

TIFAE – TRFAE with Type Inference

1 INTRODUCTION

TIFAE is a toy language for the COSE212 course at Korea University. TIFAE stands for an extension of the TRFAE language with **type inference**, and it supports the following features:

- **number (integer) values** ($0, 1, -1, 2, -2, 3, -3, \dots$)
- **arithmetic operators**: negation ($-$), addition ($+$), subtraction ($-$), multiplication ($*$), division ($/$), and modulo ($\%$)
- **first-class functions** ($=>$)
- **recursive functions** (def)
- **conditionals** (if-else)
- **boolean values** (true and false)
- **comparison operators**: equality ($==$ and $!=$) and relational ($<$, $>$, \leq , and \geq)
- **logical operators**: conjunction ($\&\&$), disjunction ($\|$), and negation ($!$)
- **static type checking with type inference**

This document is the specification of TIFAE. First, Section 2 describes the concrete syntax, and Section 3 describes the abstract syntax with the desugaring rules. Then, Section 4 describes the type system. Finally, Section 5 describes the big-step operational (natural) semantics of TIFAE.

2 CONCRETE SYNTAX

The concrete syntax of TIFAE is written in a variant of the extended Backus–Naur form (EBNF). The notation $\langle nt \rangle$ denotes a nonterminal, and " t " denotes a terminal. We use $?$ to denote an optional element and $+$ (or $*$) to denote one or more (or zero or more) repetitions of the preceding element. We use **butnot** to denote a set difference to exclude some strings from a producible set of strings. We omit some obvious terminals using the ellipsis (\dots) notation.

```
// basic elements
<digit>   ::= "0" | "1" | "2" | ... | "9"
<number>   ::= "-"? <digit>+
<alphabet> ::= "A" | "B" | "C" | ... | "Z" | "a" | "b" | "c" | ... | "z"
<idstart> ::= <alphabet> | "_"
<idcont>  ::= <alphabet> | "_" | <digit>
<keyword>  ::= "true" | "false" | "def" | "if" | "else" | "val"
<id>        ::= <idstart> <idcont>* butnot <keyword>
// expressions
<expr>     ::= <number> | "true" | "false" | <uop> <expr> | <expr> <bop> <expr>
              | "(" <expr> ")" | "{" <expr> "}"
              | "val" <id> "=" <expr> ";"? <expr> | <id>
              | "(" <id> ")" "=>" <expr> | <expr> "(" <expr> ")"
              | "def" <id> "(" <id> ")" "=" <expr> ";"? <expr>
              | "if" "(" <expr> ")" <expr> "else" <expr>
// operators
<uop>      ::= "-"
<bop>      ::= "+" | "-" | "*" | "/" | "%" | "&&" | "||"
              | "==" | "!=" | "<" | "<=" | ">" | ">="
```

For types, the arrow (\Rightarrow) operator is right-associative. For expressions, the precedence and associativity of operators are defined as follows:

Description	Operator	Precedence	Associativity
Unary	$-$, $!$	7	right
Multiplicative	$*$, $/$, $\%$	6	
Additive	$+$, $-$	5	
Relational	$<$, \leq , $>$, \geq	4	
Equality	$=$, \neq	3	
Logical Conjunction	$\&\&$	2	
Logical Disjunction	$\ $	1	

3 ABSTRACT SYNTAX

The abstract syntax of TIFAE is defined as follows:

$$\begin{array}{llll}
 \text{Expressions} & \mathbb{E} \ni e ::= n & (\text{Num}) & | \text{val } x = e; e & (\text{Val}) \\
 & | b & (\text{Bool}) & | x & (\text{Id}) \\
 & | e + e & (\text{Add}) & | \lambda x. e & (\text{Fun}) \\
 & | e * e & (\text{Mul}) & | \text{def } x(x) = e; e & (\text{Rec}) \\
 & | e / e & (\text{Div}) & | e(e) & (\text{App}) \\
 & | e \% e & (\text{Mod}) & | \text{if } (e) e \text{ else } e & (\text{If}) \\
 & | e == e & (\text{Eq}) & \\
 & | e < e & (\text{Lt}) & \\
 \\
 \text{Numbers} & n \in \mathbb{Z} & & & (\text{BigInt}) \\
 \text{Identifiers} & x \in \mathbb{X} & & & (\text{String}) \\
 \text{Booleans} & b \in \mathbb{B} = \{\text{true}, \text{false}\} & & & (\text{Boolean})
 \end{array}$$

