FUNCTIONAL PEARL

A well-known representation of monoids and its application to the function ‘vector reverse’

WOUTER SWIERSTRA

Utrecht University, Utrecht, Netherlands
(e-mail: w.s.swierstra@uu.nl)

Abstract

Vectors—or length-indexed lists—are classic example of a dependent type. Yet, most tutorials stay clear of any function on vectors whose definition requires non-trivial equalities between natural numbers to type check. This pearl shows how to write functions, such as vector reverse, that rely on monoidal equalities to be type correct without having to write any additional proofs. These techniques can be applied to many other functions over types indexed by a monoid, written using an accumulating parameter, and even be used to decide arbitrary equalities over monoids ‘for free.’

1 Introduction

Many tutorials on programming with dependent types define the type of length-indexed lists, also known as vectors. Using a language such as Agda (Norell, 2007), we can write:

```agda
data Vec (a : Set) : Nat → Set where
  Nil : Vec a Zero
  Cons : a → Vec a n → Vec a (Succ n)
```

Many familiar functions on lists can be readily adapted to work on vectors, such as concatenation:

```agda
vappend : Vec a n → Vec a m → Vec a (n + m)
vappend Nil ys = ys
vappend (Cons x xs) ys = Cons x (vappend xs ys)
```

Here, the definitions of both addition and concatenation proceed by induction on the first argument; this coincidence allows concatenation to type check, without having to write explicit proofs involving natural numbers. Programming languages such as Agda will happily expand definitions while type checking—but any non-trivial equality between natural numbers may require further manual proofs.

However, not all functions on lists are quite so easy to adapt to vectors. How should we reverse a vector? There is an obvious—but inefficient—definition.
The `snoc` function traverses a vector, adding a new element at its end. Repeatedly traversing the intermediate results constructed during reversal yields a function that is quadratic in the input vector's length. Fortunately, there is a well-known solution using an accumulating parameter, often attributed to Hughes (1986). If we try to implement this version of the reverse function on vectors, we get stuck quickly:

```haskell
slowReverse : Vec a n → Vec a n
slowReverse Nil = Nil
slowReverse (Cons x xs) y = Cons x (slowReverse xs y)
```

We can remedy this by defining a variation of addition that mimics the accumulating recursion of the `revAcc` function:

```haskell
addAcc : Nat → Nat → Nat
addAcc Zero m = m
addAcc (Succ n) m = addAcc n (Succ m)
```

Using this accumulating addition, we can define the accumulating vector reversal function directly:

```haskell
revAcc : Vec a n → Vec a m → Vec a (n + m)
revAcc Nil ys = ys
revAcc (Cons x xs) ys = revAcc xs (Cons x ys)
```

When we try to use the `revAcc` function to define the top-level `vreverse` function; however, we run into a new problem:

```haskell
vreverse : Vec a n → Vec a n
vreverse xs = revAcc Nil xs
```

```
Goal: Vec a (addAcc n Zero)
Have: Vec a (addAcc n Zero)
```
Again, the desired definition does not type check: \( \text{revAcc} \times \text{Nil} \) produces a vector of length \( \text{addAcc} \ n \ Zero \), whereas a vector of length \( n \) is required. We could try another variation of addition that pattern matches on its second argument, but this will break the first clause of the \( \text{revAcc} \) function. To complete the definition of vector reverse, we can use an explicit proof to coerce the right-hand side, \( \text{revAcc} \times \text{Nil} \), to have the desired length. To do so, we define an auxiliary function that coerces a vector of length \( n \) into a vector of length \( m \), provided that we can prove that \( n \) and \( m \) are equal:

\[
\text{coerce-length} : n \equiv m \rightarrow \text{Vec} \ a \ n \rightarrow \text{Vec} \ a \ m
\]

Using this function, we can now complete the definition of \( \text{vreverse} \) as follows:

\[
\text{vreverse} : (n : \text{Nat}) \rightarrow \text{Vec} \ a \ n \rightarrow \text{Vec} \ a \ n
\]

\[
\text{vreverse} \ n \ \text{xs} = \text{coerce-length} \ \text{proof} \ (\text{revAcc} \ \text{xs} \ \text{Nil})
\]

where

\[
\text{proof} : \text{addAcc} \ n \ Zero \ \equiv \ n
\]

We have omitted the definition of \( \text{proof} \)—but we will return to this point in the final section.

This definition of \( \text{vreverse} \) is certainly correct—but the additional coercion will clutter any subsequent lemmas that refer to this definition. To prove any property of \( \text{vreverse} \) will require pattern matching on the \( \text{proof} \) to reduce—rather than reasoning by induction on the vector directly.

Unfortunately, it is not at all obvious how to complete this definition without such proofs. We seem to have reached an impasse: how can we possibly define addition in such a way that \( \text{Zero} \) is both a left and a right identity?

