Appendix for “Calculating Correct Compilers”

PATRICK BAHR
Department of Computer Science, University of Copenhagen, Denmark

GRAHAM HUTTON
School of Computer Science, University of Nottingham, UK

A Global State

Recall the specification of the compiler:

\[
\text{exec} (\text{comp}' \ x \ c) (s, q) = \text{case eval x q of}
\]
\[
(\text{Just } n, q') \rightarrow \text{exec c (VAL } n : s, q')
\]
\[
(\text{Nothing}, q') \rightarrow \text{fail (s, q')}
\]

Below we give the full calculations for the Add and Catch cases.

\[
\text{exec (comp' (Add x y) c) (s, q)}
\]
\[
= \{ \text{ specification (12) } \}
\]
\[
\text{case eval x q of}
\]
\[
(\text{Just } n, q') \rightarrow \text{case eval y q' of}
\]
\[
(\text{Just } m, q'') \rightarrow \text{exec c (VAL } n + m : s, q'')
\]
\[
(\text{Nothing}, q'') \rightarrow \text{fail (s, q'')}
\]
\[
(\text{Nothing}, q') \rightarrow \text{fail (s, q')}
\]
\[
= \{ \text{ define: exec (ADD c) (VAL m : VAL n : s, q'') = exec c (VAL } n + m : s, q'') \}
\]
\[
\text{case eval x q of}
\]
\[
(\text{Just } n, q') \rightarrow \text{case eval y q' of}
\]
\[
(\text{Just } m, q'') \rightarrow \text{exec (ADD c) (VAL } n + m : s, q'')
\]
\[
(\text{Nothing}, q'') \rightarrow \text{fail (s, q'')}
\]
\[
(\text{Nothing}, q') \rightarrow \text{fail (s, q')}
\]
\[
= \{ \text{ define: fail (VAL } n : s, q'') = \text{fail (s, q'') } \}
\]
\[
\text{case eval x q of}
\]
\[
(\text{Just } n, q') \rightarrow \text{case eval y q' of}
\]
\[
(\text{Just } m, q'') \rightarrow \text{exec (ADD c) (VAL } n + m : s, q'')
\]
\[
(\text{Nothing}, q'') \rightarrow \text{fail (VAL } n : s, q'')
\]
\[
(\text{Nothing}, q') \rightarrow \text{fail (s, q')}
\]
\[
= \{ \text{ induction hypothesis for y } \}
\]
\[
\text{case eval x q of}
\]
\[
(\text{Just } n, q') \rightarrow \text{exec (comp' y (ADD c)) (VAL } n : s, q')
\]
\[
(\text{Nothing}, q') \rightarrow \text{fail (s, q')}
\]
\[
= \{ \text{ induction hypothesis for x } \}
\]
\[
\text{exec (comp' x (comp' y (ADD c))) (s, q)}
\]
\[
\begin{align*}
\text{exec } (\text{comp}' \ (\text{Catch } x \ h) \ c) \ (s, q) &= \{ \text{ specification (12) } \\
\text{case eval } x \ q \ \text{of} \quad &\quad \rightarrow \text{ exec } (\text{VAL } n : s, q') \\
&\quad (\text{Nothing}, q') \rightarrow \text{ case eval } h \ q' \ \text{of} \\
&\quad \quad (\text{Just } m, q'') \rightarrow \text{ exec } (\text{VAL } m : s, q'') \\
&\quad (\text{Nothing}, q'') \rightarrow \text{ fail } (s, q'') \\
&= \{ \text{ induction hypothesis for } h \} \\
\text{case eval } x \ q \ \text{of} \quad &\quad \rightarrow \text{ exec } (\text{VAL } n : s, q') \\
&\quad (\text{Nothing}, q') \rightarrow \text{ fail } (\text{HAN } (\text{comp}' \ h \ c) : s, q') \\
&= \{ \text{ define; fail } (\text{HAN } c' : s, q') = \text{ exec } c' (s, q') \} \\
\text{case eval } x \ q \ \text{of} \quad &\quad \rightarrow \text{ exec } (\text{VAL } n : \text{HAN } _{\_} : s, q') = \text{ exec } c (\text{VAL } n : s, q') \\
&\quad (\text{Nothing}, q') \rightarrow \text{ fail } (\text{HAN } (\text{comp}' \ h \ c) : s, q') \\
&= \{ \text{ induction hypothesis for } x \} \\
\text{exec } (\text{comp}' \ x \ (\text{UNMARK } c)) \ (\text{HAN } (\text{comp}' \ h \ c) : s, q) &= \{ \text{ define; exec } (\text{MARK } c'' \ c') (s, q) = \text{ exec } c' (\text{HAN } c'' : s, q') \} \\
\text{exec } (\text{MARK } (\text{comp}' \ h \ c) (\text{comp}' \ x \ (\text{UNMARK } c))) \ (s, q) &= \text{ exec } (\text{MARK } (\text{comp}' \ h \ c) (\text{comp}' \ x \ (\text{UNMARK } c))) \ (s, q)
\end{align*}
\]

