Under preparation for ICFP 2014

Auto in Agda

Programming proof search

Pepijn Kokke Wouter Swierstra

Universiteit Utrecht

pepijn.kokke@gmail.com w.s.swierstra@uu.nl

Abstract

As proofs in type theory become increasingly complex, there is a growing need to provide better proof automation. This paper shows how to implement a Prolog-style resolution procedure in the dependently typed programming language Agda. Connecting this resolution procedure to Agda's reflection mechanism provides a first-class proof search tactic for first-order Agda terms. Furthermore, the same mechanism may be used in tandem with Agda's instance arguments to implement type classes in the style of Haskell. As a result, writing proof automation tactics need not be different from writing any other program.

1. Introduction

Writing proof terms in type theory is hard and often tedious. Interactive proof assistants based on type theory, such as Agda [17] or Coq [9], take very different approaches to facilitating this process.

The Coq proof assistant has two distinct language fragments. Besides the programming language Gallina, there is a separate tactic language for writing and programming proof scripts. Together with several highly customizable tactics, the tactic language Ltac can provide powerful proof automation [7]. Having to introduce a separate tactic language, however, seems at odds with the spirit of type theory, where a single language is used for both proof and computation. Having a separate language for programming proofs has its drawbacks. Programmers need to learn another language to automate proofs. Debugging Ltac programs can be difficult and the resulting proof automation may be inefficient [5].

Agda does not have Coq's segregation of proof and programming language. Instead, programmers are encouraged to automate proofs by writing their own solvers [18]. In combination with Agda's reflection mechanism [1, 27], developers can write powerful automatic decision procedures [2]. Unfortunately, not all proofs are easily automated in this fashion. In that case, the user is forced to interact with the integrated development environment and manually construct a proof term step by step.

This paper tries to combine the best of both worlds by implementing a library for proof search *within* Agda itself. More specifically, this paper makes the several novel contributions.

- We show how to implement a Prolog interpreter in the style of Stutterheim et al. [24] in Agda (Section 3). Note that, in contrast to Agda, resolving a Prolog query need not terminate. Using coinduction, however, we can write an interpreter for Prolog that is *total*.
- Resolving a Prolog query results in a substitution that, when applied to the goal, produces a solution in the form of a term that can be derived from the given rules. We extend our interpreter to also produce a trace of the applied rules, which allow us to produce a proof term that is a witness to the validity of the resulting substitution (Section 4).
- We integrate this proof search algorithm with Agda's *reflection* mechanism (Section 5). This enables us to *quote* the type of a lemma we would like to prove, pass this term as the goal of our proof search algorithm, and finally, *unquote* the resulting proof term, thereby proving the desired lemma.
- Finally, we show how we can use our proof search together with Agda's *instance arguments* [10] to implement lightweight type classes in Agda (Section 6). This resolves one of the major restrictions of instance arguments: the lack of a recursive search procedure for their construction.

Although Agda already has built-in proof search functionality [13], we believe that exploring the first-class proof automation defined in this paper is still worthwhile. For the moment, however, we would like to defer discussing the various forms of proof automation until after we have presented our work (Section 7).

All the code described in this paper is freely available from GitHub.[1] It is important to emphasize that all our code is written in the safe fragment of Agda: it does not depend on any postulates or foreign functions; all definitions pass Agda's termination checker; and all metavariables are resolved.

2. Motivation

Before describing the *implementation* of our library, we will provide a brief introduction to Agda's reflection mechanism and illustrate how the proof automation described in this paper may be used.

Reflection in Agda

Agda has a *reflection* mechanism[2] for compile time metaprogramming in the style of Lisp [21], MetaML [26], and Template Haskell [22]. This reflection mechanisms make it possible to convert a program fragment into its corresponding abstract syntax tree

[1] See `https://github.com/pepijnkokke/AutoInAgda`.

[2] Note that Agda's reflection mechanism should not be confused with 'proof by reflection' – the technique of writing a verified decision procedure for some class of problems.

and vice versa. We will introduce Agda's reflection mechanism here with several short examples, based on the explanation in previous work [27]. A more complete overview can be found in the Agda release notes [1] and Van der Walt's thesis [28].

The type Term : Set is the central type provided by the reflection mechanism. It defines an abstract syntax tree for Agda terms. There are several language constructs for quoting and unquoting program fragments. The simplest example of the reflection mechanism is the quotation of a single term. In the definition of idTerm below, we quote the identity function on Boolean values.

```
idTerm : Term
idTerm = quoteTerm (λ (x : Bool) → x)
```

When evaluated, the idTerm yields the following value:

```
lam visible (var 0 [])
```

On the outermost level, the lam constructor produces a lambda abstraction. It has a single argument that is passed explicitly (as opposed to Agda's implicit arguments). The body of the lambda consists of the variable identified by the De Bruijn index 0, applied to an empty list of arguments.

More generally, the **quote** language construct allows users to access the internal representation of an identifier, a value of a built-in type Name. Users can subsequently request the type or definition of such names.

Dual to quotation, the **unquote** mechanism allows users to splice in a Term, replacing it with a its concrete syntax. For example, we could give a convoluted definition of the K combinator as follows:

```
const : ∀ {A B} → A → B → A
const = unquote (lam visible (lam visible (var 1 [])))
```

The language construct **unquote** is followed by a value of type Term. In this example, we manually construct a Term representing the K combinator and splice it in the definition of const.

The final piece of the reflection mechanism that we will use is the **quoteGoal** construct. The usage of **quoteGoal** is best illustrated with an example:

```
goalInHole : ℕ
goalInHole = quoteGoal g in { }0
```

In this example, the construct **quoteGoal** g binds the Term representing the *type* of the current goal, ℕ, to the variable g. When completing this definition by filling in the hole labeled 0, we may now refer to the variable g. This variable is bound to to def ℕ [], the Term representing the type ℕ.

Using proof automation

To illustrate the usage of our proof automation, we begin by defining a predicate Even on natural numbers as follows:

```
data Even : ℕ → Set where
  isEven0  : Even 0
  isEven+2 : ∀ {n} → Even n → Even (suc (suc n))
```

Next we may want to prove properties of this definition:

```
even+ : Even n → Even m → Even (n + m)
even+ isEven0        e2 = e2
even+ (isEven+2 e1) e2 = isEven+2 (even+ e1 e2)
```

Note that we omit universally quantified implicit arguments from the typeset version of this paper, in accordance with convention used by Haskell [20] and Idris [3].

As shown by Van der Walt and Swierstra [27], it is easy to decide the Even property for closed terms using proof by reflection. The interesting terms, however, are seldom closed. For instance, if we would like to use the even+ lemma in the proof below, we need to call it explicitly.

```
simple : Even n → Even (n + 2)
simple e = even+ e (isEven+2 isEven0)
```

Manually constructing explicit proof objects in this fashion is not easy. The proof is brittle. We cannot easily reuse it to prove similar statements such as Even (n + 4). If we need to reformulate our statement slightly, proving Even (2 + n) instead, we need to rewrite our proof. Proof automation can make propositions more robust against such changes.

