Soundness of Types

Ensuring that a type system is not broken

Example: Tootool 0.1 Language



Tootool is a rural community in the central east part of the Riverina [New South Wales, Australia]. It is situated by road, about 4 kilometres east from French Park and 16 kilometres west from The Rock. Tootool Post Office opened on 1 August 1901 and closed in 1966. [Wikipedia]

unsound Type System for Tootool 0.1

Pos <: Int Neg <: Int

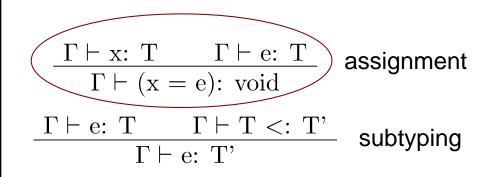
$\Gamma \vdash x$: T	$\Gamma \vdash e: T$	assignment
$\Gamma \vdash (\mathbf{x} = \mathbf{e})$: void		
$\Gamma \vdash e: T$	$\Gamma \vdash T <: T$, subtyping
$\Gamma \vdash e: T'$		Subtyping

does it type check? def intSqrt(x:Pos) : Pos = { ...} var p : Pos var q : Neg var r : Pos q = -5p = q $\Gamma = \{(p, Pos), (q, Neg), (r, Pos), (n, Pos), (n, Pos), (r, Po$

Runtime error: intSqrt invoked with a negative argument!

What went wrong in Tootool 0.1 ?

Pos <: Int Neg <: Int



does it type check? – yes def intSqrt(x:Pos) : Pos = { ...} var p : Pos var q : Neg var r : Pos q = -5p = q $\Gamma = \{(p, Pos), (q, Neg), (r, Pos), (n, Pos),$

Runtime error: intSqrt invoked with a negative argument!

x must be able to store any e can have any value from T value from T $\underline{? \quad \Gamma \vdash e: T}$ $\Gamma \vdash (x = e): \text{ void}$

Cannot use $\Gamma \vdash e: T$ to mean "x promises it can store any $e \in T$ "

Recall Our Type Derivation

Pos <: Int Neg <: Int

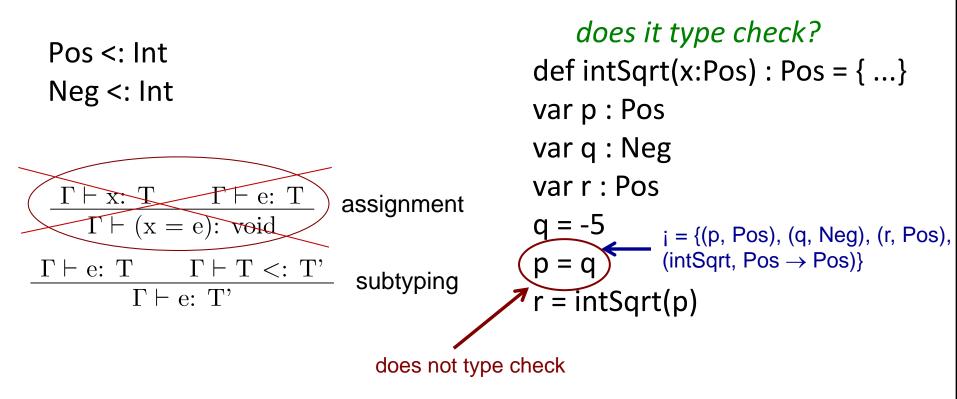
 $\begin{array}{c|c} \hline \Gamma \vdash x: \ T & \Gamma \vdash e: \ T \\ \hline \Gamma \vdash (x = e): \ void \end{array} \quad \text{assignment} \\ \hline \hline \Gamma \vdash e: \ T & \Gamma \vdash T <: \ T' \\ \hline \Gamma \vdash e: \ T' \end{array} \quad \text{subtyping} \end{array}$

does it type check? – yes def intSqrt(x:Pos) : Pos = { ...} var p : Pos var q : Neg var r : Pos q = -5 p = q $i = \{(p, Pos), (q, Neg), (r, Pos), (intSqrt, Pos <math>\rightarrow Pos)\}$ r = intSqrt(p)

Runtime error: intSqrt invoked with a negative argument!

Values from p are integers. But p did not promise to store all kinds of integers/ Only positive ones!

Corrected Type Rule for Assignment

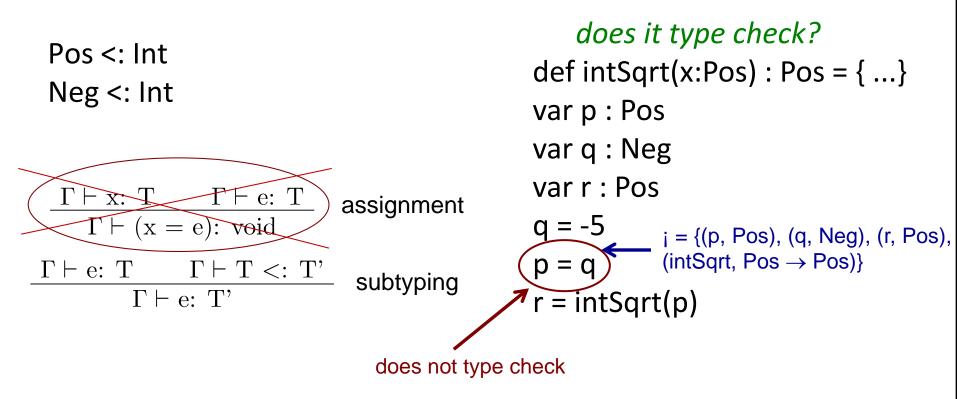


x must be able to store any value from T

$$(x,T) \in \Gamma \qquad \Gamma \vdash e: T$$
$$\Gamma \vdash (x = e): \text{ void}$$

 Γ stores declarations (promises)

Corrected Type Rule for Assignment



Is there another way to fix the type system ?

