Verification of Program Transformations with Inductive Refinement Types*

AHMAD SALIM AL-SIBAHI, University of Copenhagen, Denmark
THOMAS P. JENSEN, INRIA Rennes, France and University of Copenhagen, Denmark
ALEKSANDAR S. DIMOVSKI, Mother Teresa University, Skopje, Macedonia
ANDRZEJ WĄSOWSKI, IT University of Copenhagen, Denmark

High-level transformation languages like Rascal include expressive features for manipulating large abstract syntax trees: first-class traversals, expressive pattern matching, backtracking and generalized iterators. We present the design and implementation of an abstract interpretation tool, Rabit, for verifying inductive type and shape properties for transformations written in such languages. We describe how to perform abstract interpretation based on operational semantics, specifically focusing on the challenges arising when analyzing the expressive traversals and pattern matching. Finally, we evaluate Rabit on a series of transformations (normalization, desugaring, refactoring, code generators, type inference, etc.) showing that we can effectively verify stated properties.

CCS Concepts: • **Theory of computation → Program verification**; **Program analysis**; **Abstraction**; *Functional constructs*; *Program schemes*; *Operational semantics*; Control primitives; • **Software and its engineering → Translator writing systems and compiler generators**; Semantics;

Additional Key Words and Phrases: transformation languages, abstract interpretation, static analysis

ACM Reference Format:
Ahmad Salim Al-Sibahi, Thomas P. Jensen, Aleksandar S. Dimovski, and Andrzej Wąsowski. 2020. Verification of Program Transformations with Inductive Refinement Types. *ACM Trans. Softw. Eng. Methodol.* 1, 1, Article 1 (January 2020), 33 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Transformations play a central role in software development. Examples include desugaring, model transformations, refactoring, and code generation. The artifacts involved in transformations—e.g., structured data, domain-specific models, and code—often have large abstract syntax, spanning hundreds of syntactic elements, and a correspondingly rich semantics. Writing transformations is thus a tedious and error-prone process. Specialized languages and frameworks with high-level features have been developed to address this challenge of writing and maintaining transformations. These languages include Rascal [36], Stratego/XT [13], TXL [17], Uniplate [39] for Haskell, and

*Based on [5]

Authors' addresses: Ahmad Salim Al-Sibahi, University of Copenhagen, Department of Computer Science, Denmark, ahmad@di.ku.dk; Thomas P. Jensen, INRIA Rennes, France , University of Copenhagen, Department of Computer Science, Denmark, thomas.jensen@inria.fr; Aleksandar S. Dimovski, Mother Teresa University, Skopje, Macedonia, aleksandar.dimovski@unt.edu.mk; Andrzej Wąsowski, IT University of Copenhagen, Denmark, wasowski@itu.dk.

1049-331X/2020/1-ART1 $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Kiama [53] for Scala. For example, Rascal combines a functional core language supporting state and exceptions, with constructs for processing of large structures.

```
1 public Script flattenBlocks(Script s) {
2   solve(s) {
3     s = bottom-up visit(s) {
4       case stmtList: [*xs,block(ys),*zs] =>
5                 xs + ys + zs
6     }
7   }
8   return s;
9 }
```

Fig. 1. Transformation in Rascal that flattens all nested blocks in a statement

Figure 1 shows an example Rascal transformation program taken from a PHP analyzer.[1] This transformation program recursively flattens all blocks in a list of statements. The program uses the following core Rascal features:

- A *visitor* (`visit`) to traverse and rewrite all statement lists containing a block to a flat list of statements. Visitors support various strategies, like the `bottom-up` strategy that traverses the abstract syntax tree starting from leaves toward the root.
- An *expressive pattern matching* language is used to non-deterministically find blocks inside a list of statements. The starred variable patterns `*xs` and `*zs` match arbitrary number of elements in the list, respectively before and after the `block(ys)` element. Rascal supports non-linear matching, negative matching and specifying patterns that match deeply nested values.
- The solve-loop (`solve`) performing the rewrite until a fixed point is reached (the value of *s* stops changing).

To rule out errors in transformations, we propose a static analysis for enforcing type and shape properties, so that target transformations produce output adhering to particular shape constraints. For our PHP example, this would include:

- The transformation preserves the constructors used in the input: does not add or remove new types of PHP statements.
- The transformation produces flat statement lists, i.e., lists that do not recursively contain any block.

To ensure such properties, a verification technique must reason about shapes of inductive data—also inside collections such as sets and maps—while still maintaining soundness and precision. It must also track other important aspects, like cardinality of collections, which interact with target language operations including pattern matching and iteration.

In this paper, we address the problem of verifying type and shape properties for high-level transformations written in Rascal and similar languages. We show how to design and implement a static analysis based on abstract interpretation. Concretely, our contributions are:

(1) An abstract interpretation-based static analyzer—Rascal ABstract Interpretation Tool (Rabit)—that supports inferring types and inductive shapes for a large subset of Rascal.

[1] https://github.com/cwi-swat/php-analysis

(2) An evaluation of Rabit on several program transformations: two refactorings, desugaring, two normalization procedures, derivative calculator, modernization transformation, code generator, and language implementation of an expression language.
(3) A modular design for abstract shape domains, that allows extending and replacing abstractions for concrete element types, e.g. extending the abstraction for lists to include length in addition to shape of contents.
(4) Schmidt-style abstract *operational* semantics [49] for a significant subset of Rascal adapting the idea of *trace memoization* to support arbitrary recursive calls with input from infinite domains.

This journal paper extends our conference version [5] with the following contributions:

(1) A significant extension of the number of evaluation cases (from 5 to 9) and properties (from 15 to 24). The new properties explore better interactions with abstractions on integers, iteration, conditioning using if-statements and maps.
(2) An explication of the effect of pragmatic choices regarding trace partitioning and trace memoization in our implementation. We now run our evaluation under 2×3 configurations exploring choices on trace partitioning (on/off) and widening domain for trace memoization (simple/type-based/constructor-based). We further provide a new extended discussion on these effects in the evaluation.
(3) We now look into the statistics of the operations of trace memoization (hit/miss/widen) during execution. We compare the effects to the execution time and configuration choice.

Together, these contributions show feasibility of applying abstract interpretation for constructing analyses for expressive transformation languages and properties.

We proceed by presenting a running example in Sect. 2. We introduce the key constructs of Rascal in Sect. 3. Section 4 describes the modular construction of abstract domains. Sections 5 to 8 describe abstract semantics. We evaluate the analyzer on realistic transformations, reporting results in Sect. 9. Sections 10 and 11 discuss related papers and conclude.

2 MOTIVATION AND OVERVIEW

Verifying types and state properties such as the ones stated for the program of Fig. 1 poses the following key challenges:

- The programs use *heterogeneous inductive data types* that can contain *collections* (lists, maps and sets), and basic data (integers, strings, etc.). Modelling these types precisely requires us to construct sophisticated abstract domains that capture the mutual interaction between the different types.
- The traversal of syntax trees depends heavily on the *type and shape of input*, on a *complex program state*, and involves *unbounded recursion*. This makes it challenging to approximate required invariants in a way that is both sufficiently precise and terminating.
- Backtracking and exceptions in large programs introduce the possibility of *state-dependent non-local jumps*. This makes it difficult to statically calculate the control flow of target programs and have a compositional denotational semantics, instead of an operational one.

Figure 2 presents a small pedagogical example using visitors. The program performs expression simplification by traversing a syntax tree bottom-up and reducing multiplications by constant zero. We use this example to explain the analysis techniques contributed in this paper.

Inductive Refinement Types. Rabit is a static analysis tool that given the shape of possible inputs (any expression of type Expr in this case), infers the shape of possible outputs. It infers inductive refinement types by interpreting the simplification program abstractly, considering all possible

```
1 data Nat = zero() | suc(Nat pred);
2 data Expr = var(str nm) | cst(Nat vl)
3           | mult(Expr el, Expr er);
4
5 Expr simplify(Expr expr) =
6   bottom-up visit (expr) {
7       case mult(cst(zero()), y) => cst(zero())
8       case mult(x, cst(zero())) => cst(zero())
9   };
```

Fig. 2. The running example: eliminating multiplications by zero from expressions

paths the program can take. The result of running Rabit on this case is:

$$\text{success}\ \text{Expr}'' \qquad \text{fail}\ \text{Expr}''$$

where $\text{Expr}'' = \text{cst}\,(\text{Nat}) \wr \text{var}\,(\text{str}) \wr \text{mult}\,(\text{Expr}', \text{Expr}')$ and $\text{Expr}' = \text{cst}\,(\text{suc}\,(\text{Nat})) \wr \text{var}\,(\text{str}) \wr \text{mult}\,(\text{Expr}', \text{Expr}')$.

We briefly interpret how to read the definition of type Expr''. The bar $\wr$ denotes a choice between alternative constructors. If the input was rewritten during traversal (success, the first term) then the resulting syntax tree contains no multiplications by zero. Observe how the last alternative $\text{mult}\,(\text{Expr}', \text{Expr}')$ contains only expressions of type Expr', which in turn only allows multiplications by constants constructed using $\text{suc}\,(\text{Nat})$ (that is ≥ 1). If the traversal failed to match (fail, the second term), then the input did not contain any multiplication by zero to begin with and so does not the output, which has not been rewritten.

The success and failure cases happen to be the same for our example, but this is not necessarily always the case. For example, if we instead visited a statement $\text{return}\,(\text{Expr})$, the top-level result would be $\text{fail}\,\text{return}\,(\text{Expr}'')$ where the inner expression has been rewritten but the top-level statement would fail to match any of the cases (no success). Keeping separate result values allows retaining precision throughout the traversal, and better reflects concrete execution paths. We now proceed discussing how Rabit can infer this shape using abstract interpretation.

Abstractly Interpreting Traversals. The core idea of abstractly executing a traversal is similar to concrete execution: we recursively traverse the input structure and rewrite the values that match target patterns. However, because the input is abstracted, we must make sure to take into account all applicable paths. Figure 3 shows the execution tree of the traversal on the simplification example (Fig. 2) when it starts with shape $\text{mult}\,(\text{cst}\,(\text{Nat}), \text{cst}\,(\text{Nat}))$. Since there is only one constructor, it will initially *recurse* down to traverse the contained values (children) creating a new recursion node (yellow, light shaded) in the figure (ii) containing the left child $\text{cst}\,(\text{Nat})$, and then recurse again to create a node (iii) containing Nat. Observe here that Nat is an abstract type with two possible constructors (zero, $\text{suc}\,(\cdot)$), and during abstract interpretation the identity of these constructors are unknown. When Rabit hits a choice between alternative constructors, it explores each alternative separately creating new *partition* nodes (blue, darker). In our example we partition the Nat type into its constructors zero (node iv) and $\text{suc}\,(\text{Nat})$ (node v). The zero case now represents the first case without children and we can run the visitor operations on it. Since no pattern matches zero it will return a $\text{fail}\,\text{zero}$ result indicating that it has not been rewritten. For the $\text{suc}\,(\text{Nat})$ case it will try to recurse down to Nat (node vi) which is equal to (node iii). We observe a problem: if we continue our traversal algorithm as is, we will not terminate and get a result. To provide a terminating algorithm we will resort to using *trace memoization*.

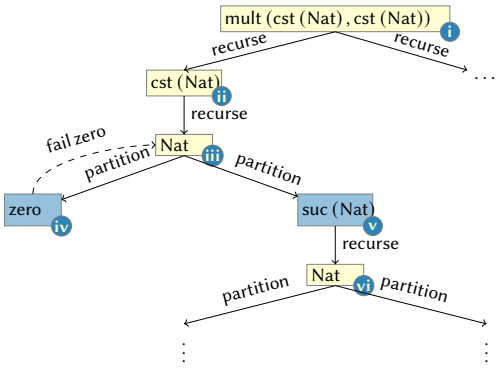

Fig. 3. Naively abstractly interpreting the simplification example from Fig. 2 with initial input mult (cst (Nat) , cst (Nat)). The procedure does not terminate because of infinite recursion on Nat.

Partition-driven Trace Memoization. The idea is to detect paths where execution recursively meets similar input, merging the new recursive node with the similar previous one, thus creating a loop in the execution graph [46, 49]. This loop is then resolved by a fixed-point iteration.

In Rabit, we propose *partition-driven trace memoization*, which works with potentially unbounded input like the inductive type refinements. We detect cycles by maintaining a *memoization map* which for each type—used for partitioning—stores the last traversed value (input) and the last result produced for this value (output). This memoization map is initialized to map all types to the bottom element ($\bot$) for both input and output. The evaluation is modified to use the memoization map, so it checks on each iteration the input i against the map:

- If the last processed refinement type representing the input i' is greater than the current input ($i' \sqsupseteq i$), then it uses the corresponding output; i.e., we found a hit in the memoization map.
- Otherwise, it will merge the last processed and current input refinement types to a new value $i'' = i' \nabla i$, update the memoization map and continue execution with i''. The operation ∇ is called a *widening*; it ensures that the result is an upper bound of its inputs, i.e., $i' \sqsubseteq i'' \sqsupseteq i$ and that the merging will eventually terminate for the increasing chain of values. The memoization map is updated to map the general type of i'' (not refined, for instance Nat) to map to a pair (i'', o), where the first component denotes the new input i'' refinement type and the second component denotes the corresponding output o refinement type; initially, o is set to $\bot$ and then changed to the result of executing input i'' repeatedly until a fixed-point is reached.

We demonstrate the trace memoization and fixed-point iteration procedures on Nat in Fig. 4, beginning with the leftmost tree. The expected result is fail Nat, meaning that no pattern has matched, no rewrite has happened, and a value of type Nat is returned, since the simplification program only introduces changes to values of type Expr.

We show the memoization map inside a framed orange box. The result of the widening is presented below the memoization map. In all cases the widening in Fig. 4 is trivial, as it happens against $\bot$. The final line in node 1 stores the value o_{prev} produced by the previous iteration of the traversal, to establish whether a fixed point has been reached ($\bot$ initially).

Trace Partitioning. We *partition* [44] the abstract value Nat along its constructors: zero and suc $(\cdot)$ (Fig. 4). This partitioning is key to maintain precision during the abstract interpretation. As in Fig. 3, the left branch fails immediately, since no pattern in Fig. 2 matches zero. The right branch

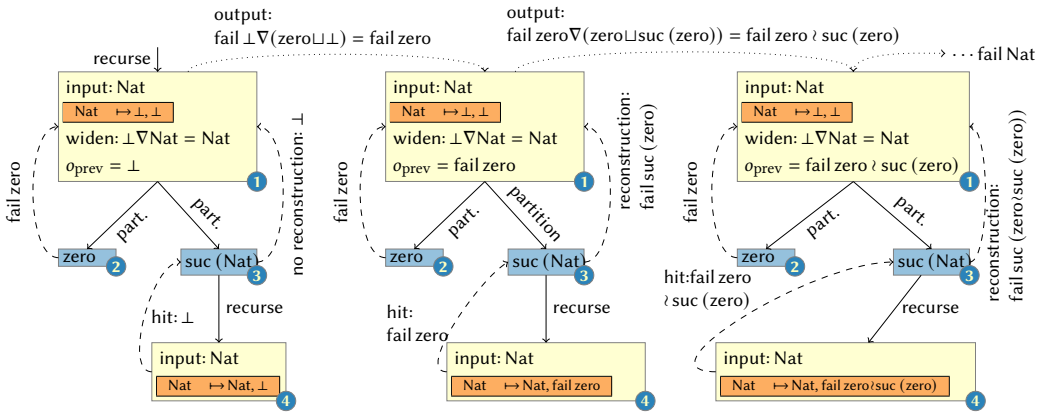

Fig. 4. Three iterations of a fixed point computation for input Nat. Iterations are separated by dotted arrows on top