Coq's proof search tactics, such as auto, can be customized with a *hint database*, a collection of related lemmas. In our example, auto would be able to prove the simple lemma, provided it the hint database contains at least the constructors of the Even data type and the even+ lemma. In contrast to the construction of explicit proof terms, changes to the theorem statement need not break the proof. This paper shows how to implement a similar tactic as an ordinary function in Agda.

Before we can use our auto function, we need to construct a hint database:

```
hints : HintDB
hints = hintdb
  (quote isEven0 :: quote isEven+2 :: quote even+ :: [])
```

To construct such a database, we **quote** any terms that we wish to include in it and pass them to the hintdb function. We defer any discussion about the hintdb function to Section 5. Note, however, that unlike Coq, the hint data base is a *first-class* value that can be manipulated, inspected, or passed as an argument to a function.

We now give an alternative proof of the simple lemma, using this hint database:

```
simple : Even n → Even (n + 2)
simple = quoteGoal g in unquote (auto 5 hints g)
```

The central ingredient is a *function* auto with the following type:

```
auto : (depth : ℕ) → HintDB → Term → Term
```

Given a maximum depth, hint database, and goal, it searches for a proof Term that witnesses our goal. If this term can be found, it is spliced back into our program using the **unquote** statement.

Of course, such invocations of the auto function may fail. What happens if no proof exists? For example, trying to prove Even n → Even (n + 3) in this style gives the following error:

```
Exception searchSpaceExhausted !=<
  Even .n -> Even (.n + 3) of type Set
```

When no proof can be found, the auto function generates a dummy term whose type explains the reason why the search has failed. In this example, the search space has been exhausted. Unquoting this term, then gives the type error message above. It is up to the programmer to fix this, either by providing a manual proof or diagnosing why no proof could be found.

The remainder of this paper will explain how this auto function is implemented.

3. Prolog in Agda

Let us set aside Agda's reflection mechanism for the moment. In this section, we will present a standalone Prolog interpreter. Subsequently, we will show how this can be combined with the reflection mechanism and suitably invoked in the definition of the auto function. The code in this section is contained in its own Agda module, parametrized by two sets.

```
module Prolog
  (TermName : Set) (RuleName : Set) where
```

Terms and Rules

The heart of our proof search implementation is the structurally recursive unification algorithm described by McBride [15]. Here the type of terms is indexed by the number of variables a given term may contain. Doing so enables the formulation of the unification algorithm by structural induction on the number of free variables. For this to work, we will use the following definition of terms:

```
data PrologTerm (n : ℕ) : Set where
  var : Fin n → PrologTerm n
  con : TermName → List (PrologTerm n)
      → PrologTerm n
```

In addition to a restricted set of variables, we will allow first-order constants encoded as a TermName with a list of arguments.

For instance, if we choose to instantiate the TermName with the following Arith data type, we can encode numbers and simple arithmetic expressions:

```
data Arith : Set where
  Suc  : Arith
  Zero : Arith
  Add  : Arith
```

The closed term corresponding to the number one could be written as follows:

```
One : PrologTerm 0
One = con Suc (con Zero [] :: [])
```

Similarly, we can use the var constructor to represent open terms, such as $x + 1$. We use the prefix operator # to convert from natural numbers to finite types:

```
AddOne : PrologTerm 1
AddOne = con Add (var (# 0) :: con One [] :: [])
```

Note that this representation of terms is untyped. There is no check that enforces addition is provided precisely two arguments. Although we could add further type information to this effect, this introduces additional overhead without adding safety to the proof automation presented in this paper. For the sake of simplicity, we have therefore chosen to work with this untyped definition.

We shall refrain from further discussion of the unification algorithm itself. Instead, we restrict ourselves to presenting the interface that we will use:

```
unify : (t₁ t₂ : PrologTerm m) → Maybe (∃ (Subst m))
```

Substitutions are indexed by two natural numbers m and n. A substitution of type Subst m n can be applied to a PrologTerm m to produce a value of type PrologTerm n. The unify function takes two terms t_1 and t_2 and tries to compute the most general unifier. As unification may fail, the result is wrapped in the Maybe monad. The number of variables in the terms resulting from the unifying substitution is not known *a priori*, hence this number is existentially quantified over. Agda allows us to write ∃ (Subst m) for the dependent pair containing a number n and a substitution Subst m n, inferring the type of the existentially quantified variable.

This unification function is defined using an accumulating parameter, representing an approximation of the final substitution. In what follows, we will use the following, more general, function:

```
unifyAcc : (t₁ t₂ : PrologTerm m) →
           ∃ (Subst m) → Maybe (∃ (Subst m))
```

Next we define Prolog rules as records containing a name and terms for its premises and conclusion:

```
record Rule (n : ℕ) : Set where
  field
    name       : RuleName
    conclusion : PrologTerm n
    premises   : List (PrologTerm n)
```

Again the data type is quantified over the number of variables used by its constituents. Note that variables are shared between the premises and conclusion.

Using our newly defined Rule we can give a simple definition of addition. In Prolog, this would be written as follows.

```
add(0, X, X).
add(suc(X), Y, suc(Z)) :- add(X, Y, Z).
```

Unfortunately, the named equivalents in our Agda implementation are a bit more verbose. Note that we have, for the sake of this example, instantiated the RuleName and TermName to String and Arith respectively.

```
AddBase : Rule 1
AddBase = record {
  name       = "AddBase"
  conclusion = con Add ( con Zero []
                       :: var (# 0)
                       :: var (# 0)
                       :: [])
  premises   = []
  }

AddStep : Rule 3
AddStep = record {
  name       = "AddStep"
  conclusion = con Add ( con Suc (var (# 0) :: [])
                       :: var (# 1)
                       :: con Suc (var (# 2) :: [])
                       :: [])
  premises   = con Add ( var (# 0)
                       :: var (# 1)
                       :: var (# 2)
                       :: [])
               :: []
  }
```

Before we can implement some form of proof search, we define a pair of auxiliary functions. During proof resolution, we will need to work with terms and rules containing a different number of variables. We will use the following pair of functions, inject and raise, to weaken bound variables, that is, map values of type Fin n to some larger finite type.

```
inject : ∀ {m} n → Fin m → Fin (m + n)
inject n zero    = zero
inject n (suc i) = suc (inject n i)

raise : ∀ m {n} → Fin n → Fin (m + n)
raise zero    i = i
raise (suc m) i = suc (raise m i)
```

We have tried to visualize the behaviour of inject and raise, embedding Fin 3 into Fin (3 + 1) in Figure 1. On the surface, the inject function appears to be the identity. When you make all the implicit arguments explicit, however, you will see that it sends the zero constructor in Fin m to the zero constructor of type Fin (m + n). Hence, the inject function maps Fin m into the *first* m elements of the type Fin (m + n). Dually, the raise function maps Fin n into

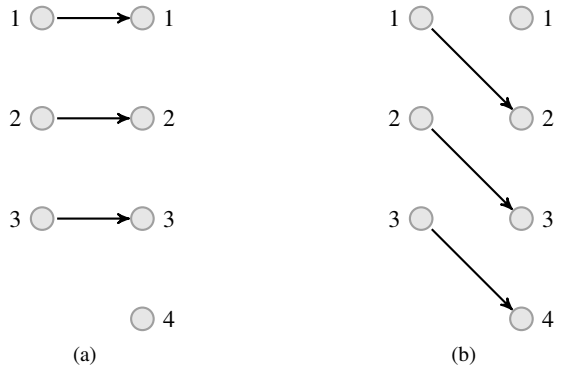

Figure 1. The graph of the inject function (a) and the raise function (b) embedding Fin 3 in Fin (3 + 1)

the *last* n elements of the type Fin (m + n) by repeatedly applying the suc constructor.

