CubeX with Laziness

Ross Tate

November 26, 2013

0.1 Laziness

This extension of the CubeX employs lazy evaluation in the language. All expressions should be evaluated lazily (only when their values are needed). In addition, we've made the following changes/simplifications (highlighted throughout the spec with "**"):

- 1. There is an added select method on the Boolean type. (See spec for details.)
- 2. Certain operators should appropriately short-circuit. (See spec for details.)
- 3. Interfaces do not implement any methods.
- 4. There is only single interface inheritance. (No class inheritance or multiple inheritance.)
- 5. There is no Intersection type $(\tau \cap \tau)$.
- 6. There are no if statements, no while loops, and no for loops.
- 7. There are no Iterables, and thus no divide, no modulo, no onwards, no through, and no Strings.
- 8. Input and output should now be of type Integer, not Iterable(String). That is, instead of void print_line(char* str, int len), int next_line_length(), and void read_line(char* buffer), you'll have int read_int() and void print_int(int output).

0.2 Grading

This extension will NOT be graded in stages. You should prepare for ALL tests to include:

- Statements
- Non-generic functions
- Generic functions
- Non-Generic classes without inheritance
- Generic classes without inheritance
- Generic classes with single interface inheritance.

No tests will include Generic classes with single interface inheritance + interface method implementations, Generic classes with class inheritance, or Generic classes with multiple interface inheritance. However, we will not test for the exclusion of these or any other CubeX language features that have been dropped specifically for this Laziness extension (e.g. We will not test to make sure a while loop does not parse/type-check/execute.) You are free to keep pre-existing features in your language so long as their inclusion does not break any tests pertaining to this modified spec. Your implementation will be graded on correctness (a single call to print_int(int output) should be made with the correct return value of the program) and performance (memory management and optimizations outlined in PA4 and PA5 should still be applied).

1 Lexing and Parsing

1.1 Core and Full Languages

 $\nu_v ::= \text{variable/function/method names}$ $\nu_c ::= \text{class/interface names}$ $\nu_p ::= \text{type-parameter names}$ $\nu_{vc} ::= \nu_v \mid \nu_c$

```
kind context \Theta ::= \nu_p, \ldots, \nu_p
       type context \Gamma
                                            ::= \nu_v: \tau, \ldots, \nu_v: \tau
               ** type \tau
                                            ::= \nu_p \mid \nu_c \langle \tau, \ldots, \tau \rangle \mid \top \mid \bot
       type scheme \sigma
                                            ::= \langle \Theta \rangle (\Gamma) : \tau
    ** expression e
                                            ::= \nu_v \mid \nu_{vc}\langle \tau, \dots, \tau \rangle(e, \dots, e) \mid e.\nu_v\langle \tau, \dots, \tau \rangle(e, \dots, e) \mid \mathbf{true} \mid \mathbf{false} \mid n \mid e ? e : e
     ** statement s
                                            ::= \{s \dots s\} \mid \nu_v \coloneqq e; | \mathbf{return} \; e;
        ** interface i
                                            ::= interface \nu_c \langle \Theta \rangle extends \tau {fun \nu_v \sigma; ... fun \nu_v \sigma; }
                                            ::= \operatorname{class} \nu_e(\Theta)(\Gamma) \operatorname{extends} \tau \{s \dots s \operatorname{super}(e, \dots, e); \operatorname{fun} \nu_v \sigma s \dots \operatorname{fun} \nu_v \sigma s \}
                      class c
               program p
                                            ::= s \mid s \dots s \mid p \mid \mathbf{fun} \ \nu_v \sigma \ s \dots \mathbf{fun} \ \nu_v \sigma \ s \mid i \mid p \mid c \mid p
function context \Delta
                                            ::=
                                                      \varnothing \mid \Delta, \nu_{vc} \sigma
                                                      \varnothing \mid \Psi, \textbf{interface} \; \nu_c \langle \Theta \rangle \; \textbf{extends} \; \tau \; \{\Delta; \nu_v, \dots, \nu_v\} \mid \Psi, \textbf{class} \; \nu_c \langle \Theta \rangle \; \textbf{extends} \; \tau \; \{\Delta\}
      class context
                                    \Psi
                                            ::=
```

Figure 1: Lazy Cubex Core Language Grammar, which gives the grammar for the Cubex core language, along with definitions for the formal constructs Δ and Ψ , which are not part of the language per se but are useful in our formalization of the type system. Note that p stands for "program" and defines a syntactically-valid program in the Cubex core language. Note also that lists, represented with an elipsis, may consist of zero, one, or more elements. So a, \ldots, a may be the empty string, a, or some list of as separated by commas.