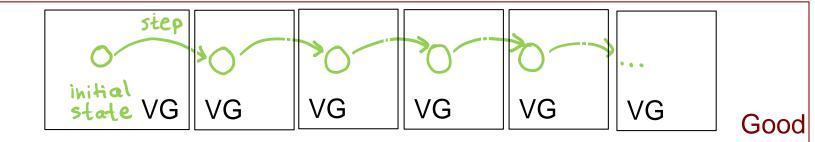
How could we ensure that some other programs will not break?

Type System Soundness

Proving Soundness of Type Systems

- Goal of a sound type system:
 - if a program type checks, it never "crashes"
 - crash = some precisely specified bad behavior
 - e.g. invoking an operation with a wrong type
 - dividing a string by another string: "cat" / "frog"
 - trying to *multiply* a Window object by a File object
 - e.g. dividing an integer by zero
- Never crashes: no matter how long it executes
 - proof is done by induction on program execution

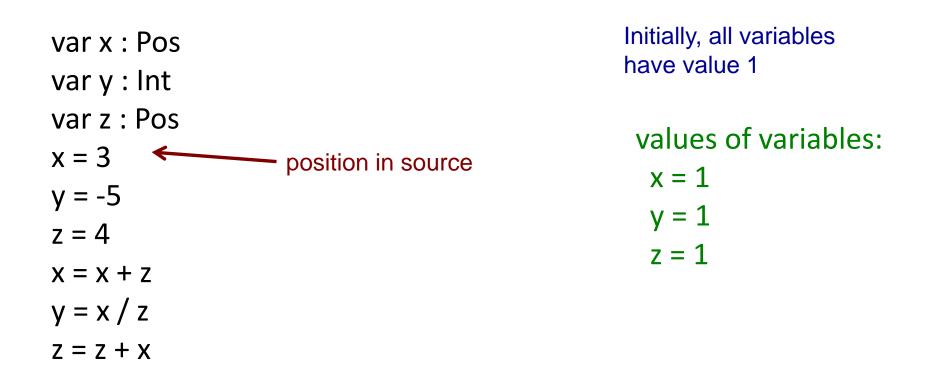
Proving Soundness by Induction

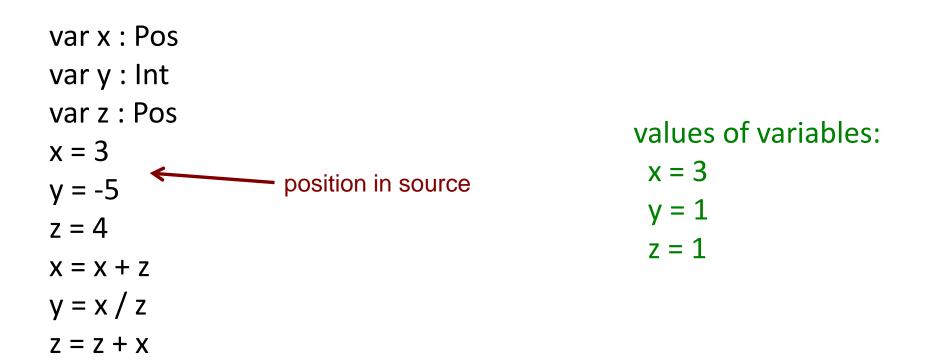


- Program moves from state to state
- Bad state = state where program is about to exhibit a bad operation ("cat" / "frog")
- Good state = state that is not bad
- To prove:

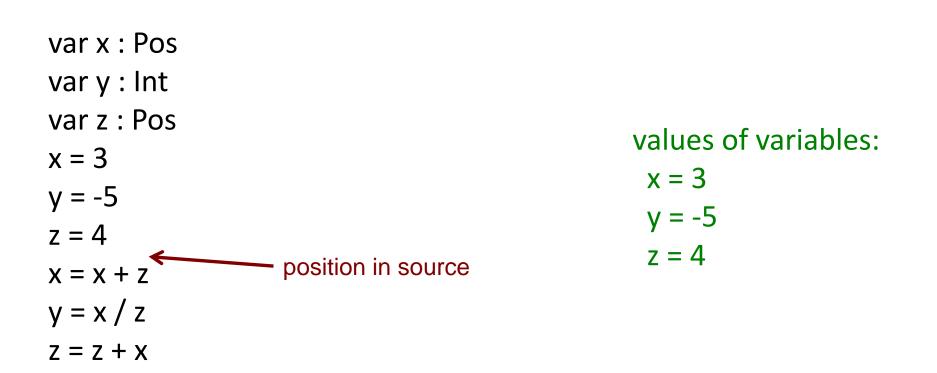
program type checks \rightarrow states in all executions are good

 Usually need a stronger inductive hypothesis; some notion of very good (VG) state such that: program type checks → program's initial state is very good state is very good → next state is also very good state is very good → state is good (not about to crash) A Simple Programming Language