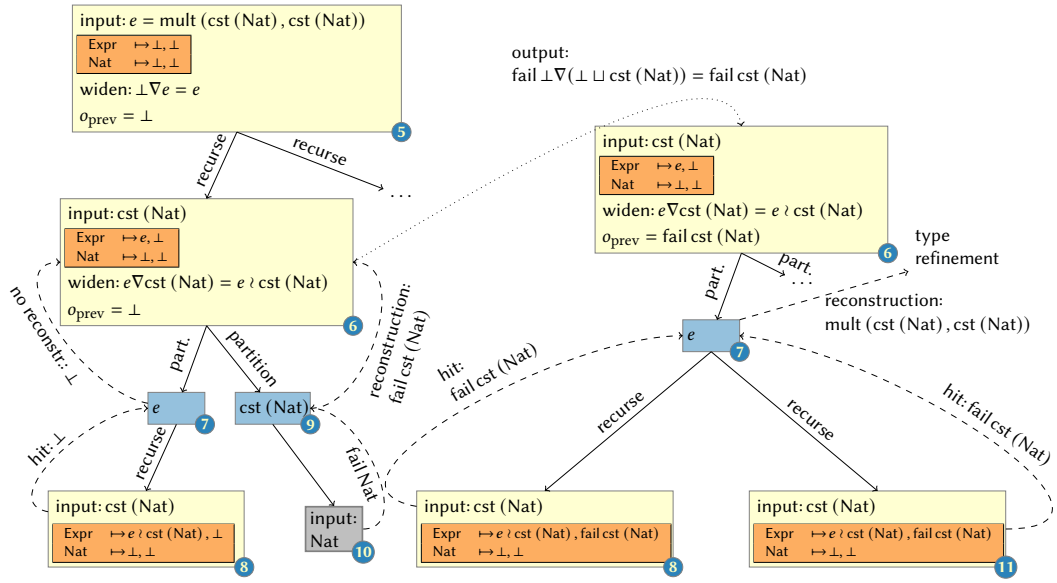

Fig. 5. A prefix of the abstract interpreter run for e = mult (cst (Nat) , cst (Nat)). Fragments of two iterations involving node 6 are shown, separated by a dotted arrow.

descends into a new recursion over Nat, with an updated memoization table. This run terminates, due to a hit in the memoization map, returning ⊥. After returning, the value of suc (Nat) should be reconstructed with the result of traversing the child Nat, but since the result is ⊥ there is no value to reconstruct with, so ⊥ is just propagated upwards. At the return to the last widening node, the values are joined, and widen the previous iteration result o_{prev} (the dotted arrow on top). This process repeats in the second and third iterations, but now the reconstruction in node 3 succeeds: the child Nat is replaced by zero and fail suc (zero) is returned (dashed arrow from 3 to 1). In the third iteration, we join and widen the following components (cf. o_{prev} and the dashed arrows

incoming into node 1 in the rightmost column):

$$[\text{zero} \wr \text{suc}\,(\text{zero}) \;\nabla\; (\text{zero} \sqcup \text{suc}\,(\text{zero} \wr \text{suc}\,(\text{zero})))] = \text{Nat}$$

Here, the used widening operator [19] accelerates the convergence by increasing the value to represent the entire type Nat. It is easy to convince yourself, by following the same recursion steps as in the figure, that the next iteration, using $o_{\text{prev}} = \text{Nat}$ will produce Nat again, arriving at a fixed point. Observe, how consulting the memoization map, and widening the current value accordingly, allowed us to avoid infinite recursion over unfoldings of Nat.

Nesting Fixed Point Iterations. When inductive shapes (e.g., Expr) refer to other inductive shapes (e.g., Nat), it is necessary to run nested fixed-point iterations to solve recursion at each level. Figure 5 returns to the more high-level fragment of the traversal of Expr starting with $\text{mult}\,(\text{cst}\,(\text{Nat})\,, \text{cst}\,(\text{Nat}))$ as in Fig. 3. We follow the recursion tree along nodes 5, 6, 7, 8, 9, 10, 9, 6 with the same rules as in Fig. 4. In node 10 we run a nested fixed point iteration on Nat, already discussed in Fig. 4, so we just include the final result.

Type Refinement. The output of the first iteration in node 6 is $\text{fail}\,\text{cst}\,(\text{Nat})$, which becomes the new o_{prev}, and the second iteration begins (to the right). After the widening the input is partitioned into e (node 7) and $\text{cst}\,(\text{Nat})$(node elided). When the second iteration returns to node 7 we have the following reconstructed value: $\text{mult}\,(\text{cst}\,(\text{Nat})\,, \text{cst}\,(\text{Nat}))$. Contrast this with lines 6-7 in Fig. 2, to see that running the abstract value against this pattern might actually produce success. In order to obtain precise result shapes, we refine the input values when they fail to match a pattern. Our abstract interpreter produces a refinement of the type, by running it through the pattern matching, giving:

$$\begin{aligned}&\text{success}\,\text{cst}\,(\text{Nat})\\&\text{fail}\,\text{mult}\,(\text{cst}\,(\text{suc}\,(\text{Nat}))\,,\; \text{cst}\,(\text{suc}\,(\text{Nat})))\end{aligned}$$

The result means, that if the pattern match succeeds then it produces an expression of type $\text{cst}\,(\text{Nat})$. More interestingly, if the matching failed neither the left nor the right argument of $\text{mult}\,(\cdot, \cdot)$ could have contained the constant zero—the interpreter captured some aspect of the semantics of the program by *refining* the input type. Naturally, from this point on the recursion and iteration continues, but we shall abandon the example, and move on to formal developments.

3 FORMAL LANGUAGE

The presented technique is meant to be general and applicable to many high-level transformation languages. To keep the presentation concise, we focus on few key constructs from Rascal [36], relying on the concrete semantics from Rascal Light [2].

We consider algebraic data types (at) and finite sets ($\text{set}\langle t\rangle$) of elements of type t. Each algebraic data type at has a set of unique constructors. Each constructor $k(\underline{t})$ has a fixed set of typed parameters. The language includes sub-typing, with void and value as bottom and top types respectively.

$$t \in \text{Type} ::= \text{void} \mid \text{set}\langle t\rangle \mid at \mid \text{value}$$

We consider the following subset of Rascal expressions: From left to right we have: variable access, assignments, sequencing, constructor expressions, set literal expressions, matching failure expression, and bottom-up visitors:

$$e ::= x \in \text{Var} \mid x = e \mid e; e \mid k(\underline{e}) \mid \{\underline{e}\} \mid \text{fail} \mid \text{visit}\; e\; \underline{cs} \qquad cs ::= \text{case}\; p \Rightarrow e$$

Visitors are a key construct in Rascal. A visitor $\text{visit}\; e\; \underline{cs}$ traverses recursively the value obtained by evaluating e (any combination of simple values, data type values and collections). During the

traversal, case expression $\underline{cs}$ are applied to the nodes, and the values matching target patterns are rewritten. We will discuss a concrete subset of patterns p further in Sect. 6. For brevity, we only discuss bottom-up visitors, but Rabit (Sect. 9) supports all strategies of Rascal.

Notation. We write $(x, y) \in f$ to denote the pair (x, y) such that $x \in \operatorname{dom} f$ and $y = f(x)$. Abstract semantic components, sets, and operations are marked with a hat: $\widehat{a}$. A sequence of $e_1, \ldots, e_n$ is contracted using an underlining $\underline{e}$. The empty sequence is written by ε, and concatenation of sequences $\underline{e_1}$ and $\underline{e_2}$ is written $\underline{e_1}, \underline{e_2}$. Notation is lifted to sequences in an intuitive manner: for example given a sequence $\underline{v}$, the value v_i denotes the ith element in the sequence, and $\underline{v : t}$ denotes the sequence $v_1 : t_1, \ldots, v_n : t_n$.

4 ABSTRACT DOMAINS

Our abstract domains are designed to compose modularly. Modularity is important for transformation languages, which manipulate a large variety of values. The design allows easily replacing abstract domains for particular types of values, as well as adding support for new types. We want to construct an abstract value domain $\widehat{vs} \in \widehat{\mathsf{ValueShape}}$ which captures inductive refinement types of form:

$$at^r = k_1(\underline{\widehat{vs}_1}) \wr \cdots \wr k_n(\underline{\widehat{vs}_n})$$

where each value $\widehat{vs}_i$ can possibly recursively refer to at^r. Below, we define abstract domains for sets, data types and recursively defined domains. The modular domain design generalizes parameterized domains [18] to follow a design inspired by the modular construction of types and domains [9, 16, 50]. The idea is to define domains parametrically—i.e. in the form $\widehat{\mathsf{F}}(\underline{\widehat{\mathsf{E}}})$—so that abstract domains for subcomponents are taken as parameters, and explicit recursion is handled separately. We use standard domain combinators [59] to combine individual domains into the abstract value domain.

Set Shape Domain. Let $\mathsf{Set}(\mathsf{E})$ denote the domain of finite sets consisting of elements taken from E. We define abstract finite sets using abstract elements $\{\widehat{e}\}_{[l;u]}$ from a parameterized domain $\widehat{\mathsf{SetShape}}(\widehat{\mathsf{E}})$. The component from the parameter domain ($\widehat{e} \in \widehat{\mathsf{E}}$) represents the abstraction of the shape of elements, and a non-negative interval component $[l; u] \in \widehat{\mathsf{Interval}^+}$ is used to abstract over the cardinality (so $l, u \in \mathbb{Z}^+$ and $l \leq u$). The abstract set element acts as a reduced product between $\widehat{e}$ and $[l; u]$ and the lattice operations follow directly.

Given a concretization function for the abstract content domain $\gamma_{\widehat{\mathsf{E}}} \in \widehat{\mathsf{E}} \to \wp(\mathsf{E})$, we can define a concretization function for the abstract set shape domain to possible finite sets of concrete elements $\gamma_{\widehat{\mathsf{SS}}} \in \widehat{\mathsf{SetShape}}(\widehat{\mathsf{E}}) \to \wp(\mathsf{Set}(\mathsf{E}))$:

$$\gamma_{\widehat{\mathsf{SS}}}(\{\widehat{e}\}_{[l;u]}) = \left\{ es \;\middle|\; es \subseteq \gamma_{\widehat{\mathsf{E}}}(\widehat{e}) \wedge |es| \in \gamma_{\widehat{\mathsf{I}}}([l; u]) \right\}$$

Example 4.1. Let $\widehat{\mathsf{Interval}}$ be a domain of intervals of integers (a standard abstraction over integers). We can concretize abstract elements from $\widehat{\mathsf{SetShape}}(\widehat{\mathsf{Interval}})$ to a set of possible sets of integers from $\wp(\mathsf{Set}(\mathbb{Z}))$ as follows:

$$\gamma_{\widehat{\mathsf{SS}}}(\{[42; 43]\}_{[1;2]}) = \{\{42\}, \{43\}, \{42, 43\}\}$$

Data Shape Domain. Inductive refinement types are defined as a generalization of refinement types [26, 47, 61] that inductively constrain the possible constructors and the content in a data structure. We use a parameterized abstraction of data types $\widehat{\mathsf{DataShape}}(\widehat{\mathsf{E}})$, whose parameter $\widehat{\mathsf{E}}$ abstracts over the shape of constructor arguments:

$$\widehat{d} \in \widehat{\mathsf{DataShape}}(\widehat{\mathsf{E}}) = \{\bot_{\widehat{\mathsf{DS}}}\} \cup \{k_1(\underline{e_1}) \wr \ldots \wr k_n(\underline{e_n}) \mid e_i \in \widehat{\mathsf{E}}\} \cup \{\top_{\widehat{\mathsf{DS}}}\}$$

We have the least element $\bot_{\widehat{\mathsf{DS}}}$ and top element $\top_{\widehat{\mathsf{DS}}}$ elements—respectively representing no data types value and all data type values—and otherwise a non-empty choice between unique (all different) constructors of the same algebraic data type $k_1(\underline{e_1}) \wr \cdots \wr k_\mathsf{n}(\underline{e_\mathsf{n}})$ (shortened $\underline{k}(\underline{e})$). We can treat the constructor choice as a finite map $[k_1 \mapsto \underline{e_1}, \ldots, k_\mathsf{n} \mapsto \underline{e_\mathsf{n}}]$, and then directly define our lattice operations point-wise.

Given a concretization function for the concrete content domain $\gamma_{\widehat{\mathsf{E}}} \in \widehat{\mathsf{E}} \to \wp(\mathsf{E})$, we can create a concretization function for the data shape domain

$$\gamma_{\widehat{\mathsf{DS}}} \in \widehat{\mathsf{DataShape}}(\widehat{\mathsf{E}}) \to \wp(\mathsf{Data}(\mathsf{E}))$$

where $\mathsf{Data}(\mathsf{E}) = \left\{k(\underline{v}) \;\middle|\; \exists \text{ a type } at.\ k(\underline{v}) \in [\![at]\!] \wedge \underline{v \in \mathsf{E}}\right\}$. The concretization is defined as follows:

$$\gamma_{\widehat{\mathsf{DS}}}(\bot_{\widehat{\mathsf{DS}}}) = \emptyset \qquad \gamma_{\widehat{\mathsf{DS}}}(\top_{\widehat{\mathsf{DS}}}) = \mathsf{Data}(\mathsf{E})$$

$$\gamma_{\widehat{\mathsf{DS}}}(k_1(\underline{e_1}) \wr \cdots \wr k_\mathsf{n}(\underline{e_\mathsf{n}})) = \left\{k_i(\underline{v}) \;\middle|\; i \in [1,\mathsf{n}] \wedge \underline{v \in \gamma_{\widehat{\mathsf{E}}}(e_i)}\right\}$$

Example 4.2. We can concretize abstract data elements $\widehat{\mathsf{DataShape}}(\widehat{\mathsf{Interval}})$ to a set of possible concrete data values $\wp(\mathsf{Data}(\mathbb{Z}))$. Consider values from the algebraic data type:

$$\mathsf{data\ errorloc} = \mathsf{repl}() \mid \mathsf{linecol}(\mathsf{int}, \mathsf{int})$$

We can concretize abstracting elements as follows:

$$\gamma_{\widehat{\mathsf{DS}}}(\mathsf{repl}() \wr \mathsf{linecol}([1;1],[3;4])) = \{\mathsf{repl}(), \mathsf{linecol}(1,3), \mathsf{linecol}(1,4)\}$$

Recursive shapes. We extend our abstract domains to cover recursive structures such as lists and trees. Given a type expression $\mathsf{F}(X)$ with a variable X, we construct the abstract domain as the solution to the recursive equation $X = \mathsf{F}(X)$ [50, 54, 59], obtained by iterating the induced map F over the singleton domain $\mathbb{1} = \{\bot\}$. The concretization function of the recursive domain follows from the concretization function of the underlying functor domain. A detailed exposition of how to solve recursive equations over abstract domains is presented in Al-Sibahi et al. [4].

Example 4.3. We can concretize abstract elements of the refinement type from our running example:

$$\gamma_{\widehat{\mathsf{DS}}}(\mathsf{Expr}^e) = \left\{ \overbrace{\mathsf{cst}(\mathsf{suc}(\mathsf{suc}(\mathsf{zero})))}^{2}, \mathsf{mult}(2,2), \mathsf{mult}(\mathsf{mult}(2,2),2), \ldots \right\}$$

where $\mathsf{Expr}^e = \mathsf{cst}(\mathsf{suc}(\mathsf{suc}(\mathsf{zero}))) \wr \mathsf{mult}(\mathsf{Expr}^e, \mathsf{Expr}^e)$ In particular, our abstract element represents the set of all multiplications of the constant 2.

Value Domains. We presented the required components for abstracting individual types, and now all that is left is putting everything together. We construct our value shape domain using choice and recursive domain equations:

$$\widehat{\mathsf{ValueShape}} = \widehat{\mathsf{SetShape}}(\widehat{\mathsf{ValueShape}}) \oplus \widehat{\mathsf{DataShape}}(\widehat{\mathsf{ValueShape}})$$

Similarly, we have the corresponding concrete shape domain:

$$\mathsf{Value} = \mathsf{Set}\,(\mathsf{Value}) \uplus \mathsf{Data}(\mathsf{Value})$$

We then have a concretization function $\gamma_{\widehat{\mathsf{VS}}} \in \widehat{\mathsf{ValueShape}} \to \wp(\mathsf{Value})$, which follows directly from the previously defined concretization functions.

4.1 Abstract State Domains

We now explain how to construct abstractions of states and results when executing Rascal programs.

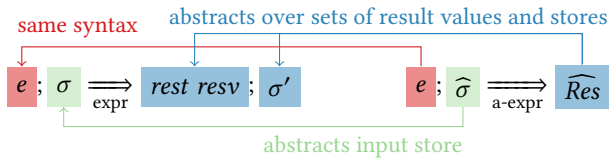

Fig. 6. Relating concrete semantics (left) to abstract semantics (right).

