1 Dependent types

Suppose we introduce immutable vectors of booleans to the lambda calculus. A vector \( \langle v_1, \ldots, v_n \rangle \) is of length \( n \), and each \( v_i \) is a boolean value. Let’s add primitive function \( \text{init} \) as a way to construct vectors. It will take a natural number and a value as arguments and produce a vector: \( \text{init} \ k \ v \) will evaluate to \( \langle v, \ldots, v \rangle \), a vector of length \( k \) where each element has the value \( v \).

Similarly, we’ll add a primitive function \( \text{index} \) to access an element of the vector: \( \text{index} \ \langle v_1, \ldots, v_n \rangle \ i \) will evaluate to \( v_{i+1} \), provided \( 0 \leq i < n \). (We could add a way to produce vectors that contain different values, but for our purposes just \( \text{init} \) and \( \text{index} \) are sufficient.)

We’ll also include natural numbers \( n \), pairs, and the unit value in the language. The syntax of expressions and values for the extended language is the following.

\[
\begin{align*}
e &::= x | \lambda x. e | e_1 \ e_2 | n | (e_1, e_2) | () | \text{true} | \text{false} \quad | \text{init} \ | \text{index} \\
v &::= \lambda x. e \ n | \langle v_1, \ldots, v_n \rangle | (v_1, v_2) | () \quad \text{true} | \text{false}
\end{align*}
\]

Evaluation rules for the new constructs are defined as follows.

\[
\begin{align*}
\text{init} \ k \ v &\rightarrow \langle v_1, \ldots, v_k \rangle \quad \forall i \in 1..k, v_i = v \\
\text{index} \ \langle v_1, \ldots, v_k \rangle \ i &\rightarrow v_i
\end{align*}
\]

We could run in to some problems when executing programs in our new language. We could try to access a vector with an inappropriate index, for example \( \text{index} \ \langle \text{false}, \text{false}, \text{true} \rangle \ 7 \).

The first problem we can avoid by using type systems like we have seen previously: give vectors a type, say \( \text{boolvec} \), and give the primitive function \( \text{index} \) the type \( \text{nat} \rightarrow \text{bool} \rightarrow \text{boolvec} \) (where \( \text{nat} \) is the type of natural numbers). But this approach does not stop us from having incorrect indices, as it would still allow the expression \( \text{index} \ \langle \text{false}, \text{false}, \text{true} \rangle \ 7 \).

1.1 First attempt at a type system

To address these problems, we are going to use a dependent type for boolean vectors, where the length of the vector is part of the type of a vector. The type \( \text{boolvec} \ e \) represents boolean vectors of length \( e \), where \( e \) is a natural number expression.

The signature for \( \text{init} \) becomes \( (n : \text{nat}) \rightarrow \text{bool} \rightarrow \text{boolvec} \ n \). That is, \( \text{init} \) takes two arguments, a natural number \( n \), and a boolean, and produces a boolean vector of length \( n \), a \( \text{boolvec} \ n \).

With this new type, we define typing rules for vectors are the following.

\[
\begin{align*}
\forall i \in 1..n. \quad \Gamma \vdash v_i : \text{bool} & \quad \Gamma \vdash e_1 : \text{nat} \quad \Gamma \vdash e_2 : \text{bool} \quad \Gamma \vdash \text{init} \ e_1 \ e_2 : \text{boolvec} \ e_1 \\
\Gamma \vdash \langle v_1, \ldots, v_n \rangle : \text{boolvec} \ n & \quad \Gamma \vdash \text{index} \ e_1 \ e_2 : \text{bool} \quad \Gamma \vdash e_2 : \text{nat} \quad e_2 \leq e_3
\end{align*}
\]

The type of a vector of length \( n \) is simply \( \text{boolvec} \ n \). The first argument to the primitive function to create vectors, \( \text{init} \), is the length of the new vector, and so the type of \( \text{init} \ e_1 \ e_2 \) is \( \text{boolvec} \ e_1 \), a vector with length \( e_1 \). The typing rule for \( \text{index} \) requires that the index \( e_2 \) is no greater than the length of the vector being accessed: \( e_2 \leq e_3 \).

Now, the type \( \text{boolvec} \ e \) is very strange! We have at least three problems with this newly proposed type \( \text{boolvec} \ e \).
1. In the type for init, \((n : \text{nat}) \to \text{bool} \to \text{boolvec} n\), the first argument is somehow bound to the variable \(n\) which occurs in the return type of the function. What does this mean?