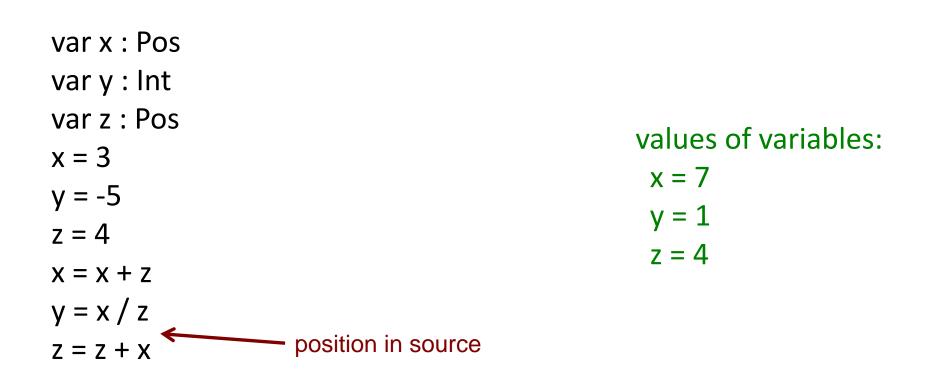








var x : Pos var y : Int var z : Pos x = 3 y = -5 z = 4 x = x + z y = x / z z = z + x values of variables: x = 7 y = -5 z = 4 x = x + z y = x / z



formal description of such program execution is called operational semantics

Operational semantics

Operational semantics gives meaning to programs by describing how the program state changes as a sequence of steps.

 <u>Small-step (or Structural) Operational Semantics (SOS)</u>: consider individual steps (e.g. z = x + y)

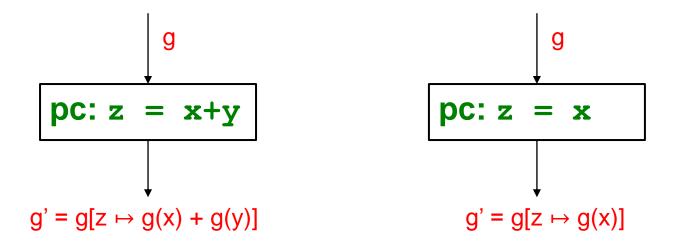
V: set of variables in the program

- pc: integer variable denoting the program counter
- g: $V \rightarrow Int$ function giving the values of program variables (g, pc) program state

Then, for each possible statement in the program we define how it changes the program state.

• Big-step semantics: consider the effect of entire blocks

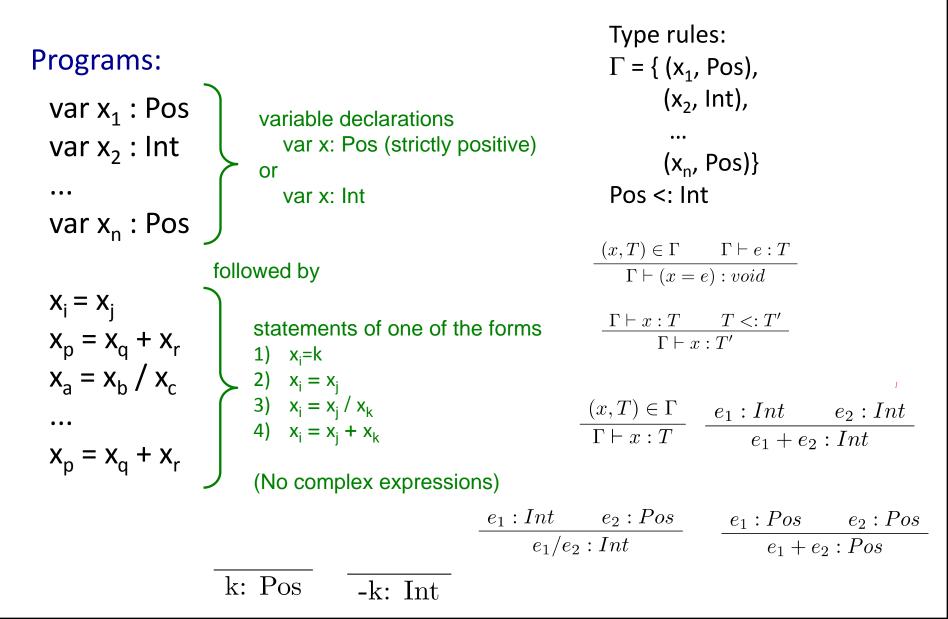
Operational semantics



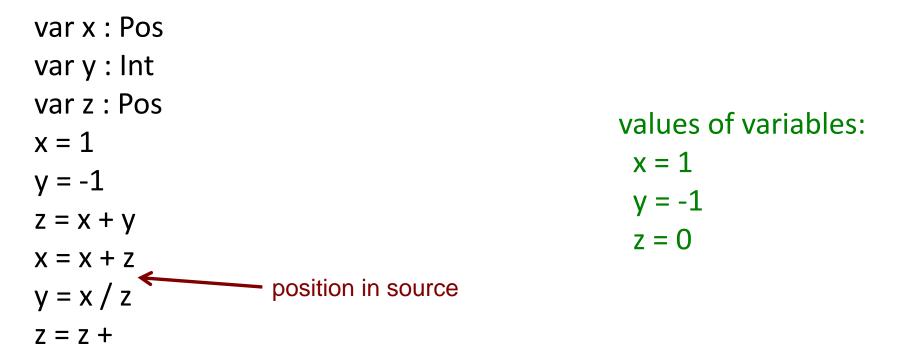
Operation semantics

- If pc: z = x + y, $(g, pc) \rightarrow (g', pc + 1)$, where $g' = g[z \mapsto g(x)+g(y)]$
- If pc: z = x, $(g, pc) \rightarrow (g', pc + 1)$, where $g' = g[z \mapsto g(x)]$

