3.3 Type Inference

Let us assume we can define a interpretation function

$$[\![.]\!]: \operatorname{values}_{\Gamma}(\texttt{}) \cup \operatorname{values}_{\Gamma}(\texttt{}) \to A$$

where Γ is a type context and A is a non-empty set. We will discuss the definition of such a function as well as the definition of A in the next section.

In this section we are more interested in the judgment $[e] \in \text{values}_{\Gamma}(T)$ for a given type T a type context Γ and an expression e. For now let us assume that the definition of A actually allows this.

3.3.1 Typing Judgments

A judgment can arise from a set of logical inference rules [Pie+02].

Definition 3.1: inference rules

Let $n \in \mathbb{N}$, P_i be sequents for all $i \in \mathbb{N}$. Let C be a sequent.

TA7- --11

We call

$$\frac{P_1 \dots P_n}{C}$$

an inference rule.

If a judgment holds, then there exists a set of logical inference rules that proves it.

We will see that the initial judgment ($\llbracket \mathtt{e} \rrbracket \in \mathrm{values}_{\Gamma}(T)$) is dependent on another context, this time for variables instead of types:

Definition 3.2: Variable Context

 $\Delta \in \mathcal{V} \nrightarrow \mathcal{T}$ is called the *variable context*.

We also introduce a new syntax for saying a value is of some type.

Definition 3.3: type of value

Let $T \in \mathcal{T}$. Let Γ be a type context. Let e be arbitary.

We say e is of type T in the context of Γ (Notation: $e :_{\Gamma} T$): \Leftrightarrow

 $e \in \text{values}_{\Gamma}(T)$

Is $\Gamma = \emptyset$ then we may write e : T instead of $e :_{\Gamma} T$.

The judgment in question can now be expressed more generally as

$$\Delta \vdash \llbracket \mathbf{e} \rrbracket :_{\Gamma} T$$

where Γ is a type context, Δ is a variable context, $e \in \text{values}_{\Gamma}(\text{sprogram}) \cup \text{values}_{\Gamma}(\text{expression})$ and T is a type. Note that we originally assumed $\Delta = \emptyset$. Also, note that the values of sprogram and expression are disjoint and therefore we will look at e : sprogram and e : expression separately.

If the type T is known then we talk about *type checking* else we call the judgment a *type inference*. For inferring a type, the result is not necessary unique, that is why we might only want to find the most general type, meaning a type T_1 such that

$$\underbrace{(\forall T_2 \in \mathcal{T} \land T_1 \sqsubseteq T_2.e :_{\Gamma} T_2)}_{T_1 \text{ is a infered type}} \land \underbrace{(\forall T_2 \in \mathcal{T} \land T_2 \sqsubseteq T_1. \exists T_3 \in \mathcal{T} \land T_2 \sqsubseteq T_3. \neg (e :_{\Gamma} T_3))}_{T_1 \text{ is sharp}}.$$

3.3.2 Auxiliary Definitions

We will need the semantics of <type>, namely a function that maps values $_{\Gamma}(<$ type>) to $\mathcal{T}.$

Definition 3.4: Semantics of <type>

Let $n \in \mathbb{N}$. Let $t, t_1, t_2:$ <type> and c: <upper-var>. Let $t_i:$ <type> for all $i \in \mathbb{N}_3^n$ and $v_i:$ <lower-var> for all $i \in \mathbb{N}_1^n$. Let C be a symbol. Let Γ be a type context. Let $Nat = \mu C.1 \mid Succ\ C$.

We define

$$\begin{split} \llbracket . \rrbracket_{\Gamma} : \text{values}_{\Gamma}(<& \text{type>}) \to \mathcal{T} \\ \llbracket \text{Bool} \rrbracket_{\Gamma} = & \mu_{-}. True \mid False \\ \llbracket \text{Int} \rrbracket_{\Gamma} = & \mu_{-}.0 \mid Pos \; Nat \mid Neg \; Nat \\ \llbracket \text{List} \rrbracket_{\Gamma} = & \forall a.\mu C. [\;] \mid Cons \; a \; C \\ \llbracket "(" \; t_{1} \; , \; t_{2} \; ")" \rrbracket_{\Gamma} = & \{1 : \llbracket t_{1} \rrbracket_{\Gamma}, 2 : \llbracket t_{2} \rrbracket_{\Gamma} \} \\ \llbracket "\{" \; v_{1} \; ": " \; t_{1} \; ", " \; ... \; ", " \; v_{n} \; ": " \; t_{n} \; "\}" \rrbracket_{\Gamma} = & \{v_{1} : \llbracket t_{1} \rrbracket_{\Gamma}, \ldots, v_{n} : \llbracket t_{n} \rrbracket_{\Gamma} \} \\ \llbracket t_{1} \; "->" \; t_{2} \rrbracket_{\Gamma} = \llbracket t_{1} \rrbracket_{\Gamma} \to \llbracket t_{2} \rrbracket_{\Gamma} \\ \llbracket c \; t_{1} \ldots t_{n} \rrbracket = [\llbracket c \rrbracket]_{\Gamma}(t_{1}, \ldots, t_{n}) \end{split}$$