We can use these inject and raise to define similar functions that work on our Rule and Term data types, by mapping them over all the variables that they contain.

Proof search

Our implementation of proof search is split into two steps. In the first step we set up an higher-order representation of the search space, where we branch over some collection of undetermined rules at every step. In the second step we flatten this abstract representation to a first-order search tree.

The distinction between these two phases keeps the nitty-gritty details involved with unification and weakening used in the first phase separate from the actual proof search. By doing so, we can implement various search strategies, such as breadth-first search, depth-first search or an heuristic-driven algorithm, by simply traversing the final search tree in a different order.

Setting up the search space

We start by defining the following type synonym to distinguish goals from regular Prolog terms:

```
Goal : ℕ → Set
Goal n = PrologTerm n
```

Next we define the data type that we will use to model the abstract search space.

```
data SearchSpace (m : ℕ) : Set where
  fail : SearchSpace m
  retn : Subst (m + δ) n → SearchSpace m
  step : (∃ Rule → ∞ (SearchSpace m))
       → SearchSpace m
```

Ignoring the indices for the moment, the SearchSpace type has three constructors: fail, retn and step. The fail constructor is used to mark branches of the search space that fail, i.e., where the selected rule is not unifiable with the current goal. In the case of retn, we have found a substitution that resolves the goal we are trying to prove. In the step constructor, we have not yet resolved the goal, and instead have a choice of which Rule to apply. Note that we do not specify *which* rules may be used; only how the choice of *any* rule determines the remainder of the search. As a search need not terminate, the SearchSpace resulting from applying a rule are marked as coinductive.

Now let us turn our attention to the indices. The variable m denotes the number of variables in the goal; δ denotes the number of fresh variables necessary to apply a rule; and n denotes the number of variables remaining after we have resolved the goal. This naming will be used consistently in subsequent definitions.

We can now define a function resolve that will be in charge of building up a value of type SearchSpace from an initial goal:

```
resolve : Goal m → SearchSpace m
resolve {m} g = resolveAcc (just (m, nil)) [g]
```

The resolve function is once again defined by calling an auxiliary function defined using an accumulating parameter. It starts with an empty substitution and a list of goals that only contains the initial goal g. The resolveAcc function will attempt to resolve a list of sub-goals, accumulating a substitution along the way:

```
resolveAcc : Maybe (∃ (λ n → Subst (m + δ) n))
  → List (Goal (m + δ)) → SearchSpace m
resolveAcc (just (n, subst)) []              = retn subst
resolveAcc nothing _                         = fail
resolveAcc (just (n, subst)) (goal :: goals) = step next
```

If we have no remaining goals, we can use the retn constructor to return the substitution we have accumulated so far. If at any point, however, the conclusion of the chosen rule was not unifiable with the next open sub-goal – and thus the accumulating parameter has become nothing – the search will fail. The interesting case is the third one. If there are remaining goals to resolve, we recursively construct a new SearchSpace. To do so, we use the step constructor and branch over the choice of rule. The locally defined next function computes the remainder of the SearchSpace after trying to apply a given rule:

```
next : ∃ Rule → ∞ (SearchSpace m)
next (δ', rule) =
  ♯ resolveAcc mgu (prems' ++ goals')
  where
    mgu : Maybe (∃ (λ n → Subst (m + (δ + δ')) n))
    mgu = unifyAcc goal' concl' subst'
      where
        goal' : PrologTerm (m + (δ + δ'))
        goal' = injectTerm δ' goal

        subst' : ∃ (Subst (m + (δ + δ')))
        subst' = (n + δ', injectSubst δ' subst)

        concl' : PrologTerm (m + (δ + δ'))
        concl' = raiseTerm (m + δ) (conclusion rule)
    goals' : List (PrologTerm (m + (δ + δ')))
    goals' = injectTermList δ' goals

    prems' : List (PrologTerm (m + (δ + δ')))
    prems' = raiseTermList (m + δ) (premises rule)
```

For the moment, try to ignore the various calls to raise and inject. Given the rule that must be applied, the next function computes most general unifier of the conclusion of rule and our current goal. The resulting substitution is passed to resolveAcc, which continues the construction of the SearchSpace. The prefix operator ♯ creates a coinductive suspension, that may later be forced. The premises of the rule are added to the list of open goals that must be resolved. The apparent complexity of the next function comes from the careful treatment of variables.

First of all, note that we pass the substitution accumulated so far to unifyAcc. This ensures that the constraints on any variables occurring in the two terms being unified are taken into account.

Next, there is the problem of avoiding variable capture. We can only unify two terms that have the same type. Therefore we must

ensure that the goal, the rule's conclusion and its premises have the same number of variables. At the same time, the substitution we are accumulating should be kept in sync with the variables used in the initial goal. Furthermore, the variables mentioned in the rules are implicitly universally quantified. We need to instantiate them with fresh variables to avoid introducing unintended constraints. This is where inject and raise come in.

Recall that injecting a variable into a larger set would keep its value the same, whereas raise maps the variable into a 'fresh' portion of the set that was previously unused. Therefore we will always take care to inject our goal terms and our accumulating substitution, whereas we raise the terms in the applied rule. This ensures that the substitution and goals are kept in sync, whereas any variables mentioned in the rule are fresh.

Note the number of free variables in the chosen rule, δ', is added to the amount of space that had to be made for previous rule applications, δ. As a result, we need to raise by more and more as the proof search proceeds.

Constructing search trees

The second step in our proof search implementation is to flatten a SearchSpace to a first-order rose tree. We do this by branching once for every rule at every step constructor. The result of this transformation shall be expressed in terms of the following data type.

```
data SearchTree (A : Set) : Set where
  fail : SearchTree A
  retn : A → SearchTree A
  fork : List (∞ (SearchTree A)) → SearchTree A
```

Note that this SearchTree is finitely branching, but potentially infinitely deep. At every fork we may branch over some finite set of rules, but there is no guarantee that we can construct the entire SearchTree in finite time.

In our case, we will instantiate the type variable A with a tuple containing a substitution together with a trace that keeps track of all the applied rules. In order to keep the code readable, let us introduce the following type synonyms:

```
Rules      = List (∃ Rule)
Result m = ∃₂ (λ δ n → Subst (m + δ) n) × Rules
```

The existential quantifier $\exists_2$ hides both the number of fresh variables that we need to introduce, δ, and the number of variables in the terms produced by the final substitution, n.