Abstract Store Domain. Tracking assignments of variables is important since matching variable patterns depends on the value being assigned in the store:

$$\widehat{\sigma} \in \widehat{\mathsf{Store}} = \mathsf{Var} \to \{\mathsf{ff}, \mathsf{tt}\} \times \widehat{\mathsf{ValueShape}}$$

For a variable x we get $\widehat{\sigma}(x) = (b, \widehat{vs})$ where b is true if x might be unassigned, and false otherwise (when x is definitely assigned). The second component, $\widehat{vs}$ is a shape approximating a possible value of x.

We lift the orderings and lattice operations point-wise from the value shape domain to abstract stores. We define the concretization function $\gamma_{\widehat{\mathsf{Store}}} \in \widehat{\mathsf{Store}} \to \wp(\mathsf{Store})$ as:

$$\gamma_{\widehat{\mathsf{Store}}}(\widehat{\sigma}) = \left\{ \sigma \;\middle|\; \begin{array}{c} \forall x, b, \widehat{vs}.\ \widehat{\sigma}(x) = (b, \widehat{vs}) \Rightarrow \\ \left((\neg b \Rightarrow x \in \mathbf{dom}\ \sigma) \wedge (x \in \mathbf{dom}\ \sigma \Rightarrow \sigma(x) \in \gamma_{\widehat{\mathsf{V}}}(\widehat{vs}))\right) \end{array} \right\}$$

Abstract Result Domain. Traditionally, abstract control flow is handled using a collecting denotational semantics with continuations, or by explicitly constructing a control flow graph. These methods are non-trivial to apply for a rich language like Rascal, especially considering backtracking, exceptions and data-dependent control flow introduced by visitors. A nice side-effect of Schmidt-style abstract interpretation is that it allows abstracting control flow directly.

We model different type of results—successes, pattern match failures, errors directly in a $\widehat{\mathsf{ResSet}}$ domain which keeps track of possible results with each its own separate store. Keeping separate stores is important to maintain precision around different paths:

$$\mathit{rest} \in \widehat{\mathsf{ResType}} ::= \mathsf{success} \mid \mathit{exres} \quad \mathit{exres} ::= \mathsf{fail} \mid \mathsf{error} \quad \widehat{resv} \in \widehat{\mathsf{ResVal}} ::= \cdot \mid \widehat{vs}$$

$$\widehat{Res} \in \widehat{\mathsf{ResSet}} = \widehat{\mathsf{ResType}} \rightharpoonup \widehat{\mathsf{ResVal}} \times \widehat{\mathsf{Store}}$$

The lattice operations are lifted directly from the target value domains and store domains. We define the concretization function $\gamma_{\widehat{\mathsf{RS}}} \in \widehat{\mathsf{ResultSet}} \to \wp(\mathsf{Result} \times \mathsf{Store})$:

$$\gamma_{\widehat{\mathsf{RS}}}(\widehat{Res}) = \left\{ (\mathit{rest}\ \mathit{resv}, \sigma) \;\middle|\; (\mathit{rest}, (\widehat{resv}, \widehat{\sigma})) \in \widehat{Res} \wedge \mathit{resv} \in \gamma_{\widehat{\mathsf{RV}}}(\widehat{resv}) \wedge \sigma \in \gamma_{\widehat{\mathsf{Store}}}(\widehat{\sigma}) \right\}$$

5 ABSTRACT SEMANTICS

A distinguishing feature of Schmidt-style abstract interpretation is that the derivation of abstract operational rules from a given concrete operational semantics is systematic and to a large extent mechanisable [11, 49]. The creative work is largely (1) providing abstract definitions for conditions, (2) abstract semantic operations like pattern matching, and (3) defining *trace memoization strategies* for non-structurally recursive operational rules.

Figure 6 relates the concrete evaluation judgment (left) to the abstract evaluation judgment (right) for Rascal expressions. Both judgements evaluate the same expression e. The abstract evaluation judgment abstracts the initial concrete store σ with an abstract store $\widehat{\sigma}$. The result of the abstract evaluation is a finite result set $\widehat{Res}$, abstracting over possibly infinitely many concrete result values

rest $\widehat{resv}$ and stores σ'. $\widehat{Res}$ maps each result type *rest* to a pair of abstract result value $\widehat{resv}$ and abstract result store $\widehat{\sigma}'$, i.e.:

$$\widehat{Res} = [rest_1 \mapsto (\widehat{resv_1}, \widehat{\sigma}_1), \ldots, rest_n \mapsto (\widehat{resv_n}, \widehat{\sigma}_n)]$$

There is an important difference in how the concrete and abstract semantic rules are used. In a concrete operational semantics a language construct is usually evaluated as soon as the premises of a rule are satisfied. When evaluating abstractly, we must consider *all* applicable rules, to soundly over-approximate the possible concrete executions. To this end, we introduce a special notation to collect all derivations with the same input i into a single derivation with output O equal to the join of the individual outputs:

$$\{\!|i \Rightarrow O|\!\} \triangleq O = \bigsqcup \{o | i \Rightarrow o\}$$

Let's use the operational rules for variable accesses to illustrate the steps in Schmidt-style translation of operational rules. The concrete semantics contains two rules for variable accesses, E-V-S for successful lookup, and E-V-Er for producing errors when accessing unassigned variables:

$$\text{E-V-S}\ \frac{x \in \text{dom}\ \sigma}{x;\sigma \underset{\text{expr}}{\Longrightarrow} \text{success}\ \sigma(x);\sigma} \qquad \text{E-V-Er}\ \frac{x \notin \text{dom}\ \sigma}{x;\sigma \underset{\text{expr}}{\Longrightarrow} \text{error};\sigma}$$

We follow three steps, to translate the concrete rules to abstract operational rules:

(1) For each concrete rule, create an abstract rule that uses a judgment for evaluation of a syntactic form, e.g., AE-V-S and AE-V-Er for variables.
(2) Replace the concrete conditions and semantic operations with the equivalent abstract conditions and semantic operations for target abstract values, e.g. $x \in \text{dom}\ \sigma$ with $\widehat{\sigma}(x) = (b, \widehat{vs})$ and a check on b. We obtain two execution rules:

$$\text{AE-V-S}\ \frac{\widehat{\sigma}(x) = (b, \widehat{vs})}{x;\widehat{\sigma} \underset{\text{a-expr-v}}{\Longrightarrow} [\text{success} \mapsto (\widehat{vs}, \widehat{\sigma})]} \qquad \text{AE-V-ER}\ \frac{\widehat{\sigma}(x) = (\text{tt}, \widehat{vs})}{x;\widehat{\sigma} \underset{\text{a-expr-v}}{\Longrightarrow} [\text{error} \mapsto (\cdot, \widehat{\sigma})]}$$

Observe when b is true, both a success and failure may occur, and we need rules to cover both cases.

(3) Create a rule that collects all possible evaluations of the syntax-specific judgment rules, e.g. AE-V for variables:

$$\text{AE-V}\ \frac{\{\!|x;\widehat{\sigma} \underset{\text{a-expr-v}}{\Longrightarrow} \widehat{Res}'|\!\}}{x;\widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow} \widehat{Res}'}$$

The possible shapes of the result value depend on the pair assigned to x in the abstract store. If the value shape of x is $\bot$, we drop the success result from the result set. The following examples illustrate the possible outcome result shapes:

Assigned Value	**Result Set**	**Rules**
$\widehat{\sigma}(x) = (\text{ff}, \bot_{\widehat{VS}})$	[]	AE-V-S
$\widehat{\sigma}(x) = (\text{ff}, [1;3])$	$[\text{success} \mapsto ([1;3], \widehat{\sigma})]$	AE-V-S
$\widehat{\sigma}(x) = (\text{tt}, \bot_{\widehat{VS}})$	$[\text{error} \mapsto (\cdot, \widehat{\sigma})]$	AE-V-S, AE-V-Er
$\widehat{\sigma}(x) = (\text{tt}, [1;3])$	$[\text{success} \mapsto ([1;3], \widehat{\sigma}), \text{error} \mapsto (\cdot, \widehat{\sigma})]$	AE-V-S, AE-V-Er

It is possible to translate the operational semantics rules for other basic expressions using the presented steps (see Appendix B). The core changes are the ones moving from checks of definiteness to checks of *possibility*. For example:

- Checking that evaluation of e has succeeded, requires that the abstract semantics uses $e; \widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow} \widehat{Res}$ and $(\text{success}, (\widehat{vs}, \widehat{\sigma}')) \in \widehat{Res}$, as compared to $e; \sigma \underset{\text{expr}}{\Longrightarrow} \text{success } v; \sigma'$ in the concrete semantics.
- Typing is now done using abstract judgments $\widehat{vs} \,\widehat{:}\, t$ and $t \mathrel{\widehat{<:}} t'$. In particular, type t is an abstract subtype of type t' ($t \mathrel{\widehat{<:}} t'$) if there is a subtype t'' of t ($t'' <: t$) that is also a subtype of t' ($t'' <: t'$). This implies that $t \mathrel{\widehat{<:}} t'$ and $t \mathrel{\widehat{\not<:}} t'$ are non-exclusive.
- To check whether a result contains a particular constructor, we use the abstract function $\widehat{\text{unfold}}(\widehat{vs}, t)$. $\widehat{\text{unfold}}(\widehat{vs}, t)$ will produces a refined set of abstract values of type t based on $\widehat{vs}$, splitting alternative constructors. Additionally, it produces error if the value is possibly not an element of t.

6 PATTERN MATCHING

Expressive pattern matching is a key feature of Rascal, and Rabit handles the full pattern language of Rascal. For brevity, we discuss a subset, including variables x, constructor patterns $k(\underline{p})$, and set patterns $\{\underline{\star p}\}$:

$$p ::= x \mid k(\underline{p}) \mid \{\underline{\star p}\} \qquad \star p ::= p \mid \star x$$

Rascal allows non-linear matching where the same variable x can be mentioned more than once: all values matched against x must have equal values for the match to succeed. Each set pattern contains a sequence of sub-patterns $\star p$; each sub-pattern in the sequence is either an ordinary pattern p matched against a single set element, or a star pattern $\star x$ to be matched against a subset of elements. Star patterns can backtrack when pattern matching fails because of non-linear variables, or when explicitly triggered by the fail expression.

This expressiveness poses challenges for developing an abstract interpreter that is both sound and sufficiently *precise*. The key aspects of Rabit in handling pattern matching is how we maintain precision by *refining* input values on pattern matching successes and failures.

6.1 Satisfiability Semantics for Patterns

We begin by defining what it means that a (concrete/abstract) value matches a pattern. Figure 7a shows the concrete semantics for patterns. In the figure, ρ is a binding environment:

$$\rho \in \mathsf{BindingEnv} = \mathsf{Var} \rightharpoonup \mathsf{Value}$$

A value v matches a pattern p ($v \models p$) iff v is accepted by the satisfiability semantics (defined in Fig. 7a) give a binding environment ρ that maps the variables in the pattern to values in $\mathbf{dom}\, \rho = \mathsf{vars}(p)$.

Constructor patterns $k(\underline{p})$ accept any well-typed value $k(\underline{v})$ of the same constructor, and whose subcomponents $\underline{v}$ match the sub-patterns $\underline{p}$ in a consistent fashion using the same binding environment ρ. A variable x matches exactly the value it is bound to in the binding environment ρ. A set pattern $\{\underline{\star p}\}$ accepts any set of values $\{\underline{v}\}$ such that an associative-commutative arrangement of the sub-values $\underline{v}$ matches the sequence of sub-patterns $\underline{\star p}$ under ρ.

A value sequence $\underline{v}$ matches a pattern sequence $\underline{\star p}$ ($\underline{v} \models^{\star} \underline{\star p}$) if there exists a binding environment ρ such that $\mathbf{dom}\ \rho = \mathsf{vars}(\underline{\star p})$ and $\underline{v} \models^{\star}_{\rho} \underline{\star p}$. An empty sequence of patterns ε accepts an empty sequence of values ε. A sequence starting $p, \underline{\star p'}$ with an ordinary pattern p matches any non-empty sequence of values $v, \underline{v'}$ where v matches p and $\underline{v'}$ matches $\underline{\star p'}$ in a consistent fashion using the

$k(\underline{v}) \models_\rho k(\underline{p})$	**iff**	$\underline{t}$ are parameter types of k **and** $\underline{v : t'}$ **and** $\underline{t' <: t}$ **and** $\underline{v} \models^\star_\rho \underline{p}$
$v \models_\rho x$	**iff**	$\rho(x) = v$
$\{\underline{v}\} \models_\rho \{\underline{\star p}\}$	**iff**	$\underline{v} \models^\star_\rho \underline{\star p}$
$\varepsilon \models^\star_\rho \varepsilon$	**always**	
$v, \underline{v'} \models^\star_\rho p, \underline{\star p'}$	**iff**	$v \models_\rho p$ **and** $\underline{v'} \models^\star_\rho \underline{\star p'}$
$\underline{v}, \underline{v'} \models^\star_\rho \star x, \underline{\star p'}$	**iff**	$\rho(x) = \{\underline{v}\}$ **and** $\underline{v'} \models^\star_\rho \underline{\star p'}$

(a) Concrete ($v \models_\rho p$ reads: v matches p with ρ)

$k(\underline{\widehat{vs}}) \mathrel{\widehat{\models}}_{\widehat{\rho}} k(\underline{p})$	**iff**	$\underline{t}$ are parameter types of k **and** $\underline{\widehat{vs} : t'}$ **and** $\underline{t' \mathrel{\widehat{<:}} t}$ **and** $\underline{\widehat{vs}} \mathrel{\widehat{\models}}^\star_{\widehat{\rho}} \underline{p}$
$\widehat{vs} \mathrel{\widehat{\models}}_{\widehat{\rho}} x$	**iff**	$\widehat{\rho}(x) \sqsubseteq \widehat{vs}$
$\{\underline{\widehat{vs}}\}_{[l;u]} \mathrel{\widehat{\models}}_{\widehat{\rho}} \{\underline{\star p}\}$	**iff**	$\widehat{vs}, [l;u] \mathrel{\widehat{\models}}^\star_{\widehat{\rho}} \underline{\star p}$
$\widehat{vs}, [0;u] \mathrel{\widehat{\models}}^\star_{\widehat{\rho}} \varepsilon$	**always**	
$\widehat{vs}, [l;u] \mathrel{\widehat{\models}}^\star_{\widehat{\rho}} p, \underline{\star p'}$	**iff**	$u > 0$ **and** $\widehat{vs} \mathrel{\widehat{\models}}_{\widehat{\rho}} p$ **and** $\widehat{vs}, [l-1; u-1] \mathrel{\widehat{\models}}^\star_{\widehat{\rho}} p, \underline{\star p'}$
$\widehat{vs}, [l;u] \mathrel{\widehat{\models}}^\star_{\widehat{\rho}} \star x, \underline{\star p'}$	**iff**	$\widehat{\rho}(x) = \{\widehat{vs}'\}_{[l';u']}$ **and** $l' \le l$ **and** $u' \le u$ **and** $\widehat{vs}' \sqsubseteq \widehat{vs}$ **and** $\widehat{vs}, [l-u'; u-l'] \mathrel{\widehat{\models}}^\star_{\widehat{\rho}} \underline{\star p'}$

(b) Abstract ($\widehat{vs} \mathrel{\widehat{\models}}_{\widehat{\rho}} \widehat{p}$ reads: $\widehat{vs}$ may match $\widehat{p}$ with $\widehat{\rho}$)

Fig. 7. Satisfiability semantics for pattern matching

same binding environment ρ. A sequence $\star x, \underline{\star p'}$ works analogously but it splits the value sequence in two $\underline{v}$ and $\underline{v'}$, such that x is assigned to $\underline{v}$ in ρ and $\underline{v'}$ matches $\underline{\star p'}$ consistently in ρ.

