Nullable Type Inference

Michel Mauny  Benoît Vaugon

Unité d’Informatique et d’Ingénierie des Systèmes (U2IS)
ENSTA-ParisTech
France

{michel.mauny,benoit.vaugon} <at> ensta-paristech.fr

We present type inference algorithms for nullable types in ML-like programming languages. Starting with a simple system, presented as an algorithm, whose only interest is to introduce the formalism that we use, we replace unification by subtyping constraints and obtain a more interesting system. We state the usual properties for both systems. This is work in progress.

1 Nullable vs. option types

Imperative programming languages, such as C or Java derivatives, make abundant use of NULL either as a value for unknown or invalid references, or as failure return values. Using NULL is rather practical, since the if statement suffices for checking NULL-ity. Of course, the downside of having NULL as a possible value is that, without further support, it could accidentally be confused with a legal value, leading to execution errors [7].

In languages using the ML type discipline, the option type

\[
\text{type } \alpha \text{ option } = \text{None | Some of } \alpha
\]

injects regular values and the nullary data constructor None, into a single type that could be thought as a nullable type. The type system guarantees that options cannot be confused with regular values, and pattern-matching is used to check and extract regular values from options. In Haskell, the “maybe” data type is heavily used for representing successes and failures. The JaneStreet Core library [9] uses the option type to show possible failures in function types: it purposely avoids exceptions, because their non-appearance in OCaml type to show possible failures in function types: it purposely avoids exceptions, because their non-appearance in OCaml type system hides the partiality of functions. In OCaml, None and Some are automatically inserted by the compiler for dealing with optional arguments [5].

Using the option data type presents the disadvantage of allocating memory blocks for representing Some(v) values, which may refrain experienced programmers from heavily using options. Avoiding those memory allocations is not really difficult: if Some(v) is represented as v itself, that is, without allocating a block tagged as Some, we need a special representation for None, distinct from the representation of v, for any v. Of course, when v is None itself, it then becomes impossible to distinguish Some(None) from None. Therefore, None is not the only special case that needs a special treatment: the whole range of Some(\(n\)) values, for \(n \geq 0\) needs a special representation such that Some(\(\text{None}\)) cannot be confused with Some(\(\text{None}\)) when \(i \neq j\).

For instance, unaligned addresses on 64-bits architectures, or a statically pre-allocated array of a sufficient size could do the job.¹

¹Although polymorphic recursion theoretically allows for unbounded depth of Some(\(n\)) (None) while this representation allows for representing only a finite number of \(n\), this limitation should never be met in practice.

Now, the compilation of Some(expr) needs to generate a test, in order to use the special representation of Some(\(n\)) (None) when expr evaluates to None, and pattern-matching against Some/None also needs to be adjusted.

This paper does not aim at opposing nullable types to option types. Options, in the Hindley-Milner type discipline, offer not only type safety, but also precision by distinguishing Some(\(n\)) (None) from None, but at the price of a memory allocation or a dynamic test for Some. On the other hand, nullable types extend any classical type \(i\) into \(?\), to include NULL. Such “nullable values” are easier to represent and compile than options, but offer less precision since it makes no sense to extend further \(?\). Also, their static inference haven’t received much attention, so far. Indeed, although quite a few recent programming languages statically check the safety of NULL [3, 2, 11, 1], none of them really performs type inference in the ML sense, but rather local inference, propagating mandatory type annotations of function parameters inside the function bodies.

