Failing Faster: Overlapping Patterns for Property-Based Testing

Jonathan Fowler(✉) and Graham Hutton

School of Computer Science, University of Nottingham, Nottingham, UK
psxjf@nottingham.ac.uk

Abstract. In property-based testing, a key problem is generating input data that satisfies the precondition of a property. One approach is to attempt to do so automatically, from the definition of the precondition itself. This idea has been realised using the technique of needed narrowing, as in the Lazy SmallCheck system, however in practice this method often leads to excessive backtracking resulting in poor efficiency. To reduce the amount of backtracking, we develop an extension to needed narrowing that allows preconditions to fail faster based on the use of overlapping patterns. We formalise our extension, show how it can be implemented, and demonstrate that it improves efficiency in many cases.

1 Introduction

Property-based testing, popularised by systems such as QuickCheck [4], is an automated approach to testing in which a program is validated against a specification. In most tools, the specification consists of properties written as programs outputting Boolean values. Input data is generated randomly or systematically, and the program is executed in an attempt to find a counterexample. In order to generate the input data, it is often required to write a custom generator. For example, consider the following simple property of a sorting function:

$$\mathit{propSort}\ n\ l = \mathit{perm}\ n\ l \implies \mathit{sort}\ l \equiv [0\,..\,(n-1)]$$

This property has two arguments, given by a number n and a list of numbers l. The property itself states that if the list l is a permutation of the numbers from 0 to $n-1$, then sorting this list will give the expected result. However, while the above definition captures a valid property, it suffers from a practical problem. In particular, if we use a standard generator for a list of numbers, then the precondition *perm n l* will rarely be met, making it difficult to generate enough test cases to adequately test the *sort* function.

To overcome this problem, a custom generator is often used. For example, the QuickCheck system [4] provides a range of type-classes and combinators for building custom generators, using which we can define a generator for properties such as *propSort*. Nevertheless, writing custom generators is time consuming and for more complex examples, such as generating well-typed terms, is the subject of ongoing research [12,17]. Furthermore, it is difficult to combine such generators,

© Springer International Publishing AG 2017
Y. Lierler and W. Taha (Eds.): PADL 2017, LNCS 10137, pp. 103–119, 2017.
DOI: 10.1007/978-3-319-51676-9_7

in the sense that two generators that are efficient in isolation may no longer be efficient when they are combined together in some way.

Another approach is to attempt to derive an efficient generator from the definition of the precondition, in our example the property *perm*. One realisation of this approach is to use the technique of *needed narrowing* [1,10] from functional logic programming. For example, Lazy Smallcheck [19] adopts a narrowing strategy and EasyCheck [2] directly uses the needed narrowing language Curry. Using this approach, a program is evaluated in a speculative manner. Beginning with a free input variable, the variable is refined by choosing a constructor when the value is required to proceed with evaluation. If evaluation ends negatively then the process backtracks, while if it ends positively then we have generated a value satisfying the condition. However, a naturally written property often does not make an efficient generator. In particular, the generator may be forced to backtrack excessively if it finds itself in a branch of the program for which the constraints are never satisfied, as we shall see in the next section.

In this paper, we explore a new technique to help reduce the amount of backtracking that is required in a needed-narrowing approach to property-based testing. The technique, which is a generalisation of the parallel conjunction approach used in several tools [3,13,19], allows evaluation of multiple branches of the program simultaneously, potentially allowing a result to be derived at an earlier stage of refinement. Particularly for commonly-used operators such as conjunction, disjunction and addition, both arguments can be evaluated in tandem. To achieve this, we use a form of *overlapping pattern matching*. The pattern matching is resolved in an order-independent fashion and overlapping patterns are allowed. More precisely, in this paper we:

– Motivate and introduce the use of overlapping patterns for needed narrowing property-based testing using our permutation example (Sect. 2).
– Define our source language and formalise the notion of overlapping patterns within this language using a needed narrowing semantics (Sect. 3).
– Give an overview of our prototype implementation (Sect. 4), and explore the benefits of overlapping patterns in two case studies (Sect. 5).
– Compare with related work (Sect. 6) and conclude (Sect. 7).

The paper is aimed at a general functional programming audience with some experience of property-based testing systems such as QuickCheck, but no specialist knowledge of needed narrowing is assumed. For the purposes of examples, we use a Haskell-like syntax and semantics. Although we only apply our technique to property-based testing, we believe it could also be used to improve the efficiency of other tools and languages based on needed narrowing.

2 Motivation and Basic Idea

To motivate the need for overlapping pattern matching we take a deeper look at the example from the introduction. We show that naively evaluating the *perm* precondition with a needed narrowing strategy results in excessive backtracking, and how the use of an overlapping conjunction operator mitigates the problem. We begin by defining the *perm* condition as follows:

$$\begin{aligned}&\mathit{perm} &&:: \mathit{Nat} \rightarrow [\mathit{Nat}] \rightarrow \mathit{Bool}\\ &\mathit{perm}\ n\ l &&= \mathit{length}\ l \equiv n\ \wedge\ \mathit{all}\ ({<}n)\ l\ \wedge\ \mathit{allDiff}\ l\end{aligned}$$

That is, a list l is a permutation of the natural numbers below a given limit n if three conditions are satisfied: the list has the correct length, all the numbers in the list are below the limit, and all the numbers in the list are different. Preconditions defined as a conjunction of constraints in this manner are a common pattern in properties. Because needed narrowing is more effective on algebraic data types than on primitive data types [16], we assume an inductive type *Nat* of natural numbers built up from basic constructors *Zero* and *Suc*.

By running a needed narrowing evaluation on the above definition for *perm*, we can generate values satisfying the constraints. Needed narrowing is based upon the idea of extending normal evaluation to include *free variables*, with the values of such variables being refined as evaluation proceeds. We give a brief overview to the technique below; for a more in-depth introduction, see [9,15].