Example 6.1. We revisit the running example to understand how the data type values are matched. We consider matching the following set of expression values:

$$\overbrace{\{\mathsf{mult}\,(\mathsf{cst}\,(\mathsf{zero})\,,\mathsf{cst}\,(\mathsf{suc}\,(\mathsf{zero})))\,,\mathsf{cst}\,(\mathsf{zero})\}}^{\underline{v}}$$

against the pattern $p = \{\mathsf{mult}\,(x,y)\,,\star w, x\}$ in the environment $\rho = [x \mapsto \mathsf{cst}\,(\mathsf{zero})\,, y \mapsto \mathsf{cst}\,(\mathsf{suc}\,(\mathsf{zero}))\,, w \mapsto \{\}]$. The matching argument is as follows:

$\{\underline{v}\} \models_\rho p$ **iff** $\underline{v} \models^\star_\rho \mathsf{mult}\,(x,y)\,,\star w, x$
iff $\mathsf{mult}\,(\mathsf{cst}\,(\mathsf{zero})\,,\mathsf{cst}\,(\mathsf{suc}\,(\mathsf{zero}))) \models_\rho \mathsf{mult}\,(x,y)$
and $\mathsf{cst}\,(\mathsf{zero}) \models^\star_\rho \star w, x$

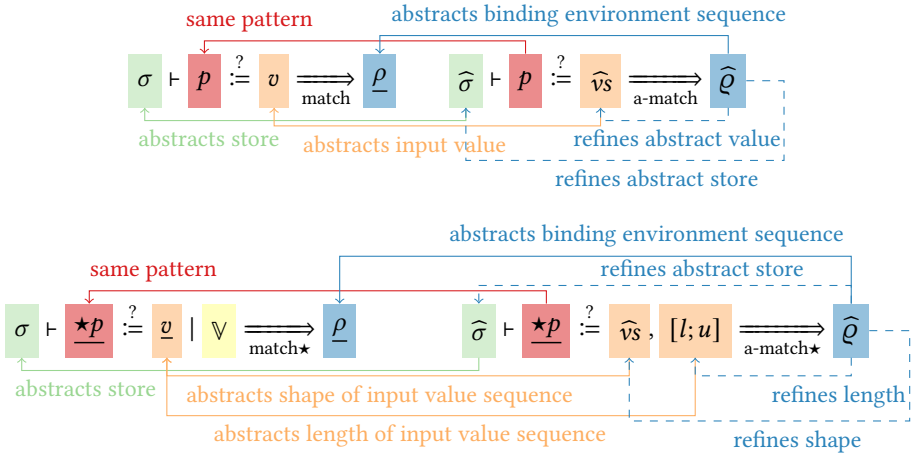

Fig. 8. Relating abstract operational semantics (left) to the concrete operational semantics (right).

We see that the first conjunct matches as follows:

$$\begin{aligned}&\mathsf{mult}\,(\mathsf{cst}\,(\mathsf{zero})\,,\mathsf{cst}\,(\mathsf{suc}\,(\mathsf{zero}))) \models_\rho \mathsf{mult}\,(x,y)\\ &\textbf{iff } \mathsf{cst}\,(\mathsf{zero})\,,\mathsf{cst}\,(\mathsf{suc}\,(\mathsf{zero})) \models^\star_\rho x,y\\ &\textbf{iff } \rho(x) = \mathsf{cst}\,(\mathsf{zero}) \;\; \textbf{and} \;\; \rho(y) = \mathsf{cst}\,(\mathsf{suc}\,(\mathsf{zero}))\end{aligned}$$

Similarly, the second matches as follows:

$$\mathsf{cst}\,(\mathsf{zero}) \models^\star_\rho \star w, x \;\textbf{iff}\; \rho(w) = \{\} \;\textbf{and}\; \rho(x) = \mathsf{cst}\,(\mathsf{zero})$$

The abstract pattern matching semantics (Fig. 7b) is analogous, but with a few noticeable differences. First, an abstract value $\widehat{vs}$ matches a pattern p ($\widehat{vs} \mathrel{\widehat{\models}} p$) if there exists a more precise value $\widehat{vs}'$ (so $\widehat{vs}' \sqsubseteq \widehat{vs}$) and an abstract binding environment $\widehat{\rho}$ with $\textbf{dom}\ \widehat{\rho} = \mathsf{vars}(p)$ so that $\widehat{vs}' \mathrel{\widehat{\models}}_{\widehat{\rho}} p$. The reason for using a more precise shape is the potential loss of information during over-approximation—a more precise value might have matched the pattern, even if the relaxed value does not necessarily. Second, sequences are abstracted by shape–lengths pairs, which needs to be taken into account by sequence matching rules. This is most visible in the very last rule, with a star pattern $\star x$, where we accept any assignment to a set abstraction $\widehat{vs}$ which has a more precise shape and a smaller length.

6.2 Computing Pattern Matches

Albeit clean, the declarative satisfiability semantics of patterns is not directly computable. In Rabit, we rely on an abstract operational semantics based on the concrete operational pattern matching semantics [2]. This semantics is translated using the technique presented in Sect. 5.

The operational judgements are presented in Fig. 8 for both the concrete and abstract rules. Consider the concrete (top-left) judgement: a value v matches a pattern p, given a store σ, producing a sequence of binding environments $\overline{\rho}$. The binding environments form a sequence, since multiple concrete environments, say ρ_1 and ρ_2, can make v match against p, i.e., $v \models_{\rho_1} p$ and $v \models_{\rho_2} p$. Backtracking using the fail-expression, allows the programmer to explore a different assignment from the sequence of environments, until no possible assignment is left. A concrete discussion of the operational rules is available in appendix A. The interesting ideas are mainly in the refining semantic operators, which we now discuss.

Semantic Operators with Refinement. Non-linear matching in Rascal requires values assigned to the same variable to be equal. We must therefore merge environments computed when matching sub-patterns to check whether a match succeeds or not. In abstract interpretation, we can refine the abstract environments when merging for each possibility. Consider when merging two abstract environments, where some variable x is assigned to $\widehat{vs}$ in one, and $\widehat{vs}'$ in the other. If $\widehat{vs}'$ is possibly equal to $\widehat{vs}$, we refine both values using this equality assumption $\widehat{vs} \mathrel{\widehat{=}} \widehat{vs}'$. Here, we have that abstract equality is defined as the greatest lower bound if the value is non-bottom, i.e. $\widehat{vs} \mathrel{\widehat{=}} \widehat{vs}' \triangleq \{\widehat{vs}'' | \widehat{vs}'' = \widehat{vs} \sqcap \widehat{vs}' \neq \bot\}$. Similarly, we can also refine both values if they are possibly non-equal $\widehat{vs} \mathrel{\widehat{\neq}} \widehat{vs}'$. Here, abstract inequality is defined using relative complements:

$$\widehat{vs} \mathrel{\widehat{\neq}} \widehat{vs}' \triangleq \left\{(\widehat{vs}'', \widehat{vs}') | \widehat{vs}'' = \widehat{vs} \setminus (\widehat{vs} \sqcap \widehat{vs}') \neq \bot\right\} \cup \left\{(\widehat{vs}, \widehat{vs}'') | \widehat{vs}'' = \widehat{vs}' \setminus (\widehat{vs} \sqcap \widehat{vs}') \neq \bot\right\}$$

In our abstract domains, the relative complement ($\setminus$) is limited[2]. We heuristically define it for interesting cases, and otherwise it degrades to identity in the first argument (no refinement). There are however useful cases, e.g., for excluding unary constructors $\mathsf{suc}\,(\mathsf{Nat}) \wr \mathsf{zero} \setminus \mathsf{zero} = \mathsf{suc}\,(\mathsf{Nat})$ or at the end points of a lattice $[1;10] \setminus [1;2] = [3;10]$.

Similarly, for matching against a constructor pattern $k(\underline{p})$, the core idea is that we should be able to partition our value space into two: the abstract values that match the constructor and those that do not. For those values that possibly match $k(\underline{p})$, we produce a refined value with k as the only choice, making sure that the sub-values in the result are refined by the sub-patterns $\underline{p}$.

Otherwise, we exclude k from the refined value. For a data type abstraction exclusion removes the pattern constructor from the possible choices

$$\widehat{\mathsf{exclude}}(k(\widehat{vs}) \wr k_1(\widehat{vs}_1) \wr \cdots \wr k_n(\widehat{vs}_n), k) = k_1(\widehat{vs}_1) \wr \ldots \wr k_n(\widehat{vs}_n)$$

and does not change the input shape otherwise.

7 TRAVERSALS

First-class traversals are a key feature of high-level transformation languages, since they enable effectively transforming large abstract syntax trees. We will focus on the challenges for bottom-up traversals, but they are shared amongst all strategies supported in Rascal. The core idea of a bottom-up traversal of an abstract value $\widehat{vs}$, is to first traverse children of the value $\widehat{\mathsf{children}}(\widehat{vs})$ possibly rewriting them, then reconstruct a new value using the rewritten children and finally traversing the reconstructed value. The main challenge is handling traversal of children, whose representation and thus execution rules depend on the particular abstract value.

Concretely, the $\widehat{\mathsf{children}}(\widehat{vs})$ function returns a set of pairs $(\widehat{vs}', \widehat{cvs})$ where the first component $\widehat{vs}'$ is a refinement of $\widehat{vs}$ that matches the shape of children $\widehat{cvs}$ in the second component. For data type values the representation of children is a heterogeneous sequence of abstract values $\underline{\widehat{vs}''}$, while for set values the representation of children is a pair $(\widehat{vs}'', [l;u])$ with the first component representing the shape of elements and the second representing their count. For example,

$$\widehat{\mathsf{children}}(\mathsf{mult}\,(\mathsf{Expr}, \mathsf{Expr}) \wr \mathsf{cst}\,(\mathsf{suc}\,(\mathsf{Nat}))) = \left\{\begin{pmatrix}(\mathsf{mult}\,(\mathsf{Expr}, \mathsf{Expr})\,, (\mathsf{Expr}, \mathsf{Expr})),\\ (\mathsf{cst}\,(\mathsf{suc}\,(\mathsf{Nat}))\,, \mathsf{suc}\,(\mathsf{Nat}))\end{pmatrix}\right\}$$

and $\widehat{\mathsf{children}}(\{\mathsf{Expr}\}_{[1;10]}) = \{(\{\mathsf{Expr}\}_{[1;10]}, (\mathsf{Expr}, [1;10]))\}$. Note how the $\widehat{\mathsf{children}}$ function maintains precision by partitioning the alternatives for data-types, when traversing each corresponding sequence of value shapes for the children.

[2]Note that it is *not* a relative complement of the abstract domain lattice.

Traversing Children. The shape of execution rules depend on the representation of children; this is consistent with the requirements imposed by Schmidt [49]. For heterogeneous sequences of value shapes $\underline{\widehat{vs}}$, the execution rules iterate through the sequence recursively traversing each element. Due to over-approximation we may re-traverse the same or a more precise value on recursion, and so we need to use trace memoization (Sect. 8) to terminate. For example the children of an expression Expr refer to itself:

$$\widehat{\mathsf{children}}(\mathsf{Expr}) = \{(\mathsf{mult}\,(\mathsf{Expr},\mathsf{Expr})\,,(\mathsf{Expr},\mathsf{Expr})), (\mathsf{cst}\,(\mathsf{Nat})\,,\mathsf{Nat}), (\mathsf{var}\,(\mathsf{str})\,,\mathsf{str})\}$$

Traversing children represented by a shape-length pair, is directed by the length interval $[l;u]$. If 0 is a possible value of the length interval, then traversal can finish, refining the input shape to be empty. Otherwise, we perform another traversal recursively on the shape of elements and recursively on a new shape-length pair which decreases the length, finally combining their values. Note, that if the length is unbounded, e.g. $[0;\infty]$, then the value can be decreased forever and trace memoization is also needed here for termination. This means that trace memoization must here be nested breadth-wise (when recursing on an unbounded sequence of children), in addition to depth-wise (when recursing on children); this can be computationally expensive, and we will discuss in Sect. 9 how our implementation handles that.

8 TRACE MEMOIZATION

Abstract interpretation and static program analysis in general perform fixed-point calculation for analysing unbounded loops and recursion. In Schmidt-style abstract interpretation, the main technique to handle recursion is *trace memoization* [46, 49]. The core idea of *trace memoization* is to detect non-structural re-evaluation of the same program element, i.e., when the evaluation of a program element is recursively dependent on itself, like a while-loop or traversal.

The main challenge when recursing over inputs from infinite domains, is to determine *when* to merge recursive paths together to correctly over-approximate concrete executions. We present an extension that is still terminating, sound and, additionally, allows calculating results with good precision. The core idea is to partition the infinite input domain using a finite domain of elements, and on recursion degrade input values using previously met input values from the same partition. We assume that all our domains are lattices with a widening operator. Consider a recursive operational semantics judgment $i \Longrightarrow o$, with i being an input from domain $\widehat{\mathsf{Input}}$, and o being the output from domain $\widehat{\mathsf{Output}}$. For this judgment, we associate a memoization map $\widehat{M} \in \widehat{\mathsf{PInput}} \to \widehat{\mathsf{Input}} \times \widehat{\mathsf{Output}}$ where $\widehat{\mathsf{PInput}}$ is a finite partitioning domain that has a Galois connection with our actual input, i.e. $\widehat{\mathsf{Input}} \underset{\alpha_{\widehat{PI}}}{\overset{\gamma_{\widehat{PI}}}{\leftrightarrows}} \widehat{\mathsf{PInput}}$. The memoization map keeps track of the previously seen input and corresponding output for values in the partition domain. For example, for input from our value domain $\widehat{\mathsf{Value}}$ we can use the corresponding type from the domain Type as input to the memoization map.[3] So for values 1 and $[2;3]$ we would use int, while for mult(Expr, Expr) we would use the defining data type Expr.

We perform a fixed-point calculation over the evaluation of input i. Initially, the memoization map $\widehat{M}$ is $\lambda pi.(\bot,\bot)$, and during evaluation we check whether there was already a value from the same partition as i, i.e., $\alpha_{\widehat{PI}}(i) \in \mathrm{dom}\ \widehat{M}$. At each iteration, there are then two possibilities:

Hit The corresponding input partition key is in the memoization map and a less precise input is stored, so $\widehat{M}(\alpha_{\widehat{PI}}(i)) = (i', o')$ where $i \sqsubseteq_{\widehat{\mathsf{Input}}} i'$. Here, the output value o' that is stored in the memoization map is returned as result.

[3] Provided that we bound the depth of type parameters of collections.

Widen The corresponding input partition key is in the memoization map, but an unrelated or more precise input is stored, i.e., $\widehat{M}(\alpha_{\widehat{PI}}(i)) = (i'', o'')$ where $i \not\sqsubseteq_{\widehat{\mathsf{Input}}} i''$. In this case we continue evaluation but with a widened input $i' = i''\nabla_{\widehat{\mathsf{Input}}}(i'' \sqcup i)$ and an updated map $\widehat{M}' = [\alpha_{\widehat{PI}}(i) \mapsto (i', o_{\text{prev}})]$. Here, o_{prev} is the output of the last iteration for the fixed-point calculation for input i'. On the initial iteration, o_{prev} is assigned $\bot$.

Intuitively, the technique is terminating because the partitioning is finite, and widening ensures that we reach an upper bound of possible inputs in a finite number of steps, eventually getting a hit. The fixed-point iteration also uses widening to calculate an upper bound, which similarly finishes in a number of steps. The technique is sound because we only use output for previous input that is less precise; therefore our function is continuous and a fixed-point exists.

9 EXPERIMENTAL EVALUATION

We demonstrate the ability of Rabit to verify inductive shape properties, using nine transformation programs, where five are classical and four are extracted from open source projects.

Negation Normal Form. (**NNF**) transformation [30, Section 2.5] is a classical rewrite of a propositional formula to combination of conjunctions and disjunctions of literals, so negations appear only next to atoms. An implementation of this transformation should guarantee the following:

P1 Implication is not used as a connective in the result
P2 All negations in the result are in front of atoms

Rename Struct Field. (**RSF**) refactoring [25] changes the name of a field in a struct, and that all corresponding field access expressions are renamed correctly as well:

P3 Structure should not define a field with the old field name
P4 No field access expression to the old field

Desugar Oberon-0. (**DSO**) transformation, translates for-loops and switch-statements to while-loops and nested if-statements, respectively. The transformation is part of the Oberon-0 [60] implementation in Rascal [8], containing all the necessary stages for compiling a structured imperative programming language.

P5 for should be correctly desugared to while
P6 switch should be correctly desugared to if
P7 No auxiliary data in output

Code Generation for Glagol. (**G2P**) a DSL for REST-like web development, translated to PHP for execution.[4] We are interested in the part of the generator that translates Glagol expressions to PHP, and the following properties:

P8 Output only simple PHP expressions for simple Glagol expression inputs
P9 No unary PHP expressions if no sign marks or negations in Glagol input

Mini Calculational Language. (**MCL**) a programming language text-book [51] implementation of a small expression language, with arithmetic and logical expressions, variables, if-expressions, and let-bindings. The implementation contains an expression simplifier (larger version of running example), type inference, an interpreter and a compiler.

P10 Simplification procedure produces an expression with no additions with 0, multiplications with 1 or 0, subtractions with 0, logical expressions with constant operands, and if-expressions with constant conditions.