The types or semantics of the remaining cases are defined with the following desugaring rules:

$$\begin{array}{lll}
 \mathcal{D}[[-e]] & = \mathcal{D}[e] * (-1) & \mathcal{D}[e_1 != e_2] = \mathcal{D}[\neg (e_1 == e_2)] \\
 \mathcal{D}[\mathbf{!} e] & = \text{if } (\mathcal{D}[e]) \text{ false else true} & \mathcal{D}[e_1 \leq e_2] = \mathcal{D}[(e_1 < e_2) \vee (e_1 == e_2)] \\
 \mathcal{D}[e_1 - e_2] & = \mathcal{D}[e_1] + \mathcal{D}[-e_2] & \mathcal{D}[e_1 > e_2] = \mathcal{D}[\neg (e_1 \leq e_2)] \\
 \mathcal{D}[e_1 \& e_2] & = \text{if } (\mathcal{D}[e_1]) \mathcal{D}[e_2] \text{ else false} & \mathcal{D}[e_1 \geq e_2] = \mathcal{D}[\neg (e_1 < e_2)] \\
 \mathcal{D}[e_1 \mid\mid e_2] & = \text{if } (\mathcal{D}[e_1]) \text{ true else } \mathcal{D}[e_2]
 \end{array}$$

The omitted cases recursively apply the desugaring rule to sub-expressions.

4 TYPE SYSTEM

This section explains type system of TIFAE, and we use the following notations:

$$\text{Type Environments} \quad \Gamma \in \mathbb{F} = \mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}^V \quad (\text{TypeEnv})$$

$$\text{Solution} \quad \psi \in \Psi = \mathbb{X}_\alpha \xrightarrow{\text{fin}} (\mathbb{T} \uplus \{\bullet\}) \quad (\text{Solution})$$

$$\text{Type Schemes} \quad \tau^\forall = \forall \alpha^*. \tau \in \mathbb{T}^\forall = \mathbb{X}_\alpha^* \times \mathbb{T} \quad (\text{TypeScheme})$$

$$\begin{array}{lll}
 \text{Types} & \mathbb{T} \ni \tau ::= \text{num} & (\text{NumT}) \quad | \tau \rightarrow \tau & (\text{ArrowT}) \quad \text{Type Variables} \quad \alpha \in \mathbb{X}_\alpha & (\text{Int}) \\
 & | \text{bool} & (\text{BoolT}) \quad | \alpha & (\text{VarT})
 \end{array}$$

We skip the \forall -quantifier in type schemes if they have no type variables. In the type system, type checking and type inference is defined with the following typing rules:

