3 Liquid types

First, we define some notations:

- \mathbb{N} are the natural numbers starting from 1.
- \mathbb{N}_0 are the natural numbers starting from 0.
- $\mathbb{N}_a^b := \{i \in \mathbb{N}_0 | a \le i \land i \le b\}$ are the natural numbers between a and b.
- We'll use "." to separate a quantifier from a statement: $\forall a.b$ and $\exists a.b$.
- Function types will be written as $a_1 \to \cdots \to a_n \to b$ instead of $a_1 \times \cdots \times a_n \to b$.
- We allow the use of lambda notation for functions: $\lambda x.x$ instead of f(x) = x for a function f.

3.1 Hindley-Milner Type System

For this thesis we will use a Hindley-Milner type system [DM82]. The main idea is to have a type system that implies an order among the types. The ordering will then allow us to infer the type of any expression.

3.1.1 Notion of Types

We will first introduce types, afterwards we will define how types relate to sets by defining the values of types as explicit finite sets. Types are split in *mono types* and *poly types*. Mono types can contain so called *type variables* that can then be bound with a quantifier as a poly type. Note that quantifiers can only occur in the outermost position, thus poly types are more general types than mono types.

Definition 3.1: Mono types, poly types, types

```
We say
```

```
T is a mono\ type:\Leftrightarrow T is a type variable  \lor T \text{ is a type application}   \lor T \text{ is a algebraic type}   \lor T \text{ is a product type}   \lor T \text{ is a function type}   \lor T \text{ is a function type}   T \text{ is a } poly\ type:\Leftrightarrow T=\forall a.T'   \text{where } T' \text{ is a mono type or poly type and } a \text{ is a symbol}   T \text{ is a } type:\Leftrightarrow T \text{ is a mono type } \lor T \text{ is a poly type}.
```

by using the following predicates.

T is a $type\ variable:\Leftrightarrow T$ is a symbol.

T is a $type \ application :\Leftrightarrow \operatorname{Let} n \in \mathbb{N}, C$ be a symbol. Let T_i be mono types for all

$$i \in \mathbb{N}_1^n$$
 in

$$C T_1 \dots T_n$$

T is a $algebraic\ type:\Leftrightarrow$ Let $n\in\mathbb{N}, k\in\mathbb{N}_1^n\to\mathbb{N}_0, C$ be a symbol. Let $T_{i,k(j)}$ be a

mono type or C for all $i \in \mathbb{N}_1^n$ and $j \in \mathbb{N}_1^{k(i)}$ in

$$\mu C.C_1 T_{1,1} \dots T_{1,k(1)} | \dots | C_n T_{n,1} \dots T_{n,k(n)}$$

such that
$$\exists i \in \mathbb{N}. \forall j \in \mathbb{N}_1^{k(i)}. T_{i,j} \neq C$$
.

T is a $product\ type:\Leftrightarrow \mathrm{Let}\ n\in\mathbb{N}_0.$ Let l_i symbols and T_i be a mono types for

all
$$i \in \mathbb{N}_1^n$$
 in

$$T = \{l_1 : T_1, \dots, l_n : T_n\}$$

T is a *function type* : $\Leftrightarrow T_1$ and T_2 be mono types in

$$T=T_1 \rightarrow T_2$$
.

Axiom 3.1

The types T_i for $i \in \mathbb{N}$ in a product type are unordered:

$${a:T_1,b:T_2,\ldots} = {b:T_2,a:T_1,\ldots}$$

for any symbols a, b and mono types T_1, T_2 .

Example 3.1

The symbol Char is a type variable. Sequence Char is a type application. They can be thought of as types, whose implementation is unknown. The interpretation of a type variable or a type application depends on its context.

Example 3.2

 $Bool = \mu$. True | False is an algebraic type.

Example 3.3

 $List = \forall a.\mu C. Empty \mid Cons \ a \ C$ is a poly type.

Example 3.4

The empty product type {} is a mono type.

