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.

In languages using the ML type discipline, the option type

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

injests 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 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.

Section 3 starts with a naïve approach, where the types \(t\) and \(\nu\) 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 \text{Some(None)} from None, but at the price of a memory allocation or a dynamic test for \text{Some}. On the other hand, nullable types extend any classical type \(t\) 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, 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 \(\tau\) and \(\nu\) needs to generate a test, in order to use the special representation of \text{Some}(\text{None}) when expr evaluates to \text{None}, and pattern-matching against \text{Some}/\text{None} also needs to be adjusted.

Figure 1: The language (that are assigned to expressions \(e\)) are pairs \((t,\nu)\) of a usual type \(t\) and a “nullability” type information \(\nu\). A type \((t,?\)) corresponds to values that may be NULL, whereas \((t,\Delta)\) 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.
3 A simple type system

We first present a rather simple type system, where the types carry a “nullability information” saying whether the value of an expression may be NULL or not.

The judgements of our language’s type system are of the form $\Phi, \Gamma \vdash e : (\tau, \nu) \triangleright \Phi'$, where $\Phi$ and $\Phi'$ are substitutions, $\Gamma$ is a type environment that maps program identifiers to type schemes (types with a prefix universal quantification of type and nullability variables). Such a judgement should be read as: given $\Phi$, under assumption $\Gamma$, the expression $e$ has type $\tau$ with substitution $\Phi'$.

The rules should be read as the different cases of an algorithm that, given $\Phi$, $\Gamma$, $e$, and $\tau$, computes substitution $\Phi'$ which, when applied to $\tau$ and $\Gamma$, assign the type $\Phi'(\tau)$ to $e$. In other words, under assumptions $\Phi'(\Gamma)$, $e$ has type $\Phi'(\tau)$.

Because we have two kinds of variables, we have two instantiation mechanisms, in distinct rules: TInstVar for universally quantified type variables, and TInst(δ) for universally quantified nullability variables.

Type equality constraints, introduced by typing rules, are solved using a set of rules displayed in figure 3. The only interesting resolution rules are the ones that introduce (EQNew) or merge (EQMerge) variable bindings in the $\Phi$ substitution.

EQNew(a), installs a binding $a = t'$ in the substitution $\Phi[a \rightarrow t']$ (that is, $\Phi$ in which free occurrences of $a$ are replaced by $t'$) when there is no previous binding about $a$ in $\Phi$. Here, $t'$ is either $\mu \alpha.t$ or $t$, depending on whether $a$ occurs or not in $t$.

EQNew(δ) does the same, in a simpler context, on nullability variables.

EQMerge(a) merges a new constraint $a = t'$ to a previous $a = t$ occurring in $\Phi$. Here, the resulting substitution $\Phi'$ is obtained by resolving the constraint $t = t'$. EQMerge does the same for nullability variables.

Correctness. The operational semantics needed for stating the correction of the type system is also classical. NULL cannot handle operations such as being called as a function, tested, etc. The typing of values, produced by correct executions, is also slightly changed: the type $(\tau, \nu)$ includes NULL as well as regular values. The correctness property can be stated as follows:

If $\Phi, \Gamma \vdash e : \tau \triangleright \Phi'$, then the evaluation of $e$ in an execution environment compatible with $\Phi'(\Gamma)$ produces a value typable with $\Phi'(\tau)$, if evaluation terminates.

Although this type system is simple and may seem practical,
it imposes a certain style of programming where NULL values are as much isolated as possible from other parts of the program. In a ML-like language, where NULL (or similar construct such as None) are not as commonly used as in C or Java, this might be acceptable. Still, one might want the following example to be typable:

```
let f b k =
  let p = k + 1 in
  if b then k else NULL in
...
```

This program cannot typecheck since the `k` parameter must at the same time have type `(int, α)` in order to be sent to `+` and `(int, ?)` in order to have the same type as NULL.

