch}(u,\ p,\ X)$. When $n > 0$, analyse the form of $(u,\ p)$.

When $p = x$, since the reduction is minimal $u = \theta p$, but this violates the assumption that $n > 0$.

When $p = a\ t_1\ \cdots\ t_m$ and $u = a\ t_1\ \cdots\ t_m$ we have $t_i \triangleright^* \theta_i p_i$ minimally for each $1 \leq i \leq m$, not containing any violations. By the induction hypothesis and (ii), for each i, $\theta_i \in \mathsf{rmatch}(t_i,\ p_i,\ X \cup \{(u,\ p)\})$ whence $\theta = \bigcup_{i=1}^m \theta_i \in \mathsf{rmatch}(t,\ p,\ X)$.

When $p \neq x$ and $u = x\ t_1\ \cdots\ t_m$ we have that the minimal reduction is necessarily of the form $u \triangleright t \triangleright^* \theta p$ with $t \triangleright^* \theta p$ minimal for some $t \in \mathsf{head}(u)$. By the induction hypothesis and (ii), $\theta \in \mathsf{rmatch}(t,\ p,\ X \cup \{(u,\ p)\})$ whence $\theta \in \mathsf{rmatch}(t,\ p,\ X)$ by definition.

When $p \neq x$ and $u = F\ t_1\ \cdots\ t_m\ v$ we have that the reduction must of the form $u \triangleright^+ t \triangleright^* \theta p$ with $t \triangleright^* \theta p$ minimal for some t occurring on the right-hand side of a rule $F\ t_1\ \cdots\ t_m\ q \longrightarrow t$. For $u \triangleright^+ t$ to reduce in this way, it is necessary that it contains a reduction $v \triangleright^* \theta' q$ for some θ' and that this reduction is itself minimal. By (ii), this reduction cannot contain a reduction from u to $\theta'' p$ for any θ'' so, by the induction hypothesis $\theta' \in \mathsf{rmatch}(v,\ q,\ X \cup \{(u,\ p)\})$. Similarly, by (ii), $t \triangleright^* \theta p$ cannot contain a reduction from u to $\theta'' p$ for any θ'', so a second application of the induction hypothesis gives $\theta \in \mathsf{rmatch}(t,\ p,\ X \cup \{(u,\ p)\})$. By definition, $\theta \in \mathsf{rmatch}(u,\ p,\ X)$ as required.

□

C. Construction of witnesses from feasible counterexamples

We assume that every variable in the image of mgci is fresh. We begin the proof by making some definitions.

We first define a transformation $t^\#$ of a labelled term t that removes all occurrences of live-symbols N^l, replacing each with an open term $N^{l\#}$ which represents the family of terms to which this occurrence of N^l can reduce.

$$t^\# = \begin{cases} \mathsf{a} & t = \mathsf{a} \\ F & t = F \text{ F is not live-symbol} \\ \mathsf{mgci}(l) & t = F^l \text{ F is a live-symbol} \\ t_0^\#\ t_1^\# & t = t_0\ t_1 \end{cases}$$

$\mathsf{rmatch}(u,\, p,\, X) =$

if $(u,\, p) \in X$ then $\emptyset$ else:

$$\begin{cases} \{\{x \mapsto u\}\} & p = x \\ \{\, \bigcup_{i=1}^{m} X_i \mid \forall i \cdot X_i \in \mathsf{rmatch}(t_i,\, p_i,\, X \cup \{(u,\, p)\})\,\} & p = a\ p_1\ \cdots\ p_m\ \&\ u = a\ t_1\ \cdots\ t_m \\ \bigcup\{\, \mathsf{rmatch}(t,\, p,\, X \cup \{(u,\, p)\}) \mid t \in \mathsf{head}(u)\,\} & p \neq x\ \&\ u = x\ t_1\ \cdots\ t_m \\ \bigcup\{\, \mathsf{rmatch}(t,\, p,\, X \cup \{(u,\, p)\}) \mid F\ x_1\ \cdots\ x_m\ q \longrightarrow t \in \mathcal{R},\, \mathsf{rmatch}(v,\, q,\, X \cup \{(u,\, p)\}) \neq \emptyset\,\} & p \neq x\ \&\ u = F\ t_1\ \cdots\ t_m\ v \end{cases}$$

Figure 4. rmatch algorithm

Note that, if t is a term in a feasible trace, then $t^{\#}$ is always well defined.

We say that a term t is *linear* just in case every variable in $\mathsf{FV}(t)$ occurs exactly once in t. We say that a set of (possibly open) terms l is *linear* just in case every term $t \in l$ is linear. We say that a labelled term t is *label-linear* just in case every label l labelling t (or any of its subterms) is linear and whenever l_1 and l_2 are distinct labels labelling t then $\bigcup\{\, \mathsf{FV}(t) \mid t \in l_1\,\} \cap \bigcup\{\, \mathsf{FV}(t) \mid t \in l_2\,\} = \emptyset$.

Notice that linearity is preserved by the resolution of a term.

Lemma 23. *If t is label linear then $\bar{t}$ is linear.*

Proof. By induction on the shape of t:

$t = \mathtt{a}$ Then $\bar{t} = \{\mathtt{a}\}$.

$t =$ **(non live-symbol)** F Then $\bar{t} = \{F\}$.

