Type Systems and Their Soundness

Viktor Kunčak

Why types are good

Prevent errors: many simple errors are caught by types

Ensure memory safety or other desired properties

Document the program (purpose of parameters)

Make it easier to change program

Make compilation more efficient: remove checks, specialize operations

An unsound (broken) type system

A type system that aims to ensure some property but, in fact, fails.

For example: suppose we have a system that aims to ensure that if parameter is of type Int, then it is only invoked with values of type Int. But we find a (tricky) program that passes the type checker and ends up invoking the function with the reference to a string. This is unsoundness.
Sometimes unsoundness is an *intentional* compromise:

- type casts in C
- covariance for function arguments and arrays

Often *unintentional* (unsoundness bugs in type systems), due to subtle interactions between e.g. subtyping, generics, mutation, higher-order functions, recursion

Java and Scala's Type Systems are Unsound *

The Existential Crisis of Null Pointers

Nada Amin
EPFL, Switzerland
nada.amin@epfl.ch

Ross Tate
Cornell University, USA
ross@cs.cornell.edu

Abstract

We present short programs that demonstrate the unsoundness of Java and Scala's current type systems. In particular, these programs provide parametrically polymorphic functions that can turn any type into any type without (down)casting. Fortunately, parametric polymorphism was not integrated into the Java Virtual Machine (JVM), so these examples do not demonstrate any unsoundness of the JVM. Nonetheless, we discuss broader implications of these findings on the field of programming languages.

ture, we often develop a minimal calculus employing that feature and then verify key properties of that calculus. But these results provide no guarantees about how the feature in question will interact with the many other common features one might expect for a full language. The unsoundness we identify results from such an interaction of features. Thus, in addition to valuing the development and verification of minimal calculi, our community should explore more ways to improve our chances of identifying abnormal interactions of features within reasonable time but without unreasonable resources and distractions. Ideally our community could pro-

1. Introduction

In 2004, Java 5 introduced generics, i.e. parametric polymorphism, to the Java programming language. In that same year, Scala was publicly released, introducing path-dependent types as a primary language feature. Upon their release 12 years ago, both languages were unsound; the examples we will present were valid even in 2004. But despite the fact that Java has been formalized repeatedly [3, 4, 6, 9, 10, 18, 26, 38], this unsoundness has not been discovered until now. It was found in Scala in 2008 [40], but the bug was deferred and its broader significance was not realized until now.

–same paper, published in November 2016

Goal of today's lecture

Explain that "expression has a type" is an *inductively defined relation*
Define precisely a small language:

- its abstract syntax (as certain math expressions)
- its operational semantics (interpreter written in math)
- its type rules

Show that our type system prevents certain kinds of errors

Background: inductively defined relations and sets

Define relation $r \subseteq \mathbb{Z} \times \mathbb{Z}$ using these rules (x, y range over $\mathbb{Z}$):

$$\frac{}{(0,0) \in r} \quad \text{(zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \quad \text{(increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \quad \text{(increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \quad \text{(decrease both)}$$

For which of the following relations r are all the above rules true?

Background: inductively defined relations and sets

Define relation $r \subseteq \mathbb{Z} \times \mathbb{Z}$ using these rules (x, y range over $\mathbb{Z}$):

$$\frac{}{(0,0) \in r} \quad \text{(zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \quad \text{(increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \quad \text{(increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \quad \text{(decrease both)}$$

For which of the following relations r are all the above rules true?

- $r = \{(x,y) \mid x = 0 \vee y = 0\}$?

Background: inductively defined relations and sets

Define relation $r \subseteq \mathbb{Z} \times \mathbb{Z}$ using these rules (x, y range over $\mathbb{Z}$):

$$\frac{}{(0,0) \in r} \quad \text{(zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \quad \text{(increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \quad \text{(increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \quad \text{(decrease both)}$$

For which of the following relations r are all the above rules true?

- $r = \{(x,y) \mid x = 0 \vee y = 0\}$? No (increase right)

Background: inductively defined relations and sets

Define relation $r \subseteq \mathbb{Z} \times \mathbb{Z}$ using these rules (x, y range over $\mathbb{Z}$):

$$\frac{}{(0,0) \in r} \quad \text{(zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \quad \text{(increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \quad \text{(increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \quad \text{(decrease both)}$$

For which of the following relations r are all the above rules true?

- $r = \{(x,y) \mid x = 0 \vee y = 0\}$? No (increase right)
- $r = \{(x,y) \mid x \leq 0 \wedge 0 \leq y\}$?

Background: inductively defined relations and sets

Define relation $r \subseteq \mathbb{Z} \times \mathbb{Z}$ using these rules (x, y range over $\mathbb{Z}$):

$$\frac{}{(0,0) \in r} \quad \text{(zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \quad \text{(increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \quad \text{(increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \quad \text{(decrease both)}$$

For which of the following relations r are all the above rules true?

- $r = \{(x,y) \mid x = 0 \vee y = 0\}$? No (increase right)
- $r = \{(x,y) \mid x \leq 0 \wedge 0 \leq y\}$? No

Background: inductively defined relations and sets

Define relation $r \subseteq \mathbb{Z} \times \mathbb{Z}$ using these rules (x, y range over $\mathbb{Z}$):

$$\frac{}{(0,0) \in r} \quad \text{(zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \quad \text{(increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \quad \text{(increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \quad \text{(decrease both)}$$

For which of the following relations r are all the above rules true?

- $r = \{(x,y) \mid x = 0 \vee y = 0\}$? No (increase right)
- $r = \{(x,y) \mid x \leq 0 \wedge 0 \leq y\}$? No
- $r = \mathbb{Z} \times \mathbb{Z}$?