Definition 3.2: Sort, Terminal

Let $n \in \mathbb{N}$, $k_j \in \mathbb{N}$, $T_{i,j}$ be a mono type, C, C_i be symbols, $t_j : T_{i,j}$ for all $j \in \mathbb{N}_1^n$, $i \in \mathbb{N}_1^n$. and $T = \mu C.C_1 T_{1,1} \dots T_{1,k_1} | \dots | C_n T_{n,1} \dots T_{n,k_n}$ be a algebraic type.

We call

- $C_i T_{i,1} \dots T_{i,k_i}$ a sort of T,
- C_i a terminal of T.

Example 3.5

The natural numbers and the integers can be defined as algebraic types using the peano axioms [Pea89]:

- 1 is a natural number.
- Every natural number has a successor.

These axioms can be used for the definition of the type application.

$$Nat ::= \mu C.1 \mid Succ \ C$$

For integers, we can use the property that they contain 0 as well as all positive and negative numbers.

$$Int ::= \mu C.0 \mid Pos\ Nat \mid Neg\ Nat$$

In this case numbers like $1, Succ\ 1$ for Nat or $Neg\ (Succ\ 1)$ for Int are sorts, where as 1 and Succ for Nat and $Neg\ Pos$ and 0 for Int are terminals.

Definition 3.3: Label

Let $n \in \mathbb{N}$. Let T_i be a type, l_i be a unique symbol for all $i \in \mathbb{N}_1^n$.

We say l_i are the *labels* of the product type $\{l_1: T_1,..,l_n: T_n\}$ for all $i \in \mathbb{N}_1^n$.

We define

$$T_1 \times \cdots \times T_n := \{1 : T_1, \dots, n : T_n\}$$

as the *ordered product type* with n components.

The most general example of a product type is a record. Tuples can be represented as ordered product types.

Definition 3.4: Bound, Free, Set of free variables

Let $n \in \mathbb{N}_0$, a be a type variable, T be a type, C be a symbol, $k \in \mathbb{N}_1^n \to \mathbb{N}_0$, T_i be a type, $T_{i,k(j)}$ be a type or a symbol and C_i be a symbol for all $i \in \mathbb{N}_1^n$ and $j \in \mathbb{N}_1^n$.

We say

- a is free in $T :\Leftrightarrow a \in free(T)$
- a is bound in $T : \Leftrightarrow a \notin free$ and a occurs in T.

where

$$\operatorname{free}(a) := \{a\}$$

$$\operatorname{free}(C \ T_1 \dots T_n) := \bigcup_{i \in \mathbb{N}_1^n} \operatorname{free}(T_i)$$

$$\operatorname{free}\begin{pmatrix} \mu C \\ C_1 \ T_{1,1} \dots \ T_{1,k(1)} \\ | \dots \\ | C_n \ T_{n,1} \dots \ T_{n,k(n)} \end{pmatrix} := \bigcup_{i \in \mathbb{N}_0^n} \bigcup_{j \in \mathbb{N}_0^{k_i}} \begin{cases} \varnothing & \text{if } T_{i,j} = C \\ \operatorname{free}(T_{i,j}) & \text{else} \end{cases}$$

$$\operatorname{free}(\{ _: T_1, \dots, _: T_n \}) := \bigcup_{i \in \mathbb{N}_1^n} \operatorname{free}(T_i)$$

$$\operatorname{free}(T_1 \to T_2) := \operatorname{free}(T_1) \cup \operatorname{free}(T_2)$$

$$\operatorname{free}(\forall a.T) := \operatorname{free}(T) \setminus \{a\}$$

Definition 3.5: Partial function

Let T_1 and T_2 be types. $f \subseteq T_1 \times T_2$

We say f is a partial function (Notation: $f: T_1 \nrightarrow T_2$): \Leftrightarrow

$$\forall x \in T_1, y \in T_2.(x, y_1) \in f \land (x, y_2) \in f \Rightarrow y_1 = y_2.$$

Definition 3.6: Sets of Types

We define

- $\mathcal{V} := \{a \mid a \text{ is a symbol}\}\$ as the set of all type variables (symbols).
- $T := \{T \mid T \text{ is a type}\}$ as the set of all types.

