

Type-checking with Grace

Kim B. Bruce*
Pomona College

Andrew Black
Portland State University

James Noble, Michael Homer
Victoria University of Wellington

February 22, 2023

1 Preliminary Definitions

Definition 1.1 *Let \mathcal{V} be an infinite collection of type variables, \mathcal{L} be an infinite collection of labels, and \mathcal{C} be a collection of type constants that includes at least the type constants **Boolean**, **Number**, **Object**, and **Done**. The simple pre-type expressions, *PreType*, and record pre-type expressions, *PreRecType*, of Grace with respect to \mathcal{V} , \mathcal{L} , and \mathcal{C} are given by the following context-free grammar. We assume $t, t_i \in \mathcal{V}$, $c \in \mathcal{C}$, and $m_i \in \mathcal{L}$ in the following.*

$$\tau \in \text{PreType} ::= t \mid c \mid \mathbf{ref} \ \tau \mid \tau_1 \times \dots \times \tau_n \rightarrow \tau \mid \forall t_1 <: \tau_1, \dots, t_n <: \tau_n. \tau \mid \{m_1: \tau_1; \dots; m_n: \tau_n\} \mid < l_1: \tau_1, \dots, l_n: \tau_n >$$

Type **Object** will stand for the type of an imperative command expression, i.e., an expression that does not return a value. The type **Object** is a supertype of all object types, and (in our current implementation) contains **asString** and **asDebugString** methods that are (implicitly) inherited by all object types.

Reference types are the types of variables. That is, if x is a variable holding values of type τ , then x has type **ref** τ . This notation allows us to distinguish between values of type τ and variables that hold values of that type.

As is standard, the type $\tau_1 \times \dots \times \tau_n \rightarrow \tau$ is the type of functions taking parameters of type τ_1 through τ_n and returning a value of type τ . The type $\forall t_1 <: \tau_1, \dots, t_n <: \tau_n. \tau$ represents bounded polymorphic functions (that is functions that take types as parameters). Unbounded polymorphic functions can be represented by terms of the form $\langle t_1 <: \mathbf{Object}, \dots, t_n <: \mathbf{Object} \rangle. \tau$. The identifiers t_i are bound by these type expressions. As usual we identify polymorphic types that are the same up to renaming of the bound variable.

The type $\{m_1: \tau_1; \dots; m_n: \tau_n\}$ represents the type of an object with public methods m_1, \dots, m_n . On the other hand the type $< l_1: \tau_1, \dots, l_n: \tau_n >$ represents a tagged variant type. A typical element, written $< l_i = e_i >$, has the type if e_i has type τ_i .

We will use the abbreviation **Block0** $[\tau, \sigma]$ to stand for the type $\{apply: \rightarrow \sigma\}$, **Block1** $[\tau, \sigma]$ to stand for the type $\{apply: \tau \rightarrow \sigma\}$, etc.

The axioms and rules for determining valid types and constructors are given with respect to a set, C , of simple type constraints, which provide information about free type variables. The definition

*Corresponding author: Kim B. Bruce, Department of Computer Science, Pomona College, Claremont, CA 91711. kim@cs.pomona.edu

of type constraints, the rules for determining valid types and constructors, and the matching and subtyping rules are mutually recursive.

Definition 1.2 *Relations of the form $t <: \tau$ and $t = \tau$, where t is a type variable and τ is a type expression, are said to be simple type constraints. A type constraint system, C , is defined as follows:*

1. *The empty set, \emptyset , is a type constraint system.*
2. *If C is a type constraint system such that $C \vdash \tau : \text{Type}$, and t is a type variable that does not appear in C , then $C \cup \{t <: \tau\}$ is a type constraint system.*
3. *If C is a type constraint system such that $C \vdash \tau <: \text{Object}$, and t is a type variable that does not appear in C or τ , then $C \cup \{t = \tau\}$ is a type constraint system.*

In Figure 1 we include axioms and rules for determining which are the legal type and constructor expressions. In the **ObjectType** rule, the types of all methods must be function or polymorphic function types.

The axioms and rules for $<:$ can be found in Figure 2. Reference types have no non-trivial subtypes. The subtyping rule for function types is contravariant in the argument type and covariant in the result type. The subtyping rules for polymorphic types support covariant changes in the return type, but no changes in the bounds of the type parameters. While the rules could be generalized to allow contravariant changes in the type bounds, the decidability of subtyping would be lost. The subtyping rule for object types allows both depth and width subtyping.