Type Rules of Simple Language

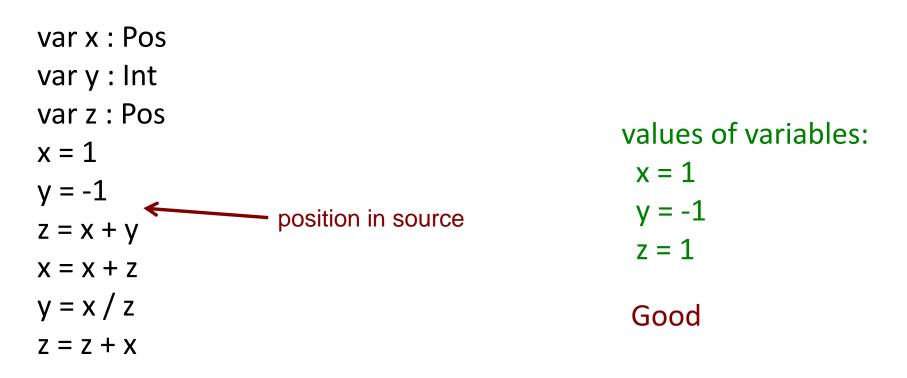


Bad State: About to Divide by Zero (Crash)



Definition: state is *bad* if the next instruction is of the form $x_i = x_j / x_k$ and x_k has value 0 in the current state.

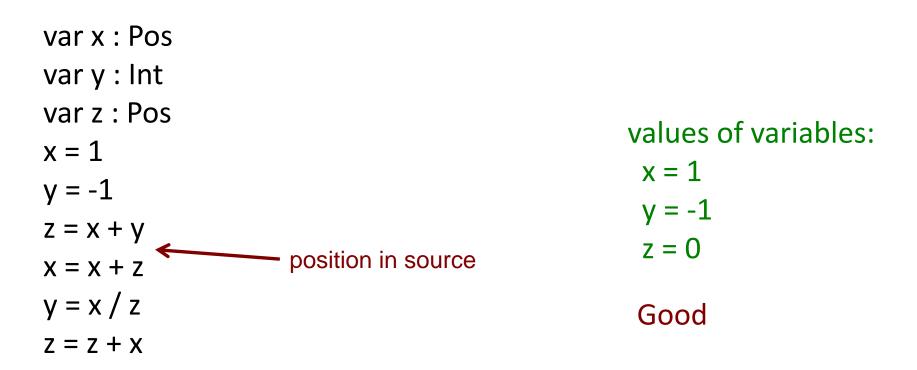
Good State: Not (Yet) About to Divide by Zero



Definition: state is *good* if it is not *bad*.

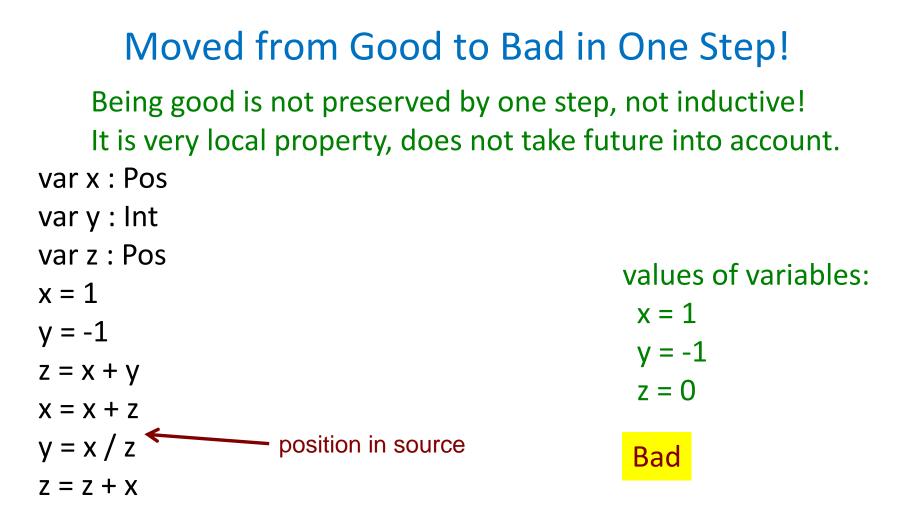
Definition: state is *bad* if the next instruction is of the form $x_i = x_i / x_k$ and x_k has value 0 in the current state.

Good State: Not (Yet) About to Divide by Zero



Definition: state is *good* if it is not *bad*.

Definition: state is *bad* if the next instruction is of the form $x_i = x_i / x_k$ and x_k has value 0 in the current state.