2. The type contains an expression \(e\). This could be an arbitrary expression. If \(e\) is a literal natural number, then the type maybe makes sense, for example \(\text{boolvec} 7\) is the type of vectors of length 7. But what do the types \(\text{boolvec} (9 + 1)\) or \(\text{boolvec} x\) mean? And what does it mean in the proposed typing rule for index to have a side condition \(e_1 \leq e_3\)?

3. The expression \(e\) in the type \(\text{boolvec} e\) should be of type \(\text{nat}\). For example, we shouldn’t let someone write code where \(e\) is not of type \(\text{nat}\), such as \(\lambda x : \text{boolvec} \text{unit}\). . . . How do we ensure that \(e\) is limited to expressions of type \(\text{nat}\).

1.2 LF

We address some of these problems by considering boolean vectors in the language LF, which stands for Logical Framework.

We give expressions types to allow us to restrict and reason about the use of expressions. Types can be thought of as describing sets of expressions. In LF, just as expressions have types, types have kinds. That is, kinds describe sets of types, and we use kinds to restrict and reason about the use of types.

The syntax of LF is given by the following grammar. Here, we use metavariable \(K\) to range over kinds.

\[
\text{Expressions} : \quad e ::= x \mid \lambda x : \tau. e \mid e_1 e_2 \mid n \mid e_1 + e_2 \mid \langle v_1, \ldots, v_n \rangle \mid \ldots
\]

\[
\text{Types} : \quad \tau ::= \text{nat} \mid \text{boolvec} \mid \text{bool} \mid \text{unit} \mid \tau e \mid (x : \tau_1) \to \tau_2
\]

\[
\text{Kinds} : \quad K ::= \text{Type} \mid (x : \tau) \Rightarrow K
\]

In LF, the literal \(n\) has type \(\text{nat}\), just as in the simply-typed lambda calculus. Also, the type of a vector \(\langle v_1, \ldots, v_n \rangle\) is \(\text{boolvec} n\). However, we will prevent the use of ill-formed types such as \(\text{boolvec} \text{true}\) by using kinds to restrict how types may be composed. The kind of \(\text{boolvec}\) will be \((x : \text{nat}) \Rightarrow \text{Type}\): it takes a natural number and produces a type.

We will have three judgments, one each for expressions, types, and kinds. The form of the judgment for expressions is \(\Gamma \vdash e : \tau\), and means that under context \(\Gamma\), expression \(e\) has type \(\tau\). We extend contexts so that they include both the types of program variables \((x : \tau)\) and the kinds of type variables \((X : K)\).

The form of the judgment for types is \(\Gamma \vdash \tau : K\), meaning that under context \(\Gamma\), type \(\tau\) has kind \(K\). In LF, kinds do not have their “types” (i.e., there is no entity that describes sets of kinds), so the judgment for kinds is of the form \(\Gamma \vdash K \text{ ok}\), which means that under context \(\Gamma\), kind \(K\) is well-formed.

Since types may contain expressions, in order to perform type checking, we may need to evaluate expressions to determine if two types are equivalent, for example the types \(\text{boolvec} 19\) and \(\text{boolvec} (12 + 7)\). Similarly, since kinds may contain types (in the production \((x : \tau) \Rightarrow K\)), we may need to evaluate expressions when checking kinds. As such, we also define another three relations that describe equivalence between expressions, types, and kinds, respectively. Relation \(\Gamma \vdash e_1 \equiv e_2 : \tau\) means that (under context \(\Gamma\)), expressions \(e_1\) and \(e_2\) are equivalent (and have type \(\tau\)). Relation \(\Gamma \vdash \tau_1 \equiv \tau_2 :: K\) means that (under context \(\Gamma\)), types \(\tau_1\) and \(\tau_2\) are equivalent (and have kind \(K\)). Finally, relation \(\Gamma \vdash K_1 \equiv K_2\) means that (under context \(\Gamma\)), kinds \(K_1\) and \(K_2\) are equivalent.

Types are similar to what we have seen previously. We have primitive types (\(\text{nat}, \text{bool}, \text{boolvec}\), etc.). The function type gives the argument type a binder: \((x : \tau_1) \to \tau_2\) allows the program variable \(x\) to appear in the type \(\tau_2\), where \(x\) represents the argument given to the expression. We also have type application \(\tau e\), which applies a type to an expression. The key example of this in our system is the application of \(\text{boolvec}\) to an expression, such as \(\text{boolvec} 7\).