## 2 Monoids and endofunctions

The solution can also be found in Hughes’s article, that explores using an alternative representation of lists known as *difference lists*. These difference lists identify a list with the partial application of the append function. Rather than work with natural numbers directly, we choose an alternative representation of natural numbers that immediately satisfies the desired monoidal equalities, representing a number as the partial application of addition.

\[
\text{DNat} : \text{Set}
\]

\[
\text{DNat} = \text{Nat} \rightarrow \text{Nat}
\]

In what follows, we will refer to these functions \( \text{Nat} \rightarrow \text{Nat} \) as *difference naturals*. We can readily define the following conversions between natural numbers and difference naturals:

\[
[\_] : \text{Nat} \rightarrow \text{DNat}
\]

\[
[n] = \lambda m \rightarrow m + n
\]

\[
\text{reify} : \text{DNat} \rightarrow \text{Nat}
\]

\[
\text{reify} \ m = m \ \text{Zero}
\]
We have some choice of how to define the reify function. As addition is defined by induction on the first argument, we define reify by applying Zero to its argument. This choice ensures that the desired ‘return trip’ property between our two representations of naturals holds definitionally:

\[ \text{reify-correct : } \forall n \rightarrow \text{reify } n \equiv n \]
\[ \text{reify-correct } n = \text{refl} \]

Note that we have chosen to use the type \( \text{Nat} \rightarrow \text{Nat} \) here, but there is nothing specific about natural numbers in these definitions. These definitions can be readily adapted to work for any monoid—an observation we will explore further in later sections. Indeed, this is an instance of Cayley’s theorem for groups (Armstrong, 1988, Chapter 8), or the Yoneda embedding more generally (Boisseau & Gibbons, 2018; Awodey, 2010), that establishes an equivalence between the elements of a group and the partial application of the group’s multiplication operation.

While this fixes the conversion between numbers and their representation using functions, we still need to define the monoidal operations on this representation. Just as for difference lists, the zero and addition operation correspond to the identity function and function composition, respectively:

\[ \text{zero : DNat} \]
\[ \text{zero } = \lambda x \rightarrow x \]
\[ \text{_⊕_ : DNat} \rightarrow \text{DNat} \rightarrow \text{DNat} \]
\[ n \oplus m = \lambda x \rightarrow m(nx) \]

Somewhat surprisingly, all three monoid laws hold definitionally using this functional representation of natural numbers:

\[ \text{zero-right : } \forall x \rightarrow \text{reify } x \equiv \text{reify } (x \oplus \text{zero}) \]
\[ \text{zero-right } = \lambda x \rightarrow \text{refl} \]
\[ \text{zero-left : } \forall x \rightarrow \text{reify } x \equiv \text{reify } (\text{zero} \oplus x) \]
\[ \text{zero-left } = \lambda x \rightarrow \text{refl} \]
\[ \oplus\text{-assoc : } \forall xyz \rightarrow \text{reify } ((x \oplus (y \oplus z)) \equiv \text{reify } ((x \oplus y) \oplus z) \]
\[ \oplus\text{-assoc } = \lambda xyz \rightarrow \text{refl} \]

As adding zero corresponds to applying the identity function and addition is mapped to function composition, the proof of these equalities follows immediately after evaluating the left- and right-hand sides of the equality.

To convince ourselves that our definition of addition is correct, we should also prove the following lemma, stating that addition on ‘difference naturals’ and natural numbers agree for all inputs:

\[ \oplus\text{-correct : } \forall n m \rightarrow n + m \equiv \text{reify } ([n] \oplus [m]) \]

After simplifying both sides of the equation, the proof boils down to the associativity of addition. Proving this requires a simple inductive argument, and does not hold definitionally. The reverse function we will construct, however, does not rely on this property.
3 Revisiting reverse

Before we try to redefine our accumulating reverse function, we need one additional auxiliary definition. Besides zero and the $\oplus$ operation on these naturals—we will need a successor function to account for new elements added to the accumulating parameter. Given that Cons constructs a vector of length Succ $n$ for some $n$, our first attempt at defining the successor operation on difference naturals becomes

$$
succ : \text{DNat} \to \text{DNat}
$$

$$
succ m = \lambda n \to \text{Succ} (m \ n)
$$

With this definition in place, we can now fix the type of our accumulating reverse function:

$$
\text{revAcc} : (m : \text{DNat}) \to \text{Vec a n} \to \text{Vec a (reify m)} \to \text{Vec a (m n)}
$$

As we want to define $\text{revAcc}$ by induction over its first argument vector, we choose that vector to have length $n$, for some natural number $n$. Attempting to pattern match on a vector of length $\text{reify m}$ creates unification problems that Agda cannot resolve: it cannot decide which constructors of the Vec datatype can be used to construct a vector of length $\text{reify m}$. As a result, we index the first argument vector by a Nat; the second argument vector has length $\text{reify m}$, for some $m : \text{DNat}$. The length of the vector returned by $\text{revAcc}$ is the sum of the input lengths—$\text{reify ( n \oplus m)}$—which simplifies to $m \ n$. We can now attempt to complete the definition as follows:

$$
\text{revAcc} m \ \text{Nil} \ \ y s = y s
$$

$$
\text{revAcc} m \ (\text{Cons x xs}) \ y s = \{ \text{revAcc (succ m) xs (Cons x ys)}\}_2
$$

Goal: Vec a (m (Succ n))

Have: Vec a (Succ (m n))

Unfortunately, the desired definition does not type check. The right-hand side produces a vector of the wrong length. To understand why, compare the types of the goal and expression we have produced. Using this definition of $\text{succ}$ creates an outermost successor constructor, hence we cannot produce a vector of the right type.