At the start of evaluation, a free variable is chosen for each argument of the program and the program is reduced until the value of a variable is required to continue. The program is then *suspended* on the variable. To allow further progress, a constructor is chosen for the suspended variable and the program is then reduced further in the same way. If evaluation fails to succeed we backtrack, choosing alternative constructors for each variable.

For example, if we consider the first constraint $\mathit{length}\ l \equiv n$ in *perm* for the case when $n = 4$, needed narrowing will yield the following solution:

$$l = [x_0, x_1, x_2, x_3]$$

Along the way we would discard lists such as $[x_0, x_1, x_2]$ which fail to satisfy the constraint. The variables x_0–x_3 are free in the solution, in the sense they can be substituted for any value while still satisfying the constraint. Continuing with evaluation of the second constraint, $\mathit{all}\ ({<}4)\ l$, we further refine the variables. We consider a *partial* solution, in which the constraint is not yet satisfied, to illustrate a situation in which excessive backtracking will occur:

$$l = [1, 1, x_2, x_3]$$

This list begins with two ones and according to the current constraint, $\mathit{all}\ ({<}4)\ l$, we have neither failed nor succeeded and as such we should carry on refining x_2 and x_3. However, when we arrive at the final constraint, $\mathit{allDiff}\ l$, this partial solution will fail, and we have to backtrack over all combinations of x_2 and x_3 before we can continue again. Moreover, as we consider generating longer permutations, the amount of backtracking increases exponentially.

Note that the problem is not resolved by reordering the constraints. For example, suppose that we swapped the order of the last two constraints:

$$\mathit{perm}\ n\ l = \mathit{length}\ l \equiv n\ \wedge\ \mathit{allDiff}\ l\ \wedge\ \mathit{all}\ ({<}n)\ l$$

Then we quickly run into a similar issue. For example, the partial solution, $l = [4, x_1, x_2, x_3]$ does not fail the *allDiff* l constraint, however it will fail the *all* (<4) l constraint but only after the remaining variables x_1–x_3 are refined while evaluating *allDiff*. In both cases, the backtracking is caused because a partial solution fails to satisfy a later constraint but this is not evident at the time due to the evaluation order. A natural way to avoid this problem is to evaluate all the constraints simultaneously, rather than sequentially.

To realise this behaviour, we replace traditional pattern matching in our language with *overlapping* pattern matching. Pattern matching in the language is then order-independent, and in each iteration of needed narrowing all relevant arguments to a pattern match are normalised. By way of example, consider the logical conjunction operator, which is traditionally defined as follows:

$$\begin{aligned} &\mathit{False} \wedge \_ = \mathit{False} \\ &\mathit{True} \wedge x = x \end{aligned}$$

Using this definition, progress can only be made by evaluating the first argument of a conjunction, because each clause of the definition depends on the value of the first argument. Instead, we re-define the operator using overlapping patterns, using a special-purpose pragma to indicate the change in intended semantics:

{-# OVERLAP (∧) #-}

$$\begin{aligned} &\mathit{False} \wedge \_ &&= \mathit{False} \\ &\_ \wedge \mathit{False} &&= \mathit{False} \\ &\mathit{True} \wedge x &&= x \\ &x \wedge \mathit{True} &&= x \end{aligned}$$

The definition has two new clauses, given by simply commuting the order of the arguments in the original definition. The idea is that a pattern match can succeed on any of the four clauses, independent of the order that they are stated in. Using this definition, progress can be made by evaluating either argument of the conjunction as the new clauses are no longer dependent on the first argument. For example, we can now reduce $x \wedge \mathit{False}$ to *False* for any expression x, which is not the case with the original definition. To take advantage of this additional power, the underlying needed narrowing mechanism must be modified to evaluate both arguments of the pattern match before it refines variables.

In *perm* example above, we considered the list $l = [1, 1, x_2, x_3]$ and found that it required a large amount of backtracking. In particular, the constraint *allDiff* l only failed once we had considered all combinations of x_2 and x_3. The new overlapping conjunction operator avoids this problem because it is not biased to the left-argument, allowing *allDiff* l to fail immediately for this example list without the need to further refine the remaining variables.

This additional efficiency is also borne out in practice. For example, using the implementation that we describe in Sect. 4, in the time it takes to generate one hundred valid permutations of length eight for the *perm* constraint defined using the traditional conjunction operator, we can generate one hundred valid permutations of length thirty using the overlapping version.

However, we have to be careful when using overlapping pattern matching not to introduce non-determinism. Consider the following dangerous function:

$$\begin{array}{l}
\{\text{-\# OVERLAP } \textit{danger} \text{ \#-}\} \\
\textit{danger False}\ \_ \quad\;\; = \textit{False} \\
\textit{danger}\ \_ \quad\;\; \textit{False} = \textit{True} \\
\textit{danger True True} = \textit{True}
\end{array}$$

Using this definition, *danger False False* can reduce to either *False* or *True*, depending on whether the first or second clause is used, and is therefore non-deterministic. To counter this, we require confluence laws which guarantee that evaluation is deterministic if all expressions are terminating.

Other logical operators such as disjunction and implication can be defined using overlapping patterns in a similar manner to conjunction, and will benefit from similar improvements in efficiency. The mechanism can also be used with other data types. For example it is straightfoward to define overlapping versions of the addition and maximum operators on natural numbers, and for the applicative operator <*> on the *Maybe* type [14]. As illustrated by the latter example, overlapping definitions are not restricted to commutative operators. The *maximum* function is defined and used in Sect. 5, and a range of other useful overlapping definitions are provided in our implementation [8].