The function that takes care of the transformation is almost trivial. For a given set of rules, we simply traverse the SearchSpace structure, where at every step we apply the continuation to every rule. For every rule, we create a new suspension using the ♯ operator; the recursive call to go forces the coinductive thunk using the ♭ operator. Since we also wish to maintain a trace of the rules that have been applied, we shall define this transformation using an auxiliary function with an accumulating parameter:

```
mkTree : Rules → SearchSpace m
         → SearchTree (Result m)
mkTree rules s = go s []
  where
  go : SearchSpace m → Rules → SearchTree (Result m)
  go fail      _   = fail
  go (retn s) acc = retn ((_, (_, s)), acc)
  go (step f) acc =
    fork (map (λ r → ♯ go (♭ f r) (acc ::ʳ r)) rules)
```

Note that we accumulate the trace of rules applied in the order in which they are applied: new rules are added to the end of the list with the snoc operator $::^{r}$.

In the implementation of mkTree, Agda's guardedness checker cannot tell that the call to map is size-preserving, and therefore the forcing of the coinductive suspension is safe. To show this definition is suitably guarded, we need to inline the definition of map and explicitly recurse over the list of rules rules.

After the transformation, we are left with a first-order tree structure, that we can traverse in search of solutions. For example, we can define a simple bounded depth-first traversal as follows:

```
dfs : (depth : ℕ) → SearchTree A → List A
dfs zero    _          = []
dfs (suc k) fail       = []
dfs (suc k) (retn x)   = return x
dfs (suc k) (fork xs)  = concatMap (λ x → dfs k (♭ x)) xs
```

It is fairly straightforward to define other traversal strategies, such as a breadth-first search. Similarly, we can also vary the rules used to construct the SearchTree. For example, you may want to define a function that constructs a 'linear' proof, where every rule is applied at most once. All these search strategies are simple variations of the solution presented here.

Putting all these pieces together, we can define a function searchToDepth, which implements proof search up to a given depth d, i.e. it constructs the SearchSpace, expands this into a SearchTree, and finally traverses the resulting tree in depth-first order up to depth d.

```
searchToDepth : ℕ → Rules → Goal m → List (Result m)
searchToDepth depth rules goal =
  dfs depth (mkTree rules (resolve goal))
```

Example

Using this implementation of proof search, together with the terms and rules defined above, we can compute, for instance, the sum $3 + 1$. First we define a query, corresponding to the Prolog query `add(3,1,x)`:

```
query : Term 1
query =
  con Add (inject 1 Three :: inject 1 One :: var (# 0) :: [])
```

Note that we must inject the terms Three and One, which are closed terms, in order to make it match the variable domain of our variable var (# 0).

Next, we use searchToDepth to search for a substitution. We use a function apply which applies a list of solutions to a goal term:

```
apply : List (Result m) → Goal m → List (Term 0)
```

The apply function applies the substitutions in the list of results to the goal, and discards any resulting terms that still contain variables. Although not useful in general, we can use it together with the searchToDepth function to illustrate the resolution of the previously defined query:

```
result : List (Term 0)
result = apply substs (var (# 0))
  where
    rules  = (1, AddBase) :: (3, AddStep) :: []
    substs = searchToDepth 5 rules query
```

Once we have this, we can show that the result of adding 1 and 3 is indeed 4.

```
test : result ≡ (Four :: [])
test = refl
```

4. Constructing proof trees

The Prolog interpreter described in the previous section returns a substitution. To use such an interpreter to produced proof terms, however, we need to do a bit more work.

Besides the resulting substitution, the Result type returned by the proof search process also contains a trace of the applied rules. In the following section we will discuss how to use this information to reconstruct a proof term. That is, we will construct a closed term of the following type:

```
data ProofTerm : Set where
  con : RuleName → List ProofTerm → ProofTerm
```

It is easy to compute the arity of every rule: we simply take the length of the list of premises. After making this observation, we can define a function to construct such a ProofTerm as a simple fold:

```
toProofTerms : Rules → List ProofTerm
toProofTerms = foldr next []
  where
    next : ∃ Rule → List ProofTerm → List ProofTerm
    next (δ, r) pfs with arity r ⩽? length pfs
    ... | no  r>p = []   -- should not occur
    ... | yes r⩽p =
      con (name r) (take (arity r) pfs) :: drop (arity r) pfs
```

The next function combines a list of proof terms, produced by recursive calls, and the single rule r that has just been applied. If the list contains enough elements, we construct a new ProofTerm node by applying the rule to the first (arity r) elements of the list. This new ProofTerm is the head of the list, replacing the children terms that previously formed the prefix of the list. Essentially, this is the 'unflattening' of a rose tree using the the arities of the individual nodes. Upon completion, toProofTerms should return a list with a single element: the proof term that witnesses the validity of the our derivation. The function, toProofTerm, returns this witness if it exists:

```
toProofTerm : Rules → Maybe ProofTerm
toProofTerm rs with toProofTerms rs
... | []        = nothing
... | p :: []   = just p
... | p :: _ :: _ = nothing
```

Of course, the toProofTerms function may fail if there are not enough elements in the list to fully apply a rule. When run on the result of our proof search functions, such as searchToDepth, however, we know that the list has the right length, even if this is not enforced by its type. While we could use a clever choice of indexed data type to show that the toProofTerms can be defined in a *total* fashion, there is little benefit in doing so. The proof search functions such as searchToDepth are already *partial* by their very nature. Adding further structure to the accumulated list of rules to guarantee totality will not change this.

5. Adding reflection

To complete the definition of our auto function, we still need to convert between Agda's built-in Term data type and the data type required by our unification and resolution algorithms, PrologTerm. This is an essential piece of plumbing, necessary to provide the desired proof automation. While not conceptually difficult, this does expose some of the limitations and design choices of the auto function.

The first thing we will need are concrete definitions for the TermName and RuleName data types, which were parameters to the development presented in the previous sections. It would be desirable to identify both types with Agda's Name type, but unfortunately Agda does not assign a name to the function space type operator, $\_\to\_$; nor does Agda assign names to locally bound variables. To address this, we define two new data types TermName and RuleName.
First, we define the TermName data type.

```
data TermName : Set where
  pname : (n : Name) → TermName
  pvar  : (i : ℕ) → TermName
  pimpl : TermName
```

The TermName data type has three constructors. The pname constructor embeds Agda's built-in Name in the TermName type. The pvar constructor describes locally bound variables, represented by their De Bruijn index. Note that the pvar constructor has nothing to do with PrologTerm's var constructor: it is not used to construct a Prolog variable, but rather to be able to refer to a local variable as a Prolog constant. Finally, pimpl explicitly represents the Agda function space.

We define the RuleName type in a similar fashion:

```
data RuleName : Set where
  rname : (n : Name) → RuleName
  rvar  : (i : ℕ) → RuleName
```

The rvar constructor is used to refer to Agda variables as rules. Its argument i corresponds to the variable's De Bruijn index – the value of i can be used directly as an argument to the var constructor of Agda's Term data type.

As we have seen in Section 2, the auto function may fail to find the desired proof. Furthermore, the conversion from Agda Term to PrologTerm may also fail for various reasons. To handle such errors, we will work in the Error monad defined below:

```
Error : (A : Set) → Set a
Error A = Either Message A
```

Upon failure, the auto function will produce an error message. The corresponding Message type simply enumerates the possible sources of failure:

```
data Message : Set where
  searchSpaceExhausted : Message
  unsupportedSyntax    : Message
  panic!               : Message
```

The meaning of each of these error messages will be explained as we encounter them in our implementation below.