Note that Pierce, pg 197, uses different rules.

$$C \vdash c: \text{TYPE}, \text{ for } c \in \mathcal{C}$$

$$C \cup \{t <: \tau\} \vdash t: \text{TYPE}$$

$$C \cup \{t = \tau\} \vdash t: \text{TYPE}$$

$$\frac{C \vdash \sigma_i: \text{TYPE for } i = 1, \dots, n \quad C \vdash \tau: \text{TYPE}}{C \vdash \sigma_1 \times \dots \times \sigma_n \rightarrow \tau: \text{TYPE}}$$

$$\frac{C \cup \{t <: \gamma\} \vdash \tau: \text{TYPE}}{C \vdash \forall t <: \gamma. \tau: \text{TYPE}}$$

$$\frac{C \vdash \tau_i: \text{TYPE for } 1 \leq n}{C \vdash \{l_i: \tau_i\}_{i \leq n}: \text{TYPE}}$$

$$\frac{C \vdash \tau_i: \text{TYPE for } 1 \leq n}{C \vdash < l_i: \tau_i >_{i \leq n}: \text{TYPE}}$$

$$\frac{C \vdash \tau: \text{TYPE}}{C \vdash \text{ref } \tau: \text{TYPE}}$$

$$\frac{C \cup \{t: \text{TYPE}\} \vdash \kappa: K}{C \vdash \text{TFunc}[t] \kappa: \text{TYPE} \Rightarrow K}$$

$$\frac{C \vdash \kappa: \text{TYPE} \Rightarrow K, \quad C \vdash \tau: \text{TYPE}}{\kappa[\tau]: K}$$

Figure 1: Type and constructor rules

$$\begin{array}{l}
\text{Ref}_{<}: \frac{C \vdash \tau : \mathbf{TYPE}}{C \vdash \tau <: \tau} \\
\\
\text{Hyp}_{<}: C \cup \{t <: \tau\} \vdash t <: \tau \\
\\
\text{Trans}_{<}: \frac{C \vdash \sigma <: \tau \quad C \vdash \tau <: \delta}{C \vdash \sigma <: \delta} \\
\\
\text{Func}_{<}: \frac{C \vdash \sigma <: \sigma' \quad C \vdash \tau' <: \tau}{C \vdash \sigma' \rightarrow \tau' <: \sigma \rightarrow \tau} \\
\\
\text{Poly}_{<}: \frac{C \cup \{u : \mathbf{TYPE}\} \vdash \tau'[t' \mapsto u] <: \tau[t \mapsto u] \quad u \notin FV(\tau') \cup FV(\tau)}{C \vdash \forall t'. \tau' <: \forall t. \tau} \\
\\
\text{SBdedPoly}_{<}: \frac{C \cup \{u <: \gamma\} \vdash \tau'[t' \mapsto u] <: \tau[t \mapsto u] \quad u \notin FV(\tau') \cup FV(\tau) \cup FV(\gamma)}{C \vdash \forall t' <: \gamma. \tau' <: \forall t <: \gamma. \tau} \\
\\
\text{ObjType}_{<}: \frac{C \vdash \tau'_i <: \tau_i \text{ for } 1 \leq i \leq n, \quad C \vdash \tau'_i : \mathbf{TYPE} \text{ for } n+1 \leq i \leq n+m}{C \vdash \{l_j : \tau'_j\}_{j \leq n+m} <: \{l_j : \tau_j\}_{j \leq n}} \\
\\
\text{VarWidth}_{<}: \frac{C \vdash \tau'_i <: \tau_i \text{ for } 1 \leq i \leq n, \quad C \vdash \tau'_i : \mathbf{TYPE} \text{ for } n+1 \leq i \leq n+m}{C \vdash \langle l_j : \tau'_j \rangle_{j \leq n} <: \langle l_j : \tau_j \rangle_{j \leq n+m}} \\
\\
\text{VariantLeft}_{<}: \frac{C \vdash \sigma <: \gamma \quad C \vdash \tau <: \gamma}{C \vdash \sigma | \tau <: \gamma} \\
\\
\text{VariantRight}_{<}: \frac{C \vdash \gamma <: \sigma \text{ or } C \vdash \gamma <: \tau}{C \vdash \gamma <: \sigma | \tau} \\
\\
\text{AndLeft}_{<}: \frac{C \vdash \sigma <: \gamma \text{ or } C \vdash \tau <: \gamma}{C \vdash \sigma \& \tau <: \gamma} \\
\\
\text{AndRight}_{<}: \frac{C \vdash \gamma <: \sigma, \quad C \vdash \gamma <: \tau}{C \vdash \gamma <: \sigma \& \tau}
\end{array}$$