2 Nullable type inference

The purpose of this work is to study type inference of nullable types, by adding them as a feature in a small functional language. The language that we consider, given in figure 1, is a classical mini-ML, extended with NULL test and creation. Section 3 starts with a naïve approach, where the types \(\tau\)

\[
\begin{align*}
  c & ::= 0 | \text{true} | \text{false} | 0 | 1 | 2 | \ldots | + | \ldots \\
  e & ::= c | x | \lambda x. e | e_1 \cdot e_2 | \text{if } e_1 \text{ then } e_2 \text{ else } e_3 \\
  & \quad \mid \text{let } x = e_1 \text{ in } e_2 \\
  & \quad \mid \text{NULL } \mid \text{case } e_1 \text{ of } \text{NULL } \rightarrow e_2 \mid x \rightarrow e_3
\end{align*}
\]

Figure 1: The language (that are assigned to expressions \(e\)) are pairs \((t, v)\) of a usual type \(t\) and a “nullability” type information \(v\). A type \((t, ?)\) corresponds to values that may be NULL, whereas \((t, \Lambda)\) denotes values that cannot be NULL. Nullability variables are written \(\delta\). This system is mainly used to introduce the formalism that we use for writing our algorithms.

Section 4 shows a translation algorithm that encode nullable values with polymorphic variants. Typing the translated programs with a unification-based mechanism suffers the same weakness as our naïve type system.

Section 5 presents a more sophisticated typing mechanism, where unification is replaced by subtyping constraints.
The operational semantics needed for stating the execution of a conditional expression was expected of type \([< 'Some of int \]).

Error: this expression has type \([> 'None \]) but an expression was expected of type \([< 'Some of int \]).

This is due to the fact that the type inference of OCaml polymorphic variants use unification, even though it emulates some form of subtyping with a rich type algebra \([5]\).

We have extended the work of Garrigue in order to have a more flexible and powerful type system for polymorphic variants. Although this is still work in progress, we show here how to apply this result to nullable types.

5 A subtyping approach

We saw in the example above that the propagation of information “backwards”, by unification in the typing environment, prevents typing some programs that could be perfectly acceptable.

Replacing unification, which comes from type equality constraints, by inequality constraints, that is, by subtyping, relaxes the programming style imposed by using type unification. While this is clearly more permissive, the resolution of inequality constraints may still fail. On the one hand, some unification constraints remain hidden as double inequalities (e.g. when trying to type \(1 + "hello"\)). On the other hand, some inequalities are clearly not satisfiable, such as those produced when typing a conditional \(\text{if } \text{NULL} \text{ then } \ldots \text{ else } \ldots\)
where one fails to prove \( a? \leq \text{bool} \), or an application \( \text{NULL}(\ldots) \), failing to prove \( a? \leq t_1 \rightarrow t_2 \). Also, many primitive operations (like \( + \)) won’t accept nullable arguments.

We start to change the type algebra of our language and introduce the syntax “?!” for nullable types, figure 10.

The new set of typing rules is given in figure 6. The essential change that we bring to our initial system is in the TAPP rule, where the domain type of the function is constrained to be “larger” than the type of the argument in order to accept it. Intuitively, a function accepting a possibly null value as argument, accepts also a provably non-null argument.

The inequality constraints, written \( t_1 \geq t_2 \), are integrated in the \( \Phi \) component of the typing rules by the set of resolution rules given in figure 7. At resolution time, newly integrated constraints are checked to be consistent with constraints existing in \( \Phi \). In particular, resolution performs the basic subtyping checks through rules \( \text{LEQBASENULL} \) and \( \text{LEQARROWNULL} \), on figure 7(b).

When a type variable \( a \) has to be “smaller” (resp. “greater”) than two types \( t_1 \) and \( t_2 \), the \( \tau \) become constrained to be “compatible” (see figure 8), that is to differ only in their (possibly internal) “?” annotations.

Generalization (figure 11) universally quantifies type variables \( a \) together with its associated constraints, written \( \Phi_a \), when the set \{ \( a \) \} \( \cup \text{FTV}(\Phi_a) \) does not intersect the set of free type variables occurring in \( \Gamma \).

At instanciation time, a fresh instance of constraints is re-injected in \( \Phi \).

Another important change in the new system concerns the conditional, case selection (and, more generally pattern-matching constructs). Instead of unifying the types of all branches to be compatible, but no more. See for instance rules \( \text{TIFTHENELSE} \) and \( \text{TCASE} \) on figure 6. This is precisely the reason why the example given at the end of section 3 is accepted by the new system.