\text{B Local State}

We now consider the local approach to combining exceptions and state, in which the current state is discarded when an exception is thrown. This idea is reflected in the type for evaluation by moving the output state ‘inside’ the Maybe type:

\[
\text{eval } : \text{Expr} \rightarrow \text{State} \rightarrow \text{Maybe } (\text{Int, State})
\]

That is, if evaluation succeeds then eval returns an integer value and a new state, and if an exception is thrown it returns Nothing. The definition for eval is similar to the previous section except there is now no state to propagate when evaluation fails, and in the case for Catch the handler uses the state from when the catch was entered:

\[
\begin{align*}
\text{eval } (\text{Val } n) \ q &= \text{ Just } (n, q) \\
\text{eval } (\text{Add } x \ y) \ q &= \text{ case eval } x \ q \ \text{of} \\
&\quad \quad \rightarrow \text{ case eval } y \ q' \ \text{of} \\
&\quad \quad \quad \quad \text{ Just } (n, q') \rightarrow \text{ Just } (n + m, q'') \\
&\quad \quad \quad \quad \text{ Nothing} \rightarrow \text{ Nothing} \\
\text{eval } \text{Throw } q &= \text{ Nothing} \\
\text{eval } (\text{Catch } x \ h) \ q &= \text{ case eval } x \ q \ \text{of} \\
&\quad \quad \rightarrow \text{ case eval } x \ q \ \text{of} \\
&\quad \quad \quad \quad \text{ Just } (n, q') \rightarrow \text{ Just } (n, q')
\end{align*}
\]
Appendix for “Calculating Correct Compilers”

Nothing → eval h q

eval Get q = Just (q,q)
eval (Put x y) q = case eval x q of
  Just (n,q’) → eval y n
  Nothing → Nothing

For the purposes of the derivation of the compilation function \( \text{comp}' :: \text{Expr} \rightarrow \text{Code} \rightarrow \text{Code} \) we use the same types as for the global state semantics:

\[
\text{exec} :: \text{Code} \rightarrow \text{Conf} \rightarrow \text{Conf}
\]

\[
\text{type Conf} = (\text{Stack}, \text{State})
\]

\[
\text{type Stack} = [\text{Elem}]
\]

\[
\text{data Elem} = \text{VAL \text{Int}}
\]

The specification for the desired behaviour of \( \text{comp}' \) is essentially the same as for global state, except that when evaluation fails we no longer have an output state to consider and hence the function \( \text{fail} \) only takes a stack as argument:

\[
\text{exec} (\text{comp}' e c) (s,q) = \text{case eval e q of}\]

\[
\text{Just (n,q')} \rightarrow \text{exec c (VAL n:s,q')}
\]

\[
\text{Nothing} \rightarrow \text{fail s}
\]