Let us not give up just yet. We can still redefine our successor operation as follows:

$$
succ : \text{DNat} \to \text{DNat}
$$

$$
succ m = \lambda n \to m \ (\text{Succ} \ n)
$$

This definition should avoid the problem that arises from the outermost $\text{Succ}$ constructor that we observed previously. If we now attempt to complete the definition of $\text{revAcc}$, we encounter a different problem:

$$
\text{revAcc} : (m : \text{DNat}) \to \text{Vec a n} \to \text{Vec a (reify m)} \to \text{Vec a (m n)}
$$

$$
\text{revAcc} m \ \text{Nil} \ y s = y s
$$

$$
\text{revAcc} m \ (\text{Cons x xs}) \ y s = \text{revAcc (succ m) xs (Cons x ys)}_3
$$

Goal: Vec a (m (Succ Zero))

Have: Vec a (Succ (m Zero))
Once again, the problem lies in the case for Cons. We would like to make a tail recursive call on the remaining list \( xs \), passing \( \text{succ} \ m \) as the length of the accumulating parameter. This call now type checks—as the desired length \( m \ (\text{Succ} \ n) \) and computed length \( (\text{succ} \ m) \ n \) coincide. The problem, however, lies in constructing the accumulating parameter to pass to the recursive call. The recursive call requires a vector of length \( m \ (\text{Succ} \ \text{Zero}) \), whereas the Cons constructor used here returns a vector of length \( \text{Succ} \ (m \ \text{Zero}) \).

We might try to define an auxiliary function, analogous to the Cons constructor:

\[
\text{cons} : (m : \text{DNat}) \to a \to \text{Vec} a \ (\text{reify} \ m) \to \text{Vec} a \ (\text{reify} \ (\text{succ} \ m))
\]

If we try to define this function directly, however, we get stuck immediately. The type requires that we produce a vector of length, \( m \ (\text{Succ} \ \text{Zero}) \). Without knowing anything further about \( m \), we cannot even decide if the vector should be empty or not. Fortunately, we do know more about the difference natural \( m \) in the definition of \( \text{revAcc} \). Initially, our accumulator will be empty—hence \( m \) will be the identity function. In each iteration of \( \text{revAcc} \), we will compose \( m \) with an additional succ until our input vector is empty.

If we assume we are provided with a \( \text{cons} \) function of the right type, we can complete the definition of vector reverse as expected:

\[
\begin{align*}
\text{revAcc} & : \forall \ m \to (\forall \{k\} \to a \to \text{Vec} a \ (m \ k) \to \text{Vec} a \ ((\text{succ} \ m) \ k)) \to \\
& \quad \text{Vec} a \ n \to \text{Vec} a \ ((\text{reify} \ m) \ n) \to \text{Vec} a \ (\text{reify} \ (\text{succ} \ m)) \\
\text{revAcc} \ m \ \text{cons} \ \text{Nil} \quad & \text{acc} = \text{acc} \\
\text{revAcc} \ m \ \text{cons} \ (\text{Cons} \ x \ xs) \ \text{acc} & = \text{revAcc} \ (\text{succ} \ m) \ \text{cons} \ xs \ (\text{cons} \ x \ \text{acc})
\end{align*}
\]

This definition closely follows our previous attempt. Rather than applying the Cons constructor, this definition uses the argument \( \text{cons} \) function to extend the accumulating parameter. Here, the \( \text{cons} \) function is assumed to commute with the successor constructor and an arbitrary difference natural \( m \). In the recursive call, the first argument vector has length \( n \), whereas the second has length \( \text{reify} \ (\text{succ} \ m) \). As the cons parameter extends a vector of length \( m \ k \) for any \( k \), we use it in our recursive call, silently incrementing the implicit argument passed to \( \text{cons} \). In this way, we count down from \( n \), the length of the first vector, whilst incrementing the difference natural \( m \) in each recursive call.

But how are we ever going to call this function? We have already seen that it is impossible to define the \( \text{cons} \) function in general. Yet we do not need to define \( \text{cons} \) for arbitrary values of \( m \)—we only ever call the \( \text{revAcc} \) function from the \( \text{vreverse} \) function with an accumulating parameter that is initially empty. As a result, we only need to concern ourselves with the case that \( m \) is zero—or rather, the identity function. When \( m \) is the identity function, the type of the \( \text{cons} \) function required simply becomes:

\[
\forall \{k\} \to a \to \text{Vec} a \ k \to \text{Vec} a \ (\text{Succ} \ k)
\]

Hence, it suffices to pass the Cons constructor to \( \text{revAcc} \) after all:

\[
\begin{align*}
\text{vreverse} & : \text{Vec} a \ n \to \text{Vec} a \ n \\
\text{vreverse} \ xs & = \text{revAcc} \ \text{Zero} \ \text{Cons} \ xs \ \text{Nil}
\end{align*}
\]

This completes the first proof-free reconstruction of vector reverse.
Correctness

Reasoning about this definition of vector reverse, however, is a rather subtle affair. Suppose we want to prove \( v_{\text{reverse}} \) is equal to the quadratic \( \text{slowReverse} \) function from the introduction:

\[
v_{\text{reverse}} \text{-correct} : (xs : \text{Vec a n}) \to v_{\text{reverse}} \, xs \equiv \text{slowReverse} \, xs
\]

If we try to prove this using induction on \( xs \) directly, we quickly get stuck in the case for non-empty vectors: we cannot use our induction hypothesis, as the definition of \( v_{\text{reverse}} \) assumes that the accumulator is the empty vector. To fix this, we need to formulate and prove a more general statement about calls to \( \text{revAcc} \) with an arbitrary accumulator, corresponding to a lemma of the following form:

\[
\text{revAcc} \, m \, \text{cons} \, xs \, ys = v_{\text{append}} (\text{slowReverse} \, xs) \, ys
\]