Finally, we will need one more auxiliary function to manipulate bound variables. The match function takes two bound variables of types Fin m and Fin n and computes the corresponding variables in Fin (m ⊔ n) – where m ⊔ n denotes the maximum of m and n:

```
match : Fin m → Fin n → Fin (m ⊔ n) × Fin (m ⊔ n)
```

The implementation is reasonably straightforward. We compare the numbers n and m, and use the inject function to weaken the appropriate bound variable. It is straightforward to use this match function to define similar operations on two terms, matchTerms, or a term and a lists of terms, matchTermAndList.

Constructing terms

We now turn our attention to the conversion of an Agda Term to a PrologTerm. There are two problems that we must address.

First of all, the Agda Term type represents all (possibly higher-order) terms, whereas the PrologTerm type is necessarily first-order. We mitigate this problem, by allowing the conversion to fail, throwing an 'exception' with the message unsupportedSyntax.

Secondly, the Agda Term data type uses natural numbers to represent variables. The PrologTerm data type, on the other hand, represents variables using a finite type Fin n, for some n. To convert between these representations, the function keeps track of the current depth, i.e. the number of Π-types it has encountered, and uses this information to ensure a correct conversion. We sketch the definition of the main function below:

```
fromTerm : ℕ → Term → Error (∃ PrologTerm)
fromTerm d (var i [])      = pure (fromVar d i)
fromTerm d (con c args)    = fromDef c ⟨$⟩ fromArgs d args
fromTerm d (def f args)    = fromDef f ⟨$⟩ fromArgs d args
fromTerm d (pi (arg visible _ (el _ t₁)) (el _ t₂))
  with fromTerm d t₁ | fromTerm (suc d) t₂
... | left msg       | _          = left msg
... | _              | left msg   = left msg
... | right (n₁, p₁) | right (n₂, p₂)
  with matchTerms p₁ p₂
... | (p₁', p₂') = let term = con pimpl (p₁' :: p₂' :: [])
                   in right (n₁ ⊔ n₂, term)
fromTerm d (pi (arg _ _ _) (el _ t₂))
  = fromTerm (suc d) t₂
fromTerm _ _             = left unsupportedSyntax
```

We define special functions, fromVar and fromDef, to convert variables and constructors or defined terms respectively. The arguments to constructors or defined terms are processed using the fromArgs function defined below. The conversion of a pi node binding an explicit argument proceeds by converting the domain and codomain. If both conversions succeed, the resulting terms are matched and a PrologTerm is constructed using pimpl. Implicit arguments and instance arguments are ignored by this conversion function. Sorts, levels, or any other Agda feature mapped to the constructor unknown of type Term triggers a failure with the message unsupportedSyntax.

The fromArgs function converts a list of Term arguments to a list of Prolog terms, by stripping the arg constructor and recursively applying the fromTerm function. We only give its type signature here, as the definition is straightforward:

```
fromArgs : ℕ → List (Arg Term)
         → Error (∃ (List ∘ PrologTerm))
```

Next, the fromDef function constructs a first-order constant from an Agda Name and list of terms:

```
fromDef : Name → ∃ (λ n → List (PrologTerm n))
        → ∃ PrologTerm
fromDef f (n, ts) = n, con (pname f) ts
```

Lastly, the fromVar function converts a natural number, corresponding to a variable name in the Agda Term type, to the corresponding PrologTerm:

```
fromVar : ℕ → ℕ → ∃ PrologTerm
fromVar n i with compare n i
fromVar ⌊_⌋ _ | greater ⌊_⌋ k = (suc k, var (# k))
fromVar ⌊_⌋ _ | equal   ⌊_⌋   = (suc 0, var (# 0))
fromVar _ ⌊_⌋ | less    ⌊_⌋ k = (0, con (pvar k) [])
```

The fromVar function compares the number of binders that have been encountered with its argument De Bruijn index. If the variable is bound within the goal type, it computes a corresponding PrologTerm variable; if the variable is bound *outside* of the goal type, however, we compute a skolem constant.

To convert between an Agda Term and PrologTerm we simply call the fromTerm function, initializing the number of binders encountered to 0.

```
toPrologTerm : Term → Error (∃ PrologTerm)
toPrologTerm = fromTerm 0
```

Constructing rules

Our next goal is to construct rules. More specifically, we need to convert a list of quoted Names to a hint database of Prolog rules. To return to our example in Section 2, the definition of even+ had the following type:

```
even+ : Even n → Even m → Even (n + m)
```

We would like to construct a value of type Rule that expresses how even+ can be used. In Prolog, we might formulate the lemma above as the rule:

```
even(add(M,N)) :- even(M), even(N).
```

In our Agda implementation, we can define such a rule manually:

```
Even+ : Rule 2
Even+ = record {
  name       = rname even+
  conclusion = con (pname Even)
                   (con (pname _+_)
                     (var (# 0) :: var (# 1) :: [])
                   :: [])
  premises   = con (pname Even)
                   (var (# 0) :: [])
               :: con (pname Even)
                   (var (# 1) :: [])
               :: []
  }
```

In the coming subsection, we will show how to generate the above definition from the Name representing even+.

This generation of rules is done in two steps. First, we will convert a Name to its corresponding PrologTerm:

```
fromName : Name → Error (∃ PrologTerm)
fromName = toPrologTerm ∘ unel ∘ type
```

The type construct converts a Name to the Agda Term representing its type; the unel function discards any information about sorts; the toPrologTerm was defined previously.

In the next step, we process this PrologTerm. The splitTerm function, defined below, splits a PrologTerm at every top-most occurrence of the function symbol pimpl. Note that it would be possible to define this function directly on Agda's Term data type, but defining it on the PrologTerm data type is much cleaner as we may assume that any unsupported syntax has already been removed.

```
splitTerm : PrologTerm n
          → ∃ (λ k → Vec (PrologTerm n) (suc k))
splitTerm (con pimpl (t₁ :: t₂ :: [])) =
  Product.map suc (_::_ t₁) (splitTerm t₂)
splitTerm t = (0, t :: [])
```

Using all these auxiliary functions, it is straightforward to define the toRule function below that constructs a Rule from an Agda Name.

```
toRule : Name → Error (∃ Rule)
toRule name with fromName name
... | left msg          = left msg
... | right (n, t)      with splitTerm t
... | (k, ts)           with initLast ts
... | (prems, concl, _) =
  right (n, rule (rname name) concl (toList prems))
```

We convert a name to its corresponding PrologTerm, which is split into a vector of terms using splitTerm. The last element of this vector is the conclusion of the rule; the initial prefix constitutes the premises.

Constructing goals

Next, we turn our attention to converting a goal Term to a PrologTerm. While we could use the toPrologTerm function to do so, there are good reasons to explore other alternatives.

Consider the example given in Section 2. The goal Term we wish to prove is Even n → Even (n + 2). Calling toPrologTerm would convert this to a PrologTerm, where the function space has been replaced by the pimpl. What we would like to do, however, is to introduce arguments as available assumptions, such as Even n, and try to resolve the remaining goal Even (n + 2).