[4] https://github.com/BulgariaPHP/glagol-dsl

P11 Arithmetic expressions with no variables have type int and no type errors
P12 Interpreting expressions with no integer constants and let's gives only Boolean values
P13 Compiling expressions with no if's produces no goto's and if instructions
P14 Compiling expressions with no if's produces no labels and does not change label counter

Mini Configuration Modernization. (**MCM**) a modernization transformation designed to migrate an imperative configuration system to a declarative constraint-based one. The transformation is inspired by a realistic case given by Danfoss [32], which is a company specialized in producing power electronics.

P15 Statements which consist only of if-else conditionals and return statements should not possibly violate the post-condition stating that all statements that can be modernized.
P16 Statements which consist only of if-else conditionals and return statements, should produce output expression consisting of only contain relevant input expressions and the ternary expression $(e_1?e_2 : e_3)$.

Extract Superclass. (**ESC**) refactoring [25] takes two classes and extract a new superclass for them that contains the field definitions they have in common.

P17 The new superclass has the common fields of the two subclasses.
P18 The new superclass has no superclass itself.
P19 The two subclasses do not have the fields of the new superclass in their definition.
P20 The two subclasses have the new superclass as their superclass.

Derivative. (**DER**) computes the derivative[5] of an additive-multiplicative expression w.r.t. a particular variable x.

P21 Additive expressions containing variables $y \neq x$ and constants should produce a constant 0 when differentiated.
P22 Expressions containing only linear multiplication should produce expressions with only constant leafs when differentiated.

Normalize PHP script. (**NPS**) transforms PHP fragments into a standardized format.[6]

P23 The discardHTML function should remove all contained HTML values from a PHP script.
P24 The useBuiltin function should replace calls to built-in functions with the functions themselves for a PHP script.

All these transformations satisfy the following criteria:

(1) They are formulated by an independent source,
(2) They can be translated in relatively straightforward manner to our subset of Rascal, and
(3) They exercise important constructs, including visitors and the expressive pattern matching

We have ported all these programs to Rascal Light.

Experimental Configuration. To evaluate the effect of our pragmatic design choices, we run our evaluation using different configurations. We test out two types of configuration options: one regarding the effect of using trace partitioning and one regarding the choice of trace memoization domain. For the trace partitioning, we have two configurations: **N** representing no trace partitioning, and **T** representing use of trace partitioning. For the domain used for widening in trace memoization, we have three configurations: **S** representing always widening on recursion, **T** representing

[5] http://tutor.rascal-mpl.org/Recipes/Common/Derivative/Derivative.html
[6] https://github.com/cwi-swat/php-analysis. We rely on a simplified and rewritten version of the script for compatability and performance reasons.

```
1  data FIn = and(FIn, FIn) | atom(str) | neg(FIn) | imp(FIn, FIn) | or(FIn, FIn)
2
3  data FOut = and(FOut, FOut) | atom(str) | neg(atom(str)) | or(FOut, FOut)
```

Fig. 9. Initial and inferred refinement types for NNF

widening on recursion with values that have the same syntactic type, and **C** representing widening on recursion when the contained inductive refinements have the same set of top-level constructors.

Threats to Validity. The programs are not selected randomly, thus it can be hard to generalize the results. We mitigated this by selecting transformations that are realistic and vary in authors and purpose. While translating the programs to Rascal Light, we have striven to minimize the amount of changes, but bias cannot be ruled out entirely.

Implementation. We have implemented the abstract interpreter in a prototype tool, Rabit[7], for all of Rascal Light following the process described in sections 5 to 8. This required handling additional aspects, not discussed in the paper:

(1) Possibly undefined values
(2) Extended result state with more Control flow constructs, backtracking, exceptions, loop control,
(3) Functions and function calls, and
(4) Fine-tuning memoization strategies to the different looping constructs and recursive calls

By default, we use the top element $\top$ specified as input, but the user can specify the initial data-type refinements, store and parameters, to get more precise results. The output of the tool is the abstract result value set of abstractly interpreting target function, the resulting store state and the set of relevant inferred data-type refinements.

The implementation extends standard regular tree grammar operations [1, 19], to handle the recursive equations for the expressive abstract domains, including base values, collections and heterogeneous data types. We use a more precise partitioning strategy for trace memoization when needed, which also takes the set of available constructors into account for data types.

Results. We ran the experiments using Scala 2.12.2 on a 10-core Xeon W-2155 Dell Workstation. Table 1 summarizes the size of the programs, the runtime of Rabit, and whether the properties have been verified. Since we verify the results on the abstract shapes, the programs are shown to be correct for all possible concrete inputs satisfying the given properties. We remark that all programs use the high-level expressive features of Rascal and are succinct compared to general purpose code.

The runtime is reasonable, and varies from less than a seconds to around a minute in most cases. All, but three, properties were successfully verified. The reason that our tool runs slower on the **DSO** and **MCM** transformations is that they contain function calls and we rely on inlining for interprocedural analysis. The **P24** property of **NPS** is much slower than average (around 5 minute time), because the useBuiltin function has complex nested loops and the PHP data-structure is represented using large mutual inductive data types containing collections.

Line 1 in Fig. 9 show the input refinement type FIn for the normalization procedure. The inferred inductive output type FOut (line 3) specifies that the implication is not present in the output (P1), and negation only allows atoms as subformulae (P2). In fact, Rabit inferred a precise formulation of negation normal form as an inductive data type.

[7] https://github.com/ahmadsalim/Rascal-Light

Transformation	LOC	Property	Configuration											
			NS		NT		NC		TS		TT		TC	
			Time	Verif.	Time	Verif.	Time	Verif.	Time	Verif.	Time	Verif.	Time	Verif.
NNF	15	P1	1.4	✗	1.4	✗	22.2	✗	0.5	✓	0.5	✓	1.3	✓
		P2		✗		✗		✗		✗		✗		✓
RSF	35	P3	0.3	✓	3.4	✓	2.5	✓	0.6	✓	2.8	✓	3.0	✓
		P4		✗		✗		✗		✗		✓		✓
DSO	125	P5	0.7	✗	14.1	✗	15.7	✗	0.8	✗	5.2	✓	11.5	✓
		P6		✗		✗		✗		✗		✓		✓
		P7		✗		✗		✗		✗		✗		✗
G2P	350	P8	0.4	✓	0.2	✓	0.2	✓	0.1	✓	0.1	✓	0.1	✓
		P9	2.8	✓	1.8	✓	2.2	✓	1.1	✓	1.4	✓	1.8	✓
MCL	298	P10	2.7	✗	0.4	✗	0.3	✗	1.5	✗	0.8	✓	0.6	✓
		P11	0.1	✗	0.1	✗	0.1	✗	0.1	✓	0.1	✓	0.1	✓
		P12	0.1	✓	0.1	✓	0.1	✓	0.1	✓	0.1	✓	0.1	✓
		P13	0.1	✓	0.1	✓	0.1	✓	0.1	✓	0.1	✓	0.1	✓
		P14		✓		✓		✓		✓		✓		✓
MCM	128	P15	1.7	✗	14.1	✓	24.3	✓	1.5	✗	32.3	✓	49.8	✓
		P16	3.9	✓	3.8	✓	2.5	✓	1.7	✓	3.2	✓	3.3	✓
ESC	64	P17	0.2	✓	0.1	✓	0.1	✓	0.1	✓	0.1	✓	0.1	✓
		P18		✓		✓		✓		✓		✓		✓
		P19		✗		✗		✗		✗		✗		✗
		P20		✓		✓		✓		✓		✓		✓
DER	38	P21	0.1	✗	0.1	✓	0.1	✓	0.1	✗	0.1	✓	0.1	✓
		P22	0.1	✗	0.1	✗	0.1	✗	0.1	✗	0.1	✗	0.1	✗
NPS	220	P23	0.3	✗	230.0	✓	340.9	✓	0.15	✗	224.8	✓	326.8	✓
		P24	0.1	✗	8.9	✗	10.2	✗	0.1	✗	31.9	✓	33.1	✓

Table 1. Time and success rate for analyzing programs and properties presented earlier this section. The runtime is presented in seconds. Configuration: initial letter specifies without (N)/with (T) trace partitioning and second letter specifies simple (S)/type-based (T)/constructor-based (C) widening.

Transf.	Prop.	Configuration																	
		NS			NT			NC			TS			TT			TC		
		H	M	W	H	M	W	H	M	W	H	M	W	H	M	W	H	M	W
NNF	P1 P2	6	54	16	28	34	0	36	88	0	6	62	16	28	46	0	214	102	0
RSF	P3 P4	2	11	4	814	512	0	814	512	0	2	11	4	814	512	0	814	512	0
DSO	P5 P6 P7	11	34	12	5179	1408	0	8875	2032	0	11	38	12	3273	928	0	6969	1552	0
G2P	P8	8	20	0	8	20	0	8	20	0	8	20	0	8	20	0	8	20	0
	P9	28	60	0	28	60	0	28	60	0	28	60	0	28	60	0	28	60	0
MCL	P10	47	80	11	23	36	2	23	36	2	45	71	9	24	40	0	24	40	0
	P11	1	5	0	1	5	0	1	5	0	1	5	0	1	5	0	1	5	0
	P12	1	15	0	1	15	0	1	15	0	1	15	0	1	15	0	1	15	0
	P13 P14	18	44	8	18	44	8	18	44	8	18	44	8	18	44	8	18	44	8
MCM	P15	38	178	112	2344	808	0	2344	808	0	38	178	104	4084	1406	0	4084	1406	0
	P16	38	178	112	8424	2276	540	21084	6356	0	38	178	104	14780	4052	972	36840	11396	0
ESC	P17 P18 P19 P20	4	0	0	4	0	0	4	0	0	4	0	0	4	0	0	4	0	0
DER	P21	4	17	4	10	13	0	10	13	0	4	17	4	10	13	0	10	13	0
	P22	2	0	0	2	0	0	2	0	0	2	0	0	2	0	0	2	0	0
NPS	P23	1	3	4	68291	41896	0	68291	41896	0	1	3	4	68291	41896	0	68291	41896	0
	P24	1	2	2	1359	928	0	1359	928	0	1	2	2	1359	928	0	1359	928	0

Table 2. Memoization statistics for running the abstract interpreter on the properties for the different configurations. Hit (H) specifies when an existing partition key is in the memoization map; miss (M) specifies when a partition key is not found in the memoization map, and then is added to the map returning $\bot$; and finally widen (W) is when the partition key is found in the memoization map, but the memoized input was more precise and widening was needed. Configuration: initial letter specifies without (N)/with (T) trace partitioning and second letter specifies simple (S)/type-based (T)/constructor-based (C) widening.

Discussion. In general, it seems that our pragmatic choices of trace partitioning and widening partitioning domain had a considerable influence on whether we were able to verify the properties we had in our evaluation. The configuration that was able to verify most of the stated properties was **TC** which represents running our analyzer with both trace partitioning on and constructor-based widening. The configuration **TT** which used type-based widening instead, was able to verify all the same properties as **TC** except P2, which points out that one could rely on the increased performance gains given by type-based widening while still be able to prove desired properties. The simple widening **TS** algorithm seemed also to be able to prove a considerable amount of target properties, but we experienced that in practice one also get false positive errors which are due to over-approximation; performance-wise it is definitely the most efficient solution. Most false positive in the **TS** configuration were potential dynamic type errors, which happens because of eager widening. Examples include: errors when reconstructing objects during traversal since $\top$ is not a valid value of the target type for a constructor field, operator type errors with $\top$ as one of the operands, impossible default cases being triggered falsely during case-analysis and traversals since $\top$ usually abstracts over more values than the programmer intended, and transitive errors like exceptions being thrown or assignments not being made. For direct type errors, it is easy for the programmer to see these false positives since often $\top$ is included as an operand in the error, while for transitive errors this can be harder to check. Error reporting could in the future be improved for transitive errors by using techniques from program slicing [58] to provide a suitable trace to debug for the programmer.

There were 8 properties which required trace partitioning (**T**) to be verified, which could not be verified in the configurations without trace partitioning (**N**). This is especially true for many properties which have complex pattern matching and traversal require trace partitioning to infer the necessary inductive refinements.

Unlike the conference version of the paper [5], we are able to verify P3 in this version. This is because of two features: first we run the transformation with a nominalized version of the key abstraction to the class map, and then we have refined the preciseness of the map operations for deletion and updating. These refinements have also been useful in proving properties for the **ESC** transformation. Additionally, **ESC** required that we performed refinement of the store on simple Boolean conditions when evaluating `if`-statements, which we have also implemented.

Table 2 presents statistics regarding about trace memoization, that shows how many hits, misses and widening when accessing the memoization map during execution of each property. In general, there is a correlation on how many misses and hits there were during trace memoization and the runtime of validating a property. Surprisingly. the more rich trace memoization partitioning domains, type-based (**T**) and constructor-based (**C**), did not require widening on sub-recursion during trace memoization; this indicates that these partitioning domains are so precise that all subrecursions for our tested properties have been more precise. The simple widening strategy (**S**) did do widenings during trace memoization, which made the runtime much faster but much less precise. There seems no considerable difference on trace memoization between the configurations that use trace partitioning (**T**) and those without (**N**).

Limitations and Potential Improvements. Even the most general configuration **TC** could not verify properties P7, P19 and P22. The main reason is that our current abstractions were not precise enough to capture these properties. P22 requires widening at a larger depth than we had implemented in our evaluator, and we believe that increasing the widening depth could mitigate the issue (albeit at a huge performance cost). We discuss the extensions based on related work that are required to verify P7 and P19 in Sect. 10. For property P10, we use a Peano encoding of natural numbers since it was easy to handle using our inductive solver, but one could have used an integer abstract

domain with constant (in)equality constraints to handle more complex cases. This is possible to do since our framework is parametric in the choice of integer domain.

Our implementation does handle functions and function calls, but it does so by effectively inlining the call. The analysis is not call-site sensitive [52, via [35]], and we merge all recursive calls to the same function. A way to improve the scalability of our tool, is to consider constructing a modular static analysis at the function-level. Doing this for a complex abstract interpretation-based system like ours is however still an open research question.

10 RELATED WORK

We start with discussing techniques that could make Rabit verify properties like P7 and P19.

For a concrete input of P7, we know that the number of auxiliary data elements decreases on each iteration, but this information is lost in our abstraction. A possible solution could be to allow *abstract attributes* that extract additional information about the abstracted structures [12, 42, 55]. A generalization of the multiset abstraction [41], could be useful to track e.g., the auxiliary statement count, and show that they decrease using multiset-ordering [24]. Other techniques [6, 15, 57] support inferring inductive relational properties for general data-types—e.g, binary tree property—but require a pre-specified structure to indicate where refinement can happen. To verify P19, we need to have a way of knowing that when a superclass has a particular field, then it must necessarily have been removed from both subclasses. Relational abstract interpreteration [40] allows specifying constraints that relate values across different variables, even inside and across substructures [15, 29, 38].

Cousot and Cousot [20] present a general framework for modularly constructing program analyses, but it requires languages with compositional control flow. Both Sagiv, Reps, and Wilhelm [48] and Toubhans, Rival and Chang [45, 56] develop modular domain designs for pointer-manipulating programs, whereas our construction focuses on abstracting pure heterogeneous data-structures.

There are similarities between our work and verification techniques based on program transformation [23, 37]. Our systematic exploration of execution rules for abstraction is similar to *unfolding*, and widening is similar to *folding*. The main difference between the two techniques is that abstract interpretation mainly focuses on capturing rich domains and performing widening at syntactic program points, whereas program transformation based techniques often rely on symbolic inputs and perform folding dynamically on the semantic execution graph during specialization. We believe that there could benefits for the communities, to explore combinations of these two approaches in the future.

Definitional interpreters have been suggested as a technique for building compositional abstract interpreters [22]. The idea is to rely on a monad transformer stack to share the implementation of the concrete and abstract interpreters. We believe that our interpreter would benefit by being written in such style[8], which complements our modular domain construction well. They rely on a *caching* algorithm to ensure termination, similarly to ordinary finite input *trace memoization* [46]. Similarly, Van Horn and Might [31] present a systematic framework to abstract higher-order functional languages. They rely on store-allocated continuations to handle recursion, which is kept finite during abstraction to ensure a terminating analysis. We focused on providing a more precise widening based on the abstract input value, which was necessary for verifying the required properties in our evaluation. We believe that it could be useful to look into abstract machine-based abstractions in the future, when higher-order languages need to be handled.