$t =$ **(live-symbol)** F^l Then $\bar{t} = \hat{l}$. If $l = \emptyset$ then $\bar{t} = \{z\}$ for fresh z. If $l \neq \emptyset$ then $\bar{t} = l$ which is by assumption linear.

$t = t_0\ t_1$ Then $\bar{t} = \{\, u\ \ v \mid u \in \bar{t_0}, v \in \bar{t_1}\,\}$. It follows from the induction hypothesis that $\bar{t_0}$ and $\bar{t_1}$ are label linear. Consequently, the terms in $\bar{t_0}$ share no variables with the terms in $\bar{t_1}$ (and each is linear) and so each $u\ v$ is linear.

□

Linearity is also preserved by mgci.

Lemma 24. *If l is linear then $\mathsf{mgci}(l)$ is linear.*

Proof. Since l is linear then every term in l is linear, hence the most general common instance of the terms in l is also linear. □

Every term in the labelled feasible counterexample is label-linear.

Lemma 25. *Let $\alpha = s_1 \leftarrow s_2 \leftarrow \cdots \leftarrow s_{n-1} \leftarrow s_n$ be a labelled, feasible counterexample. Every term in α is label-linear.*

Proof. By induction on n.

$n = 1$ Then s_1 is unlabelled (or equivalently, every live-symbol is labelled by $\emptyset$).

$n = k + 1$ Then we have $s_1 \leftarrow \cdots \leftarrow s_k \leftarrow s_{k+1}$. By the induction hypothesis, s_i from 1 to k is label-linear. Consider $s_k \leftarrow s_{k+1}$.

(i) If this reduction is an abstract $\widetilde{\mathcal{P}_{\mathcal{G}}}$-reduction then any label l occurring in s_{k+1} either arises from (1) the resolution $\bar{t}$ of a term t in s_k — hence by Lemma 23 — is linear, or (2) as the union of labels $l_1, \ldots, l_m$ in s_k and hence is linear. Now consider two labels l and l' in s_{k+1}. There are four cases:

(a) l arises due to (1) and l' arises due to (1): then $l = l'$ and they both label the head symbol of the redex of s_{k+1}.

(a) l arises due to (1) and l' arises due to (2): then l and l' do not share variables since l' arises due to the resolution of the head symbol of the contractum t in s_k and t has no ancestors in s_{k+1}.

(a) l arises due to (2) and l' arises due to (1): analogous to the previous case.

(a) l arises due to (2) and l' arises due to (2): $l = \bigcup\{l_1, \cdots, l_k\}$ labels N and there are live-symbols $N^{l_1}, \cdots, N^{l_k}$ in s_k which are descendents of this N. Similarly for l'. Since ancestors are unique, either $l = l'$ and they label the same N or the descendents of the term labelled by l and the descendents of the term labelled by l' are disjoint and hence by label-linearity of s_k l and l' share no variables.

(i) If this reduction is an abstract $\mathcal{P}$-reduction then any label l occurring in s_{k+1} either arises from (1) the application of $\mathsf{labelTm}(\theta y)(\bigcup\{l_1, \cdots, l_k\} = \{s_1, \cdots, s_m\})$ to some instance θy of a pattern-matching variable y or (2) as point (2) above. In case (1), since $s_1, \cdots, s_m$ are all linear, so any subterms of any s_i will be linear and hence if $s'_1, \ldots, s'_m$ is a collection of terms such that each s'_i is a subterm of s_i or is a fresh variable then $s'_1, \cdots, s'_m$ is also linear. Hence, by Lemma 24, l which is the most general common instance of such a set is also linear. By the same argument as in the case (i), labels arising from (2) will always be linear, and an analogous case analysis shows that, due to the uniqueness of ancestors, all distinct labels arising by these two methods will be variable-disjoint.

Hence, the term s_{k+1} is label-linear.

□

Label-linearity carries over to term-linearity under the $\#$.

Proposition 4. *If t is label linear then $t^{\#}$ is linear.*

Proposition 5. *If $l = \{u_1\ v_1, \ldots, u_k\ v_k\}$ is linear, then there exists a renaming of variables ρ such that:*

$$\mathsf{mgci}(l) = \rho(\mathsf{mgci}(\{u_1, \ldots, u_k\})\ \mathsf{mgci}(\{v_1, \ldots, v_k\}))$$

Lemma 26. *Let $s_i\|_\pi$ be an occurrence of a subterm in a feasible counterexample trace. For all terms t and linear sets of terms l and m, if $l \subseteq m$ and $s_i|_\pi = t$ and $[\![s_i\|_\pi]\!]$ is labelled according to $\mathsf{labelTm}(t)(m)$ then $[\![s_i\|_\pi]\!]^{\#}$ is an instance of $\mathsf{mgci}(l)$.*

Proof. By induction on the shape of t.

$t = \mathtt{a}$: Then $\mathtt{a}$ must be an instance of every element of m and therefore an instance of every element of l. Hence, $[\![s_i\|_\pi]\!]^{\#} = \mathtt{a}$ is, in particular, an instance of the MGCI of l.

$t = F$ (and F is not a live symbol): As above.

$t = F$ (and F is a live symbol): Then $[\![s_i\|_\pi]\!]^\# = \mathsf{mgci}(m)$ and $\mathsf{mgci}(m)$ is, in particular an instance of every term in l, so is an instance of $\mathsf{mgci}(l)$.