Here the \( v_{\text{append}} \) function refers to the append on vectors, defined in the introduction. There is a problem, however, formulating such a lemma: the \( v_{\text{append}} \) function uses the usual addition operation in its type, rather than the ‘difference addition’ used by \( \text{revAcc} \). As a result, the vectors on both sides of the equality sign have different types. To fix this, we need the following variant of \( v_{\text{append}} \), where the length of the second vector is represented by a difference natural:

\[
d_{\text{append}} : \forall \, m \to (\text{cons} : \forall \, \{k\} \to a \to \text{Vec a (m k)} \to \text{Vec a ((succ m) k)}) \to
\text{Vec a n} \to \text{Vec a (reify m)} \to \text{Vec a (m n)}
\]

\[
d_{\text{append}} \, m \, \text{cons} \, \text{Nil} \, ys = ys
\]

\[
d_{\text{append}} \, m \, \text{cons} \, (\text{Cons} \, x \, xs) \, ys = \text{cons} \, x \, (d_{\text{append}} \, m \, \text{cons} \, xs \, ys)
\]

Using this ‘difference append’ operation, we can now formulate and prove the following correctness property, stating that \( \text{revAcc} \) pushes all the elements of \( xs \) onto the accumulating parameter \( ys \):

\[
\text{revAcc-correct} : (m : \text{Nat} \to \text{Nat}) \, (xs : \text{Vec a n}) \, (ys : \text{Vec a (reify m)})
\]

\[
\text{(cons} : \forall \, \{k\} \to a \to \text{Vec a (m k)} \to \text{Vec a ((succ m) k)}) \to
\]

\[
\text{revAcc} \, m \, \text{cons} \, xs \, ys \equiv d_{\text{append}} \, m \, \text{cons} \, (\text{slowReverse} \, xs) \, ys
\]

The proof itself proceeds by induction on the vector \( xs \) and requires a single auxiliary lemma relating \( d_{\text{append}} \) and \( \text{snoC} \). Using \( \text{revAcc-correct} \) and the fact that \( \text{Nil} \) is the right-unit of \( d_{\text{append}} \), we can now complete the proof of \( v_{\text{reverse}} \text{-correct} \).

4 Using a left fold

The version of vector reverse defined in the Agda standard library uses a left fold. In this section, we will reconstruct this definition. A first attempt might use the following type for the fold on vectors:

\[
f \text{oldl} : (b \to a \to b) \to b \to \text{Vec a n} \to b
\]

\[
f \text{oldl} \, \text{step} \, \text{base} \, \text{Nil} = \text{base}
\]

\[
f \text{oldl} \, \text{step} \, \text{base} \, (\text{Cons} \times \text{xs}) = f \text{oldl} \, \text{step} \, \text{(step} \, \text{base} \times) \, \text{xs}
\]
Unfortunately, we cannot define \textit{vreverse} using this fold. The first argument, \( f \), of \textit{foldl} has type \( b \rightarrow a \rightarrow b \); we would like to pass the flip \textit{Cons} function as this first argument, but it has type \( \text{Vec} \ a \rightarrow a \rightarrow \text{Vec} \ a \ (\text{Succ} \ n) \)—which will not type check as the first argument and return type are not identical. We can solve this, by generalising the type of this function slightly, indexing the return type \( b \) by a natural number:

\[
\text{foldl} : (b : \text{Nat} \rightarrow \text{Set}) \rightarrow (\forall \{k\} \rightarrow b \ k \rightarrow a \rightarrow b \ (\text{Succ} \ k)) \rightarrow b \ \text{Zero} \rightarrow \text{Vec} \ a \ n \rightarrow b \ n
\]

\[
\text{foldl} \ b \ \text{step} \ \text{base} \ \text{Nil} = \text{base}
\]

\[
\text{foldl} \ b \ \text{step} \ \text{base} \ (\text{Cons} \times \ \text{xs}) = \text{foldl} \ (b \circ \text{Succ}) \ \text{step} \ (\text{step} \ \text{base} \times) \ \text{xs}
\]

At heart, this definition is the same as the one above. There is one important distinction: the return type changes in each recursive call by precomposing with the successor constructor.

In a way, this ‘reverses’ the natural number, as the outermost successor is mapped to the innermost successor in the type of the result. The accumulating nature of the \textit{foldl} is reflected in how the return type changes across recursive calls.

We can use this version of \textit{foldl} to define a simple vector reverse:

\[
\text{vreverse} : \text{Vec} \ a \ n \rightarrow \text{Vec} \ a \ n
\]

\[
\text{vreverse} = \text{foldl} \ (\text{Vec} \ _) \ (\lambda \ xs \ x \rightarrow \text{Cons} \times \ xs) \ \text{Nil}
\]

This definition does not require any further proofs: the calculation of the return type follows the exact same recursive pattern as the accumulating vector under construction.