Fortunately, we can reuse many of the auxiliary functions we have defined already to achieve this. We convert a Term to the corresponding PrologTerm. Using the splitTerm and initLast function, we can get our hands on the list of arguments args and the desired return type goal.

```
toGoalRules : Term → Error (∃ PrologTerm × Rules)
toGoalRules t with fromTerm′ 0 t
... | left msg          = left msg
... | right (n , p)     with splitTerm p
... | (k , ts)          with initLast ts
... | (args , goal , _) = let rs = toRules 0 args
                          in right ((n , goal) , rs)
```

The only missing piece of the puzzle is a function, toRules, that converts a list of PrologTerms to a Rules list.

```
toRules : ℕ → Vec (PrologTerm n) k → Rules
toRules i []       = []
toRules i (t :: ts) = (n , rule (rvar i) t [])
                      :: toRules (suc i) ts
```

The toRules converts every PrologTerm in its argument list to a rule, using the argument's De Bruijn index as its rule name.

There is one last technical point. In the previous version of fromTerm, an Agda Term variable was mapped to a Prolog variable. When considering the goal type, however, we want to generate *only* skolem constants for our variables, rather than Prolog variables which may still be unified. To account for this difference, we use the fromTerm' function, a slight variation of the fromTerm function described previously. The only difference between fromTerm and fromTerm' is the treatment of variables.

Reification of proof terms

Now that we can compute Prolog terms, goals and rules from an Agda Term, we are ready to call the resolution mechanism described in Section 3. The only remaining problem is to convert the witness computed by our proof search back to an Agda Term, which can be unquoted to produce the desired proof. This is done by the fromProof function that traverses its argument ProofTerm; the only interesting question is how it handles the variables and names it encounters.

The ProofTerm may contain two kinds of variables: locally bound variables, rvar i, or variables storing an Agda Name, rname n. Each of these variables is treated differently in the fromProof function. Any references to locally bound variables are mapped to the var constructor of the Agda Term data type. These variables correspond to usage of arguments to the function being defined. As we know by construction that these arguments are mapped to rules without premises, the corresponding Agda variables do not need any further arguments.

```
fromProof : ProofTerm → Term
fromProof (con (rvar i) _) = var i []
fromProof (con (rname n) ps) with definition n
... | function _    = def n ∘ toArg ∘ fromProof ⟨$⟩ ps
... | constructor′  = con n ∘ toArg ∘ fromProof ⟨$⟩ ps
... | _             = unknown
  where
    toArg = arg visible relevant
```

If the rule being applied is constructed using an rname, we do disambiguate whether the rule name refers to a function or a constructor. The definition function, defined in Agda's reflection library, returns information about how the piece of abstract syntax to which its argument Name corresponds. For the sake of brevity, we restrict the definition here to only handle defined functions and data constructors. It is easy enough to extend with further branches for postulates, primitives, and so forth.

We will also need to wrap additional lambdas around the resulting term, due to the premises that were introduced by the toGoalRules function. To do so, we define the intros function that repeatedly wraps its argument term in a lambda:

```
intros : ℕ → Term → Term
intros zero t     = t
intros (suc k) t = lam visible (intros k t)
```

Hint databases

We allow users to provide hints, rules that may be used during resolution, in the form of a *hint database*. Such a hint database is simply a list of Prolog rules:

```
HintDB : Set
HintDB = List (∃ Rule)
```

We can 'assemble' hint databases from a list of names using the function hintdb:

```
hintdb : List Name → HintDB
hintdb = concatMap (fromError ∘ toRule)
  where
    fromError : Error A → List A
    fromError = fromEither (const []) [_]
```

Once again, we have simplified the presentation slightly. If the generation of a rule fails for whatever reason, no error is raised, and the rule is simply ignored. Our actual implementation requires an implicit proof argument that all the names in the argument list can be quoted successfully. If you define such proofs to compute the trivial unit record as evidence, Agda will fill them in automatically in every call to the hintdb function on constant arguments. This simple form of proof automation is pervasive in Agda programs [19, 25].

This is the simplest possible form of hint database. In principle, there is no reason not to define alternative versions that assign priorities to certain rules or limit the number of times a rule may be applied. The only function that would need to be adapted to handle such requirements is the mkTree function in Section 3.

Furthermore, note that a hint database is a simple list of rules. It is an entirely first-class entity. We can combine hints databases, filter certain rules from a hint database, or manipulate them in any way we wish.

Error messages

Lastly, we need to decide how to report error messages. Since we are going to return an Agda Term, we need to transform the Message type we saw previously into an Agda Term. When unquoted, this term will cause an type error, reporting the reason for

failure. To accomplish this, we introduce a dependent type, indexed by a Message:

```
data Exception : Message → Set where
  throw : (msg : Message) → Exception msg
```

The message passed as an argument to the throw constructor, will be recorded in the Exception's type, as we intended.

Next, we define a function to produce an Agda Term from a Message. We could construct such terms by hand, but it is easier to just use Agda's **quoteTerm** construct:

```
quoteError : Message → Term
quoteError (searchSpaceExhausted)
  = quoteTerm (throw searchSpaceExhausted)
quoteError (unsupportedSyntax)
  = quoteTerm (throw unsupportedSyntax)
quoteError (panic!)
  = quoteTerm (throw panic!)
```

Putting it all together

Finally, we can present the definition of the auto function used in the examples in Section 2:

```
auto : (depth : ℕ) → HintDB → Term → Term
auto depth hints goalType
  with toGoalRules goalType
... | left msg  = quoteError msg
... | right ((n, goal), args)
  with searchToDepth depth (args ++ hints) goal
... | []          = quoteError searchSpaceExhausted
... | (_, trace) :: _
  with toProofTerm trace
... | nothing  = quoteError panic!
... | just p   = intros (fromProof p)
```

The auto function converts the Term to a PrologTerm, the return type of the goal, and a list of arguments that may be used to construct this term. It then proceeds by calling the searchToDepth function with the argument hint database. If this proof search succeeds, the Result is converted to an Agda Term, a witness that the original goal is inhabited. There are three places that this function may fail: the conversion to a PrologTerm may fail, for instance because of unsupported syntax; the proof search may not find any result; or the final conversion to an Agda Term may fail unexpectedly. This last case should never be triggered, provided the toProofTerm function is only called on the result of our proof search.

6. Type classes

As a final application of our proof search algorithm, we show how it can be used to implement a *type classes* in the style of Haskell. Souzeau and Oury [23] have already shown how to use Coq's proof search mechanism to construct dictionaries. Using Agda's *instance arguments* [10] and the proof search presented in this paper, we mimic their results.

We begin by declaring our 'type class' as a record containing the desired function:

```
record Show (A : Set) : Set where
  field
    show : A → String
```

We can write instances for the Show 'class' by constructing explicit dictionary objects:

```
ShowBool : Show Bool
ShowBool = record {show = ...}

Showℕ : Show ℕ
Showℕ = record {show = ...}
```

Using instance arguments, we can now call our show function without having to pass the required dictionary explicitly:

```
open Show {{...}}

example : String
example = show 3
```

The instance argument mechanism infers that the show function is being called on a natural number, hence a dictionary of type Show ℕ is required. As there is only a single value of type Show ℕ, the required dictionary is inserted automatically. If we have multiple instance definitions for the same type or omit the required instance altogether, the Agda type checker would have given an error.