The rules for the judgment \(\Gamma \vdash e : \tau\) are the following.

\[
\begin{align*}
\Gamma \vdash e : \tau \\
\Gamma \vdash x : \tau & \quad x \in \Gamma \\
\Gamma \vdash n : \text{nat} & \quad n \in \mathbb{N} \\
\Gamma \vdash e_1 : \text{nat} & \quad \Gamma \vdash e_2 : \text{nat} \\
\Gamma \vdash e_1 + e_2 : \text{nat} & \quad \text{For all } i \in 1..n, \quad \Gamma \vdash v_i : \text{bool} \\
\Gamma \vdash \langle v_1, \ldots, v_n \rangle : \text{boolvec} n
\end{align*}
\]
The rule for conversion uses the relation $\Gamma \vdash \tau_1 \equiv \tau_2 :: K$ that we will define later. We also need rules for base values, such as $()$, true, false. We omit these rules.

Note that we need to provide types for our primitive functions init and index. We can simply assume that these primitive functions are variables that are bound in every context. We will consider types for these primitive functions later.

The judgment $\Gamma \vdash \tau :: K$ is given by the following rules and axioms.

\[
\begin{align*}
\Gamma \vdash \tau :: K & \quad \Gamma \vdash K \text{ ok} \quad \Gamma \vdash X :: K \quad X : K \in \Gamma \\
\Gamma \vdash \tau :: K & \quad \Gamma, x : \tau \vdash \tau' :: Type \quad \Gamma \vdash (x : \tau) \rightarrow \tau' :: Type \\
\Gamma \vdash \tau :: (x : \tau') \Rightarrow K & \quad \Gamma \vdash e :: \tau' \\
\Gamma \vdash \tau e :: K \{e/x\} & \quad \text{Conversion} \quad \Gamma \vdash \tau :: K' \quad \Gamma \vdash K \equiv K' \\
\Gamma \vdash e :: \tau & \quad \Gamma \vdash \tau :: Type \quad \Gamma, x : \tau \vdash e :: \tau \\
\Gamma \vdash e_2 :: \tau_2 & \quad \Gamma \vdash e_2 e_1 :: \tau \quad \Gamma \vdash \lambda x : \tau. e :: (x : \tau) \rightarrow \tau'
\end{align*}
\]

The equivalence relations $\equiv$ define what it means for expressions, types, and kinds to be equivalent. In a nutshell, it is if they evaluate to the same value. The question of what equivalences to include in a dependent type system is a critical one. Many different choices have been studied, leading to very different results. Clearly we would like to, for example, consider the types boolvec 42 and boolvec $(35 + 7)$ equivalent. But what about if we are in a context where we have variables $x$ and $f$ of type nat and nat $\rightarrow$ nat, respectively, where we know that $f x = 7$? Should we consider the types boolvec ($f x$) and boolvec 7 to be equivalent?
\[ \Gamma \vdash e_1 \equiv e_2 : \tau \]

\[ \Gamma \vdash \tau_1 \equiv \tau_2 :: Type \quad \Gamma, x : \tau_1 \vdash e_1 \equiv e_2 : \tau \]

\[ \Gamma \vdash e_1 \equiv e_2 : (x : \tau) \rightarrow \tau' \quad \Gamma \vdash e'_1 \equiv e'_2 : \tau' \{e'/x\} \]

\[ \Gamma, x : \tau \vdash e : \tau' \quad \Gamma \vdash e' : \tau \]

\[ \Gamma \vdash \lambda x : \tau_1. e_1 \equiv \lambda x : \tau_2. e_2 : (x : \tau_1) \rightarrow \tau \]

\[ \Gamma \vdash (\lambda x : \tau. e) e' \equiv e' : \tau \{e'/x\} \]

\[ \Gamma \vdash e : (x : \tau) \rightarrow \tau' \quad x \not\in FV(e) \]

\[ \Gamma \vdash (\lambda x : \tau. e \ x) \equiv e : (x : \tau) \rightarrow \tau' \]

\[ \Gamma \vdash \tau_1 \equiv \tau_2 :: K \]

\[ \Gamma \vdash \tau_1 \equiv \tau_2 :: Type \quad \Gamma, x : \tau_1 \vdash \tau'_1 \equiv \tau'_2 :: Type \quad \Gamma \vdash \tau_1 \equiv \tau_2 :: (x : \tau) \rightarrow \tau' \]