6 Properties

We have a prototype implementation of this typing algorithm, and the proof of correctness of the extension of Garrigue’s typing of polymorphic variants, in which this algorithm can easily be translated. The proof is available online at https://github.com/bvaugon/variants/

7 Conclusion

We have presented two type systems and a translation algorithm aiming at inferring nullable types in ML-like languages.
The first type system, rather naive and interesting by its simplicity, is probably too restrictive to be usable by daily programmers. The translation technique using standard polymorphic variants has the same weakness. However, exchanging unification against subtyping provides us with a more expressive type system. Soundness and termination properties have been checked.

References


(a) Main comparison rules

\[
\begin{array}{c}
\text{GEQ} \\
\Phi \vdash \tau_2 \leq \tau_1 \Rightarrow \Phi' \\
\Phi \vdash \tau_1 \geq \tau_2 \Rightarrow \Phi'
\end{array}
\]

\[
\begin{array}{c}
\text{EQ} \\
\Phi \vdash \tau_1 \leq \tau_2 \Rightarrow \Phi' \quad \Phi' \vdash \tau_1 \geq \tau_2 \Rightarrow \Phi'' \\
\Phi \vdash \tau_1 = \tau_2 \Rightarrow \Phi''
\end{array}
\]

\[
\begin{array}{c}
\text{LEQNEW} \\
\text{when } \tau_1 \leq \tau_2 \not\in \Phi \\
\Phi, \tau_1 \leq \tau_2 \vdash \tau_1 \leq \tau_2 \Rightarrow \Phi'
\end{array}
\]

\[
\begin{array}{c}
\text{LEQA} \\
\Phi, \tau_1 \leq \tau_2 \vdash \tau_1 \leq \tau_2 \Rightarrow \Phi, \tau_1 \leq \tau_2
\end{array}
\]

(b) Standard comparison rules

\[
\begin{array}{c}
\text{LEQBASETY} \\
\Phi \vdash f_3 \leq f_2 \Rightarrow \Phi
\end{array}
\]

\[
\begin{array}{c}
\text{LEQARROW} \\
\Phi \vdash \tau_1 \leq \tau_2 \Rightarrow \tau_1' \leq \tau_2' \Rightarrow \phi \\
\Phi' \vdash \tau_1 \leq \tau_2 \Rightarrow \tau_1' \leq \tau_2' \Rightarrow \phi
\end{array}
\]

\[
\begin{array}{c}
\text{LEQBASENULL} \\
\Phi \vdash f_3 \leq f_2 \Rightarrow \phi
\end{array}
\]

\[
\begin{array}{c}
\text{LEQARROWNULL} \\
\Phi \vdash \tau_1 \leq \tau_2 \Rightarrow \tau_1' \leq \tau_2' \Rightarrow \phi
\end{array}
\]

(c) Type-variable comparison rule

\[
\begin{array}{c}
\text{LEQVARLEQTY} \\
\Phi, \alpha \leq \alpha \Rightarrow \Phi
\end{array}
\]

\[
\begin{array}{c}
\text{GEOVARLEQTY} \\
\Phi, \alpha \leq \alpha \Rightarrow \tau \leq \tau' \Rightarrow \alpha \leq \phi
\end{array}
\]

\[
\begin{array}{c}
\text{LEQTYLEQVAR} \\
\Phi, \alpha \leq \alpha \Rightarrow \tau \leq \tau' \Rightarrow \phi
\end{array}
\]

\[
\begin{array}{c}
\text{GEOXTYLEQVAR} \\
\Phi, \alpha \leq \alpha \Rightarrow \tau \leq \tau' \Rightarrow \phi
\end{array}
\]

\[
\begin{array}{c}
\text{LEQUVARCTPY} \\
\Phi, \alpha \leq \alpha \Rightarrow \tau \leq \tau' \Rightarrow \phi
\end{array}
\]

\[
\begin{array}{c}
\text{GEQVARCTPY} \\
\Phi, \alpha \leq \alpha \Rightarrow \tau \leq \tau' \Rightarrow \phi
\end{array}
\]

\[
\begin{array}{c}
\text{LEQVAREND} \\
\text{when } \tau' \neq \alpha \text{ and } \tau \leq \tau' \not\in \Phi \\
\Phi \vdash \alpha \leq \phi
\end{array}
\]

\[
\begin{array}{c}
\text{GEQVAREND} \\
\text{when } \tau' \neq \alpha \text{ and } \tau \leq \tau' \not\in \Phi \\
\Phi \vdash \alpha \leq \phi
\end{array}
\]