Figure 2: Subtyping rules

1.1 Grace Expression Syntax

Definition 1.3 The set, *PreTerm*, of pre-terms of Grace over a set \mathcal{B} of term constants, a set \mathcal{L} of labels, and a set \mathcal{X} of term identifiers is given by the following context-free grammar (we assume $x \in \mathcal{X}$, $b \in \mathcal{B}$, $l, m \in \mathcal{L}$ and $\sigma, \tau \in \text{PreType}$). The notation $\text{opt}(exp)$ indicates that exp is optional in the syntax.

<i>codeSequence</i>	$::= \text{codeUnit } \text{opt}(\text{codeSequence})$
<i>codeUnit</i>	$::= \text{declaration} \mid \text{statement}$
<i>declaration</i>	$::= \text{varDeclaration} \mid \text{defDeclaration} \mid \text{classDeclaration} \mid$ $\text{typeDeclaration} \mid \text{methodDeclaration}$
<i>statement</i>	$::= \text{identifier} := \text{expression} \mid \text{if}(\text{expression}) \text{ then block else block}$
<i>innerCodeSequence</i>	$::= \text{innerCodeUnit } \text{opt}(\text{innerCodeSequence})$
<i>innerCodeUnit</i>	$::= \text{innerDeclaration} \mid \text{statement}$
<i>innerDeclaration</i>	$::= \text{varDeclaration} \mid \text{defDeclaration} \mid \text{classDeclaration} \mid$ typeDeclaration
<i>varDeclaration</i>	$::= \text{var identifier} : \text{typeExpression } \text{opt}(:= \text{expression})$
<i>defDeclaration</i>	$::= \text{def identifier} : \text{typeExpression} = \text{expression}$
<i>methodDeclaration</i>	$::= \text{method identifier } \text{opt}(\text{genericFormals}) \text{ methodFormals}$ $\rightarrow \text{nonEmptyTypeExpression } \text{opt}(\text{typeConstraints})$ $\{ \text{innerCodeSequence} \}$
<i>genericFormals</i>	$::= \llbracket t_1, \dots, t_n \rrbracket$
<i>methodFormals</i>	$::= (id_1 : \text{typeExp}_1, \dots, id_n : \text{typeExp}_n)$
<i>typeConstraints</i>	$::= \text{where } t_1 < : \text{typeExp}_1, \dots, t_n < : \text{typeExp}_n$
<i>typeDeclaration</i>	$::= \text{type typeId } \text{opt}(\text{genericFormals}) = \text{typeExpression}$
<i>classDeclaration</i>	$::= \text{class cid.constrId } \text{opt}(\text{genericFormals}) \text{opt}(\text{methodFormals})$ $\rightarrow \text{typeExp } \text{opt}(\text{typeConstraints})$ $\{ \text{opt}(\text{inherit scid}(t_1 \dots t_n)(e_1, \dots, e_m))$ $\text{codeSequence} \}$
<i>objectLiteral</i>	$::= \text{object} \{ \text{opt}(\text{inherit scid}(t_1 \dots t_n)(e_1, \dots, e_m))$ $\text{codeSequence} \}$

The only difference between *declaration* and *innerDeclaration* is that *innerDeclarations* may not include method declarations. The distinction is necessary because a method declaration may not directly contain a method declaration (though it can contain an object which itself has method declarations). The type-checking axioms and rules for Grace are given in terms of a type constraint system, C , as defined earlier, and an *static type environment*, E , which assigns types to free identifiers.

Definition 1.4 A static type environment, E , (with respect to C) is a finite set of associations between identifier and type expressions of the form $x : \tau$, where each x is unique in E and $C \vdash \tau : \text{TYPE}$. If the relation $x : \tau \in E$, then we write $E(x) = \tau$.

The collection *Term* of terms of Grace with respect to C , E is the set of pre-terms that can be assigned types with respect to the type-assignment axioms and rules in Section 2.

The type-assignment rules provided in Section 2 yield expanded type constraints and assignments as well as types. These expanded type constraints assignments are used to type check the

rest of the program. Thus an assertion of the form $C, E \vdash M \diamond \langle \tau, C', E' \rangle$ indicates that if a term M is processed under the type constraint system C and syntactic type assignment E , then M has type τ and the richer syntactic type constraint C' and type assignment E' result. These richer sets will be used in type-checking later terms. If M is a command or declaration then the type τ will be **Done**.