We distinguish between the Cubex <u>core</u> language, which is specified by the grammar in Figure 1.1, and the <u>full</u> language. The full language differs from the core language in a number of ways:

- It includes the following unary, binary, and ternary operators, listed in order of precedence, which are short for method calls on arbitrary classes (their signatures are those that are given for the built-in classes at the end of this document). All operators are left-associative.
 - 1. Unary prefixes and ! short for negative and negate respectively.
 - 2. ****** Multiplicative operator ***** short for **times**, which should short-circuit if either of its operands is 0.
 - 3. Additive + and short for plus and minus respectively.
 - 4. Inequality operators <, <=, >=, and > short for lessThan with an additional Boolean parameter indicating strictness, and with order reversed for >= and >.
 - 5. Equational operators == and !=, short for equals and equals followed by negate.
 - 6. ** Boolean operators & short for and followed by | short for or, both of which should appropriately short-circuit.
- Characters in the core language which have no obvious ASCII equivalent are represented in the full language in ASCII as follows:
 - Thing for \top and Nothing for \bot .
 - & for \wedge and \mid for \vee .
 - $< \text{for } \langle \text{ and } > \text{ for } \rangle.$
- Expressions (denoted *e* in the grammar) may be surrounded by parentheses.
- If there are no type parameters (i.e. when ⟨τ,...,τ⟩ is the empty list), a class/interface/function specification can drop the ⟨⟩.
- When calling a function/method/constructor with no type parameters, the $\langle \rangle$ may be dropped.
- If a class or interface just extends \top , the **extends** clause can be omitted.
- If the call to **super** has no arguments, it can be omitted.

• If a function implementation is simply return of some expression, the return can be abbreviated to =. For example,

```
fun foo(y : Integer) : Integer return y * 2;
```

may be abbreviated to

fun foo(y : Integer) : Integer = y * 2;

Finally, there is no name hiding. If a judgement requires binding a name twice, that judgement is ill formed and does not hold. However, all the different kinds of context are different namespaces. So a name can be bound both in Δ and Γ . To clarify, Γ and $\hat{\Gamma}$ use the same namespace, so a variable must not be bound in both.

1.2 Name, Keywords, Literals, etc.

The grammar uses ν with subscripts to denote both various kinds of names. We enforce a distinction between these kinds. Variables and function/method names (ν_v) are a lower-case letter followed by letters, numbers, and/or underscores. Class/interface names (ν_c) are an upper-case letter followed by one or more letters, numbers, and/or underscores. Type parameters (ν_p) are an upper-case letter alone. In the formalism we refer to all of these collectively as just ν .

Key words (i.e. any words receiving special treatment in the core language) cannot be used as names or variables.

Integer literals (denoted in the grammar by n) are in decimal only.

Finally, note that the grammar contains various other syntactic artifacts (semicolons, brackets, etc.) whose only legal uses in the language (outside of comments and string literals) are fully specified by their role in the grammar.

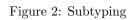
1.3 Whitespace and Comments

Unless it occurs inside a string literal, white space is disregarded except for its role of separating names and other tokens (e.g. ": =" is not the same as ":=" and only the latter can be used in a variable assignment). Unless it occurs inside a string literal, **#** starts a comment that extends to the end of the line. Unless it occurs inside a string literal, **'** starts a comment that extends to the matching ', where these comments can be nested inside each other. Comments are disregarded except for their role of separating names and other tokens.

2 Validating

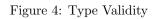
The following formalizes when a program is valid, incorporating type validity, type checking, and class/interface validation. Judgements take the form "(context) \vdash (property)". Ψ indicates the classes and interfaces in scope, what they directly inherit, and what their methods are. Θ indicates the type parameters in scope. Δ indicates the function names in scope and what their type scheme is. Γ typically indicates the immutable variable names in scope and what their type is. $\hat{\Gamma}$ typically indicates the mutable variable names in scope and what their type is. The following is a summary of the judgements used in this formalism and where their definitions can be found.