3 Generalizing and Formalizing

In this section we define the syntax and semantics of our language of overlapping patterns. We consider the normalising subset of the language and show that a confluence restriction on definitions is sufficient to guarantee that the language is deterministic. We then extend the semantics with needed narrowing, and show that the new semantics is sound and complete with respect to the original.

We use a simple functional programming core language with definitions, constructors, variables, lambda expressions and application. To simplify the theory, the language only allows one form of pattern matching: at the top-level of a function definition, interpreted in an overlapping, order-independent manner. However, other forms of pattern matching, such as **case** expressions and non-overlapping patterns, can readily be rewritten in this form.

The syntax of the language is formally defined as follows:

$$\begin{array}{lcl}
\textit{Defn}_X & ::= & \overline{\textit{Var}\ \overline{\textit{Patt}} = \textit{Expr}_X} \\
\textit{Expr}_X & ::= & \textit{Con} \mid \textit{Var} \mid X \mid \textit{Expr}_X\ \textit{Expr}_X \mid \lambda\,\textit{Var}\ .\ \textit{Expr}_X \\
\textit{Patt} & ::= & \textit{Con}\ \overline{\textit{Patt}} \mid \textit{Var}
\end{array}$$

That is, a definition is made up of a list of clauses, with a pattern for each argument on the left and an expression on the right. We use an overline to represent a list of elements, e.g. $\overline{\textit{Patt}}$ is a potentially empty list of patterns. Expressions and definitions are parameterised by a set of free variables X, which is only used in the needed narrowing semantics. The language has standard set of typing rules, which we omit for brevity. Each type has a set of constructors and the patterns used in definitions should form a covering of these constructors. Each variable should only appear once in a pattern, and the only free variables in an expression should be those that appear in the set X.

We often use f for definitions, e for expressions, c for constructors, u and v for closed variables, x and y for free variables, and p and q for patterns.

3.1 Semantics

We give a standard small-step operational semantics to the language in a contextual style. We start by defining a full reduction semantics, in which any reducible term in an expression can be reduced. This allows us to define notions of equivalence and establish confluence properties. We then define a call-by-name evaluation strategy by limiting the form of contexts that can be used, which is then used to define the needed narrowing strategy.

First we define a *local* semantics $\rightarrow_R \subseteq \mathit{Expr}_X \times \mathit{Expr}_X$ that performs basic reduction steps on expressions, which is then lifted into an evaluation context. As usual, a *substitution* is a mapping from variables to expressions, and we write $e[s]$ for the application of a substitution to each variable in an expression, $\emptyset$ for the identity substitution that maps each variable to itself, and $s;t$ for the composition of substitutions. The first local rule is the standard β-rule:

$$\frac{}{(\lambda v.e)\ e' \rightarrow_R e[v \mapsto e']}\ \textsc{sub}$$

The second rule states that we can reduce a definition if the pattern of any of its clauses matches the arguments, where $f\ \overline{p} = e\ \in\ \mathrm{defn}(f)$ means that the clause $f\ \overline{p} = e$ is part of the definition for the function f. In contrast to traditional pattern matching, the clauses of a definition may be applied in any order.

$$\frac{f\ \overline{p} = e'\ \in\ \mathrm{defn}(f) \qquad \mathtt{Matches}(\overline{p},\ \overline{e},\ s)}{f\ \overline{e} \rightarrow_R e'[s]}\ \textsc{match}$$

The predicate `Matches` used above captures the idea of a successful match of expressions against patterns, where `Match` gives the definition for a single pattern, `Matches` for a list of patterns, and s is the resulting substitution:

$$\frac{}{\mathtt{Match}(v,\ e,\ \{v \mapsto e\})} \qquad \frac{\mathtt{Matches}(\overline{p},\ \overline{e},\ s)}{\mathtt{Match}(c\ \overline{p},\ c\ \overline{e},\ s)}$$

$$\frac{}{\mathtt{Matches}(\epsilon,\ \epsilon,\ \emptyset)} \qquad \frac{\mathtt{Match}(p,\ e,\ s) \qquad \mathtt{Matches}(\overline{p},\ \overline{e},\ t)}{\mathtt{Matches}(p\ \overline{p},\ e\ \overline{e},\ s;t)}$$

In turn, a *context* is an expression with a singular hole in any location, as defined by the following set of inference rules:

$$\frac{}{[]\ \text{context}}\ \textsc{hole} \qquad \frac{\mathbf{C}\ \text{context}}{(\lambda v.\mathbf{C})\ \text{context}}\ \textsc{lam}$$

$$\frac{\mathbf{C}\ \text{context}}{(\mathbf{C}\ e)\ \text{context}}\ \textsc{app-l} \qquad \frac{\mathbf{C}\ \text{context}}{(e\ \mathbf{C})\ \text{context}}\ \textsc{app-r}$$

We use inference rules above rather than a grammer because the extra generality of this notation is used when contexts are revised later on. As usual, we write $\mathbf{C}[e]$ for the result of replacing the hole in $\mathbf{C}$ with the expression e.

Using the local semantics and the notion of contexts we can now define the full reduction semantics for expressions in our language.

Definition 1. The *full reduction* semantics, $\rightarrow \subseteq Expr_X \times Expr_X$, is given by:

$$\frac{e \rightarrow_R e' \qquad \mathbf{C}\ \text{context}}{\mathbf{C}[e] \rightarrow \mathbf{C}[e']}$$

Definition 2. $\rightarrow^*$ is the reflexive/transitive closure of $\rightarrow$.

To ensure that our language is deterministic and avoid examples such as *danger False False* from Sect. 2 that have more than one normal form, we require all definitions to satisfy a confluence property. To formalise this property we first define the notions of definitional equivalence and unification.