2 Type-checking rules

Typing rules return a triple $\langle \tau, C, E \rangle$ where τ is the type of the expression and C and E are the new type and variable environments resulting from processing the expression.

For simplicity in the following we ignore the rules for *innerCodeUnit* and *innerCodeSequences* as they duplicate those for *codeUnit* and *codeSequence*. Similarly, we generally assume that optional items are included in syntactic constructs. It is trivial to write the rules without those items.

We will also assume that all “self-inflicted” method invocations (message sends to “self”) include an explicit “self” receiver.

Rules

$$\begin{array}{l}
\textit{Subsumption} \quad \frac{C, E \vdash e \diamond \langle \tau, C', E' \rangle \quad C \vdash \tau <: \tau'}{C, E \vdash e \diamond \langle \tau', C', E' \rangle} \\
\\
\textit{CodeSeq} \quad \frac{C, E \vdash \textit{codeUnit} \diamond \langle \tau', C', E' \rangle \quad C', E' \vdash \textit{codeSequence} \diamond \langle \tau'', C'', E'' \rangle}{C, E \vdash \textit{codeUnit codeSequence} \diamond \langle \tau'', C'', E'' \rangle} \\
\\
\textit{VarDcl} \quad \frac{C \vdash \tau <: \text{Done}}{C, E \vdash \text{var } x: \tau \diamond \langle \text{Done}, C, E + \{x: \text{ref } \tau\} \rangle} \\
\\
\textit{VarDclInit} \quad \frac{C, E \vdash \textit{exp} \diamond \langle \tau, -, - \rangle}{C, E \vdash \text{var } x: \tau = \textit{exp} \diamond \langle \text{Done}, C, E + \{x: \text{ref } \tau\} \rangle} \\
\\
\textit{DefDcl} \quad \frac{C, E \vdash \textit{exp}: \langle \tau, -, - \rangle}{C, E \vdash \text{def } x: \tau = \textit{exp} \diamond \langle \text{Done}, C, E + \{x: \tau\} \rangle} \\
\\
\textit{MethodDcl} \quad \frac{C + \textit{newConstraints}, E + \textit{methodFormals} \vdash \textit{innerCodeSequence} \diamond \langle \tau, -, - \rangle}{\begin{array}{l} C, E \vdash \text{method } \textit{id} \textit{genericFormals} (\textit{methodFormals}) \rightarrow \tau \\ \textit{typeConstraints} \{ \textit{innerCodeSequence} \} \diamond \langle \text{Done}, C, E \rangle \end{array}}
\end{array}$$

In the above, if **genericFormals** is empty, then *newConstraints* is empty. Otherwise if *typeConstraints* = $[t_1 <: T_1, \dots, t_n <: T_n]$ and U is the set of types in *genericFormals* that do not occur on the left side of a type constraint in *typeConstraints*, then $\textit{newConstraints} = \{t_1 <: T_1, \dots, t_n <: T_n\} \cup \{t <: \text{Object} \mid t \in U\}$.

We require t_1, \dots, t_n from **typeConstraints** to all occur in *genericFormals*.

Note that the method *id* will be added to C when the class as a whole is type-checked rather than here.

$$\textit{TypeDef} \quad \frac{C, E \vdash \text{type } \textit{typeId} \textit{genericFormals} = \textit{typeExpression} \diamond}{\langle \text{Done}, C + \{ \textit{typeId} = \text{TFunc}(\textit{genericFormals}) \textit{typeExpression} \}, E \rangle}$$

$$\text{Object} \quad \frac{
\begin{array}{c}
C, E' \vdash \text{codeSequence} \diamond \langle -, -, - \rangle, \\
C, E \vdash \text{supObj} \diamond \langle \sigma, -, - \rangle
\end{array}
}{
C, E \vdash \text{object} \{ \text{inherits supObj} \\ \text{codeSequence} \} \diamond \langle \tau, C, E \rangle
}$$

where $\sigma' = \{m'_1: \sigma'_1, \dots, m'_p: \sigma'_p\}$ be the type formed by the declared types of all public and confidential methods of supObj . It is an extension of σ .

$\tau = \{m_1: \tau_1, \dots, m_n: \tau_n\}$ is the subtype of σ formed by adding the declared types of all new public methods.

$\tau' = \{m_1: \tau_1, \dots, m_k: \tau_k\}$ is the subtype of σ' formed by adding the declared types of all new methods, including confidential and private methods.