Additionally, we will need to introduce a pattern matching function:

$$\mathsf{match}_{\Theta} : \mathsf{value}(\langle \mathsf{type} \rangle) \times \mathsf{value}(\langle \mathsf{exp} \rangle) \to \{\mathit{True}, \mathit{False}\}$$

for a given substitution Θ . The function will be defined afterwards. For now its definition will be arbitrary.

We can already state two universal inference rules for any Hindley-Milner type system.

Definition 3.5: Instantiation, Generalization

Let $T',T\in\mathcal{T}$ and $e\in \text{values}(\text{sprogram})\cup \text{values}(\text{expression})$. Let a be a type variable. Let Δ be a variable context. Let A be a set and $\llbracket.\rrbracket: \text{values}(\text{sprogram})\cup \text{values}(\text{expression})\to A$

$$\frac{T' \sqsubseteq T \quad \Delta \vdash \llbracket \mathbf{e} \rrbracket :_{\Gamma} T'}{\Delta \vdash \llbracket \mathbf{e} \rrbracket :_{\Gamma} T} \qquad [Instantiation]$$

$$\frac{(a,_) \not\in \Delta \quad \Delta \vdash \llbracket \mathbf{e} \rrbracket :_{\Gamma} T}{\Delta \vdash \llbracket \mathbf{e} \rrbracket :_{\Gamma} \forall a.T} \quad [\textit{Generalization}]$$

The [*Instantiation*] rule says that if a type can be inferred, the same holds for a more specific type. The [*Generalization*] rule states the opposite: if a type with a free variable can be inferred, then the same holds for a poly type, binding the free variable.

3.3.3 Inference Rules for Programs

The inference rules for programs will done statement for statement. Note that every statement has one rule that can be applied, statement with optional parameters have a rule with and one without the optional parameter.

Definition 3.6: Inference rules for programs

The inference rules for programs are defined in table 1.

TConstant, TConstant2 Check if v is still free then add (v, T_1) to the variable context and evaluate the next statement.

TAlias Check if c is still free. $\{v_1, \ldots, v_2\}$ needs to be the set of all free variables in T_2 . If all checks are valid we add (v, T_1) to the type context and evaluate the next statement.

TCustomType Similar to [TAlias] we add (v, T_1) to the type context with the only difference that we explicitly define T_1 as an algebraic type.

TMain,TMain2 Evaluate e.