Definition 3. Two expressions are *definitionally equivalent*, written $e \equiv e'$, if there exists reduction sequences from e and e' to the same expression:

$$e \equiv e' \iff \exists e''.\ e \rightarrow^* e'' \wedge e' \rightarrow^* e''$$

Informally, two patterns are unifiable if there exists an expression which matches both the patterns. We can formalise this by giving a pair of substitutions which when applied to each pattern yield the common expression.

Definition 4. The *most general unifier* is defined by the inference rules below. $\texttt{Unify}(p,\ q,\ s_1,\ s_2)$ denotes the unification of patterns p and q by substitutions s_1 and s_2, and similarly for a list of patterns with `Unifies`. Note we are using the assumption that every variable appears only once in each pattern here.

$$\frac{\texttt{Unifies}(\overline{p},\ \overline{q},\ s_1,\ s_2)}{\texttt{Unify}(c\ \overline{p},\ c\ \overline{q},\ s_1,\ s_2)} \qquad \frac{}{\texttt{Unify}(v,\ p,\ \{v \mapsto p\},\ \emptyset)}$$

$$\frac{}{\texttt{Unify}(p,\ v,\ \emptyset,\ \{v \mapsto p\})}$$

$$\frac{}{\texttt{Unifies}(\epsilon,\ \epsilon,\ \emptyset,\ \emptyset)} \qquad \frac{\texttt{Unify}(p,\ q,\ s_1,\ s_2) \qquad \texttt{Unifies}(\overline{p},\ \overline{q},\ s_1',\ s_2')}{\texttt{Unifies}(p\ \overline{p},\ q\ \overline{q},\ s_1;s_1',\ s_2;s_2')}$$

This definition has the expected behaviour, that is:

$$\begin{aligned}\texttt{Unify}(p,\ q,\ s_1,\ s_2) \implies\ & p[s_1] = q[s_2] && \text{(unifier)}\\ & \wedge \forall t_1 t_2.\ p[t_1] = q[t_2].\ \exists r.\ s_1; r = t_1 \wedge s_2; r = t_2 && \text{(most general)}\end{aligned}$$

If the patterns of two clauses of a definition are unifiable then, given a suitable context, it is possible for two different MATCH reductions in our semantics to be applied. In order to maintain determinism for such clauses a confluence restriction is required. The confluence restriction states that the right-hand sides of each pair of clauses must be definitionally equivalent under their unifying substitution if one exists. For the terminating subset of the language we can check whether a definition satisfies the confluence property automatically by generating the unifiers pairwise and normalising each clause.

Definition 5. A definition satisfies the *confluence restriction* if for any pair of clauses, $f\ \overline{p} = e$ and $f\ \overline{q} = e'$, we have the following property:

$$\texttt{Unifies}(\overline{p},\ \overline{q},\ s_1,\ s_2) \implies e[s_1] \equiv e'[s_2]$$

Theorem 1. *The relation $\rightarrow^*$ is confluent if all the definitions satisfy the confluence restriction, i.e. for any reductions $e \rightarrow^* e_1$, $e \rightarrow^* e_2$, there exists an expression e' such that $e_1 \rightarrow^* e'$ and $e_2 \rightarrow^* e'$.*

Proof. By parallel reduction with special consideration for overlapping patterns. □

It follows in the standard way from the above confluence property that any expression that only has finite reduction sequences has precisely one normal form. Hence, our semantics is deterministic for such expressions. We return to the issue of expressions with infinite reduction sequences in the concluding section.

3.2 Evaluation Order

Our current semantics allows reduction rules to be applied in any context and in any order. This is convenient for defining the behavioural properties of the semantics, but in order to define the needed narrowing semantics and give an efficient implementation, we need to restrict where reduction rules are applied. To do this we define a subset of contexts called *evaluation contexts.*

Our notion of evaluation context is call-by-name, and hence only evaluates the left-hand side of an application. When the left-most expression is a definition, we evaluate the arguments until a pattern match is possible. For efficiency, we only reduce arguments that could lead to a pattern match. Sometimes more than one argument needs to be reduced to allow a pattern match, in which case we reduce the arguments in a left-biased order. The rules are defined formally below, where $\overline{\mathbf{C}}$ is a list of expressions with one context.

$$\frac{}{\bullet\ \text{evalcxt}}\ \textsc{hole} \qquad \frac{\mathbf{C}\ \text{evalcxt}}{(\mathbf{C}\ e)\ \text{evalcxt}}\ \textsc{app-l} \qquad \frac{\texttt{Subjects}(\overline{\mathbf{C}},\ f)}{(f\ \overline{\mathbf{C}})\ \text{evalcxt}}\ \textsc{args}$$

The `Subject` predicates specify the parts of the arguments that should be reduced. The contexts which form the subjects have a clause of the definition for which they are the first sub-expression blocking the pattern match. All expressions to the left of the subject should already match their respective patterns in the clause. The predicates are defined by the following rules:

$$\frac{\texttt{Subject}(\mathbf{C},\ p) \qquad \texttt{Matches}(\overline{p},\ \overline{e_0},\ \_) \qquad f\ \overline{p_0}\ p\ \overline{p_1} = \_ \in \mathrm{defn}(f)}{\texttt{Subjects}(\overline{e_0}\ \mathbf{C}\ \overline{e_1},\ f)}$$

$$\frac{\mathbf{C}\ \text{evalcxt}}{\texttt{Subject}(\mathbf{C},\ c\ \overline{p})} \qquad \frac{\texttt{Subject}(\mathbf{C},\ p) \qquad \texttt{Matches}(\overline{p_0},\ \overline{e_0},\ \_)}{\texttt{Subject}(c\ \overline{e_0}\ \mathbf{C}\ \overline{e_1},\ c\ \overline{p_0}\ p\ \overline{p_1})}$$

Definition 6. The *evaluation reduction* semantics, $\rightarrow_E$, is now defined by:

$$\frac{\mathbf{C}\ \text{evalcxt} \qquad e \rightarrow_R e'}{\mathbf{C}[e] \rightarrow_E \mathbf{C}[e']}$$

3.3 Needed Narrowing

The needed narrowing semantics reduces an expression until all the evaluation contexts are suspended on a free variable. At this point we refine the left-most suspending variable to a new value, with the resulting refinements to the free variables being stored in an accompanying substitution. We call this type of substitution a *refinement* to disambiguate it from the general notion.