$$\begin{array}{c}
 \boxed{\Gamma, \psi \vdash e : \tau, \psi} \\
 \\
 \frac{\tau\text{-Num}}{\Gamma, \psi \vdash n : \text{num}, \psi} \quad \frac{\tau\text{-Bool}}{\Gamma, \psi \vdash b : \text{bool}, \psi} \\
 \\
 \frac{\tau\text{-Add} \quad \Gamma, \psi_0 \vdash e_1 : \tau_1, \psi_1 \quad \Gamma, \psi_1 \vdash e_2 : \tau_2, \psi_2 \quad \text{unify}(\tau_1, \text{num}, \psi_2) = \psi_3 \quad \text{unify}(\tau_2, \text{num}, \psi_3) = \psi_4}{\Gamma, \psi_0 \vdash e_1 + e_2 : \text{num}, \psi_4} \\
 \\
 \frac{\tau\text{-Mul} \quad \Gamma, \psi_0 \vdash e_1 : \tau_1, \psi_1 \quad \Gamma, \psi_1 \vdash e_2 : \tau_2, \psi_2 \quad \text{unify}(\tau_1, \text{num}, \psi_2) = \psi_3 \quad \text{unify}(\tau_2, \text{num}, \psi_3) = \psi_4}{\Gamma, \psi_0 \vdash e_1 * e_2 : \text{num}, \psi_4} \\
 \\
 \frac{\tau\text{-Div} \quad \Gamma, \psi_0 \vdash e_1 : \tau_1, \psi_1 \quad \Gamma, \psi_1 \vdash e_2 : \tau_2, \psi_2 \quad \text{unify}(\tau_1, \text{num}, \psi_2) = \psi_3 \quad \text{unify}(\tau_2, \text{num}, \psi_3) = \psi_4}{\Gamma, \psi_0 \vdash e_1 / e_2 : \text{num}, \psi_4} \\
 \\
 \frac{\tau\text{-Mod} \quad \Gamma, \psi_0 \vdash e_1 : \tau_1, \psi_1 \quad \Gamma, \psi_1 \vdash e_2 : \tau_2, \psi_2 \quad \text{unify}(\tau_1, \text{num}, \psi_2) = \psi_3 \quad \text{unify}(\tau_2, \text{num}, \psi_3) = \psi_4}{\Gamma, \psi_0 \vdash e_1 \% e_2 : \text{num}, \psi_4} \\
 \\
 \frac{\tau\text{-Eq} \quad \Gamma, \psi_0 \vdash e_1 : \tau_1, \psi_1 \quad \Gamma, \psi_1 \vdash e_2 : \tau_2, \psi_2 \quad \text{unify}(\tau_1, \text{num}, \psi_2) = \psi_3 \quad \text{unify}(\tau_2, \text{num}, \psi_3) = \psi_4}{\Gamma, \psi_0 \vdash e_1 == e_2 : \text{bool}, \psi_4} \\
 \\
 \frac{\tau\text{-Lt} \quad \Gamma, \psi_0 \vdash e_1 : \tau_1, \psi_1 \quad \Gamma, \psi_1 \vdash e_2 : \tau_2, \psi_2 \quad \text{unify}(\tau_1, \text{num}, \psi_2) = \psi_3 \quad \text{unify}(\tau_2, \text{num}, \psi_3) = \psi_4}{\Gamma, \psi_0 \vdash e_1 < e_2 : \text{bool}, \psi_4} \\
 \\
 \frac{\tau\text{-Val} \quad \Gamma, \psi_0 \vdash e_1 : \tau_1, \psi_1 \quad \text{gen}(\tau_1, \Gamma, \psi_1) = \tau_1^\forall \quad \Gamma[x : \tau_1^\forall], \psi_1 \vdash e_2 : \tau_2, \psi_2}{\Gamma, \psi_0 \vdash \text{val } x = e_1; e_2 : \tau_2, \psi_2} \\
 \\
 \frac{\tau\text{-Id} \quad \Gamma(x) = \tau^\forall \quad \text{inst}(\tau^\forall, \psi) = (\tau, \psi')}{\Gamma, \psi \vdash x : \tau, \psi'} \quad \frac{\tau\text{-Fun} \quad \alpha_p \notin \psi \quad \Gamma[x : \alpha_p], \psi[\alpha_p \mapsto \bullet] \vdash e : \tau, \psi'}{\Gamma, \psi \vdash \lambda x.e : \alpha_p \rightarrow \tau, \psi'} \\
 \\
 \frac{\tau\text{-App} \quad \alpha_r \notin \psi_a \quad \text{unify}(\tau_a \rightarrow \alpha_r, \tau_f, \psi_a[\alpha_r \mapsto \bullet]) = \psi'}{\Gamma, \psi \vdash e_f(e_a) : \alpha_r, \psi'} \\
 \\
 \frac{\tau\text{-Rec} \quad \alpha_p, \alpha_r \notin \psi \quad \alpha_p \neq \alpha_r \quad \Gamma_1 = \Gamma[x_p \mapsto (\alpha_p \rightarrow \alpha_r)] \quad \Gamma_2 = \Gamma_1[x_p \mapsto \alpha_p] \quad \Gamma_2, \psi[\alpha_p \mapsto \bullet, \alpha_r \mapsto \bullet] \vdash e_b : \tau_b, \psi_b \quad \text{unify}(\tau_b, \alpha_r, \psi_b) = \psi_r \quad \Gamma_1, \psi_r \vdash e_s : \tau_s, \psi_s}{\Gamma, \psi \vdash \text{def } x_f(x_p) = e_b; e_s : \tau_s, \psi_s} \\
 \\
 \frac{\tau\text{-If} \quad \Gamma, \psi \vdash e_c : \tau_c, \psi_c \quad \Gamma, \psi_c \vdash e_t : \tau_t, \psi_t \quad \text{unify}(\tau_c, \text{bool}, \psi_e) = \psi' \quad \text{unify}(\tau_t, \tau_e, \psi') = \psi''}{\Gamma, \psi \vdash \text{if } (e_c) e_t \text{ else } e_e : \tau_t, \psi''}
 \end{array}$$

type unification is defined as a partial function:

$$\boxed{\text{unify} : (\mathbb{T} \times \mathbb{T} \times \Psi) \rightarrow \Psi}$$