$t = t_0\ t_1$: Then $[\![s_i\|_\pi]\!]^\# = [\![s_i\|_{\pi\cdot 1}]\!]^\#\ [\![s_i\|_{\pi\cdot 2}]\!]^\#$. Let $m = \{m_1, \ldots, m_k\}$. In this case, t_0 is labelled according to $\mathsf{labelTm}(t_0)(\{m_{0_1}, \cdots, m_{0_k}\} = m_0)$ and each m_{0_i} is either a fresh variable or part of some $m_{0_i}\ m_{1_i} = m_i \in m$. Let $l_0 = \{\, m_{0_i} \mid m_{0_i}\ m_{1_i} \in l \,\}$. Then $l_0 \subseteq m_0$ and so, by IH, $[\![s_i\|_{\pi\cdot 1}]\!]^\#$ is an instance of $\mathsf{mgci}(l_0)$. Construct m_1 and l_1 similarly to show that $[\![s_i\|_{\pi\cdot 2}]\!]^\#$ is an instance of $\mathsf{mgci}(l_1)$. Since l_0 and l_1 are linear (and we assume every variable in the image of mgci is fresh), the term $[\![s_i\|_{\pi\cdot 1}]\!]^\#\ [\![s_i\|_{\pi\cdot 2}]\!]^\#$ is an instance of the term $\mathsf{mgci}(l_0)\ \mathsf{mgci}(l_1)$. Since l is linear, the set $\{l_{0_1}\ l_{1_1}, \ldots, l_{0_j}\ l_{1_j}\} = l \setminus \mathcal{V}$, for $j \leq k$, is also linear. Hence, by Proposition 5, the term $\mathsf{mgci}(l_0)\ \mathsf{mgci}(l_1)$ is (a renaming of) the MGCI of this set. Hence, $\mathsf{mgci}(l_0)\ \mathsf{mgci}(l_1)$ is also the MGCI of l.

□

Lemma 27. *For all closed, label-linear, labelled terms t, there exists a variable renaming ρ such that* $\mathsf{mgci}(\bar{t}) = \rho t^\#$.

Proof. By induction on t:

- When $t = \mathsf{a}$, also $\bar{t} = \{\mathsf{a}\}$ and $t^\# = \mathsf{a}$. Hence $\mathsf{mgci}(\bar{t}) = \mathsf{a} = t^\#$.
- When $t = F$, for non-live F, $\mathsf{mgci}(\bar{F}) = \mathsf{mgci}(\{F\}) = F = F^\#$.
- When $t = F^l$, for live-symbol F^l there are two sub-cases. When $l = \emptyset$ $\mathsf{mgci}(\bar{t}) = \mathsf{mgci}(\{z_1\}) = z_2$ and $t^\# = z_3$ for some fresh variables z_2 and z_3. So let $\rho(z_3) = z_2$. When $l \neq \emptyset$ then $\mathsf{mgci}(\bar{t}) = \mathsf{mgci}(l) = t^\#$.
- When $t = t_0\ t_1$, $\overline{t_0\ t_1} = l = \{u_1\ v_1, \ldots, u_k\ v_k\}$ since the image of the resolution function is non-empty. By Proposition 5, there is a renaming ρ' such that $\mathsf{mgci}(l) = \rho'(\mathsf{mgci}(\{u_1, \ldots, u_k\})\ \mathsf{mgci}(\{v_1, \ldots, v_k\}))$ since t is label linear. It follows from the induction hypothesis that there is a renaming ρ_0 such that $\mathsf{mgci}(\{u_1, \ldots, u_k\}) = \rho_0 t_0^\#$ (and similarly for $t_1^\#$). Since t is label-linear the domains of ρ_0 and ρ_1 do not intersect, so the composite $\rho'' = \rho_0 \cup \rho_1$ is again a well-formed substitution. Hence, $\mathsf{mgci}(\overline{t_0\ t_1}) = (\rho'' \circ \rho')(t_0^\#\ t_1^\#)$ as required.

□

For the purposes of this proof, we adopt the usual concepts and notations from term rewriting. An occurrence of a subterm s in a term t is a pair $\langle s, C[]\rangle$ consisting of the subterm s and a context $C[]$ such that $C[s] = t$.

The subterm of t at position π is written $t|_\pi$ and the one hole context created by replacing the subterm of t located at position π by a hole is written $t[]_\pi$. We adopt the usual numbering scheme for term trees which is best demonstrated by the following example. Assume s is a term with a redex $F\ \theta x_1 \cdots \theta x_m\ \theta p$ at position π. Then we write $s|_\pi = F\ \theta x_1 \cdots \theta x_m\ \theta p$. Moreover, $s|_{\pi\cdot 1^{m+1}} = \theta p$, $s|_{\pi\cdot 1^{m+1-i}\cdot 2} = \theta x_i$ and $s|_{\pi\cdot 2} = \theta p$. We write the *occurrence* $\langle t|_\pi, t[]_\pi\rangle$ by $t\|_\pi$ as a short-hand.

Given a feasible counterexample

$$Main\ S = s_0 \to s_1 \to \cdots \to s_n$$

there is a labelling of the counterexample which is determined, up to variable renaming, by the procedure labelSeq and which witnesses the feasibility. Given any occurrence of a term in the counterexample, say $s_i\|_\pi$, we write $[\![s_i\|_\pi]\!]$ to mean the labelled term resulting from annotating that occurrence according to labelSeq. When the position $\pi = \epsilon$ we just write $[\![s_i]\!]$.