Formally, a refinement σ of type $X \mapsto Y$ is a function from the free variable set X to *partial values* with free variables Y, where a partial value is a term build up from constructors and variables. Composition of refinements, which we denote by $\ggg$, is defined in the standard way. The null refinement, $\mathit{return} \in X \mapsto X$, corresponds to the trivial substitution that maps each free variable to itself.

Definition 7. The *narrowing set* of a expression, $\mathit{narrowing}(e)$, is the set of refinements that replace the left-most suspended variable with a constructor of the correct type with new free variables for the fields. The narrowing set should be complete, in the sense that it contains every constructor of the type.

For example, the narrowing set for an expression suspended on a variable x of type *Nat* is (where x/c is the point refinement replacing x with c):

$$\{x/\,\texttt{Zero}, x/\,\texttt{Suc}\, y\} \qquad (y \text{ is a fresh variable})$$

We can now define the needed narrowing reduction as follows:

Definition 8. The *needed narrowing* reduction, $\rightsquigarrow \subseteq \mathit{Expr}_X \times (\mathit{Expr}_Y \times X \mapsto Y)$, is defined by the following two inference rules:

$$\frac{e \rightarrow_E e'}{e \rightsquigarrow \langle e',\ \mathit{return}\rangle} \qquad \frac{e \not\rightarrow_E \qquad \sigma \in \mathit{narrowing}(e)}{e \rightsquigarrow \langle e[\sigma],\ \sigma\rangle}$$

The first rule states that any evaluation reduction is also a needed narrowing reduction, with no refinement necessary. The second states that if no such reduction is possible then a refinement from the narrowing set should be used.

Definition 9. The natural extension to the reflexive/transitive closure of the needed narrowing reduction is given by composing the resulting refinements:

$$\frac{}{e \rightsquigarrow^* \langle e[\sigma],\ \sigma\rangle} \qquad \frac{e \rightsquigarrow \langle e,\ \sigma\rangle \qquad e' \rightsquigarrow^* \langle e'',\ \sigma'\rangle}{e \rightsquigarrow^* \langle e'',\ \sigma \ggg \sigma'\rangle}$$

Note that in the reflexive case we use an arbitrary substituton σ rather than the null refinement *return*, as this simplifies the formulation of the completeness result for our new semantics, which we now present along with soundness.

Theorem 2. *(Needed narrowing is sound.) For every needed narrowing reduction sequence there exists a corresponding reduction in the original semantics:*

$$e \rightsquigarrow^* \langle e',\ \sigma\rangle \implies e[\sigma] \rightarrow_E^* e'$$

Proof. By induction on the needed narrowing reduction chain. □

To ensure that the corresponding completeness theorem is valid, we restrict our attention to expressions that strongly normalise under any refinement, which we denote using the predicate `Norm`.

Theorem 3. *(Needed narrowing is complete.) For every reduction of a normalising expression there is a corresponding needed narrowing reduction:*

$$\texttt{Norm}(e_0) \wedge e_0[\sigma] \rightarrow_E^* e_1 \implies \exists\ e_1'.\ e_0 \rightsquigarrow^* \langle e_1',\ \sigma\rangle \wedge e_1 \equiv e_1'$$

Proof. By induction on the length of reduction sequences, which are guaranteed to be finite by the normalisation precondition. In order to complete the proof, we require a slightly generalised induction hypothesis:

$$\begin{aligned} &\texttt{Norm}(e_0)\ \wedge\ \texttt{Norm}(e_0')\ \wedge\ e_0[\sigma] \rightarrow_E^* e_1\ \wedge\ e_0 \equiv e_0' \\ &\quad \implies \exists\ e_1'.\ e_{0'} \rightsquigarrow^* \langle e_1',\ \sigma\rangle\ \wedge\ e_1 \equiv e_1' \end{aligned}$$ □

4 Implementation

In this section we give an overview of our prototype implementation of a property-based testing system based upon the ideas that we have introduced in the previous sections. The system itself is freely available on GitHub [8].

The source language used in the implementation is a core functional language with a Haskell-like syntax. The language includes algebraic data types and supports definitions with both overlapping and traditional pattern matching. Definitions are not currently checked for confluence, but as noted earlier this would be possible in a more mature implementation. For the purposes of the case studies in the next section we use Haskell syntax, which is translated into caseless monomorphic code in our implementation.

The implementation realises the needed-narrowing evaluation in a virtual machine encoded in Haskell. The result of evaluation is given by a lazy search tree, where each node comprises a free variable and sub-trees that provide constructor bindings for the variable, and the leaves are normal-form results. Different search strategies correspond to different methods of traversing the tree.

Properties in our system are functions with return type *Result*, which represents three possible outcomes: a failed precondition, in which case the test case is invalid; a successful result, where the test case satisfies the property; and a failure, where the test case is a counterexample. Properties are typically defined using a specialised implication operator $(\Longrightarrow) :: \mathit{Bool} \rightarrow \mathit{Bool} \rightarrow \mathit{Result}$.

Our system implements a random search strategy. At each node we select a random constructor according to a defined distribution, until we arrive at a result. If the precondition failed, we backtrack to try and find a valid result. It is sensible to limit the amount of backtracking as sometimes we might arrive at a state with no nearby solutions. We do this by limiting how many variables we can reverse, randomly enumerating all possible constructors at each variable in an attempt to find a continuation. For example, if the backtrack limit is set to three, and a failure occurs at a node which is twelve deep in the tree, we will backtrack to a minimum of depth nine in search of a solution.