The \textit{foldl} function on vectors is a useful abstraction for defining accumulating functions over vectors. For example, as Kidney (2019) has shown we can define the convolution of two vectors in a single pass in the style of Danvy & Goldberg (2005):

\[
\text{convolution} : \forall \ (a \ b : \text{Set}) \rightarrow (n : \text{Nat}) \rightarrow \text{Vec} \ a \ n \rightarrow \text{Vec} \ b \ n \rightarrow \text{Vec} \ (a \times b) \ n
\]

\[
\text{convolution} \ a \ b \ n = \text{foldl} \ (\lambda \ n \rightarrow \text{Vec} \ b \ n \rightarrow \text{Vec} \ (a \times b) \ n)
\]

\[
(\lambda \ \{k \times (\text{Cons} \ y \ ys) \rightarrow \text{Cons} \ (x, y) \ (k \ ys)\})
\]

\[
(\lambda \ \{\text{Nil} \rightarrow \text{Nil}\})
\]

\textbf{Monoids indexed by monoids}

A similar problem—monoidal equalities in indices—shows up when trying to prove that vectors form a monoid. Where proving the monoidal laws for natural numbers or lists is a straightforward exercise for students learning Agda, vectors pose more of a challenge. Crucially, if the lengths of two vectors are not (definitionally) equal, the statement that the vectors themselves are equal is not even \textit{type correct}. For example, given a vector \( \text{xs} : \text{Vec} \ a \ n \), we might try to state the following equality:

\[
\text{xs} \equiv \text{xs} \ # \ \text{Nil}
\]

The vector on the left-hand side of the equality has type \( \text{Vec} \ a \ n \), while the vector on the right-hand side has type \( \text{Vec} \ a \ (n + 0) \). As these two types are not the same—the vectors have different lengths—the statement of this equality is not type correct.

For \textit{difference vectors}, however, this is not the case. To illustrate this, we begin by defining the type of difference vectors as follows:

\[
\text{DVec} : \text{Set} \rightarrow \text{DNat} \rightarrow \text{Set}
\]

\[
\text{DVec} \ a \ d = \forall \{n\} \rightarrow \text{Vec} \ a \ n \rightarrow \text{Vec} \ a \ (d \ n)
\]
We can then define the usual zero and addition operations on difference vectors as follows:

\[
\begin{align*}
vzero : & \text{DVec a zero} \\
vzero &= \lambda x \rightarrow x \\
_+\_ : & \{n \in \text{DNat}\} \rightarrow (xs : \text{DVec a n}) \rightarrow (ys : \text{DVec a m}) \rightarrow \text{DVec a (n + m)} \\
xs + ys &= \lambda env \rightarrow ys(xs env)
\end{align*}
\]

Next we can formulate the monoidal equalities and establish that these all hold trivially:

\[
\begin{align*}
vzero-left : & (xs : \text{DVec a n}) \rightarrow (vzero + xs) \equiv xs \\
vzero-left xs &= \text{refl} \\
vzero-right : & (xs : \text{DVec a n}) \rightarrow (xs + vzero) \equiv xs \\
vzero-right xs &= \text{refl} \\
+_\text{-assoc} : & (xs : \text{DVec a n}) \rightarrow (ys : \text{DVec a m}) \rightarrow (zs : \text{DVec a k}) \rightarrow \\
&(xs + (ys + zs)) \equiv (xs + (ys + zs)) \\
+_\text{-assoc} xs ys zs &= \text{refl}
\end{align*}
\]

We have elided some implicit length arguments that Agda cannot infer automatically, but it should be clear that the monoidal operations on difference vectors are no different from the difference naturals we saw in Section 2.

It is worth pointing out that using the usual definitions of natural numbers and additions, the latter two definitions would not hold—\textit{a fortiori}, the statement of the properties \textit{vzero-right} and \textit{+_assoc} would not even \textit{type check}. Consider the type of \textit{vzero-right}, for instance: formulating this property using natural numbers and addition would yield a vector on the left-hand side of the equation of length \(n + 0\), whereas \(xs\) has length \(n\). As the equality type can only be used to compare vectors of equal length, the statement of \textit{vzero-right} would be type incorrect. Addressing this requires coercing the lengths of the vectors involved—as we did in the very first definition of \textit{vreverse} in the introduction—that quickly spreads throughout any subsequent definitions.

5 Indexing beyond natural numbers

In this section, we will explore another application of the Cayley representation of monoids. Instead of indexing by a natural number, this section revolves around computations indexed by lists.

We begin by defining a small language of boolean expressions:

\[
\begin{align*}
data \text{Expr} (\text{vars} : \text{List a}) : \text{Set} \text{where} \\
\text{T} & : \text{Expr \text{vars}} \\
\text{F} & : \text{Expr \text{vars}} \\
\text{Not} & : \text{Expr \text{vars}} \rightarrow \text{Expr \text{vars}} \\
\text{And} & : \text{Expr \text{vars}} \rightarrow \text{Expr \text{vars}} \rightarrow \text{Expr \text{vars}} \\
\text{Or} & : \text{Expr \text{vars}} \rightarrow \text{Expr \text{vars}} \rightarrow \text{Expr \text{vars}} \\
\text{Var} & : x \in \text{vars} \rightarrow \text{Expr \text{vars}}
\end{align*}
\]

The \text{Expr} data type has constructors for truth, falsity, negation, conjunction and disjunction. Expressions are parametrised by a list of variables, \text{vars} : \text{List a} for some type \text{a} : \text{Set}. While we could model a finite collection of variables using the well known \text{Fin} type, we choose a slightly different representation here—allowing us to illustrate how the Cayley
representation can be used for other indices beyond natural numbers. Each \texttt{Var} constructor stores a proof, \( x \in \text{vars} \), that is used to denote the particular named variable to which it is being referred. The proofs, \( x \in \text{xs} \), can be constructed using a pair of constructors, \texttt{Top} and \texttt{Pop}, that refer to the elements in the head and tail of the list, respectively:

\begin{verbatim}
data _∈_ : a → List a → Set where
  Top : x ∈ (x :: xs)
  Pop : x ∈ xs → x ∈ (y :: xs)
\end{verbatim}