However, to ensure type correctness of the specification, \( \text{fail} \) must still return a configuration, i.e. \( \text{fail} :: \text{Stack} \rightarrow \text{Conf} \). An alternative would be to supply the input state \( q \) as an argument to \( \text{fail} \), which is a valid choice that would lead to a different compiler. We start the derivation for \( \text{comp}' \) with the cases for \text{Val n, Throw} and \text{Get}, which are easy:

\[
\text{exec} (\text{comp}' (\text{Val n}) c) (s,q)
\]

\[
= \{ \text{specification (14)} \}
\]

\[
\text{exec c (VAL n:s,q)}
\]

\[
= \{ \text{define: exec (PUSH n c) (s,q) = exec c (VAL n:s,q) } \}
\]

\[
\text{exec (PUSH n c) (s,q)}
\]

\[
\text{exec (comp’ Throw c) (s,q)}
\]

\[
= \{ \text{specification (14)} \}
\]

\[
\text{fail s}
\]

\[
= \{ \text{define: exec FAIL (s,q) = fail s } \}
\]

\[
\text{exec FAIL (s,q)}
\]

\[
\text{exec (comp’ Get c) (s,q)}
\]

\[
= \{ \text{specification (14)} \}
\]

\[
\text{exec c (VAL q:s,q)}
\]

\[
= \{ \text{define: exec (LOAD c) (s,q) = exec c (VAL q:s,q) } \}
\]

\[
\text{exec (LOAD c) (s,q)}
\]

The case for \text{Add} follows the now familiar pattern:
Patrick Bahr and Graham Hutton

\[
\text{exec} \ (\text{comp}' \ (\text{Add} \ x \ y) \ c) \ (s, q) \\
= \ \{ \ \text{specification (14)} \ \}
\]

\[
\text{case eval x q of}
\]

\[
\begin{align*}
\text{Just } (n, q') & \to \text{ case eval y q' of} \\
& \quad \text{Just } (m, q'') \to \text{exec} \ c \ (\text{VAL} \ (n + m) : s, q'') \\
& \quad \text{Nothing} \to \text{fail} \ s
\end{align*}
\]

\[
\text{Nothing} \to \text{fail} \ s
\]

\[
= \ \{ \ \text{define: exec} \ (\text{ADD} \ c) \ (\text{VAL} \ m : \text{VAL} \ n : s, q'') = \text{exec} \ c \ (\text{VAL} \ (n + m) : s, q'') \ \}
\]

\[
\text{case eval x q of}
\]

\[
\begin{align*}
\text{Just } (n, q') & \to \text{ case eval y q' of} \\
& \quad \text{Just } (m, q'') \to \text{exec} \ (\text{ADD} \ c) \ (\text{VAL} \ m : \text{VAL} \ n : s, q'') \\
& \quad \text{Nothing} \to \text{fail} \ (\text{VAL} \ n : s)
\end{align*}
\]

\[
\text{Nothing} \to \text{fail} \ s
\]

\[
= \ \{ \ \text{define: fail} \ (\text{VAL} \ n : s) = \text{fail} \ s \ \}
\]

\[
\text{case eval x q of}
\]

\[
\begin{align*}
\text{Just } (n, q') & \to \text{exec} \ (\text{comp}' \ y \ (\text{ADD} \ c)) \ (\text{VAL} \ n : s, q') \\
& \quad \text{Nothing} \to \text{fail} \ s
\end{align*}
\]

\[
= \ \{ \ \text{induction hypothesis for } y \ \}
\]

\[
\text{case eval x q of}
\]

\[
\begin{align*}
\text{Just } (n, q') & \to \text{exec} \ (\text{comp}' \ (\text{ADD} \ c)) \ (s, q) \\
& \quad \text{Nothing} \to \text{exec} \ (\text{comp}' \ h \ c) \ (s, q)
\end{align*}
\]

\[
= \ \{ \ \text{define: fail} \ (\text{HAN} \ c' \ q : s) = \text{exec} \ c' \ (s, q) \ \}
\]

\[
\text{case eval x q of}
\]

\[
\begin{align*}
\text{Just } (n, q') & \to \text{exec} \ c \ (\text{VAL} \ n : s, q') \\
& \quad \text{Nothing} \to \text{exec} \ (\text{comp}' \ h \ c) \ (s, q)
\end{align*}
\]

