

Subtyping (Review)

CIS 500
 Software Foundations
 Fall 2005
 21 November

Subtype relation

- (S-REFL) $S <: S$
- (S-TRANS)
$$\frac{S <: U \quad U <: T}{S <: T}$$
- (S-RCDWIDTH) $\{l_1: T_1, \dots, l_{n+k}: T_{n+k}\} <: \{l_1: T_1, \dots, l_n: T_n\}$
- (S-RCDDEPTH)
$$\text{for each } i \quad S_i <: T_i \quad \frac{\{l_1: S_1, \dots, l_n: S_n\} <: \{l_1: T_1, \dots, l_n: T_n\}}{\{k_j: S_j, \dots, l_{n+k}: T_{n+k}\} <: \{l_1: T_1, \dots, l_n: T_n\}}$$
- (S-RCDPERM)
$$\frac{\{k_j: S_j, \dots, l_{n+k}: T_{n+k}\} <: \{l_1: T_1, \dots, l_n: T_n\}}{\{k_j: S_j, \dots, l_{n+k}: T_{n+k}\} \text{ is a permutation of } \{l_1: T_1, \dots, l_n: T_n\}}$$

Administrivia

- ◆ Homework 8 will be on website by end of today. Due Wednesday, November 30.
- ◆ Class on Wednesday, but no recitations due to Thanksgiving.
- ◆ No office hours on Thursday or Friday.
- ◆ View your exams with Cheryl Hickey (Levine 502). (Or Kamilla Mauro tomorrow.)
- ◆ Regrade requests by Monday. Pick up exams next week.

Subtyping and References

$$\frac{S_1 < T_1 \quad T_1 < S_1}{\text{Ref } S_1 < \text{Ref } T_1}$$

(S-REF)

I.e., **Ref** is **not** a covariant (nor a contravariant) type constructor.
Why?

Subtyping and References

$$\frac{S_1 < T_1 \quad T_1 < S_1}{\text{Ref } S_1 < \text{Ref } T_1}$$

(S-REF)

I.e., **Ref** is **not** a covariant (nor a contravariant) type constructor.
Why?

◆ When a reference is **read**, the context expects a T_1 , so if $S_1 < T_1$ then an S_1 is ok.

Subtyping and Lists

$$\frac{S_1 < T_1}{\text{List } S_1 < \text{List } T_1}$$

(S-LIST)

I.e., **List** is a covariant type constructor.

Subtyping references

This is regarded (even by the Java designers) as a mistake in the design.

$$\frac{S_1 < T_1 \quad \text{Array } S_1 < \text{Array } T_1}{\text{(S-ARRAY)}} \quad \text{Similarly...}$$

$$\frac{S_1 < T_1 \quad T_1 < S_1 \quad \text{Array } S_1 < \text{Array } T_1}{\text{(S-ARRAY)}}$$

Subtyping and Arrays

Observation: a value of type **Ref T** can be used in two different ways: as a **source** for values of type **T** and as a **sink** for values of type **T**.

References again

I.e., **Ref** is **not** a covariant (nor a contravariant) type constructor. Why?

- ◆ When a reference is **read**, the context expects a **T₁**, so if **S₁ < T₁** then an **S₁** is ok.
- ◆ When a reference is **written**, the context provides a **T₁** and if the actual type of the reference is **Ref S₁**, someone else may use the **T₁** as an **S₁**. So we need **T₁ < S₁**.

$$\frac{S_1 < T_1 \quad T_1 < S_1 \quad \text{Ref } S_1 < \text{Ref } T_1}{\text{(S-REF)}}$$

Subtyping and References

Similarly...

$$\frac{S_1 < T_1 \quad T_1 < S_1 \quad \text{Array } S_1 < \text{Array } T_1}{\text{(S-ARRAY)}}$$

Subtyping and Arrays

Subtyping rules

$$\frac{S_1 < T_1}{\text{Source } S_1 < \text{Source } T_1} \quad (\text{S-SOURCE})$$

$$\frac{T_1 < S_1}{\text{Sink } S_1 < \text{Sink } T_1} \quad (\text{S-SINK})$$

$$\text{Ref } T_1 < \text{Source } T_1 \quad (\text{S-RESOURCE})$$

$$\text{Ref } T_1 < \text{Sink } T_1 \quad (\text{S-REFSINK})$$

References again

Observation: a value of type **Ref T** can be used in two different ways: as a **source** for values of type **T** and as a **sink** for values of type **T**.