It is more interesting to consider parametrized instances, such as the Either instance given below.

```
ShowEither : Show A → Show B → Show (Either A B)
ShowEither ShowA ShowB = record {show = showE}
  where
    showE : Either A B → String
    showE (left x)  = "left " ++ show x
    showE (right y) = "right " ++ show y
```

Unfortunately, instance arguments do not do any recursive search for suitable instances. Trying to call show on a value of type Either ℕ Bool, for example, will not succeed: the Agda type checker will complain that it cannot find a suitable instance argument. At the moment, the only way to resolve this is to construct the required instances manually:

```
ShowEitherBoolℕ : Show (Either Bool ℕ)
ShowEitherBoolℕ = ShowEither ShowBool Showℕ
```

Writing out such dictionaries is rather tedious.

We can, however, use the auto function to construct the desired instance argument automatically. We start by putting the desired instances in a hint database:

```
ShowHints : HintDB
ShowHints = hintdb (quote ShowEither
                    :: quote ShowBool
                    :: quote Showℕ :: [])
```

The desired dictionary can now be assembled for us by calling the auto function:

```
example : String
example = show (left 4) ++ show (right true)
  where
    instance = quoteGoal g
               in unquote (auto 5 ShowHints g)
```

Note that the type of the locally bound instance record is inferred in this example. Using this type, the auto function assembles the desired dictionary. When show is called on different types, however, we may still need to provide the type signatures of the instances we desire. While deceptively simple, this example illustrates how *useful* it can be to have even a little automation.

7. Discussion

The auto function presented here is far from perfect. This section not only discusses its limitations, but compares it to existing proof automation techniques in interactive proof assistants.

Performance First of all, the performance of the auto function is terrible. Any proofs that require a depth greater than ten are intractable in practice. This is an immediate consequence of Agda's

poor compile-time evaluation. The current implementation is call-by-name and does no optimization whatsoever. While a mature evaluator is beyond the scope of this project, we believe that it is essential for Agda proofs to scale beyond toy examples. Simple optimizations, such as the erasure of the natural number indexes used in unification [4], would certainly help speed up the proof search.

Restrictions The auto function can only handle first-order terms. Even though higher-order unification is not decidable in general, we believe that it should be possible to adapt our algorithm to work on second-order goals. Furthermore, there are plenty of Agda features that are not supported or ignored by our quotation functions, such as universe polymorphism, instance arguments, and primitive functions.

Even for definitions that seem completely first-order, our auto function can fail unexpectedly. Consider the following definition of the product type, taken from Agda's standard library:

```
_×_ : (A B : Set) → Set
A × B = Σ A (λ _ → B)
```

Here a (non-dependent) pair is defined as a special case of the type Σ, representing dependent pairs. We can define the obvious Show instance for such pairs:

```
Show× : Show A → Show B → Show (A × B)
```

Somewhat surprisingly, trying to use this rule to create an instance of the Show 'class' fails. The **quoteGoal** construct always returns the goal in normal form, which exposes the higher-order nature of A × B. Converting the goal Show (A × (λ _ → B)) to a PrologTerm will raises the 'exception' unsupportedSyntax; the goal type contains a lambda which we cannot handle.

Furthermore, there are some limitations on the hints that may be stored in the hint database. At the moment, we construct every hint by quoting an Agda Name. Not all useful hints, however, have a such a Name, such as any variables locally bound in the context by pattern matching or function arguments. For example, the following call to the auto function fails to produce the desired proof:

```
simple : Even n → Even (n + 2)
simple e = quoteGoal g in unquote (auto 5 hints g)
```

The variable e, necessary to complete the proof is not part of the hint database. We hope that this could be easily fixed by providing a variation of the **quoteGoal** construct that returns both the term representing to the current goal and a list of the terms bound in the local context.

Refinement The auto function returns a complete proof term or fails entirely. This is not always desirable. We may want to return an incomplete proof, that still has open holes that the user must complete. This difficult with the current implementation of Agda's reflection mechanism: it cannot generate an incomplete Term.

In the future, it may be interesting to explore how to integrate proof automation, as described in this paper, better with Agda's IDE. If the call to auto were to generate the concrete syntax for a (possibly incomplete) proof term, this could be replaced with the current goal quite easily. An additional advantage of this approach would be that reloading the file does no longer needs to recompute the proof terms.

Metatheory The auto function is necessarily untyped because the interface of Agda's reflection mechanism is untyped. Defining a well-typed representation of dependent types in a dependently typed language remains an open problem, despite various efforts in this direction [6, 8, 11, 16]. If we had such a representation, however, we might be able to use the type information to prove that when the auto function succeeds, the resulting term has the correct type. As it stands, proving soundness of the auto function is non-trivial: we would need to define the typing rules of Agda's Term data type and prove that the Term we produce witnesses the validity of our goal Term. It may be slightly easier to ignore Agda's reflection mechanism and instead verify the metatheory of the Prolog interpreter: if a proof exists at some given depth, searchToDepth should find it; any Result returned by searchToDepth should correspond to a valid derivation.

Related work

There are several other interactive proof assistants, dependently typed programming languages, and alternative forms of proof automation in Agda. In the remainder of this section, we will briefly compare the approach taken in this paper to these existing systems.

Coq Coq has rich support for proof automation. The Ltac language and the many primitive, customizable tactics are extremely powerful [7]. Despite Coq's success, it is still worthwhile to explore better methods for proof automation. Recent work on Mtac [29] shows how to add a typed language for proof automation on top of Ltac. Furthermore, Ltac itself is not designed to be a general purpose programming language. It can be difficult to abstract over certain patterns and debugging proof automation is not easy. The programmable proof automation, written using reflection, presented here may not be as mature as Coq's Ltac language, but addresses these issues.

Idris The dependently typed programming language Idris also has a collection of tactics, inspired by some of the more simple Coq tactics, such as rewrite, intros, or exact. Each of these tactics is built-in and implemented as part of the Idris system. There is a small Haskell library for tactic writers to use that exposes common commands, such as unification, evaluation, or type checking. Furthermore, there are library functions to help handle the construction of proof terms, generation of fresh names, and splitting subgoals. This approach is reminiscent of the HOL family of theorem provers [12] or Coq's plug-in mechanism. An important drawback is that tactic writers need to write their tactics in a different language to the rest of their Idris code; furthermore, any changes to tactics requires a recompilation of the entire Idris system.

Agsy Agda already has a built-in 'auto' tactic that outperforms the auto function we have defined here [13]. It is nicely integrated with the IDE and does not require the users to provide an explicit hint database. It is, however, implemented in Haskell and shipped as part of the Agda system. As a result, users have very few opportunities for customization: there is limited control over which hints may (or may not) be used; there is no way to assign priorities to certain hints; and there is a single fixed search strategy. In contrast to the proof search presented here, where we have much more fine grained control over all these issues.