5 Case Studies

In this section we consider two examples of using our system in practice. The first example involves the generation of ordered trees and demonstrates how overlapping patterns can be used to encode bespoke size constraints. The second generates typed expressions for a simple language and demonstrates a useful technique for writing efficient narrowing generators. In both examples we focus on the generation of data satisfying a constraint.

Our aim in each case is to find a definition of the constraint that eliminates the need for backtracking (apart from rebinding of a single constructor). We say that such a constraint *fails fast*. Formally, a constraint fails fast if when testing any partial value against the constraint it either directly fails or there is a refinement of the value that succeeds. The needed narrowing generator formed by a constraint which fails fast is generally efficient. All our examples, together with more detailed performance results, are available on GitHub [8].

5.1 Ordered Trees

Consider a type of binary trees with natural numbers stored in the nodes, with the additional constraint that numbers within the tree are ordered:

$$\begin{array}{l} \textbf{data}\ \mathit{Tree} = \mathit{Leaf} \mid \mathit{Node}\ \mathit{Tree}\ \mathit{Nat}\ \mathit{Tree} \\ \begin{array}{lcl} \mathit{ordered} & :: & \mathit{Tree} \to \mathit{Bool} \\ \mathit{ordered}\ \mathit{Leaf} & = & \mathit{True} \\ \mathit{ordered}\ (\mathit{Node}\ t0\ a\ t1) & = & \mathit{allTree}\ (\leqslant a)\ t0\ \wedge\ \mathit{allTree}\ (\geqslant a)\ t1 \\ & & \quad \wedge\ \mathit{ordered}\ t0\ \wedge\ \mathit{ordered}\ t1 \end{array} \end{array}$$

Note that *ordered* uses the overlapping version of conjunction to ensure that it can be tested in an efficient manner, and is defined using an auxiliary function *allTree* that checks if every element in a tree satisfies a given condition.

Now consider a function to delete a number from a tree while still maintaining the ordering invariant, as captured by the following property [15,16]:

$$\begin{array}{lcl} \mathit{propDelete} & :: & \mathit{Nat} \to \mathit{Tree} \to \mathit{Bool} \\ \mathit{propDelete}\ a\ t & = & \mathit{ordered}\ t\ \Longrightarrow\ \mathit{ordered}\ (\mathit{delete}\ a\ t) \end{array}$$

Unfortunately, if we test this property in its current form it will often fail to halt, because randomly generated values of recursively defined types such as trees are often infinite. To resolve this problem, narrowing-based testers usually limit the size of the solution by depth or by the number of constructors [15,16,19]. However, these metrics are often too simple to avoid backtracking.

To avoid bactracking in our example we need to limit only the depth of the tree and not the depth of elements (limiting the depth of elements limits their value but they have a minimum value dictated by the preceding elements.) Overlapping patterns allow us to limit the size with bespoke constraints. A function to calculate the *depth* of a tree can be defined as follows:

$$\begin{array}{l} \mathit{depthTree} :: \mathit{Tree} \to \mathit{Nat} \\ \mathit{depthTree}\ \mathit{Leaf} = \mathit{Zero} \\ \mathit{depthTree}\ (\mathit{Node}\ t1\ \_\ t2) = \mathit{Suc}\ (\mathit{max}\ (\mathit{depthTree}\ t1)\ (\mathit{depthTree}\ t2)) \end{array}$$

In turn, our property can then be refined to include a depth limit:

$$\mathit{propDelete}\ n\ a\ t = \mathit{ordered}\ t\ \wedge\ \mathit{depthTree}\ t \leqslant n\ \Longrightarrow\ \mathit{ordered}\ (\mathit{delete}\ a\ t)$$

The use of overlapping patterns is crucial in two ways for this kind of example. Firstly, in the new definition of *propDelete*, the sizing constraint relies on overlapping conjunction to be visible while the ordering constraint is being tested. Secondly, and more interestingly, the definition for *depthTree* relies on an overlapping version of the maximum function for natural numbers:

$$\begin{array}{l} \text{\{-\# OVERLAP } \mathit{max} \text{ \#-\}} \\ \mathit{max} :: \mathit{Nat} \to \mathit{Nat} \to \mathit{Nat} \\ \begin{array}{llcl} \mathit{max}\ \mathit{Zero} & \mathit{Zero} & = & \mathit{Zero} \\ \mathit{max}\ (\mathit{Suc}\ x) & y & = & \mathit{Suc}\ (\mathit{max}\ x\ (\mathit{pred}\ y)) \\ \mathit{max}\ x & (\mathit{Suc}\ y) & = & \mathit{Suc}\ (\mathit{max}\ (\mathit{pred}\ x)\ y) \end{array} \end{array}$$

The auxiliary function *pred* decrements a natural number, stopping at zero. A traditional maximum function would be left-biased and so the right branch of the tree could become arbitrarily large without triggering the size limit.

The constraint $\mathit{ordered}\ t \wedge \mathit{depthTree}\ t \leqslant n$ fails fast for any n. Although on initial testing overlapping patterns may not seem to give a performance benefit, analysis of the results show that without overlapping patterns the distribution is heavily skewed towards trivial small trees. If the shape of the tree is given, overlapping patterns offer a significant performance improvement.