Lemma 28. *Let $C[F\ \theta x_1 \cdots \theta x_m\ \theta p] \to C[\theta t^\dagger]$ be a reduction in a feasible counterexample. For all positions π_1, π_2 and pattern-matching variables y such that $p|_{\pi_1} = t|_{\pi_2} = y$, there is a substitution σ and:*

$$[\![C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{C[]\cdot 2\cdot \pi_1}]\!]^\# = \sigma[\![C[\theta t^\dagger]\|_{C[]\cdot \pi_2}]\!]^\#$$

Proof. Let $\{V_y^{l_1}, \ldots, V_y^{l_k}\}$ be the set of labelled pattern symbols in $[\![C[\theta t^\dagger]]\!]$ created by the contraction. Since $t|_{\pi_2} = y$, $[\![C[\theta t^\dagger]\|_{\pi_2}]\!] = V_y^{l_i}$ for some $i \in [1..k]$. According to labelSeq 2., the labelled term $[\![C[F\ \ \theta x_1 \cdots \theta x_m\ \ \theta p]\|_{C[]\cdot 2\cdot \pi_1}]\!]$ (which is a labelling of the term θy) must have been labelled according to $\mathsf{labelTm}(\theta y)(\bigcup\{l_1, \ldots, l_k\})$, since the counterexample is feasible. Since $[\![C[\theta t^\dagger]]\!]$ is, by Lemma 25, label-linear, so the sets $\bigcup\{l_1, \ldots, l_k\}$ and l_i are linear. Hence, it follows from Lemma 26 that $[\![C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{C[]\cdot 2\cdot \pi_1}]\!]^\#$ is an instance of $\mathsf{mgci}(l_i) = [\![C[\theta t^\dagger]\|_{C[]\cdot \pi_2}]\!]^\#$. □

Lemma 29. *Let $C[F\ \theta x_1 \cdots \theta x_m\ \theta p] \to C[\theta t^\dagger]$ be a reduction in a feasible counterexample. For all terms t and positions π_1, π_2 such that $C[\theta t^\dagger]\|_{\pi_2}$ is a descendent of $C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{\pi_1}$ and both are occurrences of t, there is a substitution σ and:*

$$[\![C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{\pi_1}]\!]^\# = \sigma[\![C[\theta t^\dagger]\|_{\pi_2}]\!]^\#$$

Proof. By induction on the structure of t:

- When $t = \mathsf{a}$ then
$$[\![C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{\pi_1}]\!]^\# = \mathsf{a} = [\![C[\theta t^\dagger]\|_{\pi_2}]\!]^\#$$
- When $t = F$ and F is not a live symbol then
$$[\![C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{\pi_1}]\!]^\# = F = [\![C[\theta t^\dagger]\|_{\pi_2}]\!]^\#$$
- When $t = F$ and F is a live symbol then
$$[\![C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{\pi_1}]\!]^\# = (F^{l_1})^\# = \mathsf{mgci}(l_1)$$
for some label l_1 and
$$[\![C[\theta t^\dagger]\|_{\pi_2}]\!]^\# = (F^{l_2})^\# = \mathsf{mgci}(l_2)$$
for some label l_2. According to labelSeq 2., since the letter is a descendent of the former, $l_2 \subseteq l_1$. Hence, $\mathsf{mgci}(l_1)$ is an instance of $\mathsf{mgci}(l_2)$, as required.
- When $t = u\ v$ we have
$$[\![C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{\pi_1}]\!]^\# =$$
$$[\![C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{\pi_1\cdot 1}]\!]^\#\ [\![C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{\pi_1\cdot 2}]\!]^\#$$
It follows from the induction hypothesis that there are substitutions σ_1 and σ_2 such that
$$[\![C[F\ \theta x_1 \cdots \theta x_m\ \theta p]\|_{\pi_1\cdot i}]\!]^\# = \sigma_i[\![C[\theta t^\dagger]\|_{\pi_2\cdot i}]\!]^\#$$
for $i \in [1, 2]$. By Lemma 25, $[\![C[\theta t^\dagger]\|_{\pi_2}]\!]$ is label linear so, by Proposition 4, $[\![C[\theta t^\dagger]\|_{\pi_2}]\!]^\#$ is linear and the domains of σ_1 and σ_2 are disjoint. Hence, the composite $\sigma = \sigma_1 \cup \sigma_2$ is again a substitution, as required.

□

Lemma 30. *Let s be a term in a feasible counterexample and let $s|_\pi = F\ \theta x_1 \cdots \theta x_m\ \theta p$ be a redex. For all closed substitutions σ, $\sigma[\![s\|_\pi]\!]^\#$ is again a redex.*

Proof. By definition, we have that:

$$\sigma[\![s\|_\pi]\!]^\# = F\ \sigma[\![s\|_{\pi\cdot 1^{m+1-i}\cdot 2}]\!]^\# \cdots \sigma[\![s\|_{\pi\cdot 1\cdot 2}]\!]^\#\ \sigma[\![s\|_{\pi\cdot 2}]\!]^\#$$

so it remains only to show that there is some substitution τ such that $\sigma[\![s\|_{\pi\cdot 2}]\!]^\# = \tau p$. More precisely, we show that for all positions π: if $s|_\pi = p$ then there is a substitution τ such that $\sigma[\![s\|_\pi]\!]^\# = \tau p$. We proceed by induction on the shape of p.