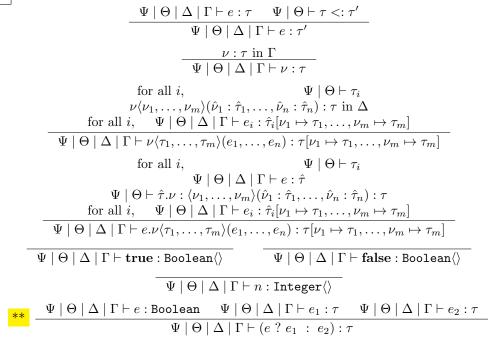
	Judgement	Meaning	Figure
$\Psi \mid \Theta$	$\vdash \tau <: \tau'$	au is a subtype of $ au'$	2
Ψ	$\vdash \hspace{0.2cm} \nu \langle \Theta \rangle \hspace{0.2cm} \textbf{extends} \hspace{0.2cm} \tau$	generic class/interface $\nu \langle \Theta \rangle$ directly inherits τ	2
$\Psi \mid \Theta$	$\vdash \Gamma <: \Gamma'$	Γ is a subcontext of Γ'	2
$\Psi \mid \Theta$	$\vdash \ \sigma \approx \sigma'$	σ and σ' are equivalent type schemes	
$\Psi \mid \Theta$	$\vdash \tau.\nu:\sigma$	method ν of type τ has type scheme σ	3
Ψ	$\vdash \hat{\nu} \langle \Theta \rangle . \nu : \sigma$	generic class/interface $\hat{\nu} \langle \Theta \rangle$ has method ν with type scheme σ	3
$\Psi \mid \Theta$	$\vdash au$	au is a valid type	4
$\Psi \mid \Theta$	$\vdash_{\hat{\tau}} \tau$	τ is an inheritable type with constructable component $\hat{\tau}$	4
$\Psi \mid \Theta$	$\vdash \Gamma$	Γ is a valid type context	4
$\Psi \mid \Theta$	$\vdash \sigma$	σ is a valid type scheme	4
$\Psi \mid \Theta \mid .$	$\Delta \mid \Gamma \vdash \ e : \tau$	expression e has type τ	5
$\Psi \mid \Theta \mid A$	$\Delta \mid \Gamma \mid \hat{\Gamma} \vdash s : \hat{\Gamma}'$	mutable variables $\hat{\Gamma}'$ are available after valid statement s	6
$\Psi \mid \Theta \mid A$	$\Delta \mid \Gamma \mid \hat{\Gamma} \vdash^b_{\tau} s : \hat{\Gamma}'$	like above and all returns have type τ with $b = \mathbf{true}$ guaranteeing a return	7
$\Psi \mid \Theta$	$\vdash \ au:\Delta$	Δ are all the methods of τ	8
Ψ	$\vdash \hat{\nu} \langle \Theta angle : \Delta$	Δ are the direct methods of generic class/interface $\hat{\nu}\langle\Theta\rangle$	8
$\Psi \mid \Theta$	$\vdash \ \tau:\nu$	$ au$ provides an implementation of method ν	9
$\Psi \mid \Delta \mid$	$\Gamma \qquad \vdash \ i: \Psi'$	i is a valid interface with signature Ψ'	10
$\Psi \mid \Delta \mid$	$\Gamma \qquad \vdash \ c: \Psi' \mid \Delta'$	c is a valid class with signature Ψ' and constructor Δ'	10
$\Psi \mid \Delta \mid$	$\Gamma \vdash p$	p is a valid program	11
	$\vdash p$	p is a valid program in the initial environment	11



$$\begin{array}{c|c} \underline{\Psi \mid \Theta \vdash \tau.\nu:\sigma} & \underline{\Psi \vdash \nu\langle\Theta\rangle.\nu:\sigma} \\ \hline & \underline{\Psi \mid \Theta \vdash \bot.\nu:\sigma} & \underline{\Psi \vdash \hat{\nu}\langle\nu_1,\ldots,\nu_n\rangle.\nu:\sigma} \\ \hline & \underline{\Psi \mid \Theta \vdash \hat{\nu}\langle\tau_1,\ldots,\tau_n\rangle.\nu:\sigma[\nu_1 \mapsto \tau_1,\ldots,\nu_n \mapsto \tau_n]} \\ \hline & \underline{\operatorname{interface} \,\hat{\nu}\langle\Theta\rangle \operatorname{extends} \tau \ \{\ldots;\nu\sigma;\ldots\} \operatorname{in} \Psi \\ \hline & \underline{\Psi \vdash \hat{\nu}\langle\Theta\rangle.\nu:\sigma} & \underline{\operatorname{class} \,\hat{\nu}\langle\Theta\rangle \operatorname{extends} \tau \ \{\ldots,\nu\sigma,\ldots\} \operatorname{in} \Psi \\ \hline & \underline{\Psi \vdash \hat{\nu}\langle\Theta\rangle.\nu:\sigma} \end{array}$$







The parentheses above have only been included to prevent confusion between the two colons.