## 4 Encoding nullability as variants

<table>
<thead>
<tr>
<th>EqSplit</th>
<th>EqBase</th>
<th>EqArrow</th>
<th>EqTrivial(α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Phi \vdash t_1 = t_2 \rightarrow \Phi')</td>
<td>(\Phi \vdash \lambda x. v \rightarrow \Phi')</td>
<td>(\Phi \vdash \tau_1 \rightarrow \tau_2 \rightarrow \tau' \rightarrow \Phi'')</td>
<td>(\Phi \vdash \alpha = \alpha \rightarrow \Phi)</td>
</tr>
<tr>
<td>(\Phi \vdash (t_1, v_1) = (t_2, v_2) \rightarrow \Phi'')</td>
<td>(\Phi \vdash \lambda x. v \rightarrow \Phi')</td>
<td>(\Phi \vdash \tau_1 \rightarrow \tau_2 \rightarrow \tau_1' \rightarrow \Phi'')</td>
<td>(\Phi \vdash \alpha = \alpha \rightarrow \Phi)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EqNew(α)</th>
<th>EqMerge(α)</th>
<th>EqTrivial(δ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Phi \vdash t \rightarrow \Phi[\alpha \mapsto t'] \oplus \alpha = t')</td>
<td>(\Phi \vdash \delta \rightarrow \delta \rightarrow \Phi)</td>
<td>(\Phi \vdash \delta = \delta \rightarrow \Phi)</td>
</tr>
<tr>
<td>(\Phi \vdash \delta \rightarrow \Phi[\delta \mapsto v] \oplus \delta = v)</td>
<td>(\Phi \vdash \delta \rightarrow \delta \rightarrow \Phi')</td>
<td>(\Phi \vdash \delta = \delta \rightarrow \Phi)</td>
</tr>
</tbody>
</table>

### Figure 5: Resolution rules

As expected, the translation of the above example does not pass OCaml typechecking:

```ocaml
let f = Some (fun b k ->
  let p = (match (+) with
    | Some f_0 -> f_0) k ('Some 1) in
  if match b with Some b_0 -> b_0
  then k else None) in ...
```

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 [6].

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 if NULL then ... else ..., where one fails to prove `a? ∈ bool`, or an application NULL(...), failing to prove `a? ≤ τ_1 → τ_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 [9]. The essential change that we bring to our initial system is in the TAPP
null expressions. Intuitively, a function accepting a possibly null value as argument, accepts also a provably non-null argument.

The inequality constraints, written $\tau_1 \triangleright \tau_2$, 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 $\alpha$ has to be “smaller” (resp. “greater”) than two types $\tau_1$ and $\tau_2$, the $\tau_i$ 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 $\alpha$ together with its associated constraints, written $\Phi_{\alpha}$, when the set $\{ \alpha \} \cup \text{FTV}(\Phi_{\alpha})$ does not intersect the set of free type variables occurring in $\Gamma$.

At instantiation 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, each of them is constrained to be “smaller” than the type of the construct itself. This forces all types of the 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.
(a) Main comparison rules

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEQ</td>
<td>$\Phi \vdash \tau_2 \leq \tau_1 \Rightarrow \Phi'$</td>
</tr>
<tr>
<td>EQ</td>
<td>$\Phi \vdash \tau_1 \leq \tau_2 \Rightarrow \Phi'$</td>
</tr>
<tr>
<td>LEQNEW</td>
<td>when $\tau_1 \leq \tau_2 \not\leq \Phi$  $\Phi, \tau_1 \leq \tau_2 \vdash \tau_1 \leq \tau_2 \Rightarrow \Phi'$</td>
</tr>
</tbody>
</table>

(b) Standard comparison rules

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEQBASETY</td>
<td>$\Phi \vdash \tau \leq \tau \Rightarrow \Phi$</td>
</tr>
<tr>
<td>LEQARROW</td>
<td>$\Phi \vdash \tau_1 \leq \tau_2 \Rightarrow \tau_2 \leq \tau_2 \Rightarrow \Phi'$</td>
</tr>
<tr>
<td>LEQBASENULL</td>
<td>$\Phi \vdash \tau \leq \tau \Rightarrow \Phi'$</td>
</tr>
<tr>
<td>LEQARROWNULL</td>
<td>$\Phi \vdash \tau_1 \leq \tau_2 \Rightarrow \tau_2 \leq \tau_2 \Rightarrow \Phi'$</td>
</tr>
</tbody>
</table>