- When $p = y$, then set $\tau(y) = \sigma[\![s\|_\pi]\!]^\#$.
- When $p = \mathsf{a}\ p_1 \cdots\ p_k$ then

  $$\sigma[\![s\|_\pi]\!]^\# = \mathsf{a}\ \sigma[\![s\|_{\pi\cdot 1^{k-1}\cdot 2}]\!]^\# \cdots \sigma[\![s\|_{\pi\cdot 2}]\!]^\#$$

  It follows from the induction hypothesis that, for each i, there is some substitution τ_i such that $\sigma[\![s\|_{\pi\cdot 1^{k-i}\cdot 2}]\!]^\# = \tau_i\ p_i$. Since patterns are linear, the composite $\tau = \bigcup\{\tau_1, \ldots, \tau_k\}$ is again a substitution, as required.

□

Lemma 31. *Let $s_i \to s_{i+1}$ be a reduction in a feasible counterexample, let $s_i|_{\pi_1} = F\ \theta x_1 \cdots \theta x_m\ \theta p$ be the redex that is contracted and $s_{i+1}|_{\pi_2} = \theta t^\dagger$ be the contractum. For all closed substitutions σ, there is a closed substitution ρ such that $\sigma[\![s_i\|_{\pi_1}]\!]^\#$ and $\rho[\![s_{i+1}\|_{\pi_2}]\!]^\#$ are a valid redex, contractum pair.*

Proof. It follows from Lemma 30 that $\sigma[\![s\|_{\pi_1}]\!]^\#$ is a redex $F\ \tau x_1 \cdots \tau x_m\ \tau p$ with, for all variables $z = s_i|_{\pi_1\cdot\pi_z}$, $\tau(z) = \sigma[\![s_i\|_{\pi_1\cdot\pi_z}]\!]^\#$. It remains to show that, for all t and for all π_2 such that $s_{i+1}\|_{\pi_2} = \theta t^\dagger$: there is a substitution ρ such that $\tau t = \rho[\![s_{i+1}\|_{\pi_2}]\!]^\#$. By induction on t.

- When $t = x_i$, then $\tau t = \sigma[\![s_i\|_{\pi_1\cdot\pi_{x_i}}]\!]^\#$. Since $s_{i+1}\|_{\pi_2}$ is a descendent of $s_i\|_{\pi_1\cdot\pi_{x_i}}$, it follows from Lemma 29 that there is a substitution σ' such that $[\![s_i\|_{\pi_1\cdot\pi_{x_i}}]\!]^\# = \sigma'[\![s_{i+1}\|_{\pi_2}]\!]^\#$. Consequently, $\sigma[\![s_i\|_{\pi_1\cdot\pi_{x_i}}]\!]^\# = \sigma(\sigma'[\![s_{i+1}\|_{\pi_2}]\!]^\#)$ and hence $\rho = \sigma \circ \sigma'$.
- When $t = y \in \mathsf{FV}(p)$, then $\tau t = \sigma[\![s_i\|_{\pi_1\cdot\pi_y}]\!]^\#$. It follows from Lemma 28 that there is a σ' such that $[\![s_i\|_{\pi_1\cdot\pi_y}]\!]^\# = \sigma'[\![s_{i+1}\|_{\pi_2}]\!]^\#$. Consequently, $\sigma[\![s_i\|_{\pi_1\cdot\pi_y}]\!]^\# = \sigma(\sigma'[\![s_{i+1}\|_{\pi_2}]\!]^\#)$ and hence $\rho = \sigma \circ \sigma'$.
- When $t = \mathsf{a}$, then, for all ρ, $\tau t = \mathsf{a} = \rho[\![s_{i+1}\|_{\pi_2}]\!]^\#$.
- When $t = F$ (necessarily not live), then for all ρ, $\tau t = F = \rho[\![s_{i+1}\|_{\pi_2}]\!]^\#$.
- When $t = t_1\ t_2$, then $\tau t = \tau t_1\ \tau t_2$. Since $s_{i+1}|_{\pi_2} = t$, $s_{i+1}|_{\pi_2\cdot 1} = t_1$ and $s_{i+1}|_{\pi_2\cdot 2} = t_2$. Hence, it follows from the induction hypothesis that there are substitutions ρ_1 and ρ_2 such that $\tau t_i = \rho_i[\![s_{i+1}\|_{\pi_2\cdot i}]\!]^\#$ for $i \in [1, 2]$. Since, by Lemma 25, s_{i+1} is label-linear then, by Lemma 4, $[\![s_{i+1}]\!]^\#$ is linear and so there is no intersection between the domains of each ρ_i. Hence the composite $\rho = \rho_1 \cup \rho_2$ is again a substitution. Hence:

  $$\tau t = \rho_1[\![s_{i+1}\|_{\pi_2\cdot 1}]\!]^\#\ \rho_2[\![s_{i+1}\|_{\pi_2\cdot 2}]\!]^\# = \rho[\![s_{i+1}\|_{\pi_2}]\!]^\#$$

  as required.