Indexing expressions by the list of variables they may contain allows us to write a \textit{total} evaluation function. The key idea is that our evaluator is passed an environment assigning a boolean value to each variable in our list:

\begin{verbatim}
data Env : List a → Set where
  Nil : Env []
  Cons : Bool → Env xs → Env (x :: xs)
\end{verbatim}

The evaluator itself is easy enough to define; it maps each constructor of the \texttt{Expr} data type to its corresponding operation on booleans.

\begin{verbatim}
eval : Expr vars → Env vars → Bool
eval T env = True
eval F env = False
eval (Not e) env = ¬ (eval e env)
eval (And e₁ e₂) env = eval e₁ env ∧ eval e₂ env
eval (Or e₁ e₂) env = eval e₁ env ∨ eval e₂ env
eval (Var x) env = lookup env x
\end{verbatim}

The only interesting case is the one for variables, where we call an auxiliary lookup function to find the boolean value associated with the given variable.

For a large fixed expression, however, we may not want to call \texttt{eval} over and over again. Instead, it may be preferable to construct a \textit{decision tree} associated with a given expression. The decision tree associated with an expression is a perfect binary tree, where each node branches over a single variable:

\begin{verbatim}
data DecTree : List a → Set where
  Node : DecTree vars → (x : a) → DecTree vars → DecTree (x :: vars)
  Leaf : Bool → DecTree []
\end{verbatim}

Given any environment, we can still ‘evaluate’ the boolean expression corresponding to the tree, using the environment to navigate to the unique leaf corresponding to the series of true-false choices for each variable:

\begin{verbatim}
treeval : DecTree xs → Env xs → Bool
treeval (Leaf x) Nil = x
treeval (Node l x r) (Cons True env) = treeval l env
treeval (Node l x r) (Cons False env) = treeval r env
\end{verbatim}

We would now like to write a function that converts a boolean expression into its decision tree representation, while maintaining the scope hygiene that our expression data type
enforces. We could imagine trying to do so by induction on the list of free variables, repeatedly substituting the variables one by one:

\[
\text{makeDecTree} : (\text{vars} : \text{List a}) \rightarrow \text{Expr vars} \rightarrow \text{DecTree vars}
\]

\[
\text{makeDecTree} [] e = \text{Leaf (eval e empty)}
\]

\[
\text{makeDecTree} (x :: \text{vars}) e =
\]

\[
\begin{align*}
\text{let} \quad l &= \text{makeDecTree vars (subst T x e)} \text{ in} \\
\text{let} \quad r &= \text{makeDecTree vars (subst F x e)} \text{ in} \\
\text{Node} \ l \ r
\end{align*}
\]

But this is not entirely satisfactory: to prove this function correct, we would need to prove various lemmas relating substitution and evaluation; furthermore, this function is inefficient, as it repeatedly traverses the expression to perform substitutions.

Instead, we would like to define an accumulating version of \text{makeDecTree}, that carries around a (partial) environment of those variables on which we have already branched. As we shall see, this causes problems similar to those that we saw previously for reversing a vector. A first attempt might proceed by induction on the free variables in our expression, that have not yet been captured in our environment:

\[
\text{makeDecTreeAcc} : (\text{xs ys} : \text{List a}) \rightarrow \text{Expr (revAcc xs ys)} \rightarrow \text{Env ys} \rightarrow \text{DecTree xs}
\]

\[
\text{makeDecTreeAcc} [] \ ys \ expr \ env = \text{Leaf (eval expr env)}
\]

\[
\text{makeDecTreeAcc} (x :: \text{xs}) \ ys \ expr \ env = \text{Node} \ l \ r
\]

\[
\begin{align*}
\text{where} \\
\quad l &= \text{makeDecTreeAcc xs (x :: \text{ys}) expr (Cons True env)} \\
\quad r &= \text{makeDecTreeAcc xs (x :: \text{ys}) expr (Cons False env)}
\end{align*}
\]

This definition, however, quickly gets stuck. In the recursive calls, the environment has grown, but the variables in the expression and environment no longer line up. The situation is similar to the very first attempt at defining the accumulating vector reverse function: the usual definition of addition is unsuitable for defining functions using an accumulating parameter. Fortunately, the solution is to define a function \text{revAcc}, akin to the one defined for vectors, that operates on lists:

\[
\text{revAcc} : \text{List a} \rightarrow \text{List a} \rightarrow \text{List a}
\]

\[
\text{revAcc} [] \ ys = \text{ys} \\
\text{revAcc} (x :: \text{xs}) \ ys = \text{revAcc} \text{xs (x :: ys)}
\]