Figure 5: Type Checking Expressions

$$\begin{array}{c|c} \underline{\Psi \mid \Theta \mid \Delta \mid \Gamma \mid \Gamma \vdash s : \Gamma} \\ \hline \underline{\Psi \mid \Theta \mid \Delta \mid \Gamma \mid \hat{\Gamma} \vdash s : \hat{\Gamma}' & \Psi \mid \Theta \vdash \hat{\Gamma}' <: \hat{\Gamma}'' \\ \hline \Psi \mid \Theta \mid \Delta \mid \Gamma \mid \hat{\Gamma} \vdash s : \hat{\Gamma}' & \Psi \mid \Theta \mid \Delta \mid \Gamma \mid \hat{\Gamma} \vdash s : \hat{\Gamma}', \nu : \tau, \nu' : \tau', \hat{\Gamma}'' \\ \hline \underline{\Psi \mid \Theta \mid \Delta \mid \Gamma \mid \hat{\Gamma} \vdash s : \hat{\Gamma}', \nu' : \tau', \nu : \tau, \hat{\Gamma}'' \\ \hline \underline{for all \ i, \quad \Psi \mid \Theta \mid \Delta \mid \Gamma \mid \hat{\Gamma}_{0} \vdash \{s_{1} \dots s_{n}\} : \hat{\Gamma}_{n} \\ \hline \underline{\Psi \mid \Theta \mid \Delta \mid \Gamma \mid \hat{\Gamma} \vdash \nu := e; : \hat{\Gamma}, \nu : \tau & \Psi \mid \Theta \mid \Delta \mid \Gamma \mid \hat{\Gamma}, \nu : \tau, \hat{\Gamma}' \vdash e : \tau' \\ \hline \Psi \mid \Theta \mid \Delta \mid \Gamma \mid \hat{\Gamma} \vdash \nu := e; : \hat{\Gamma}, \nu : \tau & \Psi \mid \Theta \mid \Delta \mid \Gamma \mid \hat{\Gamma}, \nu : \tau, \hat{\Gamma}' \vdash \nu := e; : \hat{\Gamma}, \nu : \tau', \hat{\Gamma}' \\ \hline \end{array}$$

Figure 6: Type Checking Statements

Figure 7: Type Checking Returns

$$\Psi \mid \Theta \vdash \tau : \Delta \qquad \Psi \vdash \nu \langle \Theta \rangle : \Delta$$

$$\begin{array}{ccc} \Psi \vdash \hat{\nu} \langle \nu_{1}, \dots, \nu_{n} \rangle : \Delta & \Psi \vdash \hat{\nu} \langle \nu_{1}, \dots, \nu_{n} \rangle \text{ extends } \tau & \Psi \mid \nu_{1}, \dots, \nu_{n} \vdash \tau : \Delta' \\ \hline \text{for all } \nu\sigma \text{ in } \Delta, & \text{for all } \nu'\sigma' \text{ in } \Delta', & \nu = \nu' & \text{implies} & \Psi \mid \nu_{1}, \dots, \nu_{n} \vdash \sigma \approx \sigma' \\ \hline \Psi \mid \Theta \vdash \hat{\nu} \langle \tau_{1}, \dots, \tau_{n} \rangle : (\Delta \cup \Delta') [\nu_{1} \mapsto \tau_{1}, \dots, \nu_{n} \mapsto \tau_{n}] \\ \hline \frac{\text{interface } \hat{\nu} \langle \Theta \rangle \text{ extends } \tau \ \{\Delta\} \text{ in } \Psi}{\Psi \vdash \hat{\nu} \langle \Theta \rangle : \Delta} & \frac{\text{class } \hat{\nu} \langle \Theta \rangle \text{ extends } \tau \ \{\Delta\} \text{ in } \Psi}{\Psi \vdash \hat{\nu} \langle \Theta \rangle : \Delta} \end{array}$$

Figure 8: Method-Context Lookup

 $\Psi \mid \Theta \vdash \tau : \nu$

class $\hat{\nu}\langle \nu_1, \dots, \nu_n \rangle$ extends $\tau \{\dots, \nu\sigma, \dots\}$ in Ψ $\Psi \vdash \hat{\nu}\langle \tau_1, \dots, \tau_n \rangle : \nu$

Figure 9: Method-Implemented Check

$$\Psi \mid \Delta \mid \Gamma \vdash i : \Psi \qquad \Psi \mid \Delta \mid \Gamma \vdash c : \Psi \mid \Delta$$