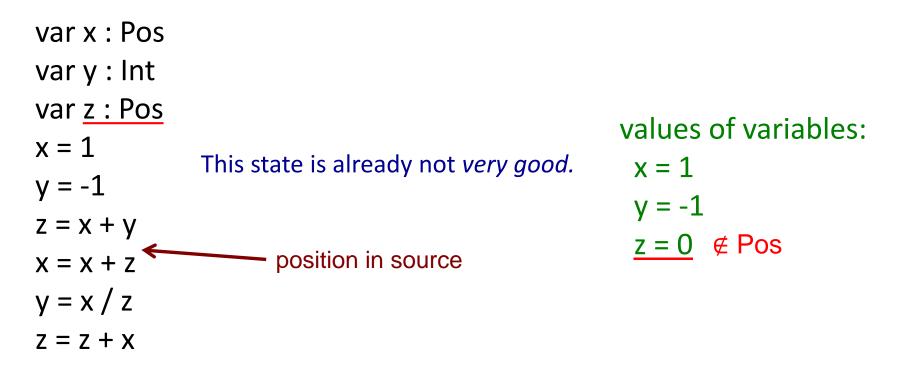


Definition: state is *good* if it is not *bad*.

Definition: state is *bad* if the next instruction is of the form $x_i = x_i / x_k$ and x_k has value 0 in the current state.

Being Very Good: A Stronger Inductive Property

Pos = { 1, 2, 3, ... }



Definition: state is *good* if it is not about to divide by zero. Definition: state is *very good* if each variable belongs to the domain determined by its type (if z:Pos, then z is strictly positive).

Proving Soundness - Intuition

We want to show if a program *type checks:*

- It will be *very good* at the start
- if it is very good in the current step, it will remain very good in the next step
- If it is *very good*, it will not *crash*

Hence, please type check your program, and it will never crash!

Soundnes proof = defining "very good" and checking the properties above.

Proving Soundness in Our Case

Holds: in initial state, variables are =1

• If a program *type checks* :

- $1 \in Pos$ $1 \in Int$
- $\sqrt{-}$ It will be *very good* from at start.
 - if it is very good in the current step, it will remain very good in the next
- \checkmark If it is *very good*, it will not *crash*.

If next state is x / z, type rule ensures z has type Pos Because state is very good, it means $z \in Pos$ so z is not 0, and there will be no crash.

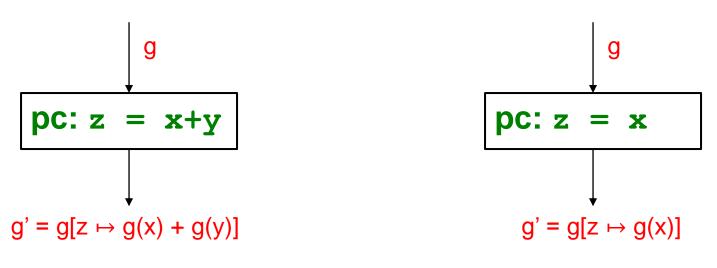
Definition: state is *very good* if each variable belongs to the domain determined by its type (if z:Pos, then z is strictly positive).

Proving that "very goodness" is preserved by state transition

- How do we prove
 - if you are very good, then you will remain very good in the next step
 - Irrespective of the actual program
- We could use SOS small step operational semantics here.

Proving that "very goodness" is preserved by state transition

Hypothesize that g is very good

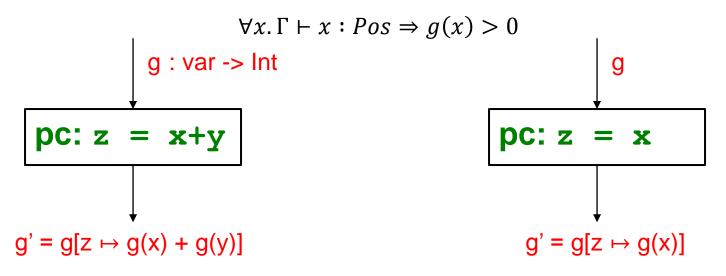


Prove that g' is very good When the program type checks

Do this for every possible "step" of the operational semantics

Proving this for our little type system

Hypothesize that the following holds in g For all vars x, x:Pos => x is strictly positive



Prove that the following holds in g' For all vars x, x:Pos => x is strictly positive

 $\forall x. \Gamma \vdash x : Pos \Rightarrow g'(x) > 0$

- Can we prove this ?
 - Only if we are given that the program type checks

Recall the Type Rules

Pos <: Int

$$\begin{array}{ccc} (x,T) \in \Gamma & \Gamma \vdash e:T \\ \hline \Gamma \vdash (x=e): void \\ \\ \hline \hline \Gamma \vdash x:T & T <:T' \\ \hline \Gamma \vdash x:T' \end{array}$$

$(x,T)\in \Gamma$	$e_1:Int$ $e_2:Int$
$\overline{\Gamma \vdash x:T}$	$\boxed{e_1 + e_2 : Int}$

k: Pos	-k: Int	$e_1: Int \qquad e_2: Pos$	$e_1:Pos$ $e_2:Pos$
		$e_1/e_2:Int$	$e_1 + e_2 : Pos$