A type can be substituted by replacing a bounded type variable with a mono type:

Definition 3.7: Type substitution

Let $n \in \mathbb{N}$, $\Theta : \mathcal{V} \nrightarrow \mathcal{T}$, $a \in \mathcal{V}$. Let $T, T_1, T_2, k \in \mathbb{N}_1^n \to \mathbb{N}_0$ and $T_{i,k(j)} \in \mathcal{T}$ for all $i \in \mathbb{N}_1^n$ and $j \in \mathbb{N}_1^n$.

We define the substitute of a type $[.]_{\Theta}: \mathcal{T} \to \mathcal{T}$ as

$$[a]_{\Theta} := \begin{cases} S & \text{if } (a,S) \in \Theta \\ a & \text{else} \end{cases}$$

$$\begin{bmatrix} \mu C. \\ C_1 T_{1,1} \dots T_{1,k(1)} \\ | \dots \\ | C_n T_{n,1} \dots T_{n,k(n)} \end{bmatrix}_{\Theta} := \begin{cases} C_1 [T_{1,1}]_{\Theta} \dots [T_{1,k_1}]_{\Theta} \\ | \dots \\ | C_n [T_{n,1}]_{\Theta} \dots [T_{n,k_n}]_{\Theta} \end{cases}$$

$$[\{l_1 : T_1, \dots, l_n : T_n\}]_{\Theta} := \{l_1 : [T_1]_{\Theta}, \dots, l_n : [T_n]_{\Theta}\}$$

$$[T_1 \to T_2]_{\Theta} := [T_1]_{\Theta} \to [T_2]_{\Theta}$$

$$[\forall b.T]_{\Theta} := \begin{cases} [T]_{\Theta} & \text{if } \exists (b, _) \in \Theta \\ \forall b.[T]_{\Theta} & \text{else.} \end{cases}$$

 Θ is called the set of substitutions.

The type substitution gives raise to a partial order \sqsubseteq :

Definition 3.8: Type Order

Let $n, m \in \mathbb{N}$, $T_1, T_2 \in \mathcal{T}$, a_i for all $i \in \mathbb{N}_0^n$ and $b_i \in \mathcal{V}$ for all $i \in \mathbb{N}_0^m$.

We define the partial order \sqsubseteq as

$$\forall a_1 \dots \forall a_n. T_1 \sqsubseteq \forall b_1 \dots \forall b_m. T_2 : \Leftrightarrow \exists \Theta = \{(a_i, S_i) \mid i \in \mathbb{N}_1^n \land a_i \in \mathcal{V} \land S_i \in \mathcal{T}\}.$$

$$T_2 = [T_1]_{\Theta} \land \forall i \in \mathbb{N}_0^m. b_i \notin \text{free}(\forall a_1 \dots \forall a_n. T)$$

The rule can be read as follows:

- First replace all bounded variables with types.
- Next rebound any new variables (variables that were previously not free).

Example 3.6

 $\forall a.a$ is the smallest type in the type system. The partial order forms a tree structure with $\forall a.a$ at the root and different branches for $\forall a. \forall b. (a,b), \forall a. C\ a, \forall a. \forall b. a \rightarrow b$ and so on. The leaves of the tree are all possible mono types.

3.1.2 Interpretation of types

Before we interprete a type, we will first introduce a set of labeled elements as a record.

Definition 3.9: Record

Let n in \mathbb{N} , l_i be a symbol, t_i arbitary for all i in \mathbb{N}_1^n .

We define

$$\{l_1 = t_1, \dots, l_n = t_n\} : \{l_1, \dots, l_n\} \to \{t_1, \dots, t_n\}$$

 $\{l_1 = t_1, \dots, l_n = t_n\}(l) := t \text{ such that } \exists i \in \mathbb{N}_1^n.l = l_i \land t = t_i$

Note that values of a ordered product type are equivant to tuples:

$$\forall i \in \mathbb{N}_1^n \{1 = t_1, \dots, n = t_n\}(i) = (t_1, \dots, t_n).i$$

Thus we will use the notation of tuples for values of a ordered product type.

Definition 3.10: Application Constructor

Let $n \in \mathbb{N}_0$. Let a_i be a symbol for all $i \in \mathbb{N}_1^n$.