$$\text{unify}(\tau_1, \tau_2, \psi) = \begin{cases} \psi & \text{if } \tau'_1 = \text{num} \wedge \tau'_2 = \text{num} \\ \psi & \text{if } \tau'_1 = \text{bool} \wedge \tau'_2 = \text{bool} \\ \text{unify}(\tau_{1,r}, \tau_{2,r}, \text{unify}(\tau_{1,p}, \tau_{2,p}, \psi)) & \text{if } \tau'_1 = (\tau_{1,p} \rightarrow \tau_{1,r}) \wedge \tau'_2 = (\tau_{2,p} \rightarrow \tau_{2,r}) \\ \psi & \text{if } \tau'_1 = \alpha = \tau'_2 \\ \psi[\alpha \mapsto \tau'_2] & \text{if } \tau'_1 = \alpha \wedge \neg\text{occur}(\alpha, \tau'_2) \\ \psi[\alpha \mapsto \tau'_1] & \text{if } \tau'_2 = \alpha \wedge \neg\text{occur}(\alpha, \tau'_1) \end{cases}$$

where $\tau'_1 = \text{resolve}(\tau_1, \psi)$ and $\tau'_2 = \text{resolve}(\tau_2, \psi)$.

type resolving and occurrence checking are defined as following functions:

$$\boxed{\text{resolve} : (\mathbb{T} \times \Psi) \rightarrow \mathbb{T}}$$

$$\text{resolve}(\tau, \psi) = \begin{cases} \text{resolve}(\tau', \psi) & \text{if } \tau = \alpha \wedge \psi(\alpha) = \tau' \\ \text{resolve}(\tau_p, \psi) \rightarrow \text{resolve}(\tau_r, \psi) & \text{if } \tau = (\tau_p \rightarrow \tau_r) \\ \tau & \text{otherwise} \end{cases}$$

$$\boxed{\text{occur} : (\mathbb{X}_\alpha \times \mathbb{T} \times \Psi) \rightarrow \text{bool}}$$

$$\text{occur}(\alpha, \tau) = \begin{cases} \text{true} & \text{if } \tau = \alpha \\ \text{occur}(\alpha, \tau_p) \vee \text{occur}(\alpha, \tau_r) & \text{if } \tau = (\tau_p \rightarrow \tau_r) \\ \text{false} & \text{otherwise} \end{cases}$$

type generalization and instantiation are defined as following functions:

$$\boxed{\text{gen} : (\mathbb{T} \times \Gamma \times \Psi) \rightarrow \mathbb{T}^\vee}$$

$$\text{gen}(\tau, \Gamma, \psi) = \forall \alpha_1, \dots, \alpha_m. \tau \quad \text{where} \quad \text{free}_\tau(\tau, \psi) \setminus \text{free}_\Gamma(\Gamma, \psi) = \{\alpha_1, \dots, \alpha_m\}$$

$$\boxed{\text{inst} : (\mathbb{T}^\vee \times \Psi) \rightarrow (\mathbb{T} \times \Psi)}$$

$$\text{inst}(\forall \alpha_1, \dots, \alpha_m. \tau, \psi) = (\text{resolve}(\tau, \psi[\alpha_1 \mapsto \alpha'_1, \dots, \alpha_m \mapsto \alpha'_m]), \psi[\alpha'_1 \mapsto \bullet, \dots, \alpha'_m \mapsto \bullet])$$

$$\text{where } \alpha'_1, \dots, \alpha'_m \notin \psi \wedge \forall 1 \leq i < j \leq m. \alpha'_i \neq \alpha'_j$$

free type variables are defined as following functions:

$$\boxed{\text{free}_\tau : (\mathbb{T} \times \Psi) \rightarrow \mathcal{P}(\mathbb{X}_\alpha)}$$

$$\text{free}_\tau(\tau, \psi) = \begin{cases} \text{free}_\tau(\tau', \psi) & \text{if } \tau = \alpha \wedge \psi(\alpha) = \tau' \\ \{\alpha\} & \text{if } \tau = \alpha \wedge \psi(\alpha) = \bullet \\ \text{free}_\tau(\tau_p, \psi) \cup \text{free}_\tau(\tau_r, \psi) & \text{if } \tau = (\tau_p \rightarrow \tau_r) \\ \emptyset & \text{otherwise} \end{cases}$$

$$\boxed{\text{free}_{\tau^\vee} : (\mathbb{T}^\vee \times \Psi) \rightarrow \mathcal{P}(\mathbb{X}_\alpha)}$$

$$\text{free}_{\tau^\vee}(\forall \alpha_1, \dots, \alpha_m. \tau, \psi) = \text{free}_\tau(\tau, \psi) \setminus \{\alpha_1, \dots, \alpha_m\}$$

$$\boxed{\text{free}_\Gamma : (\Gamma \times \Psi) \rightarrow \mathcal{P}(\mathbb{X}_\alpha)}$$

$$\text{free}_\Gamma([x_1 : \tau_1^\vee, \dots, x_n : \tau_n^\vee], \psi) = \text{free}_{\tau^\vee}(\tau_1^\vee, \psi) \cup \dots \cup \text{free}_{\tau^\vee}(\tau_n^\vee, \psi)$$