Table 1: Inference rules for programs

$$(v,_) \not\in \Delta \quad \Delta \vdash \llbracket e \rrbracket :_{\Gamma} T_{1} \\ \Delta \cup \{(v,T_{1})\} \vdash \llbracket s \quad mt \text{ "main } = \text{ " } me \rrbracket :_{\Gamma} T_{2} \\ \Delta \vdash \llbracket v \text{ " } = \text{ " } e \text{ " }; \text{ " } s \quad mt \text{ "main } = \text{ " } me \rrbracket :_{\Gamma} T_{2} \\ (v,_) \not\in \Delta \quad \Delta \vdash \llbracket e \rrbracket :_{\Gamma} T_{1} \quad \llbracket t \rrbracket_{\Gamma} = T_{1} \\ \Delta \cup \{(v,T_{1})\} \vdash \llbracket s \quad mt \text{ "main } = \text{ " } me \rrbracket :_{\Gamma} T_{2} \\ \overline{\Delta \vdash \llbracket v \text{ " } : " } t \text{ " }; \text{ " } v \text{ " } = \text{ " } e \text{ " }; \text{ " } s \quad mt \text{ "main } = \text{ " } me \rrbracket :_{\Gamma} T_{2} \\ \hline (c,_) \not\in \Gamma \quad (c,_) \not\in \Delta \\ \llbracket t \rrbracket_{\Gamma} = T_{1} \quad T_{2} \text{ is a mono type} \\ \{v_{1} \dots v_{n}\} = \text{free}(T_{2}) \quad \forall v_{1} \dots \forall v_{n}. T_{2} = T_{1} \\ \Delta \cup \{(c,T_{1})\} \vdash \llbracket s \quad mt \text{ "main } = \text{ " } me \rrbracket :_{\Gamma \cup \{(c,(T_{1}))\}} T_{3} \\ \overline{\Delta \vdash \llbracket \text{"type alias" } c \ v_{1} \dots v_{n} \text{ " } = \text{" } t \text{ " }; \text{" } s \quad mt \text{ "main } = \text{ " } me \rrbracket :_{\Gamma} T_{3} \\ \hline (c,_) \not\in \Gamma \quad (c,_) \not\in \Delta \quad \{v_{1} \dots v_{n}\} = \text{free}(T_{2}) \quad \forall v_{1} \dots \forall v_{n}. T_{2} = T_{1} \\ \mu C.c_{1} \llbracket t_{1,1} \rrbracket_{\Gamma} \dots \llbracket t_{1,k_{1}} \rrbracket_{\Gamma} \mid \dots \mid c_{m} \llbracket t_{m,1} \rrbracket_{\Gamma} \dots \llbracket t_{m,k_{m}} \rrbracket_{\Gamma} = T_{2} \\ \Delta \cup \{(c,T_{1})\} \vdash \llbracket s \quad mt \text{ "main } = \text{ " } me \rrbracket :_{\Gamma \cup \{(c,(T_{1}))\}} T_{3} \\ \hline \Delta \vdash \llbracket \text{"type" } c \ v_{1} \dots v_{n} \text{" } = \\ \neg t_{1} \mapsto m \text{ " } \neg t_{1} \mapsto m \text{ " } \neg t_{1} \mapsto m \text{ } \mapsto m \text{ } \neg t_{1} \mapsto m \text{$$

3.3.4 Inference Rules for Expressions

In the inference rules [TConstant], [TConstant2] and [Main], [Main2] we used a judgment for expressions. We will now give the corresponding inference rules. As before, each expression has one or two rules depending on optional parameters.

Definition 3.7: Inference rules for expressions

The inference rules for expressions can be found in table 2.

Table 2: Inference rules for expressions

$$\frac{(v,T) \in \Delta}{\Delta \vdash \llbracket v \rrbracket :_{\Gamma} T} \qquad [TVariable]$$

$$\frac{\Gamma,\Delta \vdash \mathsf{match}_{\Theta}(T_{1},p) \quad \Delta \cup \Theta \vdash \llbracket e \rrbracket :_{\Gamma} T_{2}}{\Delta \vdash \llbracket v \rrbracket :_{\Gamma} T_{1} \quad \forall e_{2} \rrbracket :_{\Gamma} T_{1} \rightarrow T_{2}} \qquad [TLambda]$$

$$\frac{\llbracket e_{1} \rrbracket_{\Gamma} : T_{1} \quad \llbracket e_{2} \rrbracket_{\Gamma} : T_{2}}{\Delta \vdash \llbracket v \rrbracket :_{\Gamma} \cdot H_{2} \quad \exists_{\Gamma} T_{1} \quad \exists_{\Gamma} T_{2}} \qquad [TTuple]$$

$$\Delta \vdash \llbracket v \rrbracket :_{\Gamma} \vdash \forall a.List \ a \qquad [TEmptyList]$$

$$\frac{\Delta \vdash \llbracket e \rrbracket :_{\Gamma} T}{\Delta \vdash \llbracket v \rrbracket :_{\Gamma} \cdot u \quad e^{-1} \rrbracket :_{\Gamma} \cdot List \ T} \qquad [TSingleList]$$

$$\frac{\Delta \vdash \llbracket e \rrbracket :_{\Gamma} T \quad \Delta \vdash \llbracket v \rrbracket :_{\Gamma} \cdot u \quad e^{-1} \rrbracket :_{\Gamma} \cdot List \ T}{\Delta \vdash \llbracket v \rrbracket :_{\Gamma} \cdot u \quad e^{-1} \rrbracket :_{\Gamma} \cdot List \ T} \qquad [TList]$$

$$\frac{e : \langle \operatorname{int} \rangle}{\Delta \vdash \llbracket e \rrbracket :_{\Gamma} \cdot T \quad t} \qquad [TInt]$$

$$\frac{e : \langle \operatorname{bool} \rangle}{\Delta \vdash \llbracket e \rrbracket :_{\Gamma} \cdot Bool} \qquad [TBool]$$

$$\frac{\Delta \vdash \llbracket e_{1} \rrbracket :_{\Gamma} T_{1} \rightarrow T_{2} \quad \Delta \vdash \llbracket e_{2} \rrbracket :_{\Gamma} T_{1}}{\Delta \vdash \llbracket e_{1} \cdot e_{2} \rrbracket :_{\Gamma} T_{2}} \qquad [TCall]$$