We call

$$f: \underbrace{\mathcal{T} \to \cdots \to \mathcal{T}}_{n \text{ times}} \to \mathcal{T}$$

 $f(T_1, ..., T_n) := [\forall a_1 ... a_n. T]_{\{(a_1, T_1), ..., (a_n, T_n)\}}$ for a given mono type T.

an application constructor.

We define $\mathcal{C}=\{f|f \text{ is a application constructor}\}$ as the set of all application constructors.

Definition 3.11: Context

 $\Gamma: \mathcal{V} \nrightarrow \mathcal{C}$ is a *context* : \Leftrightarrow

$$\Gamma = \{\}$$

 $\vee \Gamma = \Delta \cup \{(a,T)\}$ where T is a mono type, Δ is a context and a is a type variable $\vee \Gamma = \Delta \cup \{(C,f)\}$ where $T = C \ T_1 \dots T_n$ is a type application and T_i is a mono type for $i \in \mathbb{N}_1^n$ and f is a application constructor such that $f(T_1,\dots,T_n)$ is a mono type and Δ is a context

Definition 3.12: Values

Let S the class of all finite sets, $n \in \mathbb{N}$, $\Theta : \mathcal{V} \to \mathcal{T}$, $a \in \mathcal{V}$, $T, T_1, T_2, S \in \mathcal{T}$, $k \in \mathbb{N}_1^n \to \mathbb{N}_0$ and $T_{i,k(j)} \in \mathcal{T}$ for all $i \in \mathbb{N}_1^n$ and $j \in \mathbb{N}_1^n$. Let Γ be a context.

We define

$$\operatorname{values}_{\Gamma}: \mathcal{V} \to \mathcal{S}$$

$$\operatorname{values}_{\Gamma}(a) := \operatorname{values}_{\Gamma}(\Gamma(a))$$

$$\operatorname{values}_{\Gamma}(C \ T_1 \ \dots \ T_n) := \operatorname{values}_{\Gamma}(\Gamma(C)(T_1, \dots, T_n))$$

$$\operatorname{values}_{\Gamma} \begin{pmatrix} \mu C \\ | C_1 \ T_{1,1} \dots \ T_{1,k(1)} \\ | \dots \\ | C_n \ T_{n,1} \dots \ T_{n,k(n)} \end{pmatrix} := \bigcup_{i \in \mathbb{N}_0} \operatorname{rvalues}_{\Gamma} \begin{pmatrix} \mu C \\ i, & | C_1 \ T_{1,1} \dots \ T_{1,k(1)} \\ | \dots \\ | C_n \ T_{n,1} \dots \ T_{n,k(n)} \end{pmatrix}$$

$$\operatorname{values}_{\Gamma}(\{l_1 : T_1, \dots, l_n : T_n\}) := \{\{l_1 = t_1, \dots, l_n = t_n\} \mid \forall i \in \mathbb{N}_1^n. t_i \in \operatorname{values}_{\Gamma}(T_i)\}$$

$$\operatorname{values}_{\Gamma}(T_1 \to T_2) := \{f \mid f : \operatorname{values}_{\Gamma}(T_1) \to \operatorname{values}_{\Gamma}(T_2)\}$$

$$\operatorname{values}_{\Gamma}(\forall a.T) := \lambda b. \operatorname{values}_{\{(a,b)\} \cup \Gamma}(T) \text{ where the symbol } b \text{ does not occur in } T.$$

using the following helper function.

$$\operatorname{Let} l \in \mathbb{N}, T := \mu C. \mid C_1 T_{1,1} \dots T_{1,k(1)} \mid \dots \mid C_n T_{n,1} \dots T_{n,k(n)} \text{ in}$$

$$\operatorname{rvalues}_{\Gamma}(0,T) := \left\{ \begin{array}{c} C_i \ v_1 \dots v_n \ \middle| \ i \in \mathbb{N}_1^n \\ \land \forall j \in \mathbb{N}_1^{k(i)}.T_{i,j} \neq C \land v_j \in \operatorname{values}_{\Gamma}(T_{i,j}) \end{array} \right\}$$

$$\operatorname{rvalues}_{\Gamma}(l+1,T) := \left\{ \begin{array}{c} C_i \ v_1 \dots v_n \ \middle| \ i \in \mathbb{N}_1^n \\ \land \forall j \in \mathbb{N}_1^{k(i)}.v_j \in \begin{cases} \operatorname{rvalues}_{\Gamma}(l,T) & \text{if } T_{i,j} = C \\ \operatorname{values}_{\Gamma}(T_{i,j}) & \text{else} \end{cases} \right\}$$