Figure 7: Comparison rules
(a) Main compatibility rules

\[
\begin{align*}
\text{CPTNew} & : \frac{\tau_1 \not\in \Phi \land \Phi \vdash \tau_1 \rightarrow \tau_2 \land \Phi' \vdash \tau_2 \rightarrow \Phi'}{\Phi \vdash \tau_1 \rightarrow \tau_2 \rightarrow \Phi'} \\
\text{CPTAlreadyProved} & : \frac{\Phi \vdash \tau_1 \rightarrow \tau_2 \rightarrow \Phi'}{\Phi \vdash \tau_1 \rightarrow \tau_2 \rightarrow \Phi'}
\end{align*}
\]

(b) Standard compatibility rules

\[
\begin{align*}
\text{CPTBaseTy} & : \frac{\Phi \vdash b \rightarrow \Phi}{\Phi \vdash b \rightarrow \Phi} \\
\text{CPTArrow} & : \frac{\frac{\Phi \vdash \alpha \rightarrow \chi \rightarrow \Psi \rightarrow \Phi'}{\Phi' \vdash \tau_1 \rightarrow \tau_2 \rightarrow \Phi'}}{\Phi \vdash \tau_1 \rightarrow \tau_2 \rightarrow \Phi' \rightarrow \Phi''} \\
\text{CPTArrowNull} & : \frac{\Phi \vdash \tau_1 \rightarrow \tau_2 \rightarrow (\tau_1 \rightarrow \tau_2)' \rightarrow \Phi''} {\Phi \vdash \tau_1 \rightarrow \tau_2 \rightarrow (\tau_1 \rightarrow \tau_2)' \rightarrow \Phi''} \\
\text{CPTBaseNull} & : \frac{\Phi \vdash b \rightarrow \Phi}{\Phi \vdash b \rightarrow \Phi}
\end{align*}
\]

(c) Type-variable compatibility rules

\[
\begin{align*}
\text{CPTSameVar} & : \frac{\Phi \vdash \alpha 
\end{align*}
\]

Figure 8: Compatibility rules

\[
\begin{align*}
[\text{NULL}] & = \text{'None} \\
[x] & = x \\
[c] & = \text{'Some}(c) \\
[\text{if } c_1 \text{ then } c_2 \text{ else } c_3] & = \begin{cases} c_1 \text{ with } \text{'Some}(b) \rightarrow b & \text{if } c_1 \end{cases} & \begin{cases} c_2 \text{ else } \text{[c1]} \\
[\lambda x. c] & = \text{'Some}(\lambda x. c) \\
[\text{case } c_0 \text{ of } \text{NULL} \rightarrow c_2 \mid x \rightarrow c_3] & = \begin{cases} \text{match } [c_1] \text{ with } & \text{[c1]} \text{ with } \text{'Some}(f) \rightarrow f \rightarrow c_2 \end{cases} \text{[c2]} \\
[c_1, c_2] & = \begin{cases} \text{match } [c_1] \text{ with } & \text{match } [c_1] \text{ with } \text{'Some}() \rightarrow [c_2] \end{cases} \text{[c2]} & \text{'Some}(x) \rightarrow (\lambda x. [c_2]) ('\text{Some}(x))
\end{cases}
\end{align*}
\]

Figure 9: Encoding of NULL as variants

\[
\begin{align*}
t & ::= t_0 \mid a \mid t_1 \rightarrow t_2 \quad \tau ::= t \mid t_1 \mid t_2 \\
\end{align*}
\]

Figure 10: Nullable types