$$\frac{\Delta \vdash \llbracket e_{1} \rrbracket :_{\Gamma} T_{1} \quad \Gamma, \Delta \vdash \operatorname{match}_{\Theta}(T_{1}, \llbracket p \rrbracket) \quad \Delta \cup \Theta \vdash \llbracket e_{2} \rrbracket :_{\Gamma} T_{2}}{\Delta \vdash \llbracket v \rrbracket :_{\Gamma} \cdot T_{2}} \qquad [TSingleCaseOf]$$

$$\begin{array}{c} \Delta \vdash \llbracket e_1 \rrbracket : _T T_1 \quad \Gamma, \Delta \vdash \mathsf{match}_{\Theta}(T_1, \llbracket p \rrbracket) \quad \Delta \cup \Theta \vdash \llbracket e_2 \rrbracket :_T T_2 \\ \Delta \vdash \llbracket \mathsf{mease}^n e_1 \quad \mathsf{mof}^n \quad \mathsf{me}^n e_1 \quad e_2 \quad \mathsf{me}^n \\ \exists \vdash \mathsf{mease}^n e_1 \quad \mathsf{mof}^n \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n \\ \Delta \vdash \llbracket \mathsf{mease}^n e_1 \quad \mathsf{mof}^n \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n \\ \exists \vdash \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \\ \Delta \vdash \llbracket \mathsf{met}^n \quad \mathsf{me}^n e_1 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \\ \Delta \vdash \llbracket \mathsf{met}^n \quad \mathsf{me}^n e_1 \quad \mathsf{me}^n e_1 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \\ \Delta \vdash \llbracket \mathsf{me}^n e_2 \quad \mathsf{me}^n e_1 \quad \mathsf{me}^n e_2 \\ \Delta \vdash \llbracket \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \\ \Delta \vdash \llbracket \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \\ \mathsf{me}^n e_2 \quad \mathsf{me}^n e_2 \\ \mathsf{me}^n e_2 \\$$

```
\vdash \llbracket "(::)" \rrbracket : \forall a.a \to List \ a \to List \ a \\ \vdash \llbracket "foldl" \rrbracket : \forall a. \forall b. (a \to b \to b) \to b \to List \ a \to b \\ \llbracket TFoldl \rrbracket
```

TVariable Find the type in the context.

TLambda Elm allows the parameters of a function to be pattern matched. Therefore, we first need to find a matching type T_1 and can then infer the type of e by including the additional bindings Θ to the context.

TTuple Find the types of e_1 and e_2 , then construct the tuple.

TEmptyList The empty list is a literal for every list, therefore we can infer the list poly type.

TSingleList, **TList** Recursively we check that every element has the same type.

TInt, TBool The type of literals can be inferred without any restrictions.

TCall The first expression needs to be a function that the second type can be passed to.

TSingleCaseOf, TCaseOf First match the type of the expression e_1 to the pattern, then use the additional bindings Θ to obtain the type of e_2 . As all patterns need to have the same type, we can then recursively check the other patterns as well.

TLetIn,TLetIn2 The variable v may not have a value assined in the conext Γ . If so, we can infer the type T_1 of e_1 and add (v,T_1) to the context before we evaluate e_2 . For [TLetIn2] we already the type is already given as t. Note that t can be more specific as the type we would usually infer.

TGetter The second variable v_2 is a label of the record, that is bound to v_1 .

TSingleSetter,TSetter Setters can not change the type in Elm. But we still need to ensure that the fields are also correctly typed.

TEImptyRecord The empty record can be directly infered, as it has only one element.

TRecord Each field and its value must be given at the same time. That is why we can not use a recursive definition.

TIFEISE The first expression e_1 needs to be a boolean and the branches e_2 , e_3 must have the same type.

TComposition, TPipe The pipe applies the first expression to the second. The composition is similar to the pipe, but results in a function.

TOr, TAnd, TNot, TEqual, TDivide, TMultiply, TMinus, TPlus, TCons, TFoldl These functions can be seen as lambda function literals.

Example 3.1

In example ?? we have looked at the syntax for a list reversing function. We can now prove the typing of the reverse function for $\Gamma=\varnothing$, $\Delta=\varnothing$ and $T=\forall a.List\ a\to List\ a.$

References

[Pie+02] B.C. Pierce et al. *Types and Programming Languages*. The MIT Press. MIT Press, 2002. ISBN: 9780262162098. URL: https://books.google.at/books?id=ti6zoAC9Ph8C.