Closure

The proof automation presented in this paper is not as mature as some of these alternative systems. Yet we strongly believe that this style of proof automation is worth pursuing further.

The advantages of using reflection to program proof tactics should be clear: we do not need to learn a new programming language to write new tactics; we can use existing language technology to debug and test our tactics; and we can use all of Agda's expressive power in the design and implementation of our tactics. If a particular problem domain requires a different search strategy, this can be implemented by writing a new traversal over a SearchTree. Hint databases are first-class values. There is never any built-in magic; there are no compiler primitives beyond Agda's reflection mechanism.

The central philosophy of Martin-Löf type theory is that the construction of programs and proofs is the same activity. Any external language for proof automation renounces this philosophy. This paper demonstrates that proof automation is not inherently at odds with the philosophy of type theory. Paraphrasing Martin-Löf [14], it no longer seems possible to distinguish the discipline of *programming* from the *construction* of mathematics.

Acknowledgements We would like to thank the Software Technology Reading Club at the Universiteit Utrecht for their helpful feedback.

References

[1] Agda developers. Agda release notes, regarding reflection. The Agda Wiki: `http://wiki.portal.chalmers.se/agda/agda.php?n=Main.Version-2-2-8` and `http://wiki.portal.chalmers.se/agda/agda.php?n=Main.Version-2-3-0`, 2013. [Online; accessed 9-Feb-2013].

[2] Guillaume Allais. Proof automatization using reflection (implementations in Agda). MSc Intern report, University of Nottingham, 2010.

[3] Edwin Brady. Idris, a general-purpose dependently typed programming language: Design and implementation. *Journal of Functional Programming*, 23:552–593, 9 2013. doi: 10.1017/S095679681300018X.

[4] Edwin Brady, Conor McBride, and James McKinna. Inductive families need not store their indices. In *Types for Proofs and Programs*, pages 115–129. Springer, 2004.

[5] Thomas Braibant. Emancipate yourself from Ltac. Available online `http://gallium.inria.fr/blog/your-first-coq-plugin/`, 2012.

[6] James Chapman. *Type checking and normalisation*. PhD thesis, University of Nottingham, 2009.

[7] Adam Chlipala. *Certified programming with dependent types*. MIT Press, 2013.

[8] Nils Anders Danielsson. A formalisation of a dependently typed language as an inductive-recursive family. In *Types for Proofs and Programs*, volume 4502 of *Lecture Notes in Computer Science*. Spring Verlag, 2006.

[9] The Coq development team. The Coq proof assistant reference manual. Logical Project, 2004.

[10] Dominique Devriese and Frank Piessens. On the bright side of type classes: Instance arguments in Agda. In *Proceedings of the 16th ACM SIGPLAN International Conference on Functional Programming*, ICFP '11, pages 143–155. ACM, 2011. doi: 10.1145/2034773.2034796.

[11] Dominique Devriese and Frank Piessens. Typed syntactic meta-programming. In *Proceedings of the 2013 ACM SIGPLAN International Conference on Functional Programming (ICFP 2013)*. ACM, September 2013. doi: 10.1145/2500365.2500575.

[12] M.J.C Gordon and T.F. Melham. *Introduction to HOL: a theorem proving environment for higher order logic*. Cambridge University Press, 1993.

[13] Fredrik Lindblad and Marcin Benke. A tool for automated theorem proving in Agda. In *Proceedings of the 2004 International Conference on Types for Proofs and Programs*, pages 154–169. Springer-Verlag, 2004.

[14] Per Martin-Löf. Constructive mathematics and computer programming. In *Proceedings of a discussion meeting of the Royal Society of London on Mathematical logic and programming languages*, pages 167–184. Prentice-Hall, Inc., 1985.

[15] Conor McBride. First-order unification by structural recursion. *Journal of Functional Programming*, 13:1061–1075, 11 2003. ISSN 1469-7653. doi: 10.1017/S0956796803004957. URL `http://journals.cambridge.org/article_S0956796803004957`.

[16] Conor McBride. Outrageous but meaningful coincidences: Dependent type-safe syntax and evaluation. In *Proceedings of the 6th ACM SIGPLAN Workshop on Generic Programming*, WGP '10, pages 1–12, New York, NY, USA, 2010. ACM. doi: 10.1145/1863495.1863497.

[17] Ulf Norell. *Towards a practical programming language based on dependent type theory*. PhD thesis, Department of Computer Science and Engineering, Chalmers University of Technology, 2007.

[18] Ulf Norell. Playing with Agda. Invited talk at TPHOLS, 2009.

[19] Nicolas Oury and Wouter Swierstra. The power of pi. In *Proceedings of the 13th ACM SIGPLAN International Conference on Functional Programming*, ICFP '08, pages 39–50, 2008. doi: 10.1145/1411204.1411213.

[20] Simon Peyton Jones, editor. *Haskell 98 language and libraries: the revised report*. Cambridge University Press, 2003.

[21] Kent M. Pitman. Special forms in Lisp. In *Proceedings of the 1980 ACM conference on LISP and functional programming*, pages 179–187. ACM, 1980.

[22] Tim Sheard and Simon Peyton Jones. Template meta-programming for Haskell. In *Proceedings of the 2002 ACM SIGPLAN workshop on Haskell*, pages 1–16, 2002. doi: 10.1145/581690.581691.

[23] Matthieu Sozeau and Nicolas Oury. First-class type classes. In *Theorem Proving in Higher Order Logics*, pages 278–293. Springer, 2008.

[24] Jurriën Stutterheim, Wouter Swierstra, and Doaitse Swierstra. Forty hours of declarative programming: Teaching Prolog at the Junior College Utrecht. In *Proceedings First International Workshop on Trends in Functional Programming in Education, University of St. Andrews, Scotland, UK, 11th June 2012*, volume 106 of *Electronic Proceedings in Theoretical Computer Science*, pages 50–62, 2013.

[25] Wouter Swierstra. More dependent types for distributed arrays. *Higher-order and symbolic computation*, 23(4):489–506, 2010.

[26] Walid Taha and Tim Sheard. Multi-stage programming with explicit annotations. In *Proceedings of the 1997 ACM SIGPLAN symposium on Partial evaluation and semantics-based program manipulation*, PEPM '97, 1997. doi: 10.1145/258993.259019.

[27] Paul van der Walt and Wouter Swierstra. Engineering proof by reflection in Agda. In Ralf Hinze, editor, *Implementation and Application of Functional Languages*, Lecture Notes in Computer Science, pages 157–173. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-41581-4. doi: 10.1007/978-3-642-41582-1_10.

[28] Paul van der Walt. Reflection in Agda. Master's thesis, Department of Computer Science, Utrecht University, Utrecht, The Netherlands, 2012. Available online, `http://igitur-archive.library.uu.nl/student-theses/2012-1030-200720/UUindex.html`.

[29] Beta Ziliani, Derek Dreyer, Neelakantan R. Krishnaswami, Aleksandar Nanevski, and Viktor Vafeiadis. Mtac: A monad for typed tactic programming in coq. In *Proceedings of the 18th ACM SIGPLAN International Conference on Functional Programming*, ICFP '13, pages 87–100, 2013. doi: 10.1145/2500365.2500579.