(c) Type-variable comparison rule

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEQSAMEVAR</td>
<td>$\Phi \vdash \alpha \leq \alpha \Rightarrow \Phi$</td>
</tr>
<tr>
<td>GEQVARLEQTY</td>
<td>when $\tau \not\equiv \alpha$  and  $\tau \not\equiv \tau' \not\equiv \alpha$  $\Phi, \alpha \leq \tau, \tau \leq \tau' \Rightarrow \alpha \leq \tau' \Rightarrow \Phi'$</td>
</tr>
<tr>
<td>GEQTYLEQVAR</td>
<td>when $\tau \not\equiv \alpha$  and  $\tau \not\equiv \tau' \not\equiv \alpha$  $\Phi, \alpha \leq \tau, \tau \equiv \tau' \Rightarrow \alpha \leq \tau' \Rightarrow \Phi'$</td>
</tr>
<tr>
<td>LEQTYLEQVAR</td>
<td>when $\tau \not\equiv \alpha$  and  $\tau \not\equiv \tau' \not\equiv \alpha$  $\Phi, \alpha \leq \tau, \tau \equiv \tau' \Rightarrow \alpha \leq \tau' \Rightarrow \Phi'$</td>
</tr>
<tr>
<td>LEQVARCPTTY</td>
<td>when $\tau \not\equiv \alpha$  and  $\tau \not\equiv \tau' \not\equiv \alpha$  $\Phi, \alpha \equiv \tau, \tau \equiv \tau' \Rightarrow \alpha \leq \tau' \Rightarrow \Phi'$</td>
</tr>
<tr>
<td>GEQVARCPTTY</td>
<td>when $\tau \not\equiv \alpha$  and  $\tau \not\equiv \tau' \not\equiv \alpha$  $\Phi, \alpha \equiv \tau, \tau \equiv \tau' \Rightarrow \alpha \leq \tau' \Rightarrow \Phi'$</td>
</tr>
</tbody>
</table>

Figure 7: Comparison rules
(a) Main compatibility rules

\[
\begin{array}{c|c|c}
\text{CPTNEW} & \text{CPTALREADYPROVED} \\
\hline
\text{when } \tau_1 \approx \tau_2 \notin \Phi & \Phi, \tau_1 \approx \tau_2 \Rightarrow \Phi' \\
\hline
\Phi, \tau_1 \approx \tau_2 \Rightarrow \Phi' \\
\hline
\end{array}
\]

(b) Standard compatibility rules

\[
\begin{array}{c|c|c}
\text{CptBaseTy} & \text{CptArrow} & \text{CptBaseNull} \\
\hline
\Phi \vdash t_3 \approx t_3 \Rightarrow \Phi & \Phi \vdash t_1 \Rightarrow \tau_1 \approx \tau_2 \Rightarrow \tau_1 \approx \tau_2 \Rightarrow \Phi' & \Phi \vdash \emptyset \Rightarrow \Phi \\
\hline
\Phi \vdash t_1 \Rightarrow \tau_1 \approx \tau_1 \Rightarrow \tau_1 \approx \tau_1 \Rightarrow \Phi' & \Phi \vdash \tau \Rightarrow \tau_2 \Rightarrow \tau \Rightarrow \Phi' & \phi \vdash \emptyset \Rightarrow \Phi \\
\hline
\end{array}
\]

(c) Type-variable compatibility rules

\[
\begin{array}{c|c|c}
\text{CptSameVar} & \text{CptVarEqTy} & \text{CptVarCptTy} \\
\hline
\Phi, \alpha \Rightarrow \alpha \Rightarrow \Phi & \text{when } \tau \neq \alpha \text{ and } \tau' \notin \Phi & \Phi, \alpha \Rightarrow \tau, \tau \Rightarrow \tau \Rightarrow \tau \Rightarrow \Phi' \\
\hline
\Phi, \alpha \Rightarrow \tau, \tau \Rightarrow \alpha \Rightarrow \tau' \Rightarrow \Phi' & \text{when } \tau \neq \alpha \text{ and } \tau \Rightarrow \tau' \notin \Phi & \Phi, \alpha \Rightarrow \tau, \tau \Rightarrow \alpha \Rightarrow \tau' \Rightarrow \Phi' \\
\hline
\Phi, \alpha \Rightarrow \tau, \tau \Rightarrow \alpha \Rightarrow \tau' \Rightarrow \Phi' & \text{when } \tau \neq \alpha & \Phi, \alpha \Rightarrow \tau \Rightarrow \alpha \Rightarrow \tau' \Rightarrow \Phi' \\
\hline
\Phi, \alpha \Rightarrow \tau \Rightarrow \alpha \Rightarrow \tau' \Rightarrow \Phi' \\
\hline
\end{array}
\]

Figure 8: Compatibility rules

References