The base case of this recursive function is in rvalues (0, T) for a given T.

Theorem 3.1: rvalue is nested

Let
$$n \in \mathbb{N}_0$$
. Let $T := \mu C$. $\mid C_1 T_{1,1} \dots T_{1,k(1)} \mid \dots \mid C_n T_{n,1} \dots T_{n,k(n)}$.
$$\bigcup_{i \in \mathbb{N}_0^n} \operatorname{rvalues}_{\Gamma}(i,T) = \operatorname{rvalues}_{\Gamma}(n,T)$$

Proof. Its sufficient to prove by induction over n that

$$\forall n \in \mathbb{N}_0.\text{rvalues}_{\Gamma}(n,T) \subseteq \text{rvalues}_{\Gamma}(n+1,T).$$

Base case: We'll show $\operatorname{rvalues}_{\Gamma}(0,T) \subseteq \operatorname{rvalues}_{\Gamma}(1,T)$.

$$\operatorname{rvalues}_{\Gamma}(1,T)$$

$$= \left\{ \begin{array}{c|c} C_{i} v_{1} \dots v_{n} & i \in \mathbb{N}_{1}^{n} \\ \land \forall j \in \mathbb{N}_{1}^{k(i)} . v_{j} \in \left\{ \begin{array}{c} \operatorname{rvalues}_{\Gamma}(0,T) & \text{if } T_{i,j} = C \\ \operatorname{values}_{\Gamma}(T_{i,j}) & \text{else} \end{array} \right. \right\}$$

$$\supseteq \left\{ \begin{array}{c|c} C_{i} v_{1} \dots v_{n} & i \in \mathbb{N}_{1}^{n} \\ \land \forall j \in \mathbb{N}_{1}^{k(i)} . T_{i,j} \neq C \land v_{j} \in \operatorname{values}_{\Gamma}(T_{i,j}) \end{array} \right\}$$

$$= \operatorname{rvalues}_{\Gamma}(0,T).$$

Inductive step: Assuming $\operatorname{rvalues}_{\Gamma}(n,T) \subseteq \operatorname{rvalues}_{\Gamma}(n+1,T)$, we'll prove

$$\operatorname{rvalues}_{\Gamma}(n+1,T) \subseteq \operatorname{rvalues}_{\Gamma}(n+2,T).$$

$$\operatorname{rvalues}_{\Gamma}(n+2,T) = \left\{ \begin{array}{c|c} C_{i} v_{1} \dots v_{n} & i \in \mathbb{N}_{1}^{n} \\ \land \forall j \in \mathbb{N}_{1}^{k(i)}.v_{j} \in \left\{ \begin{array}{c} \operatorname{rvalues}_{\Gamma}(n+1,T) & \text{if } T_{i,j} = C \\ \operatorname{values}_{\Gamma}(T_{i,j}) & \text{else} \end{array} \right\} \\ \supseteq \left\{ \begin{array}{c|c} C_{i} v_{1} \dots v_{n} & i \in \mathbb{N}_{1}^{n} \\ \land \forall j \in \mathbb{N}_{1}^{k(i)}.v_{j} \in \left\{ \begin{array}{c} \operatorname{rvalues}_{\Gamma}(n,T) & \text{if } T_{i,j} = C \\ \operatorname{values}_{\Gamma}(T_{i,j}) & \text{else} \end{array} \right\} \\ = \operatorname{rvalues}_{\Gamma}(n+1,T) \end{array} \right.$$

As an example we can now prove that the values of Nat from example are isomorphic to the natural numbers.