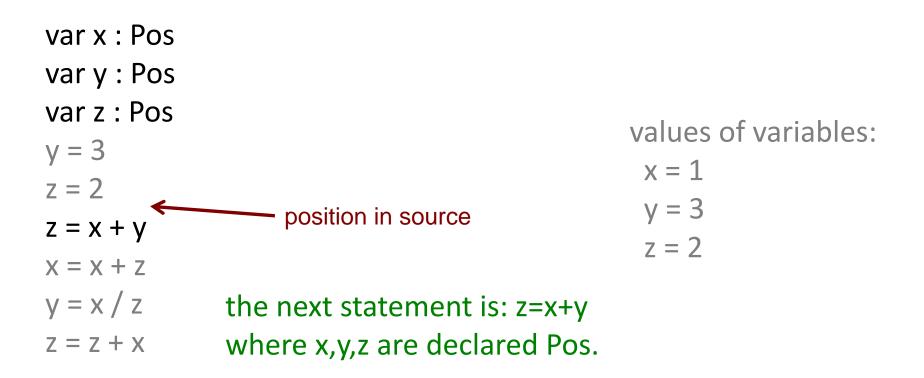
Back to the start

k: Pos -k: Int
$\begin{tabular}{ccc} \hline \Gamma \vdash x:T & \Gamma \vdash e:T \\ \hline \Gamma \vdash (x=e):void \end{tabular}$
$\frac{\Gamma \vdash x:T \qquad T <: T'}{\Gamma \vdash x:T'}$
$\frac{(x,T)\in\Gamma}{\Gamma\vdash x:T}$
$\begin{array}{cc} e_1:Int & e_2:Int \\ \hline e_1+e_2:Int \end{array}$
$\begin{array}{cc} e_1:Int & e_2:Pos\\ \hline e_1/e_2:Int \end{array}$
$\begin{array}{cc} e_1:Pos & e_2:Pos \\ \hline e_1+e_2:Pos \end{array}$

Does the proof still work?

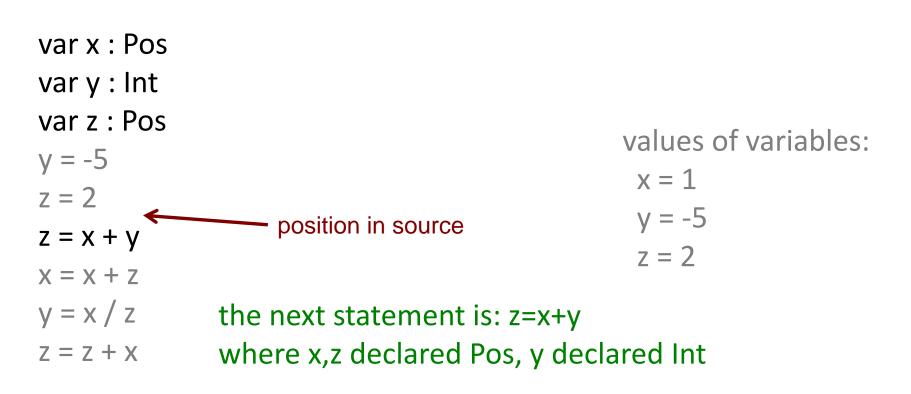
If not, where does it break?

Let's type check some programs Example 1



Goal: provide a type derivation for the program

Example 2



Goal: prove that the program type checks impossible, because z=x+y would not type check How do we know it could not type check?

Must Carefully Check Our Type Rules

Type rules:

		турсти	163.	
		$\Gamma = \{ (\mathbf{x}_1) \}$, Pos),	
		-	, Int),	
var x : Pos		(**2	<u>,</u> ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	
var y : Int	Conclude that the only	····		
var z : Pos	types we can derive are:	• 1	, Pos)}	
	x : Pos, x : Int	Pos <: ii	nt	
y = -5	,			
z = 2	y : Int		$\frac{\Gamma \vdash e:T}{}$	
	x + y : Int	$\Gamma \vdash (x =$	= e): void	
z = x + y	, , , , , , , , , , , , , , , , , , ,	$\Gamma \vdash x \cdot T$	T < T'	
x = x + z	Cannot type check	$\frac{1+x\cdot 1}{\Gamma \vdash}$	$\frac{T <: T'}{x : T'}$	
y = x / z	Cannot type check			
8 -	z = x + y in this environmer	It. $(T T) \subset \Gamma$	T. I	T. I
z = z + x		$\frac{(x, I) \in I}{\Gamma + m + T}$	$\frac{e_1:Int}{e_1+e_2}$	$e_2:Int$
		$1 \vdash x : I$	$e_1 + e_2$	$_2:Int$
	$e_1: In$	$t e_2: Pos$	$e_1: Pos$	$e_2: Pos$
	$\overline{e_1}$	$1/e_2:Int$	$e_1 +$	

k: Pos -k: Int

We would need to check all cases (there are many, but they are easy)

Remark

• We used in examples Pos <: Int

• Same examples work if we have

```
class Int { ... }
class Pos extends Int { ... }
```

and is therefore relevant for OO languages

What if we want more complex types?

```
class A { }
                      • Should it type check?
class B extends A
                     <sup>{</sup> • Does this type check in Java?
  void foo() { }
                         • can you run it?
}
                      • Does this type check in Scala?
class Test {
  public static void main(String[]
args) {
    B[] b = new B[5];
    A[] a;
    a = b;
    System.out.println("Hello,");
    a[0] = new A();
    System.out.println("world!");
    b[0].foo();
```

What if we want more complex types?

Suppose we add to our language a reference type: class Ref[T](var content : T)

Programs:

```
var x<sub>1</sub> : Pos
var x<sub>2</sub> : Int
var x<sub>3</sub> : Ref[Int]
var x<sub>4</sub> : Ref[Pos]
```

x = y

- x = y + z
- x = y / z
- x = y + z.content

x.content = y

Exercise 1:

Extend the type rules to use with Ref[T] types. Show your new type system is sound.

Exercise 2:

Can we use the subtyping rule? If not, where does the proof break?