Recently, Keidel and Erdweg [34] have developed a framework, Sturdy, that allows modularly constructing abstract interpreters for program transformations. Sturdy separates the concerns of

[8]We only learned about this related work at a late stage

interpreting a program transformation language from the choice of abstract domains. We believe writing our interpreter in this style could be beneficial for maintainability and trust in its soundness. However, there are some pragmatic challenges that need to be addressed before we could port our interpreter. In particular, trace partitioning seemed to have a big influence in our evaluation yet it is unclear how to apply this technique in their framework. Furthermore, their analyses deal with simpler domains than ours and it is unclear how we would integrate a domain like inductive refinements which require widening in their framework.

Garrigue [27, 28] presents algorithms for typing pattern matching on polymorphic variant types in OCaml, where the set of constructors for a data type is not fixed in advance. The theory is useful since it supports inferring simple recursive shapes of programs, but it has its limitations: inference is syntactic and exact, and it is unclear how to generalize it to work with the rich pattern matching constructs and heterogeneous visitors. Haskell supports analysing coverage of its pattern matching language, that includes generalized algebraic data types (GADTs) and Boolean constraints [33]. While general Haskell function calls can occur in the Boolean constraints, the analysis treats them shallowly as function symbols; some covering pattern matches that depend on particular semantics of called functions, will be marked falsely as non-exhaustive. Modern SMT solvers supports reasoning with inductive functions defined over algebraic data-types [43]. The properties they can verify are very expressive, but they do not scale to large programs like transformations. Possible constructor analysis [7] has been used to calculate the actual dependencies of a predicate and make flow-sensitive analyses more precise. This analysis works with complex data-types and arrays, but only captures the prefix of the target structures.

Techniques for model transformation verification on static analysis [21] have been suggested, but are on verification of types and undefinedness properties. Symbolic execution has previously been suggested [3] as a way to validate high-level transformation programs, but it targets test generation rather than verification of properties. Semantic typing [10, 14] has been used to infer recursive type and shape properties for language with high-level constructs for querying and iteration. However, they only consider small calculi compared to Rascal Light, and our evaluation is more extensive.

11 CONCLUSION

Our goal was to use abstract interpretation to give a solid semantic foundation for analyzing programs in modern high-level transformation languages. We designed and implemented a Schmidt-style abstract interpreter, Rabit, including *partition-driven trace memoization* that supports infinite input domains. This worked well for Rascal, and can be adapted for similar languages with complex control flow. The modular construction of abstract domains was vital for handling a language of this scale and complexity. We evaluated Rabit on classical and open source transformations, by verifying a series of sophisticated shape properties for them.

ACKNOWLEDGMENTS

We thank Paul Klint, Tijs van der Storm, Jurgen Vinju and Davy Landman for discussions on Rascal and its semantics. We thank Rasmus Møgelberg and Jan Midtgaard for discussions on correctness of the recursive shape abstractions. We thank Sebastian Erdweg for discussions on the evaluation, prompting us to investigate the influence of pragmatics. We thank the anonymous reviewers for their comments which allowed us to improve the paper. This material is based upon work supported by the Danish Council for Independent Research under Grant No. 0602-02327B and Innovation Fund Denmark under Grant No. 7039-00072B.

REFERENCES

[1] Alexander Aiken and Brian R. Murphy. 1991. Implementing Regular Tree Expressions. In *FPLCA 1991*. 427–447. https://doi.org/10.1007/3540543961_21

[2] Ahmad Salim Al-Sibahi. 2017. The Formal Semantics of Rascal Light. *CoRR* abs/1703.02312 (2017). http://arxiv.org/abs/1703.02312

[3] Ahmad Salim Al-Sibahi, Aleksandar S. Dimovski, and Andrzej Wasowski. 2016. Symbolic execution of high-level transformations. In *SLE 2016*. 207–220. http://dl.acm.org/citation.cfm?id=2997382

[4] Ahmad Salim Al-Sibahi, Thomas Jensen, Rasmus Ejlers Møgelberg, and Andrzej Wąsowski. 2020. *Galois Connections for Recursive Types*. Springer International Publishing, Cham, 105–131. https://doi.org/10.1007/978-3-030-41103-9_4

[5] Ahmad Salim Al-Sibahi, Thomas P. Jensen, Aleksandar S. Dimovski, and Andrzej Wasowski. 2018. Verification of high-level transformations with inductive refinement types. In *Proceedings of the 17th ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences, GPCE 2018, Boston, MA, USA, November 5-6, 2018*, Eric Van Wyk and Tiark Rompf (Eds.). ACM, 147–160. https://doi.org/10.1145/3278122.3278125

[6] Aws Albarghouthi, Josh Berdine, Byron Cook, and Zachary Kincaid. 2015. Spatial Interpolants. In *ESOP 2015*. 634–660. https://doi.org/10.1007/978-3-662-46669-8_26

[7] Oana Fabiana Andreescu, Thomas Jensen, and Stéphane Lescuyer. 2015. Dependency Analysis of Functional Specifications with Algebraic Data Structures. In *ICFEM 2015*. 116–133. https://doi.org/10.1007/978-3-319-25423-4_8

[8] Bas Basten, Jeroen van den Bos, Mark Hills, Paul Klint, Arnold Lankamp, Bert Lisser, Atze van der Ploeg, Tijs van der Storm, and Jurgen J. Vinju. 2015. Modular language implementation in Rascal - Experience Report. *Sci. Comput. Program.* 114 (2015), 7–19. http://dx.doi.org/10.1016/j.scico.2015.11.003

[9] Marcin Benke, Peter Dybjer, and Patrik Jansson. 2003. Universes for Generic Programs and Proofs in Dependent Type Theory. *Nordic Journal of Computing* 10, 4 (2003), 265–289.

[10] Véronique Benzaken, Giuseppe Castagna, Kim Nguyen, and Jérôme Siméon. 2013. Static and dynamic semantics of NoSQL languages. In *POPL 2013*. 101–114. https://doi.org/10.1145/2429069.2429083

[11] Martin Bodin, Thomas Jensen, and Alan Schmitt. 2015. Certified Abstract Interpretation with Pretty-Big-Step Semantics. In *CPP 2015*. 29–40. https://doi.org/10.1145/2676724.2693174

[12] Ahmed Bouajjani, Cezara Dragoi, Constantin Enea, and Mihaela Sighireanu. 2012. Abstract Domains for Automated Reasoning about List-Manipulating Programs with Infinite Data. In *VMCAI 2012*. 1–22. https://doi.org/10.1007/978-3-642-27940-9_1

[13] Martin Bravenboer, Karl Trygve Kalleberg, Rob Vermaas, and Eelco Visser. 2008. Stratego/XT 0.17. A language and toolset for program transformation. *Sci. Comput. Program.* 72, 1-2 (2008), 52–70. https://doi.org/10.1016/j.scico.2007.11.003

[14] Giuseppe Castagna and Kim Nguyen. 2008. Typed iterators for XML. In *ICFP 2008*. 15–26. https://doi.org/10.1145/1411204.1411210

[15] Bor-Yuh Evan Chang and Xavier Rival. 2008. Relational Inductive Shape Analysis. In *POPL 2008*. 247–260. https://doi.org/10.1145/1328438.1328469

[16] James Chapman, Pierre-Évariste Dagand, Conor McBride, and Peter Morris. 2010. The gentle art of levitation. In *ICFP 2010*. 3–14. https://doi.org/10.1145/1863543.1863547

[17] James R. Cordy. 2006. The TXL source transformation language. *Sci. Comput. Program.* 61, 3 (2006), 190–210. https://doi.org/10.1016/j.scico.2006.04.002

[18] Patrick Cousot. 2003. Verification by Abstract Interpretation. In *Verification: Theory and Practice, Essays Dedicated to Zohar Manna on the Occasion of His 64th Birthday*. 243–268. https://doi.org/10.1007/978-3-540-39910-0_11

[19] Patrick Cousot and Radhia Cousot. 1995. Formal Language, Grammar and Set-Constraint-Based Program Analysis by Abstract Interpretation. In *FPCA 1995*. 170–181. http://doi.acm.org/10.1145/224164.224199

[20] Patrick Cousot and Radhia Cousot. 2002. Modular Static Program Analysis. In *CC 2002*. 159–178. https://doi.org/10.1007/3-540-45937-5_13

[21] Jesús Sánchez Cuadrado, Esther Guerra, and Juan de Lara. 2017. Static Analysis of Model Transformations. *IEEE Trans. Software Eng.* 43, 9 (2017), 868–897. https://doi.org/10.1109/TSE.2016.2635137

[22] David Darais, Nicholas Labich, Phuc C. Nguyen, and David Van Horn. 2017. Abstracting definitional interpreters (functional pearl). *PACMPL* 1, ICFP (2017), 12:1–12:25. https://doi.org/10.1145/3110256

[23] Emanuele De Angelis, Fabio Fioravanti, Alberto Pettorossi, and Maurizio Proietti. 2014. Program verification via iterated specialization. *Sci. Comput. Program.* 95 (2014), 149–175. https://doi.org/10.1016/j.scico.2014.05.017

[24] Nachum Dershowitz and Zohar Manna. 1979. Proving Termination with Multiset Orderings. *Commun. ACM* 22, 8 (1979), 465–476. https://doi.org/10.1145/359138.359142

[25] Martin Fowler. 1999. *Refactoring - Improving the Design of Existing Code*. Addison-Wesley. http://martinfowler.com/books/refactoring.html

[26] Timothy S. Freeman and Frank Pfenning. 1991. Refinement Types for ML. In *PLDI 1991*. 268–277. http://doi.acm.org/10.1145/113445.113468
[27] Jacques Garrigue. 1998. Programming with polymorphic variants. In *ML Workshop*, Vol. 13.
[28] Jacques Garrigue. 2004. Typing deep pattern-matching in presence of polymorphic variants. In *JSSST Workshop on Programming and Programming Languages*.
[29] Nicolas Halbwachs and Mathias Péron. 2008. Discovering properties about arrays in simple programs. In *PLDI 2008*. 339–348. https://doi.org/10.1145/1375581.1375623
[30] John Harrison. 2009. *Handbook of Practical Logic and Automated Reasoning*. Cambridge University Press.
[31] David Van Horn and Matthew Might. 2010. Abstracting abstract machines. In *Proceeding of the 15th ACM SIGPLAN international conference on Functional programming, ICFP 2010, Baltimore, Maryland, USA, September 27-29, 2010*, Paul Hudak and Stephanie Weirich (Eds.). ACM, 51–62. https://doi.org/10.1145/1863543.1863553
[32] Alexandru Florin Iosif-Lazăr, Ahmad Salim Al-Sibahi, Aleksandar S. Dimovski, Juha Erik Savolainen, Krzysztof Sierszecki, and Andrzej Wąsowski. 2015. Experiences from Designing and Validating a Software Modernization Transformation (E). In *30th IEEE/ACM International Conference on Automated Software Engineering, ASE 2015, Lincoln, NE, USA*. 597–607. https://doi.org/10.1109/ASE.2015.84
[33] Georgios Karachalias, Tom Schrijvers, Dimitrios Vytiniotis, and Simon L. Peyton Jones. 2015. GADTs meet their match: pattern-matching warnings that account for GADTs, guards, and laziness. In *ICFP 2015*, Kathleen Fisher and John H. Reppy (Eds.). ACM, 424–436. https://doi.org/10.1145/2784731.2784748
[34] Sven Keidel and Sebastian Erdweg. 2020. A Systematic Approach to Abstract Interpretation of Program Transformations. In *Verification, Model Checking, and Abstract Interpretation - 21st International Conference, VMCAI 2020, New Orleans, LA, USA, January 16-21, 2020, Proceedings (Lecture Notes in Computer Science)*, Dirk Beyer and Damien Zufferey (Eds.), Vol. 11990. Springer, 136–157. https://doi.org/10.1007/978-3-030-39322-9_7
[35] Se-Won Kim, Xavier Rival, and Sukyoung Ryu. 2018. A Theoretical Foundation of Sensitivity in an Abstract Interpretation Framework. *ACM Trans. Program. Lang. Syst.* 40, 3 (2018), 13:1–13:44. https://doi.org/10.1145/3230624
[36] Paul Klint, Tijs van der Storm, and Jurgen Vinju. 2011. EASY Meta-programming with Rascal. In *GTTSE III*, JoãoM. Fernandes, Ralf Lämmel, Joost Visser, and João Saraiva (Eds.). 222–289. https://doi.org/10.1007/978-3-642-18023-1_6
[37] Alexei P. Lisitsa and Andrei P. Nemytykh. 2015. Finite Countermodel Based Verification for Program Transformation (A Case Study). In *Proceedings of the Third International Workshop on Verification and Program Transformation, VPT@ETAPS 2015, London, United Kingdom, 11th April 2015. (EPTCS)*, Alexei Lisitsa, Andrei P. Nemytykh, and Alberto Pettorossi (Eds.), Vol. 199. 15–32. https://doi.org/10.4204/EPTCS.199.2
[38] Jiangchao Liu and Xavier Rival. 2017. An array content static analysis based on non-contiguous partitions. *Computer Languages, Systems & Structures* 47 (2017), 104–129. https://doi.org/10.1016/j.cl.2016.01.005
[39] Neil Mitchell and Colin Runciman. 2007. Uniform boilerplate and list processing. In *Haskell 2007, Freiburg, Germany*. 49–60. https://doi.org/10.1145/1291201.1291208
[40] Alan Mycroft and Neil D. Jones. 1985. A relational framework for abstract interpretation. In *Programs as Data Objects*. 156–171. https://doi.org/10.1007/3-540-16446-4_9
[41] Valentin Perrelle and Nicolas Halbwachs. 2010. An Analysis of Permutations in Arrays. In *VMCAI 2010*. 279–294. https://doi.org/10.1007/978-3-642-11319-2_21
[42] Tuan-Hung Pham and Michael W. Whalen. 2013. An Improved Unrolling-Based Decision Procedure for Algebraic Data Types. In *VSTTE 2013*. 129–148. https://doi.org/10.1007/978-3-642-54108-7_7
[43] Andrew Reynolds and Viktor Kuncak. 2015. Induction for SMT Solvers. In *VMCAI 2015*. 80–98. https://doi.org/10.1007/978-3-662-46081-8_5
[44] Xavier Rival and Laurent Mauborgne. 2007. The trace partitioning abstract domain. *ACM Trans. Program. Lang. Syst.* 29, 5 (2007), 26. https://doi.org/10.1145/1275497.1275501
[45] Xavier Rival, Antoine Toubhans, and Bor-Yuh Evan Chang. 2014. Construction of Abstract Domains for Heterogeneous Properties. In *ISoLA 2014*. 489–492. https://doi.org/10.1007/978-3-662-45231-8_40
[46] Mads Rosendahl. 2013. Abstract Interpretation as a Programming Language. In *Semantics, Abstract Interpretation, and Reasoning about Programs: Essays Dedicated to David A. Schmidt on the Occasion of his Sixtieth Birthday*. 84–104. https://doi.org/10.4204/EPTCS.129.7
[47] John M. Rushby, Sam Owre, and Natarajan Shankar. 1998. Subtypes for Specifications: Predicate Subtyping in PVS. *IEEE Trans. Software Eng.* 24, 9 (1998), 709–720. https://doi.org/10.1109/32.713327
[48] Shmuel Sagiv, Thomas W. Reps, and Reinhard Wilhelm. 2002. Parametric shape analysis via 3-valued logic. *ACM Trans. Program. Lang. Syst.* 24, 3 (2002), 217–298. https://doi.org/10.1145/514188.514190
[49] David A. Schmidt. 1998. Trace-Based Abstract Interpretation of Operational Semantics. *Lisp and Symbolic Computation* 10, 3 (1998), 237–271.
[50] Dana S. Scott. 1976. Data Types as Lattices. *SIAM J. Comput.* 5, 3 (1976), 522–587. http://dx.doi.org/10.1137/0205037
[51] Peter Sestoft and Niels Hallenberg. 2017. *Programming language concepts*. Springer.