Theorem 3.2

Let $Nat := \mu C.1 | Succ C$ values $(Nat) \cong \mathbb{N}$

Proof.

$$\begin{split} \text{values}(\mu C.1|Succ\ C) &= \bigcup_{i \in \mathbb{N}_0} \text{rvalues}(i, \mu C.1|Succ\ C) \\ &= \lim_{n \to \infty} \bigcup_{i \in \mathbb{N}_0^n} \text{rvalues}(i, \mu C.1|Succ\ C) \\ &= \lim_{n \to \infty} \text{rvalues}(n, \mu C.1|Succ\ C). \end{split}$$

We'll now show by induction over $n \in \mathbb{N}_0$ that

$$rvalues(n, \mu C.1 | Succ\ C) = \{1, \underbrace{Succ\ 1, \dots, Succ\ \dots\ Succ\ 1}_{n\ times}\}.$$

8

Base case:
$$\operatorname{rvalues}(0,\mu C.1|Succ\ C) = \{1\}$$
 Inductive step:
$$\operatorname{Assuming\ rvalues}(n,\mu C.1|Succ\ C) = \{1,\underbrace{Succ\ 1,\ldots,Succ\ \ldots\ Succ\ 1}_{n\ \text{times}}\},\ \text{we'll}$$

$$\operatorname{prove\ rvalues}(n+1,\mu C.1|Succ\ C) = \{1,\underbrace{Succ\ 1,\ldots,Succ\ \ldots\ Succ\ 1}_{n+1\ \text{times}}\}.$$

$$\begin{aligned} &\operatorname{rvalues}(n+1,\mu C.1|Succ\ C) \\ &= \left\{ \begin{array}{ll} c_i\ v_1 \dots v_n & i \in \mathbb{N}_1^n \\ & \land \forall j \in \mathbb{N}_1^{k(i)}.v_j \in \left\{ \begin{array}{ll} \operatorname{rvalues}_{\Gamma}(n,T) & \text{if } T_{i,j} = C \\ & \text{values}_{\Gamma}(T_{i,j}) & \text{else} \end{array} \right. \\ &= \left\{ \begin{array}{ll} c_i\ v_1 \dots v_n & i \in \mathbb{N}_1^n \\ & \land \forall j \in \mathbb{N}_1^{k(i)}.v_j \in \left\{ \begin{array}{ll} \{1, \underbrace{Succ\ 1, \dots, Succ\ \dots\ Succ\ 1}\} & \text{if } T_{i,j} = C \\ & values_{\Gamma}(T_{i,j}) & \text{else} \end{array} \right. \\ &= \left\{ 1 \right\} \cup \left\{ Succ\ v \middle| v \in \left\{ 1, \underbrace{Succ\ 1, \dots, Succ\ \dots\ Succ\ 1} \right\} \right\} \\ &= \left\{ 1, \underbrace{Succ\ 1, \dots, Succ\ \dots\ Succ\ 1} \right\} \end{aligned}$$

We define a order on

$$values(Nat) = \lim_{n \to \infty} \{1, \underbrace{Succ\ 1, \dots, Succ\ \dots\ Succ\ 1}_{n \text{ times}}\} = \{1, Succ\ 1, Succ\ Succ\ 1, \dots\}$$

by $v_1 < v_2 :\Leftrightarrow v_2 = Succ \dots Succ \ v_1 \ \text{for} \ v_1, v_2 \in \text{values}(Nat)$. This is a well-ordering, thus the set is isomorphic to \mathbb{N} .

References

- [DM82] Luís Damas and Robin Milner. "Principal Type-Schemes for Functional Programs". In: Conference Record of the Ninth Annual ACM Symposium on Principles of Programming Languages, Albuquerque, New Mexico, USA, January 1982. 1982, pp. 207–212. doi: 10.1145/582153.582176. url: https://doi.org/10.1145/582153.582176.
- [Pea89] G. Peano. *Arithmetices principia: nova methodo*. Trans. by Vincent Verheyen. Fratres Bocca, 1889. url: https://github.com/mdnahas/Peano_Book/blob/master/Peano.pdf.