□

Lemma 32. *Let $s_i \to_{\widetilde{\mathcal{P}_\mathcal{G}}} s_{i+1}$ be a reduction in a feasible counterexample, let $s_i|_{\pi_1} = F\ \theta x_1 \cdots \theta x_m$ be the redex that is contracted and $s_{i+1}|_{\pi_2} = t\ \theta x_1 \cdots \theta x_m$ be the corresponding contractum. For all closed substitutions σ, there is a closed substitution ρ such that $\sigma[\![s_i\|_{\pi_1}]\!]^\# = \rho[\![s_i\|_{\pi_2}]\!]^\#$.*

Proof. By definition, the following identities are true.

$$[\![s_i\|_{\pi_1}]\!]^\# = [\![s_i\|_{\pi_1\cdot 1^m}]\!]^\#\ [\![s_i\|_{\pi_1\cdot 1^{m-1}\cdot 2}]\!]^\# \cdots [\![s_i\|_{\pi_1\cdot 2}]\!]^\#$$

$$[\![s_{i+1}\|_{\pi_2}]\!]^\# = [\![s_{i+1}\|_{\pi_2\cdot 1^m}]\!]^\#\ [\![s_{i+1}\|_{\pi_2\cdot 1^{m-1}\cdot 2}]\!]^\# \cdots [\![s_{i+1}\|_{\pi_2\cdot 2}]\!]^\#$$

Since, for each $j \in [1..k]$, $s_{i+1}\|_{\pi_2\cdot 1^{m-j}\cdot 2}$ is a descendent of $s_i\|_{\pi_1\cdot 1^{m-j}\cdot 2}$, it follows from Lemma 29 that there is a substitution σ_j such that:

$$[\![s_i\|_{\pi_1\cdot 1^{m-j}\cdot 2}]\!]^\# = \sigma_j[\![s_{i+1}\|_{\pi_2\cdot 1^{m-j}\cdot 2}]\!]^\#$$

Hence, it remains only to show that there is a substitution σ' such that

$$[\![s_i\|_{\pi_1\cdot 1^m}]\!]^\# = \sigma'[\![s_{i+1}\|_{\pi_1\cdot 1^m}]\!]^\#$$

By labelSeq 1., the labelled term $[\![s_i\|_{\pi_1\cdot 1^m}]\!]$ is annotated with the label $\overline{s_{i+1}\|_{\pi_2\cdot 1^m}}$. Hence, it follows from Lemma 27 that there is a substitution σ' such that:

$$[\![s_i\|_{\pi_1\cdot 1^m}]\!]^\# = \mathsf{mgci}(\overline{s_{i+1}\|_{\pi_2\cdot 1^m}}) = \sigma'[\![s_{i+1}\|_{\pi_2\cdot 1^m}]\!]^\#$$

Since, by Lemma 25 and Proposition 4, the term $[\![s_{i+1}]\!]^\#$ is linear, the composite mapping $\rho = \sigma' \cup \sigma_1 \cup \cdots \cup \sigma_m$ is again a closed substitution, as required. □

Lemma 33. *For all $n \in \mathbb{N}$, if*

$$Main\ S = s_0 \to \cdots \to s_n$$

is a feasible reduction sequence in $\widetilde{\mathcal{P}_\mathcal{G}}$ then, for each $i \in [0..n]$ there exists a closed substitution ρ_i such that:

$$\rho_0[\![s_0]\!]^\# \Rightarrow^{0,1} \cdots \Rightarrow^{0,1} \rho_n[\![s_n]\!]^\#$$

is a valid reduction sequence in $\mathcal{P}$ starting from a term in $\mathcal{L}(\mathcal{G})$, where $u \Rightarrow^{0,1} v$ iff $u = v$ or $u \Rightarrow v$.

Proof. By induction on n.

- When $n = 0$, $s_0 = Main\ S$ and, since $Main$ is not live:

  $$[\![s_0]\!]^\# = Main\ (S^l)^\# = Main\ \mathsf{mgci}(l)$$

  for some label l. Since the reduction sequence is feasible, there is a closed instance t of the MGCI of l in $\mathcal{L}(\mathcal{G})$, i.e. $t = \rho_0\ \mathsf{mgci}(l) \in \mathcal{L}(\mathcal{G})$ and the singleton sequence $Main\ t$ is indeed a valid reduction sequence in $\mathcal{P}$ starting from a term in $\mathcal{L}(\mathcal{G})$.
- When $n = k + 1$, then the feasible reduction sequence is of the form:

  $$Main\ S \to \cdots \to s_k \to s_{k+1}$$

  It follows from the induction hypothesis that there is a valid $\mathcal{P}$ sequence (starting from a term t in $\mathcal{L}(\mathcal{G})$):

  $$Main\ t \Rightarrow^{0,1} \cdots \Rightarrow^{0,1} \rho_k[\![s_k]\!]^\#$$

  We analyse the form of the $\widetilde{\mathcal{P}_\mathcal{G}}$ reduction $s_k \to s_{k+1}$.
  (i) When $s_k = C[F\ \theta x_1 \cdots \theta x_m\ \theta p] \to_\mathcal{P} C[\theta t^\dagger] = s_{k+1}$, then, by Lemma 31, there exists a substitution ρ_{k+1} and the $\mathcal{P}$-reduction sequence can be extended by the reduction $\rho_k[\![s_k]\!]^\# \Rightarrow \rho_{k+1}[\![s_{k+1}]\!]^\#$.
  (ii) When $s_k = C[F\ \theta x_1 \cdots \theta x_m] \to_{\widetilde{\mathcal{P}_\mathcal{G}}} C[t\ \theta x_1 \cdots \theta x_m] = s_{k+1}$ then, by Lemma 32, there exists a substitution ρ_{k+1} such that $\rho_k[\![s_k]\!]^\# = \rho_{k+1}[\![s_{k+1}]\!]^\#$.