[52] Micha Sharir and Amir Pnueli. 1981. Two Approaches to Interprocedural Data Flow Analysis. In *Program Flow Analysis: Theory and Applications*. Chapter 7.
[53] AnthonyM. Sloane. 2011. Lightweight Language Processing in Kiama. In *GTTSE III*, JoãoM. Fernandes, Ralf Lämmel, Joost Visser, and João Saraiva (Eds.). Lecture Notes in Computer Science, Vol. 6491. Springer Berlin Heidelberg, 408–425. https://doi.org/10.1007/978-3-642-18023-1_12
[54] Michael B. Smyth and Gordon D. Plotkin. 1982. The Category-Theoretic Solution of Recursive Domain Equations. *SIAM J. Comput.* 11, 4 (1982), 761–783. http://dx.doi.org/10.1137/0211062
[55] Philippe Suter, Mirco Dotta, and Viktor Kuncak. 2010. Decision procedures for algebraic data types with abstractions. In *POPL 2010*, Manuel V. Hermenegildo and Jens Palsberg (Eds.). ACM, 199–210. https://doi.org/10.1145/1706299.1706325
[56] Antoine Toubhans, Bor-Yuh Evan Chang, and Xavier Rival. 2013. Reduced Product Combination of Abstract Domains for Shapes. In *VMCAI 2013*. 375–395. https://doi.org/10.1007/978-3-642-35873-9_23
[57] Niki Vazou, Patrick Maxim Rondon, and Ranjit Jhala. 2013. Abstract Refinement Types. In *ESOP 2013*. 209–228. https://doi.org/10.1007/978-3-642-37036-6_13
[58] Mark Weiser. 1984. Program Slicing. *IEEE Trans. Software Eng.* 10, 4 (1984), 352–357. https://doi.org/10.1109/TSE.1984.5010248
[59] Glynn Winskel. 1993. Information Systems. MIT Press, Chapter 12.
[60] Niklaus Wirth. 1996. *Compiler Construction*. Addison-Wesley.
[61] Hongwei Xi and Frank Pfenning. 1998. Eliminating Array Bound Checking Through Dependent Types. In *PLDI 1998*. 249–257. https://doi.org/10.1145/277650.277732

A OPERATIONAL PATTERN MATCHING

Computing Pattern Matching. For an ordinary pattern p (top) the abstraction relation is direct: an abstract store $\widehat{\sigma}$ abstracts a concrete store σ and a value shape $\widehat{vs}$ abstracts a concrete value v. The notable change is that the abstract semantics uses a set of abstract binding environments $\widehat{\varrho} \subseteq \widehat{\mathsf{Store}} \times \widehat{\mathsf{ValueShape}} \times \widehat{\mathsf{BindingEnv}}_{\perp}$ that not only abstracts over the sequence of concrete binding environments $\underline{\rho}$, but also, for each abstract binding environment stores the corresponding refinement of the input abstract store $\widehat{\sigma}$ and the corresponding refinement of the matched value shape $\widehat{vs}$ according to the matched pattern.

For sequences of set sub-patterns $\star_{\underline{I}} p$, the sequence of concrete values $\underline{v}$ is abstracted by two components: the shape of values $\widehat{vs}$ and an interval approximating the length of the value sequence $[l; u]$. Both of these values are refined as a result of the matching, which is captured by the abstract binding environment $\widehat{\varrho}$ (of the same type as for the simple patterns), since we treat the value refined as the abstract set containing the values of the given shape and of given cardinality. The concrete semantics of set sub-patterns also contains a backtracking state $\mathbb{V}$ which is not used in the abstract semantics, because the abstraction of set elements is coarse and we thus abstractly consider all possible subset assignments at the same time (joining instead of backtracking).

Operational Rules. We will show how refinement is calculated by the abstract operational semantics by presenting some of key rules for abstract pattern matching. Rascal also allows non-linear pattern matching against assigned store variables, and it is possible to use this information for refining the input store and abstract value. In the AP-V-U rule we match the variable to the value shape and restrict the shape abstraction for the variable value to match the pattern. The binding environment does not change as the name is already bound in the store. In the AP-V-F rule, the matching fails ($\perp$), and then we learn that the value shape in the store should be refined to something that does not match.

$$\text{AP-V-U}\ \frac{\widehat{\sigma}(x) = (b, \widehat{vs}') \qquad \widehat{vs}' \neq \perp_{\widehat{\mathsf{VS}}} \qquad \widehat{vs}'' \in (\widehat{vs} \mathbin{\widehat{=}} \widehat{vs}') \quad \widehat{\sigma}' = \widehat{\sigma}[x \mapsto (\mathsf{ff}, \widehat{vs}'')]}{\widehat{\sigma} \vdash x \overset{?}{:=} \widehat{vs} \underset{\text{a-match-v}}{\Longrightarrow} (\widehat{\sigma}', \widehat{vs}'', [])}$$

$$\text{AP-V-F}\ \frac{\widehat{\sigma}(x) = (b, \widehat{vs}') \qquad \widehat{vs}' \neq \bot_{\widehat{\mathrm{VS}}} \qquad (\widehat{vs}'', \widehat{vs}''') \in (\widehat{vs} \mathbin{\widehat{\neq}} \widehat{vs}') \quad \widehat{\sigma}' = \widehat{\sigma}[x \mapsto (\mathsf{ff}, \widehat{vs}''')]}{\widehat{\sigma} \vdash x \overset{?}{\coloneqq} \widehat{vs} \underset{\text{a-match-v}}{\Longrightarrow} (\widehat{\sigma}', \widehat{vs}'', \bot)}$$

We also show the AP-V-B (abstract pattern-variable-bind) rule which simply binds the variable in the binding environment, assuming that it is possibly not assigned in the store (a free name).

$$\text{AP-V-B}\ \frac{\widehat{\sigma}(x) = (\mathsf{tt}, \widehat{vs}')}{\widehat{\sigma} \vdash x \overset{?}{\coloneqq} \widehat{vs} \underset{\text{a-match-v}}{\Longrightarrow} (\widehat{\sigma}[x \mapsto (\mathsf{tt}, \bot_{\widehat{\mathrm{VS}}})], \widehat{vs}, [x \mapsto \widehat{vs}])}$$

If our matched abstract value possibly contains the pattern constructor k (AP-C-S rule: abstract pattern-constructor-success) we produce an abstract value with k containing the sub-values refined against constructor sub-patterns:

$$\text{AP-C-S}\ \frac{\begin{array}{c}\mathsf{data}\ at = \cdots \mid k(\underline{t}) \mid \ldots \\ (\mathsf{success}\ k(\underline{\widehat{vs}'})) \in \widehat{\mathsf{unfold}}(\widehat{vs}, at) \\ \widehat{\sigma} \vdash p_1 \overset{?}{\coloneqq} \widehat{vs}'_1 \underset{\text{a-match}}{\Longrightarrow} \widehat{\varrho}_1 \ldots \widehat{\sigma} \vdash p_{\mathrm{n}} \overset{?}{\coloneqq} \widehat{vs}'_{\mathrm{n}} \underset{\text{a-match}}{\Longrightarrow} \widehat{\varrho}_{\mathrm{n}} \\ (\widehat{\sigma}'_1, \widehat{vs}'_1, \widehat{\rho}^{?}_1) \in \widehat{\varrho}_1 \ldots (\widehat{\sigma}'_{\mathrm{n}}, \widehat{vs}'_{\mathrm{n}}, \widehat{\rho}^{?}_{\mathrm{n}}) \in \widehat{\varrho}_{\mathrm{n}}\end{array}}{\widehat{\sigma} \vdash k(\underline{p}) \overset{?}{\coloneqq} \widehat{vs} \underset{\text{a-match-cons}}{\Longrightarrow} (\textstyle\bigsqcap_i \widehat{\sigma}_i, k(\underline{\widehat{vs}'}), \widehat{\mathsf{merge}}(\underline{\widehat{\rho}^{?}}))}$$

The total function $\widehat{\mathsf{merge}}$ unifies assignments from two binding environments point-wise by names, taking the greatest lower bound of shapes to combine bindings for a name. It yields bottom for the entire result if at least one of the point-wise meets yields bottom (shapes for at least one name are not reconcilable). Otherwise, we try to refine the matched value to exclude the pattern constructor in the AP-C-F rules:

$$\text{AP-C-F1}\ \frac{\begin{array}{c}\mathsf{data}\ at = \cdots \mid k(\underline{t}) \mid \ldots \\ (\mathsf{success}\ k'(\underline{\widehat{vs}'})) \in \widehat{\mathsf{unfold}}(\widehat{vs}, at) \quad k' \neq k\end{array}}{\widehat{\sigma} \vdash k(\underline{p}) \overset{?}{\coloneqq} \widehat{vs} \underset{\text{a-match-cons}}{\Longrightarrow} (\widehat{\sigma}, \widehat{\mathsf{exclude}}(\widehat{vs}, k), \bot)}$$

$$\text{AP-C-F2}\ \frac{\mathsf{data}\ at = \cdots \mid k(\underline{t}) \mid \ldots \quad \mathsf{error} \in \widehat{\mathsf{unfold}}(\widehat{vs}, at)}{\widehat{\sigma} \vdash k(\underline{p}) \overset{?}{\coloneqq} \widehat{vs} \underset{\text{a-match-cons}}{\Longrightarrow} (\widehat{\sigma}, \widehat{\mathsf{exclude}}(\widehat{vs}, k), \bot)}$$

For set patterns, the refinement happens by pattern matching set sub-patterns.

$$\text{AP-S-S}\ \frac{\begin{array}{c}\mathsf{success}\ \{\widehat{vs}'\}_{[l;u]} \in \widehat{\mathsf{unfold}}(\widehat{vs}, \mathsf{set}\langle\mathsf{value}\rangle) \\ \widehat{\sigma} \vdash \underline{\star p} \overset{?}{\coloneqq} \widehat{vs}, [l;u] \underset{\text{a-match}\star}{\Longrightarrow} \widehat{\varrho}\end{array}}{\widehat{\sigma} \vdash \{\underline{\star p}\} \overset{?}{\coloneqq} \widehat{vs} \underset{\text{a-match-set}}{\Longrightarrow} \widehat{\varrho}}$$

For example, when it is possible that the abstracted value sequence $(\widehat{vs}, [l;u])$ is empty ($l = 0$) and patterned matched against an empty set sub-pattern sequence, we can refine the result to be

the empty abstract set $\{\bot\}_0$ (rule APL-E-B).

$$\text{APL-E-B}\ \frac{l \leq u \quad l = 0}{\widehat{\sigma} \vdash \varepsilon \overset{?}{:=} \widehat{vs}, [l;u] \underset{\text{a-match}\star\text{-1}}{\Longrightarrow} (\widehat{\sigma}, \{\bot_{\widehat{VS}}\}_0, \{[]\})}$$

A more complex example is the one where we try to pattern match a potentially non-empty value sequence against a set sub-pattern sequence $p, \underline{\star p'}$ starting with an ordinary pattern (APL-M-P). Here we pattern match against p and the rest of the sequence $\underline{\star p'}$ and combine the refined results of these matches producing a refinement of the containing set value by combining the refined shapes and increasing the refinement of the length by the set sub-pattern sequence by one.

$$\text{APL-M-P}\ \frac{\begin{array}{c} l \leq u \qquad u \neq 0 \qquad \widehat{\sigma} \vdash p \overset{?}{:=} \widehat{vs} \underset{\text{a-match}}{\Longrightarrow} \widehat{\varrho_R}' \\ \widehat{\sigma} \vdash \underline{\star p} \overset{?}{:=} \widehat{vs}, [l-1;u-1] \underset{\text{a-match}\star}{\Longrightarrow} \widehat{\varrho_R}'' \\ (\widehat{\sigma}', \widehat{vs}', \widehat{\varrho}') \in \widehat{\varrho_R}' \quad (\widehat{\sigma}'', \{\widehat{vs}''\}_{[l'';u'']}, \widehat{\varrho}'') \in \widehat{\varrho_R}'' \\ \widehat{\varrho_R}''' = \left\{ \begin{array}{l} (\widehat{\sigma}' \sqcap \widehat{\sigma}'', \{\widehat{vs}' \sqcup \widehat{vs}''\}_{[l''+1,u''+1]}, \\ \widehat{\text{merge}}(\widehat{\varrho}', \widehat{\varrho}'')) \end{array} \right\} \end{array}}{\widehat{\sigma} \vdash p, \underline{\star p} \overset{?}{:=} \widehat{vs}, [l;u] \underset{\text{a-match}\star\text{-1}}{\Longrightarrow} \widehat{\varrho_R}'''}$$

B ABSTRACT SEMANTIC RULES

Figures 12 and 13 shows the formal rules for executing the bottom-up visit-expression; we have omitted the collecting rules and some error handling rules to avoid presenting unnecessary details. We will further discuss the ideas behind the rules in a high-level fashion.

Executing visitors. The evaluation rule for the visit-expression itself is mainly concerned with evaluating the target expression e to be traversed to a value, and then using a separate traversal relation to rewrite the value recursively with the sequence of cases $\underline{cs}$. The main item to notice is how it uses the value refined by the case patterns in case of failure (AE-VT-F), turning the result into successful execution (like in our running example in Sect. 2).

Evaluating Cases. During traversal, the target value will be rewritten with a sequence of cases. The evaluation of a case sequence is straight-forward, iterating through the possible cases, pattern matching against each pattern and executing the corresponding expression when applicable. The main idea is that, when the abstract value fails to match a pattern, the refined value is used to match against the rest of the cases (ACS-M-F). This ensures that the order of patterns influences the refinement, leading to a more precise abstract shape that better matches the set of concrete shapes during execution.

Expressions (General)

$$\text{AE-A}\ \frac{\{\!|x = e; \widehat{\sigma} \xRightarrow[\text{a-expr-asgn}]{} \widehat{Res}|\!\}}{x = e; \widehat{\sigma} \xRightarrow[\text{a-expr}]{} \widehat{Res}} \qquad \text{AE-Sq}\ \frac{\{\!|e_1; e_2; \widehat{\sigma} \xRightarrow[\text{a-expr-seq}]{} \widehat{Res}|\!\}}{e_1; e_2; \widehat{\sigma} \xRightarrow[\text{a-expr}]{} \widehat{Res}}$$

$$\text{AE-C}\ \frac{\{\!|k(\underline{e}); \widehat{\sigma} \xRightarrow[\text{a-expr-cons}]{} \widehat{Res}|\!\}}{k(\underline{e}); \widehat{\sigma} \xRightarrow[\text{a-expr}]{} \widehat{Res}} \qquad \text{AE-St}\ \frac{\{\!|\{\underline{e}\}; \widehat{\sigma} \xRightarrow[\text{a-expr-set}]{} \widehat{Res}|\!\}}{\{\underline{e}\}; \widehat{\sigma} \xRightarrow[\text{a-expr}]{} \widehat{Res}}$$

$$\text{AE-Fl}\ \frac{}{\mathsf{fail}; \widehat{\sigma} \xRightarrow[\text{a-expr}]{} [\mathsf{fail} \mapsto (\cdot, \widehat{\sigma})]} \qquad \text{AES}\ \frac{\{\!|\underline{e}; \widehat{\sigma} \xRightarrow[\text{a-expr}\star\text{-1}]{} \widehat{Res}|\!\}}{\underline{e}; \widehat{\sigma} \xRightarrow[\text{a-expr}\star]{} \widehat{Res}}$$

Assignment Expression

$$\text{AE-A-S}\ \frac{\begin{array}{c}\mathsf{local}\ t\ x \vee \mathsf{global}\ t\ x \quad e; \widehat{\sigma} \xRightarrow[\text{a-expr}]{} \widehat{Res} \\ (\mathsf{success}, (\widehat{vs}, \widehat{\sigma}')) \in \widehat{Res} \quad \widehat{vs} \mathbin{\widehat{:}} t' \quad t' \mathbin{\widehat{<:}} t\end{array}}{x = e; \widehat{\sigma} \xRightarrow[\text{a-expr-asgn}]{} [\mathsf{success} \mapsto (\widehat{vs}, \widehat{\sigma}'[x \mapsto (\mathsf{ff}, \widehat{vs})])]}$$

$$\text{AE-A-Er}\ \frac{\begin{array}{c}\mathsf{local}\ t\ x \vee \mathsf{global}\ t\ x \quad e; \widehat{\sigma} \xRightarrow[\text{a-expr}]{} \widehat{Res} \\ (\mathsf{success}, (\widehat{vs}, \widehat{\sigma}')) \in \widehat{Res} \quad \widehat{vs} \mathbin{\widehat{:}} t' \quad t' \mathbin{\widehat{\not<:}} t\end{array}}{x = e; \widehat{\sigma} \xRightarrow[\text{a-expr-asgn}]{} [\mathsf{error} \mapsto (\cdot, \widehat{\sigma}')]}$$

$$\text{AE-A-Ex}\ \frac{e; \widehat{\sigma} \xRightarrow[\text{a-expr}]{} \widehat{Res} \quad (exres, (\widehat{resv}, \widehat{\sigma}')) \in \widehat{Res}}{x = e; \widehat{\sigma} \xRightarrow[\text{a-expr-asgn}]{} [exres \mapsto (\widehat{resv}, \widehat{\sigma}')]}$$