We can now attempt to construct the desired decision tree, using the \text{revAcc} function in the type indices, as follows:

\[
\text{makeDecTreeAcc} : (\text{xs ys} : \text{List a}) \rightarrow \text{Expr (revAcc xs ys)} \rightarrow \text{Env ys} \rightarrow \text{DecTree xs}
\]

\[
\text{makeDecTreeAcc} [] \ ys \ expr \ env = \text{Leaf (eval expr env)}
\]

\[
\text{makeDecTreeAcc} (x :: \text{xs}) \ ys \ expr \ env = \text{Node} \ l \ r
\]

\[
\begin{align*}
\text{where} \\
\quad l &= \text{makeDecTreeAcc xs (x :: \text{ys}) expr (Cons True env)} \\
\quad r &= \text{makeDecTreeAcc xs (x :: \text{ys}) expr (Cons False env)}
\end{align*}
\]
Although this definition now type checks, just as we saw for one of our previous attempts for \textit{revAcc}, the problem arises once we try to call this function with an initially empty environment:

\[
\text{makeDecTree} : (xs : \text{List } a) \rightarrow \text{Expr } xs \rightarrow \text{DecTree } xs
\]

\[
\text{makeDecTree} \, xs \, \text{expr} = \text{makeDecTreeAcc } xs \, [] \{ \text{expr} \}_6 \, \text{Nil}
\]

\textbf{Goal:} \text{Expr} (\text{revAcc} \, xs \, [])

\textbf{Have:} \text{Expr} \, xs

Calling the accumulating version fails to produce a value of the desired type—in particular, it produces a tree branching over the variables \text{revAcc} \, xs \, [] rather than \text{xs}. To address this problem, however, we can move from an environment indexed by a regular lists to one indexed by a difference list, accumulating the values of the variables we have seen so far:

\[
\text{DEnv} : (\text{List } a \rightarrow \text{List } a) \rightarrow \text{Set}
\]
\[
\text{DEnv } f = \forall \{ \text{vars} \} \rightarrow \text{Env } \text{vars} \rightarrow \text{Env } (f \, \text{vars})
\]

Note that we use the Cayley representation of monoids in both the \textit{type} index of and the \textit{value} representing environments.

We can now complete our definition as expected, performing induction without ever having to prove a single equality about the concatenation of lists:

\[
\text{makeDecTreeAcc} : (xs : \text{List } a) \rightarrow (ys : \text{List } a \rightarrow \text{List } a) \rightarrow
\]
\[
\text{DEnv } ys \rightarrow \text{Expr } (ys \, xs) \rightarrow \text{DecTree } xs
\]
\[
\text{makeDecTreeAcc } [] \, ys \, \text{denv} \, \text{expr} = \text{Leaf } (\text{eval } \text{expr} \, (\text{denv } \text{Nil}))
\]
\[
\text{makeDecTreeAcc } (x :: xs) \, ys \, \text{denv} \, \text{expr} = \text{Node } l \, x \, r
\]

where

\[
l = \text{makeDecTreeAcc } xs \, (ys \, o \, (x :: _)) \, (\text{denv } o \, \text{Cons } \text{True}) \, \text{expr}
\]
\[
r = \text{makeDecTreeAcc } xs \, (ys \, o \, (x :: _)) \, (\text{denv } o \, \text{Cons } \text{False}) \, \text{expr}
\]

Finally, we can kick off our accumulating function with a pair of identity functions, corresponding to the zero elements of the list of variables that have been branched on and the difference environment:

\[
\text{makeDecTree} : (xs : \text{List } a) \rightarrow \text{Expr } xs \rightarrow \text{DecTree } xs
\]
\[
\text{makeDecTree} \, xs \, \text{e} = \text{makeDecTreeAcc } xs \, \text{id} \, \text{id} \, \text{e}
\]

Interestingly, the type signature of this top-level function does not mention the ‘difference environment’ or ‘difference lists’ at all.

Can we verify that definition is correct? The obvious theorem we may want to prove states the eval and treeval functions agree on all possible expressions:

\[
\text{correctness} : \forall \text{vars} \, (e : \text{Expr } \text{vars}) \, (\text{env} : \text{Env } \text{vars}) \rightarrow
\]
\[
\text{eval } \text{e} \, \text{env} \equiv \text{treeval} \, (\text{makeDecTree } \text{vars} \, \text{e}) \, \text{env}
\]

A direct proof by induction quickly fails, as we cannot use our induction hypothesis; we can, however, prove a more general lemma that implies this result:
A well-known representation of monoids

6 Solving any monoidal equation

In this last section, we show how this technique of mapping monoids to their Cayley representation can be used to solve equalities between any monoidal expressions. To generalise the constructions we have seen so far, we define the following Agda record representing monoids:

```agda
class Monoid (a : Set) : Set where
  field zero : a
  _⊕_ : a → a → a
  zero-left : ∀ x → (zero ⊕ x) ≡ x
  zero-right : ∀ x → (x ⊕ zero) ≡ x
  ⊕-assoc : ∀ x y z → (x ⊕ (y ⊕ z)) ≡ ((x ⊕ y) ⊕ z)
```

We can represent expressions built from the monoidal operations using the following data type, MExpr:

```agda
data MExpr (a : Set) : Set where
  Add : MExpr a → MExpr a → MExpr a
  Zero : MExpr a
  Var : a → MExpr a
```

If we have a suitable monoid in scope, we can evaluate a monoidal expression, MExpr, in the obvious fashion:

```agda
eval : MExpr a → a
eval (Add e₁ e₂) = eval e₁ ⊕ eval e₂
eval (Zero) = zero
eval (Var x) = x
```

This is, however, not the only way to evaluate such expressions. As we have already seen, we can also define a pair of functions converting a monoidal expression to its Cayley representation and back:

```agda
reify : (MExpr a → MExpr a) → MExpr a
reify f = f Zero
```