□

Given a feasible, abstract counterexample $Main\ S \to^* \theta q$ (in which θq is the error witness) to a property $\varphi = \mathcal{L}(\mathcal{A})$, it follows from Lemma 33 that there is some closed substitution ρ, a closed constructor term $t \in \mathcal{L}(\mathcal{G})$ and a $\mathcal{P}$-reduction sequence

$Main\ t \Rightarrow^* \rho[\![\theta q]\!]^{\#}$. It follows from the proof of Lemma 30 that $\rho[\![\theta q]\!]^{\#}$ is a term of the form τq for some closed substitution τ. Since, by definition, the pattern q does not match any prefix of any tree in $\mathcal{L}(\mathcal{A})$, this $\mathcal{P}$-reduction sequence is a witness to the absurdity of $\vDash (\mathcal{P}, \mathcal{G}, \mathcal{A})$.

D. Semi-completeness of abstraction-refinment

D.1 Proof of Lemma 9

Let $\mathcal{P} = \langle \Sigma, \mathcal{N}, \mathcal{R}, Main \rangle$ be a PMRS and $\widetilde{\mathcal{P}_\mathcal{G}}$ be the abstraction of $\mathcal{P}$ (starting from terms in $\mathcal{L}(\mathcal{G})$). Let α be a counterexample trace of $\widetilde{\mathcal{P}_\mathcal{G}}$ which is spurious with $\mathsf{Failures}$ set S. Let d be the depth profile with domain $\mathcal{N}$ defined by:

$$d(F) = \mathsf{dp}(\mathcal{R})(F) + \bigsqcup_{\mathbb{N}} \{\, \mathsf{depth}(t) \mid (F,\, P) \in S,\, t \in P \,\}$$

and let $\mathcal{P}'$ be the d-unfolding of $\mathcal{P}$. Then α is not a reduction sequence in the abstraction $\widetilde{\mathcal{P}'_\mathcal{G}}$ of $\mathcal{P}'$.

Proof. Since α is a spurious trace, necessarily there is some $(F,\, X) \in \mathsf{Failures}$. For this to be the case, there must be some redex, contractum pair $F\ \theta x_1\ \cdots\ \theta x_m\ \ \theta p,\ \theta t^\dagger$ with $\mathsf{FV}(p) \cap \mathsf{FV}(t) \neq \emptyset$. We will show that either $F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ is not a redex in $\mathcal{G}'$ or that $\theta t^\dagger$ is no longer its contractum. In $\widetilde{\mathcal{P}'_\mathcal{G}}$, the depth of the largest pattern in a defining rule for F is at least $\delta = \mathsf{depth}(p) + \mathsf{max}(\{\, \mathsf{depth}(x) \mid x \in X \,\})$. We can distinguish two cases:

(1) When $\mathsf{depth}(\theta p) > \delta$, there is a rule $F\ x_1\ \cdots\ x_m\ \sigma p \longrightarrow (\sigma t)^\dagger$ in $\widetilde{\mathcal{P}'_\mathcal{G}}$ with $\mathsf{depth}(\sigma p) = \delta$. Hence, θ can be decomposed into $\tau \circ \sigma$ for some τ. In this case, $F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ is still a redex, but its contractum is $\tau(\sigma t)^\dagger \neq (\tau \circ \sigma) t^\dagger$ since t has free pattern variables and σ is non-trivial.

(2) When $\mathsf{depth}(\theta p) \leq \delta$, we can distinguish two further cases. When $\theta p \in T(\Sigma)$, there is a new rule in $\widetilde{\mathcal{P}'_\mathcal{G}}$ which is exactly $F\ x_1\ \cdots\ x_m\ \theta p \longrightarrow \theta t$ and hence $F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ is still a redex in $\widetilde{\mathcal{P}'_\mathcal{G}}$ but it contracts to $\theta t \neq \theta(t^\dagger)$ since t has free pattern variables. When $\theta p \notin T(\Sigma)$, then $F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ is no longer a redex since, in any family of patterns generated by $\mathsf{pats}_b(n)$ for some n, no pattern of depth $< n$ contains free variables.

□

D.2 Proof of Lemma 10

Fix $n \in \mathbb{N}$. Then given any PMRS $\mathcal{P}$ and input grammar $\mathcal{G}$, there is a depth-profile δ such that, if $\widetilde{\mathcal{P}'_\mathcal{G}}$ is the abstraction of the d-unfolding of $\mathcal{P}$, then all length-$m \leq n$ reduction sequences in $\widetilde{\mathcal{P}'_\mathcal{G}}$ are feasible.