5.2 Well-Typed Expressions

In this example we generate typed expressions for a simple language. We use this example to demonstrate a technique for building constraints that fail fast which combines well with the use of overlapping patterns. The language has addition, conditional expressions, natural numbers, and logical values:

$$\textbf{data}\ \mathit{Expr} = \mathit{Add}\ \mathit{Expr}\ \mathit{Expr} \mid \mathit{If}\ \mathit{Expr}\ \mathit{Expr}\ \mathit{Expr} \mid N\ \mathit{Nat} \mid B\ \mathit{Bool}$$

A useful property for this language states that for any well-typed expression up to a given depth, evaluating the expression will not produce an error:

$$\mathit{propEval}\ n\ e = \mathit{typed}\ e\ \wedge\ \mathit{depthExpr}\ e \leqslant n \implies \mathit{notError}\ (\mathit{eval}\ e)$$

We will focus on the *typed* condition. This condition has a simple definition in terms of a more general function *typeof* that attempts to determine the type of an expression, which may be either *Nat* or *Bool*, with the *Maybe* mechanism being used handle the possibility that an expression may be ill-typed:

```
data Type = Nat | Bool
typeof               :: Expr → Maybe Type
typeof (Add e e')    = case (typeof e, typeof e') of
                           (Just Nat, Just Nat) → Just Nat
                             _                  → Nothing
typeof (If e e' e'') = case (typeof e, typeof e', typeof e'') of
                           (Just Bool, Just t', Just t'') | t' ≡ t'' → Just t'
                           _                                         → Nothing
typeof (N _)         = Just Nat
typeof (B _)         = Just Bool
```

However, the function *typeof* has an inefficient narrowing semantics. For example, an expression of the form $\mathit{If}\ (\mathit{Add}\ u\ v)\ w\ x$ is ill-typed for any u, v, w, x, because it is already evident that the first argument is not a logical value, but a version of *typed* defined using the function *typeof* would not be able to deduce this until specific expressions had been filled in for the variables u and v. In other words, the *typed* condition does not fail fast.

To solve this problem we define an alternative constraint, $\mathit{hastype} :: \mathit{Expr} \to \mathit{Type} \to \mathit{Bool}$, in which the type of the expression is taken as an argument rather then returned as a result. In this manner, the type is refined during the narrowing process alongside the expression itself.

$$\begin{array}{lll}
\mathit{hastype}\ (\mathit{Add}\ e\ e')\ \mathit{Nat} & = & \mathit{hastype}\ e\ \mathit{Nat}\ \wedge\ \mathit{hastype}\ e'\ \mathit{Nat} \\
\mathit{hastype}\ (\mathit{If}\ e\ e'\ e'')\ t & = & \mathit{hastype}\ e\ \mathit{Bool}\ \wedge\ \mathit{hastype}\ e'\ t\ \wedge\ \mathit{hastype}\ e''\ t \\
\mathit{hastype}\ (N\ \_)\ \mathit{Nat} & = & \mathit{True} \\
\mathit{hastype}\ (B\ \_)\ \mathit{Bool} & = & \mathit{True} \\
\mathit{hastype}\ \_\ \_ & = & \mathit{False}
\end{array}$$

If we reconsider our example expression, $\mathit{If}\ (\mathit{Add}\ u\ v)\ w\ x$, then we can see our new typing constraint identifies this as being ill-typed:

$$\begin{array}{l}
\mathit{hastype}\ (\mathit{If}\ (\mathit{Add}\ u\ v)\ w\ x)\ t \\
= \mathit{hastype}\ (\mathit{Add}\ u\ v)\ \mathit{Bool}\ \wedge\ \mathit{hastype}\ w\ t\ \wedge\ \mathit{hastype}\ x\ t \\
= \mathit{False}\ \wedge\ \mathit{hastype}\ w\ t\ \wedge\ \mathit{hastype}\ x\ t \\
= \mathit{False}
\end{array}$$

The *hastype* program does not fail fast but satisfies a similar weaker condition: any partial value formed by evaluating the constraint with free arguments either directly fails when applied to the constraint, or there is a refinement of the value that succeeds. Using the *hastype* constraint, our original property concerning well-typed expressions up to a given depth can now be reformulated to include the type of the expression as an additional narrowing variable:

$$\mathit{propEval}\ n\ t\ e = \mathit{hastype}\ e\ t\ \wedge\ \mathit{depthExpr}\ t \leqslant n \implies \mathit{noError}\ (\mathit{eval}\ e)$$

Note that the typing variable has no effect on the validity of the property, and is only used to make the narrowing process more efficient. Without the use of overlapping conjunction, attempting to generate expressions that satisfy both of these constraints simultaneously would typically fail to terminate, whereas the above definition can generate such expressions in an efficient manner.

5.3 Other Examples

We have also considered two more sophisticated examples, in the form of red-black trees and simply-typed lambda expressions. In both cases we were able to create generators that are both practical in terms of efficiency and modular in terms of how they are writen. For example, in the case of red-black trees, the required constraint is obtained simply by combining separate constraints for red nodes, black nodes, the ordering of elements, and the depth of the tree. Our final red-black tree implementation is similar to that used the Reach system [15], except that the additional efficiency that arises from using overlapping patterns results in the consistent finding of a bug which this system struggles to find.

6 Related Work

The functional logic language Curry [11] implements needed narrowing, and supports the use of overlapping patterns in definitions. However, the semantics is different to our system. In particular, our overlapping patterns are deterministic, with evaluation proceeding along a single branch, whereas in Curry such patterns are non-determistic, with evaluation considering every matching branch.

The form of overlapping patterns that we use in our system is similar to that proposed by Cockx [5,6], who develops the idea in the context of dependent type theory and the Agda programming language. However, the intended purpose is different, with our aim being to improve the performance of property-based testing under a needed-narrowing semantics, and Cockx seeking to simply the development of proofs in a dependently-typed setting.