\[ \Gamma \vdash e : \tau \quad \Gamma \vdash e_1 \equiv e_2 : \tau \quad \Gamma \vdash e_2 \equiv e_1 : \tau \]

\[ \Gamma \vdash \tau \equiv \tau :: K \quad \Gamma \vdash \tau_1 \equiv \tau_2 :: K \quad \Gamma \vdash \tau_2 \equiv \tau_1 :: K \quad \Gamma \vdash \tau_2 \equiv \tau_3 :: K \quad \Gamma \vdash \tau_1 \equiv \tau_3 :: K \]

\[ \Gamma \vdash K_1 \equiv K_2 \]

\[ \Gamma \vdash \tau_1 \equiv \tau_2 :: Type \quad \Gamma, x : \tau_1 \vdash K_1 \equiv K_2 \]

\[ \Gamma \vdash (x : \tau_1) \Rightarrow K_1 \equiv (x : \tau_2) \Rightarrow K_2 \]

\[ \Gamma \vdash \tau \rightarrow \tau :: K \quad \Gamma \vdash \tau_1 \equiv \tau_2 :: K \quad \Gamma \vdash \tau_2 \equiv \tau_3 :: K \]

So, what has LF gained us? The use of kinds ensure that we cannot mis-use a type constructor. For example, it will rule out any program that contains \texttt{boolvec} \( e \) where \( e \) does not have type \texttt{nat}.

LF also provides us with a clearly defined meaning for binding program variables in types. The type we gave for \texttt{init} now makes sense: \( (n : \texttt{nat}) \rightarrow \texttt{bool} \rightarrow \texttt{boolvec} \ n \). We can also give precise types to other interesting functions, such as \texttt{join}, a function that takes two vectors on length \( n \) and \( k \) respectively, and combines them to form a vector of length \( n + k \) could be given the following type:

\( (n : \texttt{nat}) \rightarrow (k : \texttt{nat}) \rightarrow \texttt{boolvec} \ n \rightarrow \texttt{boolvec} \ k \rightarrow \texttt{boolvec} \ n + k \)

What about a type for \texttt{index}? That is, how do we ensure that in expression \texttt{index} \( e_1 \ e_2 \) that if \( e_1 \) has type \texttt{boolvec} \( n \), how do we ensure that \( e_2 \) is a natural number that is less than \( n \)? And, suppose we wanted a primitive function \texttt{asPairs} that takes a vector, and returns a representation of the vector using pairs: \texttt{asPairs} \( (v_1, \ldots, v_n) \) evaluates to \( (v_1, (v_2, \ldots (v_n, (\ldots \ldots)) \ldots)) \). LF does not provide us with a way to write down an appropriate type for \texttt{asPairs}.

We can address these remaining issues by consider the calculus of constructions (CoC). The calculus of constructions is the language behind the proof assistant Coq.
We have seen functions from expressions to expressions (which are just the standard abstractions, \( \lambda x. e \)); polymorphic lambda calculus gave us functions from types to terms (\( \Lambda X. e \)); dependent types are functions from expressions to types (e.g., \texttt{boolvec} is a function from expressions of type \texttt{nat} to types). To express the type of \texttt{asPairs}, we need functions from types to types. That is, the type of \texttt{asPairs} \( e \) depends on the type of \( e \).

In part, the power of Coq comes from viewing dependent types through the Curry-Howard Isomorphism, whereby we can equate types with propositions, and expressions with proofs. With dependent types, expressions can appear in proofs. One of the ways that we can think about that, is that dependent types are allowing us to express properties (predicates) of values that the program manipulates. That is, one way to consider a dependent type \( (x : \tau_1) \to \tau_2 \) is that it corresponds to universal quantification in first-order logic: for any value (individual) \( x \) of type \( \tau_1 \), proposition \( \tau_2 \) (which mentions \( x \)) is true. A function with type \( (x : \tau_1) \to \tau_2 \) is a proof of this universal quantification, because if you give it any value \( v \) of type \( \tau_1 \), then it will give back a proof that \( \tau_2 \{v/x\} \) holds.

Due to time constraints, we did not get to discuss CoC in detail. If you are interested, a good introduction can be found in Chapter 2 of Advanced Topics in Types and Programming Languages, edited by Benjamin C. Pierce, MIT Press, 2005.