 $\Psi \mid \Theta \vdash_{\top} \tau$ $\Psi' =$ **interface** $\nu \langle \Theta \rangle$ **extends** $\tau \{\nu_1 \sigma_1, \dots, \nu_n \sigma_n\}$ $\Psi, \Psi' \mid \Theta \vdash \sigma_i$ for all i, $\Psi, \Psi' \mid \Theta \vdash \nu \langle \Theta \rangle : \dot{\Delta'}$ for all $\bar{\nu}\langle\bar{\Theta}\rangle(\bar{\Gamma}):\bar{\tau}$ in Δ' , for all $\bar{\nu}'$ in $\bar{\Theta}$, $\bar{\nu}'$ not in Θ $\Psi \mid \Delta \mid \Gamma \vdash \text{interface } \nu\langle\Theta\rangle \text{ extends } \tau \text{ {fun }} \nu_1\sigma_1;\ldots;\text{fun }\nu_n\sigma_n;\text{ }}:\Psi'$ $\Psi \mid \Theta \vdash_{\hat{\tau}} \tau$ $\begin{array}{l} \Delta' = \nu \langle \Theta \rangle (\hat{\Gamma}) : \nu \langle \Theta \rangle \\ \Psi, \Psi' \mid \Theta \vdash \sigma_i \end{array}$ $\Psi' =$ class $\nu \langle \Theta \rangle$ extends $\tau \{ \nu_1 \sigma_1; \ldots; \nu_n \sigma_n; \}$ for all i, for all i, $\Psi, \Psi' \mid \Theta \mid \Delta, \Delta' \mid \Gamma \mid \hat{\Gamma}_{i-1} \vdash s_i : \hat{\Gamma}_i$ $\Psi, \Psi' \mid \Theta \vdash \hat{\Gamma} \qquad \hat{\Gamma}_0 = \hat{\Gamma}$ or $\Psi, \Psi' \mid \Theta \mid \Delta, \Delta' \mid \Gamma, \hat{\Gamma}_m \vdash \hat{\tau}(e_1, \dots, e_k) : \hat{\tau}$ $\hat{\tau}(e_1,\ldots,e_k) = \top()$ $\Psi, \Psi' \mid \Theta \vdash \nu \langle \Theta \rangle : \Delta''$ for all $\bar{\nu} \langle \bar{\Theta} \rangle (\bar{\Gamma}) : \bar{\tau} \text{ in } \Delta'',$ for all $\bar{\nu}'$ in $\bar{\Theta}$, $\bar{\nu}'$ not in Θ for all i, $\sigma_i = \langle \Theta_i \rangle(\Gamma_i) : \tau_i \quad \Psi, \Psi' \mid \Theta, \Theta_i \mid \Delta, \Delta', \Delta'' \mid \Gamma, \hat{\Gamma}_m \mid \Gamma_i \vdash_{\tau_i}^{\mathbf{true}} \hat{s}_i : \Gamma'_i$ for all $\hat{\nu}\sigma$ in Δ'' , $\Psi, \Psi' \mid \Theta \vdash \nu \langle \Theta \rangle : \hat{\nu}$

 $\Psi \mid \Delta \mid \Gamma \vdash \mathbf{class} \; \nu \langle \Theta \rangle(\hat{\Gamma}) \; \mathbf{extends} \; \tau \; \{s_1 \dots s_m \; \mathbf{super}(e_1, \dots, e_k); \; \mathbf{fun} \; \nu_1 \sigma_1 \; \hat{s}_1 \; \dots \; \mathbf{fun} \; \nu_n \sigma_n \; \hat{s}_n\} : \Psi' \mid \Delta'$

Figure 10: Class and Interface Checking

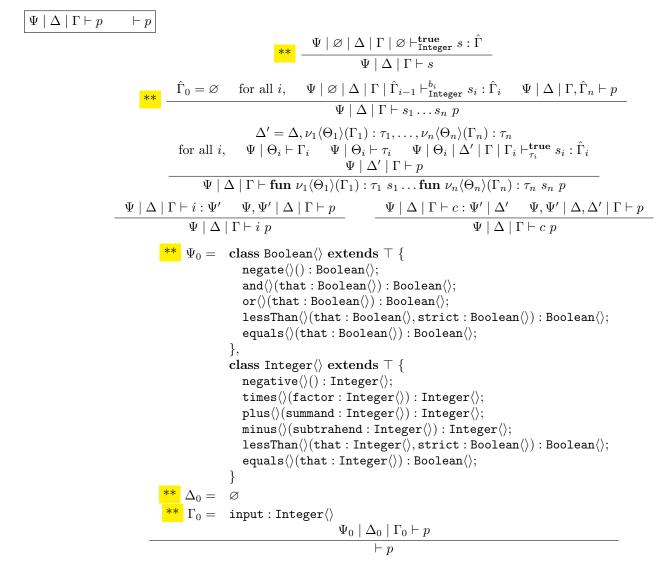


Figure 11: Program Checking