A number of narrowing-based testing tools use the notion of parallel conjunction. The idea originates in Lindblad's work on data generation [13] and Lazy Smallcheck [19], both of which use an enumerative style of testing. Subsequently, parallel conjunction has been used by Claessen et al. [3] to randomly generate data with a uniform distribution. Parallel conjunction is equivalent to overlapping conjunction, but whereas previous testing work using this operator has been more practically focused, we have given a precise narrowing semantics for a general form of overlapping definitions. The research of Claessen et al. is the most similar to our work, in that they also use a narrowing-style for random testing. However, their aim of producing a uniform distribution, via the use of Feat [7], makes backtracking hard to avoid for many problems.

7 Conclusion and Future Work

In this article we have motivated and formalised an extension to needed narrowing to allow overlapping patterns in definitions. We use the needed narrowing evaluation to generate data satisfying a constraint from a program specifying a constraint. Overlapping patterns allow us to achieve this in an efficient manner using composable constraints. Below we discuss some limitations of our approach, and suggest some possible directions for further work.

While overlapping patterns can improve the performance of property-based testing, the use of narrowing can lead to subtle performance issues, as we saw in Sect. 5 with the *typeof* constraint. To avoid performance issues close attention must be paid to possible sources of backtracking. Overlapping patterns help by making it easier to define constraints with limited backtracking, but they are no silver bullet, and further research is required to establish appropriate methodologies for identifying and limiting sources of backtracking.

The use of an overlapping conjunction operator is ubiquitous and performance critical in our examples, but it is not yet clear whether the more general notion of overlapping patterns is necessary. For example, in the case studies that we have considered the use of other overlapping functions, such as *max*, can be replaced by additional narrowing variables. However, the resulting function will usually be less general than its overlapping counterpart.

The interaction between other language features, narrowing and overlapping patterns is also an interesting topic for further work. Adding the capability to refine and narrow first and higher-order functions is one area for which the trie representation of partial functions used in the extended Lazy Smallcheck [18] offers a starting direction. We are also keen to explore how our approach can be extended to handle coinductive types and dependent types.

To demonstrate the practicality of our approach, we developed a prototype implementation in Haskell. It would be interesting to add overlapping patterns to a more established tool, either a property based testing library such Lazy Smallcheck [19], or a functional logic language such as Curry [11]. An alternative approach to enable practical use would be to extend the implementation to automatically translate a precondition into a QuickCheck generator [4].

References

1. Antoy, S., Echahed, R., Hanus, M.: A needed narrowing strategy. J. ACM **47**(4), 776–822 (2000)
2. Christiansen, J., Fischer, S.: EasyCheck — test data for free. In: Garrigue, J., Hermenegildo, M.V. (eds.) FLOPS 2008. LNCS, vol. 4989, pp. 322–336. Springer, Heidelberg (2008). doi:10.1007/978-3-540-78969-7_23
3. Claessen, K., Duregård, J., Pałka, M.H.: Generating constrained random data with uniform distribution. In: Codish, M., Sumii, E. (eds.) FLOPS 2014. LNCS, vol. 8475, pp. 18–34. Springer, Heidelberg (2014). doi:10.1007/978-3-319-07151-0_2
4. Claessen, K., Hughes, J.: QuickCheck: a lightweight tool for random testing of Haskell programs. In: International Conference on Functional Programming (2000)
5. Cockx, J.: Overlapping and Order-Independent Patterns in Type Theory. Ph.D. thesis, Master thesis, KU Leuven (2013)
6. Cockx, J., Piessens, F., Devriese, D.: Overlapping and order-independent patterns. In: Shao, Z. (ed.) ESOP 2014. LNCS, vol. 8410, pp. 87–106. Springer, Heidelberg (2014). doi:10.1007/978-3-642-54833-8_6
7. Duregård, J., Jansson, P., Wang, M.: Feat: functional enumeration of algebraic types. In: Haskell Symposium, vol. 47, no. 12 (2012)
8. Fowler, J.: The overlap check system for property-based testing (2016). https://github.com/JonFowler/OverlapCheck
9. Fowler, J., Huttom, G.: Towards a theory of reach. In: Serrano, M., Hage, J. (eds.) TFP 2015. LNCS, vol. 9547, pp. 22–39. Springer, Heidelberg (2016). doi:10.1007/978-3-319-39110-6_2
10. Hanus, M.: A unified computation model for functional and logic programming. In: Symposium on Principles of Programming Languages (1997)
11. Hanus, M.: Curry - An Integrated Functional Logic Language. Technical report (2016)
12. Hritcu, C., Hughes, J., Pierce, B.C., Spector-Zabusky, A., Vytiniotis, D., Azevedo de Amorim, A., Lampropoulos, L.: Testing noninterference, quickly. In: ACM SIGPLAN Notices, vol. 48 (2013)
13. Lindblad, F.: Property directed generation of first-order test data. In: Symposium on the Trends in Functional Programming (2007)
14. McBride, C., Paterson, R.: Applicative programming with effects. J. Funct. Program. **18**(1), 1–13 (2008)
15. Naylor, M., Runciman, C.: Finding inputs that reach a target expression. In: International Conference on Source Code Analysis and Manipulation (2007)
16. Naylor, M.F.: Hardware-Assisted and Target-Directed Evaluation of Functional Programs. Ph.D. thesis. University of York (2008)

17. Pałka, M.H., Claessen, K., Russo, A., Hughes, J.: Testing an optimising compiler by generating random lambda terms. In: International Workshop on Automation of Software Test (2011)
18. Reich, J.S., Naylor, M., Runciman, C.: Advances in lazy smallcheck. In: Hinze, R. (ed.) IFL 2012. LNCS, vol. 8241, pp. 53–70. Springer, Heidelberg (2013). doi:10.1007/978-3-642-41582-1_4
19. Runciman, C., Naylor, M., Lindblad, F.: SmallCheck and lazy smallcheck automatic exhaustive testing for small values. In: Symposium on Haskell (2008)