Sequencing Expression

$$\text{AE-Sq-S}\ \frac{e_1, e_2; \widehat{\sigma} \xRightarrow[\text{a-expr}\star]{} \widehat{Res\star} \quad (\mathsf{success}, ((\widehat{vs}_1, \widehat{vs}_2), \widehat{\sigma}')) \in \widehat{Res\star}}{e_1; e_2; \widehat{\sigma} \xRightarrow[\text{a-expr-seq}]{} [\mathsf{success} \mapsto (\widehat{vs}_2, \widehat{\sigma}')]} \qquad \text{AE-Sq-Ex}\ \frac{e_1, e_2; \widehat{\sigma} \xRightarrow[\text{a-expr}\star]{} \widehat{Res\star} \quad (exres, (\widehat{resv}, \widehat{\sigma}')) \in \widehat{Res\star}}{e_1; e_2; \widehat{\sigma} \xRightarrow[\text{a-expr-seq}]{} [exres \mapsto (\widehat{resv}, \widehat{\sigma}')]}$$

Constructor Expression

$$\text{AE-C-S}\ \frac{\begin{array}{c}\mathsf{data}\ at = \ldots |k(\underline{t})| \ldots \quad \underline{e}; \widehat{\sigma} \xRightarrow[\text{a-expr}\star]{} \widehat{Res\star} \\ (\mathsf{success}, (\underline{\widehat{vs}}, \widehat{\sigma}')) \in \widehat{Res\star} \quad \underline{\widehat{vs} \mathbin{\widehat{:}} t'} \quad \underline{t' \mathbin{\widehat{<:}} t}\end{array}}{k(\underline{e}); \widehat{\sigma} \xRightarrow[\text{a-expr}]{} [\mathsf{success} \mapsto (k(\underline{\widehat{vs}}), \widehat{\sigma}')]}$$

$$\text{AE-C-Er}\ \frac{\begin{array}{c}\mathsf{data}\ at = \ldots |k(\underline{t})| \ldots \quad \underline{e}; \widehat{\sigma} \xRightarrow[\text{a-expr}\star]{} \widehat{Res\star} \\ (\mathsf{success}, (\underline{\widehat{vs}}, \widehat{\sigma}')) \in \widehat{Res\star} \quad \underline{\widehat{vs} \mathbin{\widehat{:}} t'} \quad \exists i. t_i' \mathbin{\widehat{\not<:}} t_i\end{array}}{k(\underline{e}); \widehat{\sigma} \xRightarrow[\text{a-expr}]{} [\mathsf{error} \mapsto (\cdot, \widehat{\sigma}')]}$$

$$\text{AE-C-Ex}\ \frac{\underline{e}; \widehat{\sigma} \xRightarrow[\text{a-expr}\star]{} \widehat{Res\star} \quad (exres, (\widehat{resv}, \widehat{\sigma}')) \in \widehat{Res\star}}{k(\underline{e}); \widehat{\sigma} \xRightarrow[\text{a-expr}]{} [exres \mapsto (\widehat{resv}, \widehat{\sigma}')]}$$

Fig. 10. Abstract Operational Semantics Rules for Expressions

Set Expression

$$\text{AE-St-S}\ \frac{\underline{e};\widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res\star} \quad (\text{success}, (\underline{vs}, \widehat{\sigma'})) \in \widehat{Res\star}}{\{\underline{e}\};\widehat{\sigma} \underset{\text{a-expr-set}}{\Longrightarrow\!\!\!\!\Longrightarrow} [\text{success} \mapsto (\{\bigsqcup_i \widehat{vs_i}\}_{[0;|\underline{vs}|]}, \widehat{\sigma'})]}$$

$$\text{AE-St-Ex}\ \frac{\underline{e};\widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res\star} \quad (exres, (\widehat{resv}, \widehat{\sigma'})) \in \widehat{Res\star}}{\{\underline{e}\};\widehat{\sigma} \underset{\text{a-expr-set}}{\Longrightarrow\!\!\!\!\Longrightarrow} [exres \mapsto (\widehat{resv}, \widehat{\sigma'})]}$$

Visit Expression

$$\text{AE-Vt-S}\ \frac{e;\widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res} \qquad (\text{success}, (\widehat{vs}, \widehat{\sigma''})) \in \widehat{Res} \qquad \underline{cs};\widehat{vs};\widehat{\sigma''} \underset{\text{a-bu-visit}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res}' \quad (\text{success}, (\widehat{vs}', \widehat{\sigma'})) \in \widehat{Res}'}{\text{visit } e\ \underline{cs};\widehat{\sigma} \underset{\text{a-expr-visit}}{\Longrightarrow\!\!\!\!\Longrightarrow} [\text{success} \mapsto (\widehat{vs}', \widehat{\sigma'})]}$$

$$\text{AE-Vt-F}\ \frac{e;\widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res} \qquad (\text{success}, (\widehat{vs}, \widehat{\sigma''})) \in \widehat{Res} \qquad \underline{cs};\widehat{vs};\widehat{\sigma''} \underset{\text{a-bu-visit}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res}' \quad (\text{fail}, (\widehat{vs}', \widehat{\sigma'})) \in \widehat{Res}'}{\text{visit } e\ \underline{cs};\widehat{\sigma} \underset{\text{a-expr-visit}}{\Longrightarrow\!\!\!\!\Longrightarrow} [\text{success} \mapsto (\widehat{vs}', \widehat{\sigma'})]}$$

$$\text{AE-Vt-Ex1}\ \frac{e;\widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res} \quad (exres, (\widehat{resv}, \widehat{\sigma'})) \in \widehat{Res}}{\text{visit } e\ \underline{cs};\widehat{\sigma} \underset{\text{a-expr-visit}}{\Longrightarrow\!\!\!\!\Longrightarrow} [exres \mapsto (\widehat{resv}, \widehat{\sigma'})]}$$

$$\text{AE-Vt-Ex2}\ \frac{e;\widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res} \qquad (\text{success}, (\widehat{vs}, \widehat{\sigma''})) \in \widehat{Res} \qquad \underline{cs};\widehat{vs};\widehat{\sigma''} \underset{\text{a-bu-visit}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res}' \quad (\text{error}, (\widehat{resv}, \widehat{\sigma'})) \in \widehat{Res}'}{\text{visit } e\ \underline{cs};\widehat{\sigma} \underset{\text{a-expr-visit}}{\Longrightarrow\!\!\!\!\Longrightarrow} [\text{error} \mapsto (\widehat{resv}, \widehat{\sigma'})]}$$

Expression Sequences

$$\text{AES-Em}\ \frac{}{\varepsilon;\widehat{\sigma} \underset{\text{a-expr}\star\text{-1}}{\Longrightarrow\!\!\!\!\Longrightarrow} [\text{success} \mapsto (\varepsilon, \widehat{\sigma})]}$$

$$\text{AES-Mr}\ \frac{e;\widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res} \quad (\text{success}, (\widehat{vs}, \widehat{\sigma''})) \in \widehat{Res} \quad \underline{e'};\widehat{\sigma''} \underset{\text{a-expr}\star}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res\star}' \quad (\text{success}, (\underline{\widehat{vs}'}, \widehat{\sigma'})) \in \widehat{Res\star}'}{e, \underline{e'};\widehat{\sigma} \underset{\text{a-expr}\star\text{-1}}{\Longrightarrow\!\!\!\!\Longrightarrow} [\text{success} \mapsto ((\widehat{vs}, \underline{\widehat{vs}'}), \widehat{\sigma'})]}$$

$$\text{AES-Ex}\ \frac{e;\widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res} \quad (exres, (\widehat{resv}, \widehat{\sigma'})) \in \widehat{Res}}{e, \underline{e'};\widehat{\sigma} \underset{\text{a-expr}\star\text{-1}}{\Longrightarrow\!\!\!\!\Longrightarrow} [exres \mapsto (\widehat{resv}, \widehat{\sigma'})]}$$

$$\text{AES-Ex}\ \frac{e;\widehat{\sigma} \underset{\text{a-expr}}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res} \quad (\text{success}, (\widehat{vs}, \widehat{\sigma''})) \in \widehat{Res} \quad \underline{e'};\widehat{\sigma''} \underset{\text{a-expr}\star}{\Longrightarrow\!\!\!\!\Longrightarrow} \widehat{Res\star}' \quad (exres, (\widehat{resv}, \widehat{\sigma'})) \in \widehat{Res\star}'}{e, \underline{e'};\widehat{\sigma} \underset{\text{a-expr}\star\text{-1}}{\Longrightarrow\!\!\!\!\Longrightarrow} [exres \mapsto (\widehat{resv}, \widehat{\sigma'})]}$$

Fig. 11. Abstract Operational Semantics Rules for Expressions (Cont.)

Bottom-up Traversal of Single Value

$$\text{ABU-S}\ \frac{(\widehat{vs}'', \widehat{cvs}) \in \widehat{\text{children}}(\widehat{vs}) \quad \underline{cs}; \widehat{cvs}; \widehat{\sigma} \underset{\text{a-bu-visit}\star}{\Longrightarrow} \widehat{Res\star} \quad (\text{success}, (\widehat{cvs}', \widehat{\sigma}')) \in \widehat{Res\star} \quad \widehat{\text{recons}}\ \widehat{vs}''\ \text{using}\ \widehat{cvs}'\ \text{to}\ \widehat{RCRes} \quad (\text{success}, \widehat{vs}') \in \widehat{RCRes} \quad \underline{cs}; \widehat{vs}'; \widehat{\sigma}' \underset{\text{a-cases}}{\Longrightarrow} \widehat{Res}'}{\underline{cs}; \widehat{vs}; \widehat{\sigma} \underset{\text{a-bu-visit-go}}{\Longrightarrow} \widehat{Res}'}$$

$$\text{ABU-F}\ \frac{(\widehat{vs}'', \widehat{cvs}) \in \widehat{\text{children}}(\widehat{vs}) \qquad \underline{cs}; \widehat{cvs}; \widehat{\sigma} \underset{\text{a-bu-visit}\star}{\Longrightarrow} \widehat{Res\star} \quad (\text{fail}, (\widehat{cvs}', \widehat{\sigma}')) \in \widehat{Res\star} \quad \widehat{\text{recons}}\ \widehat{vs}''\ \text{using}\ \widehat{cvs}'\ \text{to}\ [\text{success} \mapsto \widehat{vs}'] \quad \underline{cs}; \widehat{vs}'; \widehat{\sigma}' \underset{\text{a-cases}}{\Longrightarrow} \widehat{Res}'}{\underline{cs}; \widehat{vs}; \widehat{\sigma} \underset{\text{a-bu-visit-go}}{\Longrightarrow} \widehat{Res}'}$$

Bottom-up Traversal of Children

$$\text{ABUC-E}\ \frac{}{\underline{cs}; \varepsilon; \widehat{\sigma} \underset{\text{a-bu-visit}\star\text{-go}}{\Longrightarrow} [\text{fail} \mapsto (\varepsilon, \widehat{\sigma})]}$$

$$\text{ABUC-M}\ \frac{\underline{cs}; \widehat{vs}; \widehat{\sigma} \underset{\text{a-bu-visit}}{\Longrightarrow} \widehat{Res} \quad (\widehat{vfres}, (\widehat{vs}'', \widehat{\sigma}'')) \in \widehat{Res} \quad \underline{cs}; \widehat{\underline{vs}'}; \widehat{\sigma}'' \underset{\text{a-bu-visit}\star}{\Longrightarrow} \widehat{Res\star}' \quad (\widehat{vfres}', (\widehat{\underline{vs}''}, \widehat{\sigma}')) \in \widehat{Res\star}' \quad \widehat{Res}'' = \widehat{\text{vcombine}}(\widehat{vfres}, \widehat{vs}'', \widehat{vfres}', \widehat{\underline{vs}'''}, \widehat{\sigma}')}{\underline{cs}; \widehat{vs}, \widehat{\underline{vs}'}; \widehat{\sigma} \underset{\text{a-bu-visit}\star\text{-go}}{\Longrightarrow} \widehat{Res}''}$$

$$\text{ABUS-E}\ \frac{}{\underline{cs}; (\widehat{vs}, [0; u]); \widehat{\sigma} \underset{\text{a-bu-visit}\star\text{-go}}{\Longrightarrow} [\text{fail} \mapsto ((\bot, 0), \widehat{\sigma})]}$$

$$\text{ABUS-M}\ \frac{u > 0 \qquad \underline{cs}; \widehat{vs}; \widehat{\sigma} \underset{\text{a-bu-visit}}{\Longrightarrow} \widehat{Res} \qquad (\widehat{vfres}, (\widehat{vs}'', \widehat{\sigma}'')) \in \widehat{Res} \qquad \underline{cs}; (\widehat{vs}, [l-1; u-1]); \widehat{\sigma}'' \underset{\text{a-bu-visit}\star}{\Longrightarrow} \widehat{Res\star}' \quad (\widehat{vfres}', ((\widehat{vs}''', [l'; u']), \widehat{\sigma}')) \in \widehat{Res\star}' \quad \widehat{Res}'' = \widehat{\text{vcombine}}(\widehat{vfres}, \widehat{vs}'', \widehat{vfres}', (\widehat{vs}''', [l'; u']), \widehat{\sigma}')}{\underline{cs}; (\widehat{vs}, [l; u]); \widehat{\sigma} \underset{\text{a-bu-visit}\star\text{-go}}{\Longrightarrow} \widehat{Res}''}$$

Fig. 12. Selected Abstract Operational Semantics Rules for Traversal

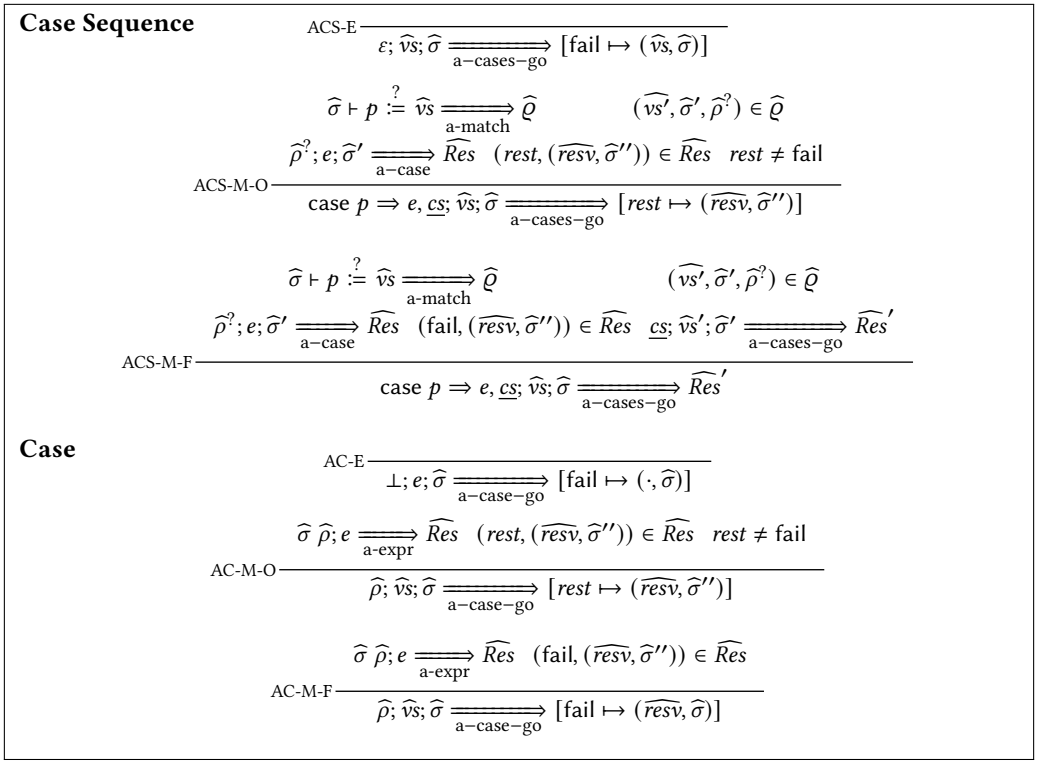

Fig. 13. Selected Abstract Operational Semantic Rules for Traversal (Cont.)