 $\frac{T <: T'}{Ref[T] <: Ref[T']}$

Extending the type system

: Pos

Pos <: Int

$$(x,T) \in \Gamma \qquad \Gamma \vdash e:T$$

$$\Gamma \vdash (x=e): void$$

$$\frac{\Gamma \vdash x:T \qquad T <:T'}{\Gamma \vdash x:T'}$$

$$T) \in \Gamma \qquad e_1: Int \qquad e_1$$

 $\begin{array}{c} (x,T) \in \Gamma \\ \hline \Gamma \vdash x:T \end{array} \qquad \begin{array}{c} e_1:Int \\ e_1 + e_2:Int \end{array}$

k: Pos -k: Int

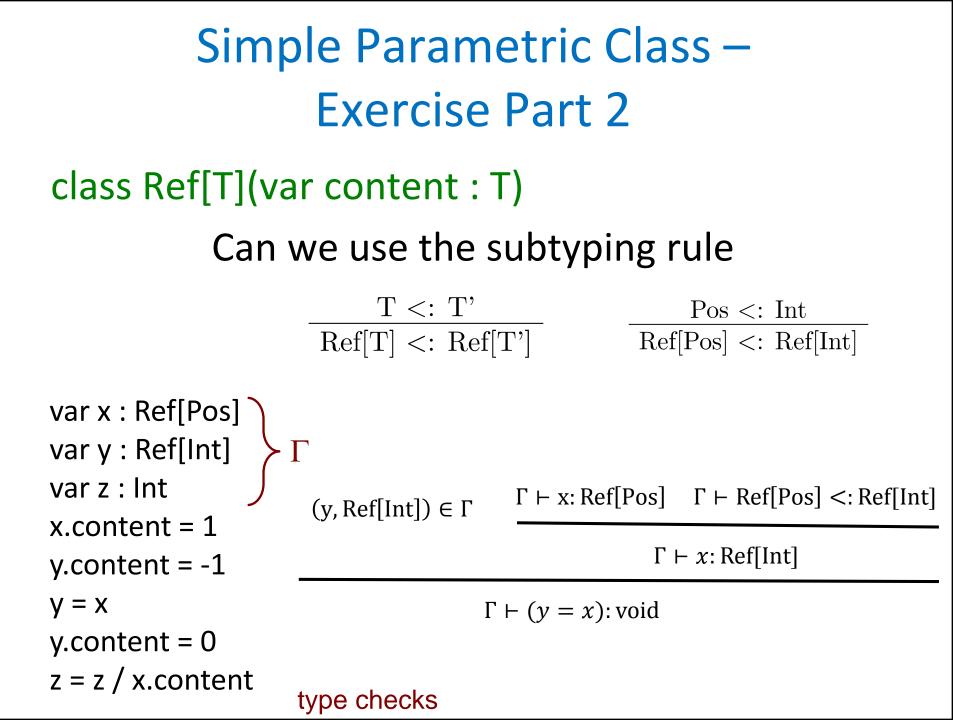
$$\begin{array}{c|ccc} e_1:Int & e_2:Pos \\ \hline e_1/e_2:Int & \hline e_1+e_2:Pos \\ \end{array}$$

 $\Gamma \vdash e : \operatorname{Ref}[T]$

 $\Gamma \vdash e.content : T$

$$(v, \operatorname{Ref}[T]) \in \Gamma \quad \Gamma \vdash e : T$$

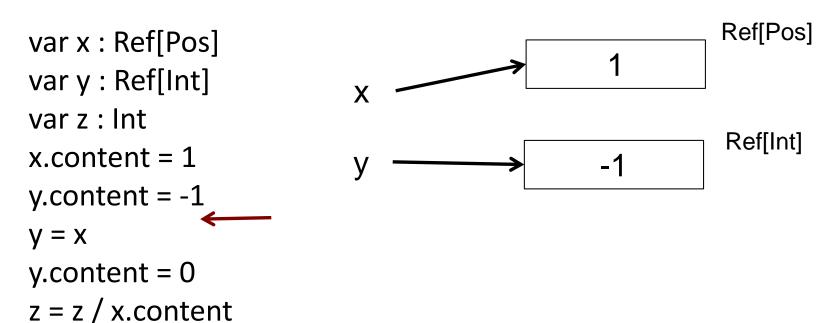
$$\Gamma \vdash (v. content = e): T$$



Simple Parametric Class

class Ref[T](var content : T) Can we use the subtyping rule

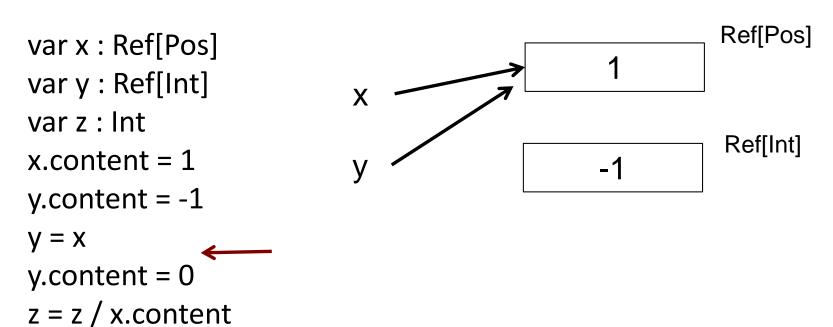
 $\frac{T <: T'}{Ref[T] <: Ref[T']}$



Simple Parametric Class

class Ref[T](var content : T) Can we use the subtyping rule

 $\frac{T <: T'}{Ref[T] <: Ref[T']}$



Simple Parametric Class class Ref[T](var content : T) Can we use the subtyping rule Ref **T** $\operatorname{Ref}[T]$ Ref[Pos] var x : Ref[Pos] ()var y : Ref[Int] Χ var z : Int Ref[Int] x.content = 1y.content = -1y = xy.content = 0**CRASHES** z = z / x.content