Background: inductively defined relations and sets

Define relation $r \subseteq \mathbb{Z} \times \mathbb{Z}$ using these rules (x, y range over $\mathbb{Z}$):

$$\frac{}{(0,0) \in r} \quad \text{(zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \quad \text{(increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \quad \text{(increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \quad \text{(decrease both)}$$

For which of the following relations r are all the above rules true?

- $r = \{(x,y) \mid x = 0 \vee y = 0\}$? No (increase right)
- $r = \{(x,y) \mid x \leq 0 \wedge 0 \leq y\}$? No
- $r = \mathbb{Z} \times \mathbb{Z}$? Yes

Background: inductively defined relations and sets

Define relation $r \subseteq \mathbb{Z} \times \mathbb{Z}$ using these rules (x, y range over $\mathbb{Z}$):

$$\frac{}{(0,0) \in r} \quad \text{(zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \quad \text{(increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \quad \text{(increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \quad \text{(decrease both)}$$

For which of the following relations r are all the above rules true?

- $r = \{(x,y) \mid x = 0 \vee y = 0\}$? No (increase right)
- $r = \{(x,y) \mid x \leq 0 \wedge 0 \leq y\}$? No
- $r = \mathbb{Z} \times \mathbb{Z}$? Yes

What is the **smallest** r (wrt. $\subseteq$) for which rules hold? $\emptyset$?

Background: inductively defined relations and sets

Define relation $r \subseteq \mathbb{Z} \times \mathbb{Z}$ using these rules (x, y range over $\mathbb{Z}$):

$$\frac{}{(0,0) \in r} \quad \text{(zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \quad \text{(increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \quad \text{(increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \quad \text{(decrease both)}$$

For which of the following relations r are all the above rules true?

- $r = \{(x,y) \mid x = 0 \vee y = 0\}$? No (increase right)
- $r = \{(x,y) \mid x \leq 0 \wedge 0 \leq y\}$? No
- $r = \mathbb{Z} \times \mathbb{Z}$? Yes

What is the **smallest** r (wrt. $\subseteq$) for which rules hold? $\emptyset$? No.

Background: inductively defined relations and sets

Define relation $r \subseteq \mathbb{Z} \times \mathbb{Z}$ using these rules (x, y range over $\mathbb{Z}$):

$$\frac{}{(0,0) \in r} \quad \text{(zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \quad \text{(increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \quad \text{(increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \quad \text{(decrease both)}$$

For which of the following relations r are all the above rules true?

- $r = \{(x,y) \mid x = 0 \vee y = 0\}$? No (increase right)
- $r = \{(x,y) \mid x \leq 0 \wedge 0 \leq y\}$? No
- $r = \mathbb{Z} \times \mathbb{Z}$? Yes

What is the **smallest** r (wrt. $\subseteq$) for which rules hold? $\emptyset$? No. $r = \{(x,y) \mid x \leq y\}$

Example derivation of $(-3,-1) \in r$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{(0,0) \in r}{(0,1) \in r}}{(0,2) \in r}}{(-1,1) \in r}}{(-2,0) \in r}}{(-3,-1) \in r}$$

$$\frac{}{(0,0) \in r} \text{ (zero)}$$

$$\frac{(x,y) \in r}{(x,y+1) \in r} \text{ (increase right)}$$

$$\frac{(x,y) \in r}{(x+1,y+1) \in r} \text{ (increase both)}$$

$$\frac{(x,y) \in r}{(x-1,y-1) \in r} \text{ (decrease both)}$$

Proof that our rules define $\{(x, y) \mid x \leq y\}$

Establish two directions:

- if there exists a derivation, then $x \leq y$
  Strategy: induction on derivation, go through each rule

- if $x \leq y$ then there exists a derivation
  Strategy (problem-specific): we can find an algorithm that given x, y finds derivation tree (what is the algorithm?)

Proof that our rules define $\{(x,y) \mid x \leq y\}$

Establish two directions:

- if there exists a derivation, then $x \leq y$
  Strategy: induction on derivation, go through each rule

- if $x \leq y$ then there exists a derivation
  Strategy (problem-specific): we can find an algorithm that given x, y finds derivation tree (what is the algorithm?)

  Example algorithm: start from $(0,0)$, then
  derive $(0, y-x)$ in $y-x$ steps of "increase right",
  then depending on whether $x<0$ or $x>0$ apply "increase both" or "decrease both" rule $|x|$ times.

Context-Free Grammars as Inductively Defined Relations

Inductive definitions work on multiple relations as well
Context-free grammars: mutually defined sets of strings (sets are relations)
Each non-terminal corresponds to a set of strings. Let $A = \{a, b\}$

context-free grammar rule	inductive rule ($S, N \subseteq A^*$)
$S ::= aN$	$\dfrac{w \in N}{aw \in S}$
$N ::= \varepsilon$	$\dfrac{}{\varepsilon \in N}$
$N ::= aNNb$	$\dfrac{w_1 \in N, w_2 \in N}{aw_1 w_2 b \in N}$

Sets of first symbols for each non-terminal is also an inductively definable relation

Inductively defined relations

We can use inductive rules to define type systems, grammars, interpreters, ...
We define a relation r using **rules** of the form

$$\frac{t_1(\bar{x}) \in r, \ldots, t_n(\bar{x}) \in r}{t(\bar{x}) \in r}$$

where $t_i(\bar{x}) \in r$ are assumptions and $t(\bar{x}) \in r$ is the conclusion.
When $n = 0$ (no assumptions), the rule is called an axiom.

A derivation tree has nodes marked by tuples $t(\bar{a})$ for some specific values $\bar{a}$ of $\bar{x}$.
We define relation r as the set of all tuples for which there exists a derivation tree.
One can prove (in general case) that tuples for which there exists a derivation tree give us precisely the smallest relation that satisfies the rules!