The proof is reassuringly simple; it has the same accumulating structure as the inductive definitions we have seen.
Finally, we can *normalise* any expression by composing these two functions:

\[
\text{normalise} : \text{MExpr } a \rightarrow \text{MExpr } a \\
\text{normalise } m = \text{reify } [ m ]
\]

Crucially, we can prove that this normalise function preserves the (monoidal) semantics of our monoidal expressions:

\[
\text{soundness } : \forall (x : \text{MExpr } a) \rightarrow \text{eval } (\text{normalise } x) \equiv \text{eval } x
\]

Where the cases for `Zero` and `Var` are straightforward, the addition case is more interesting. This final case requires a pair of auxiliary lemmas that rely on the monoid equalities:

\[
\forall xy \rightarrow \text{eval } (\text{normalise } (\text{Add } x y)) \equiv \text{eval } ([ x ] y) \\
\forall xy \rightarrow \text{eval } ([ x ] y) \equiv \text{eval } (\text{Add } x y)
\]

Using transitivity, we can complete this last case of the proof.

Finally, we can use this soundness result to prove that two expressions are equal under evaluation, provided their corresponding normalised expressions are equal under evaluation:

\[
\text{solve } : \forall (xy : \text{MExpr } a) \rightarrow \text{eval } (\text{normalise } x) \equiv \text{eval } (\text{normalise } y) \rightarrow \text{eval } x \equiv \text{eval } y
\]

What have we gained? On the surface, these general constructions may not seem particularly useful or exciting. Yet the solve function establishes that to prove any equality between two monoidal expressions, it suffices to prove that their normalised forms are equal. Yet—as we have seen previously—the monoidal equalities hold definitionally in our Cayley representation. As a result, the only ‘proof obligation’ we need to provide to the solve function will be trivial.

Let’s consider a simple example to drive home this point. Once we have established that lists are a monoid, we can use the solve function to prove the following equality:

\[
\text{example } : (xs ys zs : \text{List } a) \rightarrow ((xs \# []) \# (ys \# zs)) \equiv ((xs \# ys) \# zs)
\]

\[
\text{example xs ys zs } = \\
\text{let } e_1 = \text{Add } (\text{Add } (\text{Var } xs) \text{Zero}) (\text{Add } (\text{Var } ys) (\text{Var } zs)) \text{ in} \\
\text{let } e_2 = \text{Add } (\text{Add } (\text{Var } xs) (\text{Var } ys)) (\text{Var } zs) \text{ in} \\
\text{solve } e_1 e_2 \text{ refl}
\]

To complete the proof, we only needed to find monoidal expression representing the left- and right-hand sides of our equation—and this can be automated using Agda’s metaprogramming features (Van Der Walt & Swierstra, 2012). The only remaining proof obligation—that is, the third argument to the solve function—is indeed trivial. In this style, we can automatically solve any equality that relies exclusively on the three defining properties of any monoid.

We can also show that natural numbers form a monoid under `addAcc` and `Zero`. Using the associated solver, we can construct the proof obligations associated with the very first version of vector reverse from our introduction:
A well-known representation of monoids

\[ \text{vreverse} : (n : \text{Nat}) \rightarrow \text{Vec} \ a \ n \rightarrow \text{Vec} \ a \ n \]
\[
\text{vreverse} \ n \ \text{xs} = \text{coerce-length proof} \ (\text{revAcc} \ \text{xs} \ \text{Nil})
\]

where

\[
\text{proof} = \text{solve} \ (\text{Add} \ (\text{Var} \ n) \ \text{Zero}) \ (\text{Var} \ n) \ \text{refl}
\]

Even if the proof constructed here is a simple call to one of the monoidal identities, automating this proof lets us come full circle.

7 Discussion

I first learned of that monoidal identities hold definitionally for the Cayley representation of monoids from a message Alan Jeffrey (2011) sent to the Agda mailing list. Since then, this construction has been used (implicitly) in several papers (McBride, 2011; Jaber et al., 2016; Allais et al., 2017) and developments (Kidney, 2020; Ko, 2020)—but the works cited here are far from complete. The observation that the Cayley representation can be used to normalise monoidal expressions dates back at least to Beylin & Dybjer (1995), although it is an instance of the more general technique of normalisation by evaluation (Berger & Schwichtenberg, 1991).

The two central examples from this paper, reversing vectors and constructing trees, share a common structure. Each function uses an accumulating parameter, indexed by a monoid, but relies on the monoid laws to type check. To avoid using explicit equalities, we use the Cayley representation of monoids in the index of the accumulating parameter. In the base case, this ensures that we can safely return the accumulating parameter; similarly, when calling the accumulating function with an initially empty argument, the Cayley representation ensures that the desired monoidal property holds by definition. In our second example, we also use the Cayley representation in the value of the accumulating parameter; we could also use this representation in the definition of vreverse, but it does not make things any simpler. In general, this technique works provided we only rely on the monoidal properties. As soon as the type indices contain richer expressions, we will need to prove equalities and coerce explicitly—or better yet, find types and definitions that more closely follow the structure of the functions we intend to write.

Acknowledgements

I would like to thank Guillaume Allais, Joris Dral, Jeremy Gibbons, Donnacha Oisin Kidney and the anonymous reviewers for their insightful feedback.

Conflicts of Interest

None.

Supplementary materials

For supplementary material for this article, please visit https://doi.org/10.1017/S0956796822000065