Analogously

class Ref[T](var content : T) Can we use the converse subtyping rule T <: T' $\operatorname{Ref}[T'] <: \operatorname{Ref}[T]$ Ref[Pos] var x : Ref[Pos] 1 var y : Ref[Int] Х var z : Int Ref[Int] x.content = 1()y.content = -1 $\mathbf{x} = \mathbf{y}$ y.content = 0**CRASHES** z = z / x.content

Mutable Classes do not Preserve Subtyping class Ref[T](var content : T) Even if T <: T', Ref[T] and Ref[T'] are unrelated types

var x : Ref[T] var y : Ref[T']

 $x = y \leftarrow type$ checks only if T=T'

...

Same Holds for Arrays, Vectors, all mutable containers

Even if T <: T', Array[T] and Array[T'] are unrelated types

```
var x : Array[Pos](1)
var y : Array[Int](1)
var z : Int
x[0] = 1
y[0] = -1
y = x
y[0] = 0
z = z / x[0]
```

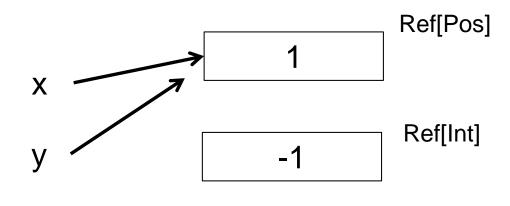
Case in Soundness Proof Attempt

class Ref[T](var content : T) Can we use the subtyping rule

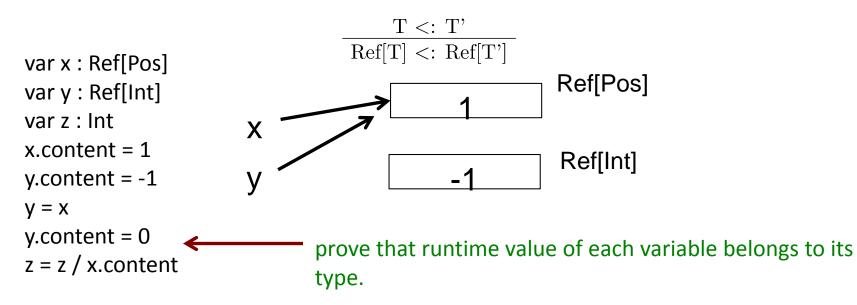
 $\frac{T <: T'}{Ref[T] <: Ref[T']}$

var x : Ref[Pos]
var y : Ref[Int]
var z : Int
x.content = 1
y.content = -1
y = x
y.content = 0
z = z / x.content

prove that runtime value of each variable belongs to its type.



Soundness Proof Attempt [Cont.]



- Need to have an operational semantics for the language
- State g : (Var U Addr) -> (Int U Addr)
- A very good property that we need :
 - $\forall x. \Gamma \vdash x : \operatorname{Ref}[\operatorname{Pos}] \Rightarrow g(g(x)) > 0$
 - Cannot prove this property is preserved because "y.content = 0" may change the value of "x.context", and hence break x'es type if it is Ref[Pos].
 - Proof will not work for any stronger properties also because we have a counter-example

Mutable vs Immutable Containers

• Immutable container, Coll[T]

- has methods of form e.g. get(x:A) : T
- if T <: T', then Coll[T'] has get(x:A) : T'</pre>
- we have (A → T) <: (A → T') covariant rule for functions, so Coll[T] <: Coll[T']</p>
- Write-only data structure have
 - setter-like methods, set(v:T) : B
 - if T <: T', then Container[T'] has set(v:T') : B</pre>
 - would need (T' → B) <: (T → B) contravariance for arguments, so Coll[T'] <: Coll[T]</p>
- Read-Write data structure need both. That is coll[T] is *invariant* in T

A cool exercise – Physical Units as Types

- Define a "unit type" by the following grammar
- $u \rightarrow b \mid u^{-1} \mid u \ast u$
- $b \rightarrow kg \mid m \mid s \mid A \mid K \mid mole \mid cd$
- We use the syntactic sugar
 - u^n to denote u multiplied with u n-times
 - $\frac{u_1}{u_2}$ to denote $u_1 * u_2^{-1}$
- Give the type rules for the arithmetic operations +,*, /, *sqrt*, sin, *abs*.
- Trigonometric functions take argument without units
- An expression has no units if $\Gamma \vdash e: 1$

Physical Units as Types Part 2

- The unit expressions are strings, so
- $\frac{S^2m^2}{m^2s}$ and s will not be considered as same types though they have same units
- How can we modify the type rules so that they type check expressions, whenever their units match as per physics?

Physical Units as Types Part 3

Determine the type of T in the following code fragment.

- val x: <m> = 800
- val y: <m> = 6378
- val g: <m/(s*s)> = 9.8
- val R = x + y
- val w = sqrt(g/R)
- val T = (2 * Pi) / w

Physical Units as Types Part 4

Suppose you want to use the unit *feet* in addition to the SI units. How can you extend your type system to accommodate for this? (Assume that 1m = 3.28084 feet.)