Proof. Let $n \in \mathbb{N}$. Given any abstract reduction sequence of length n (in any abstraction), there is a maximum size of data structure that can be constructed, i.e. there is an upper-bound on the depth of any subterm of any term appearing in the sequence. Since, in any wPMRS abstraction, each terminal symbol *qua* constructor a appears only in the body of the defining rule for K_a, this upper-bound can be crudely taken to be n. Let depth-profile δ be the constant function $\delta(F) = n + 1$ for all F and consider the abstraction $\widetilde{\mathcal{P}'_\mathcal{G}}$ of the δ-unfolding of $\mathcal{P}$. Moreover, consider any redex $F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ which is contracted during a length-$m \leq n$ reduction sequence α in $\widetilde{\mathcal{P}'_\mathcal{G}}$. Since the upper-bound on the depth of any constructor subterm of a term in α is n, in particular $\mathsf{depth}(\theta p) \leq n$. By definition of $\mathsf{pats}_b(n+1)$, necessarily $\mathsf{FV}(p) = \emptyset$ and, assuming the corresponding defining rule for F in the d-unfolding of $\mathcal{P}$ is $F\ x_1\ \cdots\ x_m\ p \longrightarrow t$, the contractum of $F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ is exactly θt. Since every contractum in α is of this form, there are no occurrences of pattern-symbols anywhere in α and hence $\mathsf{labelSeq}(\alpha)$ will always succeed with no failures. □

D.3 Proof of Semi-completeness

Lemma 34. *Assume that the model-checker always reports shortest counterexamples and fix $n \in \mathbb{N}$. Eventually either the algorithm terminates or produces a d-unfolding $\mathcal{P}'$ of the input PMRS $\mathcal{P}$ such that there are no length $m \leq n$ spurious traces in the abstraction $\widetilde{\mathcal{P}'_\mathcal{G}}$ of $\mathcal{P}'$.*

Proof. Assume for contradiction that the algorithm neither terminates nor produces such an unfolding $\mathcal{P}'$. Since the algorithm does not terminate, it will eventually produce an unfolding $\mathcal{P}''$, an abstraction $\widetilde{\mathcal{P}''_\mathcal{G}}$ of $\mathcal{P}''$ and a shortest counterexample α in $\widetilde{\mathcal{P}''_\mathcal{G}}$ which share the following characteristics:

(i) By Lemma 10, $\mathcal{P}''$ must have a depth profile d'' such that $\delta \not\leq d''$.

(ii) Since there are only finitely many depth profiles $\leq \delta$, there must be some $F \in \mathcal{N}$ such that $d''(F) > \delta(F)$ and after $\mathsf{labelSeq}(\alpha)$, $(F, X) \in \mathsf{Failures}(\alpha)$ for some non-empty X.

(iii) By assumption, there exist spurious counterexample traces in $\widetilde{\mathcal{P}''_\mathcal{G}}$ of length $m \leq n$.

However, by (ii), it is necessary that there is some redex $F\ \theta x_1\ \cdots\ \theta x_m\ \theta p$ in α such that $\mathsf{depth}(\theta p) > d''(F) > \delta(F) = n$. A redex of this form could only occur in a (shortest) counterexample of length $> n$, contradicting (iii). □

The semi-completeness is stated as follows. Assume that the model checker always produces shortest counterexamples and let $(\mathcal{P},\, \mathcal{G},\, \mathcal{A})$ be a no-instance of the verification problem. Then the algorithm terminates with a feasible counterexample trace.

Proof. Let $(\mathcal{P},\, \mathcal{G},\, \mathcal{A})$ be a no-instance. So there is a constructor term $s \in \mathcal{L}(\mathcal{G})$ and a $\mathcal{P}$-reduction sequence $\alpha = Main\ s \Rightarrow^* \theta t$ where pattern t is not a prefix of any tree in $\mathcal{L}(\mathcal{A})$. Then there is a largest size of constructor term that occurs in α, say n.

Assume for contradiction that the abstraction-refinement loop does not terminate. Then let us organise the iterations of the loop into an infinite sequence of triples $(\mathcal{P}_i,\, \widetilde{\mathcal{P}_{\mathcal{G}_i}},\, \beta_i)_{i \in \omega}$ comprising the PMRS which is given as input to the iteration, the corresponding wPMRS abstraction and the abstract counterexample arising from that abstraction respectively. By reasoning similar to that in the proof of Lemma 10, for all i, $\delta \not\leq \mathsf{dp}(\mathcal{R}_i)$ where $\delta(F) = n+1$ for all F. Moreover, we can be sure that there is an iteration k such that:

1. For all $i \geq k$, $(F,\, X) \in \mathsf{Failures}(\beta_i)$ implies $\mathsf{dp}(\mathcal{R}_i)(F) > \delta(F)$, since there are only a finite number of depth profiles $\leq \delta$.
2. By the soundness of the abstraction (Theorem 2) there is an abstract $\widetilde{\mathcal{P}_{\mathcal{G}k}}$-counterexample witnessing the no-instance, say of length m.

By (i), the same length-m conterexample will exist in every abstraction $\widetilde{\mathcal{P}_{\mathcal{G}_i}}$ for $i \geq k$. But, it follows from Lemma 34 that eventually a PMRS $\mathcal{P}_j$ is produced such that there are no length $l \leq m$ spurious traces arising from $\widetilde{\mathcal{P}_{\mathcal{G}_j}}$ and hence the length of β_j is strictly larger than m. But this contradicts the assumption that the model-checker produces shortest counterexamples since $\mathcal{P}_{\mathcal{G}_j}$ includes the witnessing counterexample which is of length m. □