The case for Catch is more interesting this time. In the calculation for the global state semantics it was straightforward to bring the configuration arguments into the right form to apply the induction hypotheses. With local state, however, when an exception handler is invoked we require access to the state that was in place when the enclosing Catch was entered, which information we communicate via the stack:

\[
\text{exec} \ (\text{comp}' \ (\text{Catch} \ x \ h) \ c) \ (s, q) \\
= \ \{ \ \text{specification (14)} \ \}
\]

\[
\text{case eval x q of}
\]

\[
\begin{align*}
\text{Just } (n, q') & \to \text{exec} \ c \ (\text{VAL} \ n : s, q') \\
& \quad \text{Nothing} \to \text{case eval h q of} \\
& \quad \quad \text{Just } (m, q'') \to \text{exec} \ c \ (\text{VAL} \ m : s, q'') \\
& \quad \quad \text{Nothing} \to \text{fail} \ s
\end{align*}
\]

\[
= \ \{ \ \text{induction hypothesis for } h \ \}
\]

\[
\text{case eval x q of}
\]

\[
\begin{align*}
\text{Just } (n, q') & \to \text{exec} \ c \ (\text{VAL} \ n : s, q') \\
& \quad \text{Nothing} \to \text{exec} \ (\text{comp}' \ h \ c) \ (s, q)
\end{align*}
\]

\[
= \ \{ \ \text{define: fail} \ (\text{HAN} \ c' \ q : s) = \text{exec} \ c' \ (s, q) \ \}
\]

\[
\text{case eval x q of}
\]

\[
\begin{align*}
\text{Just } (n, q') & \to \text{exec} \ c \ (\text{VAL} \ n : s, q')
\end{align*}
\]
Nothing \rightarrow \text{fail} (\text{HAN} (\text{comp}' h c) q : s)

= \{ \text{define: exec (UNMARK c) (VAL n: HAN } \neg_{-} : s, q') = \text{exec c (VAL n: s, q')} \}

\text{case} \text{eval} x q \text{ of}

Just (n, q') \rightarrow \text{exec (UNMARK c) (VAL n: HAN (comp' h c) q : s, q')}

Nothing \rightarrow \text{fail} (\text{HAN} (\text{comp}' h c) q : s)

= \{ \text{induction hypothesis for } x \}

\text{exec (comp' x (UNMARK c)) (HAN (comp' h c) q : s, q)}

= \{ \text{define: exec (MARK } c'' c'') (s, q) = \text{exec c' (HAN } c'' q : s, q) \}

\text{exec (MARK (comp' h c) (comp' x (UNMARK c))) (s, q)}

Note that the new constructor HAN added to the Elem type within this calculation now has two arguments: one for the handler code (as in previous calculations), and one for the state to be used if the handler is invoked (for local state). We conclude the calculation with the case for Put, which proceeds in the same manner as for global state:

\text{exec (comp' (Put x y) c) (s, q)}

= \{ \text{specification [14]} \}

\text{case} \text{eval} x q \text{ of}

Just (n, q') \rightarrow \text{case eval y n of}

Just (m, q'') \rightarrow \text{exec c (VAL m : s, q'')}

Nothing \rightarrow \text{fail } s

Nothing \rightarrow \text{fail } s

= \{ \text{induction hypothesis for } y \}

\text{case} \text{eval} x q \text{ of}

Just (n, q') \rightarrow \text{exec (comp' y c) (s, n)}

Nothing \rightarrow \text{fail } s

= \{ \text{define: exec (SAVE c') (VAL n : s, q') = exec c' (s, n) } \}

\text{case} \text{eval} x q \text{ of}

Just (n, q') \rightarrow \text{exec (SAVE (comp' y c)) (VAL n : s, q')}

Nothing \rightarrow \text{fail } s

= \{ \text{induction hypothesis for } x \}

\text{exec (comp' x (SAVE (comp' y c))) (s, q)}

In summary, collecting together everything that we have learned in the process of the above calculations, we obtained the following definitions.