$E' = E + \{\text{self}: \tau', \text{super}: \sigma'\}$

Note that τ and τ' include the types of the implicit methods generated by the public and confidential defs and variable declarations. If **self** is passed out as a parameter then its type must be restricted to τ . Keyword **super** may only be used as a receiver of method requests.

A class of the form

```

class classId genericFormals methodFormals
  → typeExp typeConstraints
  {inherits supObj
   codeSequence}

```

can be translated as

```

method classId genericFormals methodFormals
  → typeExp typeConstraints
  object {inherits supObj
          codeSequence}
}

```

and type-checked via the translation.

$$\text{MessageSend} \quad \frac{
\begin{array}{c}
C, E \vdash \text{obj} \diamond \langle \{m: \langle t_1 <: \tau_1, \dots, t_n <: \tau_n \rangle \gamma_1 \times \dots \times \gamma_k \rightarrow \tau \}, -, - \rangle \\
C \vdash \sigma_i <: \tau_i \text{ for } 1 \leq i \leq n \\
C, E \vdash e_j \diamond \langle \gamma_j[t_i \mapsto \sigma_i], -, - \rangle \text{ for } 1 \leq j \leq k
\end{array}
}{
C, E \vdash \text{obj.m} \langle \sigma_1, \dots, \sigma_n \rangle (e_1, \dots, e_k) \diamond \langle \tau[t_i \mapsto \sigma_i], C, E \rangle
}$$

NumberLit $C, E \vdash \text{numberLiteral}: \text{Number}$

StringLit $C, E \vdash \text{stringLiteral}: \text{String}$

True $C, E \vdash \text{true}: \text{Boolean}$

False

$C, E \vdash \text{false} : \text{Boolean}$

Cond

$$\frac{\begin{array}{l} C, E \vdash \text{cond} \diamond \langle \text{Boolean}, C', E' \rangle \\ C', E' \vdash \text{blk} \diamond \langle \text{Block0}[\tau], C'', E'' \rangle \\ C'', E'' \vdash \text{blk}' \diamond \langle \text{Block0}[\tau'], C''', E''' \rangle \end{array}}{C, E \vdash \text{if } (\text{cond}) \text{ then } \text{blk} \text{ else } \text{blk}' \diamond \langle \tau | \tau', C''', E''' \rangle}$$

Identifier

$C, E \vdash x : \tau \quad \text{if } E(x) = \tau$

Assn

$$\frac{C, E \vdash x \diamond \langle \text{ref } \tau, C', E' \rangle \quad C', E' \vdash M \diamond \langle \tau', C'', E'' \rangle \quad C', E' \vdash \tau' <: \tau}{C, E \vdash x := M \diamond \langle \text{Done}, C'', E'' \rangle}$$

R – Value

$$\frac{C, E \vdash M \diamond \langle \text{ref } \tau, C', E' \rangle}{C, E \vdash M \diamond \langle \tau, C', E' \rangle}$$

Perhaps we should just mark in the abstract syntax when we need the *r-value* rather than the *l-value* of a variable.

Recall that blocks are abbreviations for objects with a single method `apply`. The derived rule would look like the following.

Block

$$\frac{C, E + \{x_1 : \tau_1, \dots, x_n : \tau_n\} \vdash M : \tau}{C, E \vdash \{x_1 : \tau_1, \dots, x_n : \tau_n \rightarrow M\} : \{apply : \tau_1 \times \dots \times \tau_n \rightarrow \tau\}}$$

Match

$$\frac{\begin{array}{l} C, E \vdash v \diamond \langle \tau, C_0, E_0 \rangle \\ C_0, E_0 \vdash \text{blk}_1 \diamond \langle \text{Block1}[\sigma_1, \tau_1], C_1, E_1 \rangle \\ \dots \\ C_{n-1}, E_{n-1} \vdash \text{blk}_n \diamond \langle \text{Block1}[\sigma_n, \tau_n], C_n, E_n \rangle \\ C \vdash \tau <: \sigma_1 | \dots | \sigma_n \end{array}}{C, E \vdash \text{match}(v) \text{ case } \text{blk}_1 \dots \text{case } \text{blk}_n \diamond \langle \tau_1 | \dots | \tau_n, C_n, E'_n \rangle}$$

Deal with cases like $\{7 \rightarrow \text{"hello"}\}$ later.

Not done yet:

- try-catch, should have the type obtained by variant of the try and catch blocks
- module, should be like object, but has imports to be dealt with
- array (should be line-ups)
- generics
- import statement
- return: method with return should return the type of the value returned — last expression.
- dialect