5 SEMANTICS

We use the following notations in the semantics:

$$\begin{array}{lll} \text{Values} & \mathbb{V} \ni v ::= n & (\text{NumV}) \\ & | b & (\text{BoolV}) \\ & | \langle \lambda x. e, \sigma \rangle & (\text{CloV}) \\ \text{Environments} & \sigma \in \mathbb{X} \xrightarrow{\text{fin}} \mathbb{V} & (\text{Env}) \end{array}$$

The big-step operational (natural) semantics of TIFAE is defined as follows:

$$\boxed{\sigma \vdash e \Rightarrow v}$$

$$\begin{array}{c} \text{Num} \quad \frac{}{\sigma \vdash n \Rightarrow n} \quad \text{Bool} \quad \frac{}{\sigma \vdash b \Rightarrow b} \\ \text{Add} \quad \frac{\sigma \vdash e_1 \Rightarrow n_1 \quad \sigma \vdash e_2 \Rightarrow n_2}{\sigma \vdash e_1 + e_2 \Rightarrow n_1 + num_2} \quad \text{Mul} \quad \frac{\sigma \vdash e_1 \Rightarrow n_1 \quad \sigma \vdash e_2 \Rightarrow n_2}{\sigma \vdash e_1 * e_2 \Rightarrow n_1 * n_2} \\ \text{Div} \quad \frac{\sigma \vdash e_1 \Rightarrow n_1 \quad \sigma \vdash e_2 \Rightarrow n_2 \quad n_2 \neq 0}{\sigma \vdash e_1 / e_2 \Rightarrow n_1 \div n_2} \quad \text{Mod} \quad \frac{\sigma \vdash e_1 \Rightarrow n_1 \quad \sigma \vdash e_2 \Rightarrow n_2 \quad n_2 \neq 0}{\sigma \vdash e_1 \% e_2 \Rightarrow n_1 \% n_2} \\ \text{Eq} \quad \frac{\sigma \vdash e_1 \Rightarrow n_1 \quad \sigma \vdash e_2 \Rightarrow n_2}{\sigma \vdash e_1 == e_2 \Rightarrow n_1 = n_2} \quad \text{Lt} \quad \frac{\sigma \vdash e_1 \Rightarrow n_1 \quad \sigma \vdash e_2 \Rightarrow n_2}{\sigma \vdash e_1 < e_2 \Rightarrow n_1 < n_2} \\ \text{Val} \quad \frac{\sigma \vdash e_1 \Rightarrow v_1 \quad \sigma[x \mapsto v_1] \vdash e_2 \Rightarrow v_2}{\sigma \vdash \text{val } x = e_1; e_2 \Rightarrow v_2} \quad \text{Id} \quad \frac{x \in \text{Domain}(\sigma)}{\sigma \vdash x \Rightarrow \sigma(x)} \\ \text{Fun} \quad \frac{}{\sigma \vdash \lambda x. e \Rightarrow \langle \lambda x. e, \sigma \rangle} \quad \text{Rec} \quad \frac{\sigma' = \sigma[x_0 \mapsto \langle \lambda x_1. e_2, \sigma' \rangle] \quad \sigma' \vdash e_3 \Rightarrow v_3}{\sigma \vdash \text{def } x_0(x_1) = e_2; e_3 \Rightarrow v_3} \\ \text{App} \quad \frac{\sigma \vdash e_0 \Rightarrow \langle \lambda x. e_2, \sigma' \rangle \quad \sigma \vdash e_1 \Rightarrow v_1 \quad \sigma'[x \mapsto v_1] \vdash e_2 \Rightarrow v_2}{\sigma \vdash e_0(e_1) \Rightarrow v_2} \\ \text{If}_T \quad \frac{\sigma \vdash e_0 \Rightarrow \text{true} \quad \sigma \vdash e_1 \Rightarrow v_1}{\sigma \vdash \text{if } (e_0) e_1 \text{ else } e_2 \Rightarrow v_1} \quad \text{If}_F \quad \frac{\sigma \vdash e_0 \Rightarrow \text{false} \quad \sigma \vdash e_2 \Rightarrow v_2}{\sigma \vdash \text{if } (e_0) e_1 \text{ else } e_2 \Rightarrow v_2} \end{array}$$