Target language:

\textbf{data} \ Code = \text{HALT} | \text{PUSH Int Code} | \text{ADD Code} | \text{FAIL} | \text{MARK Code Code} | \text{UNMARK Code} | \text{LOAD Code} | \text{SAVE Code} \\

Compiler:

\text{comp} :: \text{Expr} \rightarrow \text{Code} \\
\text{comp} x = \text{comp' x HALT} \\
\text{comp'} :: \text{Expr} \rightarrow \text{Code} \rightarrow \text{Code} \\
\text{comp' (Val n) c} = \text{PUSH n c}
\[\text{comp}'(\text{Add } x \ y) \ c = \text{comp}' \ x \ (\text{comp}' \ y \ (\text{ADD } c))\]
\[\text{comp}' \ \text{Throw } c = \text{FAIL}\]
\[\text{comp}'(\text{Catch } x \ h) \ c = \text{MARK} \ (\text{comp}' \ h \ c) \ (\text{comp}' \ x \ (\text{UNMARK } c))\]
\[\text{comp}' \ \text{Get } c = \text{LOAD } c\]
\[\text{comp}'(\text{Put } x \ y) \ c = \text{comp}' \ x \ (\text{SAVE} \ (\text{comp}' \ y \ c))\]

**Virtual machine:**

\[
\text{data} \quad \text{Elem} \quad = \text{VAL } \text{Int} \mid \text{HAN } \text{Code} \mid \text{State} \]

\[
\text{exec} \quad = \text{Code} \rightarrow \text{Conf} \rightarrow \text{Conf} \]

\[
\text{exec } \text{HALT } (s,q) = (s,q) \]

\[
\text{exec } (\text{PUSH } n \ c ) \ (s,q) = \text{exec } c \ (\text{VAL } n : s,q) \]

\[
\text{exec } (\text{ADD } c ) \ (\text{VAL } m : \text{VAL } n : s,q) = \text{exec } c \ (\text{VAL } (n + m) : s,q) \]

\[
\text{exec } \text{FAIL } (s,q) = \text{fail } s \]

\[
\text{exec } (\text{MARK } h \ c ) \ (s,q) = \text{exec } c \ (\text{HAN } h \ q : s,q) \]

\[
\text{exec } (\text{UNMARK } c ) \ (\text{VAL } n : \text{HAN } _-_- : s,q) = \text{exec } c \ (\text{VAL } n : s,q) \]

\[
\text{exec } (\text{LOAD } c ) \ (s,q) = \text{exec } c \ (\text{VAL } q : s,q) \]

\[
\text{exec } (\text{SAVE } c ) \ (\text{VAL } n : s,q) = \text{exec } c \ (s,n) \]

\[
\text{fail} \quad = \text{Stack} \rightarrow \text{Conf} \]

\[
\text{fail } [] = ([],0) \]

\[
\text{fail } (\text{VAL } n : s) = \text{fail } s \]

\[
\text{fail } (\text{HAN } h \ q : s) = \text{exec } h \ (s,q) \]

Note that, as previously, we added an equation to \text{fail} for the case when the stack is empty in order to make the definition complete. Because \text{fail} does not take a state as an argument, we can only give a fixed output state as the result, for which purposes we simply return the value 0. As before, the choice for this additional equation has no impact on the correctness of the above calculations because they do not depend on this choice.