Idea: Split **Ref T** into three parts:

- ◆ **Source T**: reference cell with “read cabability”
- ◆ **Sink T**: reference cell with “write cabability”
- ◆ **Ref T**: cell with both capabilities

Metatheory of Subtyping

Modified Typing Rules

$$\frac{\Gamma \mid \Sigma \vdash t_1 : \text{Source } T_1}{\Gamma \mid \Sigma \vdash t_1 : T_1} \quad (\text{T-DEREF})$$

$$\frac{\Gamma \mid \Sigma \vdash t_1 : \text{Sink } T_1 \quad \Gamma \mid \Sigma \vdash t_2 : T_1}{\Gamma \mid \Sigma \vdash t_1 := t_2 : \text{Unit}} \quad (\text{T-ASSIGN})$$

Syntax-directed rules

In the simply typed lambda-calculus (without subtyping), each rule can be “read from bottom to top” in a straightforward way:

(T-APP)

$$\frac{\Gamma \vdash t_1 : T_1 \rightarrow T_2 \quad \Gamma \vdash t_2 : T_1}{\Gamma \vdash t_1 t_2 : T_2}$$

If we are given some Γ and some t of the form $t_1 t_2$, we can try to find a type for t by

1. finding (recursively) a type for t_1
2. checking that it has the form $T_1 \rightarrow T_2$
3. finding (recursively) a type for t_2
4. checking that it is the same as T_1

Syntax-directed sets of rules

The second important point about the simply typed lambda-calculus is that the set of typing rules is syntax-directed, in the sense that, for every “input” Γ and t , there one rule that can be used to derive typing statements involving t . E.g., if t is an application, then we must proceed by trying to use T-APP. If we succeed, then we have found a type (indeed, the unique type) for t . If it fails, then we know that t is not typable.

→ no backtracking!

Non-syntax-directedness of typing

When we extend the system with subtyping, both aspects of syntax-directedness get broken.

1. The set of typing rules now includes two rules that can be used to give a type to terms of a given shape (the old one plus T-SUB)

$$\text{(T-SUB)} \quad \frac{\Gamma \vdash t : T \quad S < T}{\Gamma \vdash t : S}$$

2. Worse yet, the new rule T-SUB itself is not syntax directed: the inputs to the left-hand subgoal are exactly the same as the inputs to the main goal! (Hence, if we translated the typing rules naively into a typechecking function, the case corresponding to T-SUB would cause divergence.)

Syntax-directed rules

In the simply typed lambda-calculus (without subtyping), each rule can be “read from bottom to top” in a straightforward way:

(T-APP)

$$\frac{\Gamma \vdash t_1 : T_1 \rightarrow T_2 \quad \Gamma \vdash t_2 : T_1}{\Gamma \vdash t_1 t_2 : T_2}$$

If we are given some Γ and some t of the form $t_1 t_2$, we can try to find a type for t by

1. finding (recursively) a type for t_1
2. checking that it has the form $T_1 \rightarrow T_2$
3. finding (recursively) a type for t_2
4. checking that it is the same as T_1

Syntax-directed sets of rules

The second important point about the simply typed lambda-calculus is that the set of typing rules is syntax-directed, in the sense that, for every “input” Γ and t , there one rule that can be used to derive typing statements involving t . E.g., if t is an application, then we must proceed by trying to use T-APP. If we succeed, then we have found a type (indeed, the unique type) for t . If it fails, then we know that t is not typable.

→ no backtracking!

2. Worse yet, the new rule T-SUB itself is not syntax directed: the inputs to the left-hand subgoal are exactly the same as the inputs to the main goal! (Hence, if we translated the typing rules naively into a typechecking function, the case corresponding to T-SUB would cause divergence.)

- ◆ For the input positions, all metavariables appearing in the premises also appear in the conclusion (so we can calculate inputs to the “subgoals” from the subexpressions of inputs to the main goal)
- ◆ For the output positions, all metavariables appearing in the conclusions also appear in the premises (so we can calculate outputs from the main goal from the outputs of the subgoals)

(T-APP)

$$\frac{\Gamma \vdash t_1 t_2 : T_2}{\Gamma \vdash t_1 : T_1 \rightarrow T_2 \quad \Gamma \vdash t_2 : T_1}$$

1. Observation: We don't **need** 1000 ways to prove a given typing or subtyping statement — one is enough.
- Think more carefully about the typing and subtyping systems to see where we can get rid of excess flexibility
2. Use the resulting intuitions to formulate new “algorithmic” (i.e., syntax-directed) typing and subtyping relations
3. Prove that the algorithmic relations are “the same as” the original ones in an appropriate sense.

What to do?

is badly non-syntax-directed: the premises contain a metavariable (in an “input position”) that does not appear at all in the conclusion.

To implement this rule naively, we'd have to **guess** a value for **u!**

$$\frac{S <: T}{S <: u \quad u <: T} \text{ (S-TRANS)}$$

1. There are **lots** of ways to derive a given subtyping statement.
 2. The transitivity rule
- Moreover, the subtyping relation is not syntax directed either.

Non-syntax-directedness of subtyping

Developing an algorithmic subtyping relation

What to do?

- For a given subtyping statement, there are multiple rules that could be used last in a derivation.
1. S-RCD-WIDTH, S-RCD-DEPTH, and S-RCD-PERM overlap with each other
 2. S-REFL and S-TRANS overlap with everything

Issues

Idea: combine all three record subtyping rules into one “macro rule” that captures all of their effects

Step 1: simplify record subtyping

$$(S-RCD) \frac{\{k_j : s_j\}_{j \in I..n} < \{l_i : T_i\}_{i \in I..n}}{\{l_i : T_i\}_{i \in I..n} \subseteq \{k_j\}_{j \in I..n} \quad k_j = l_i \text{ implies } s_j < T_i}$$

$$(S-REFL) \quad S < S$$

$$(S-TRANS) \quad \frac{S < U \quad U < T}{S < T}$$

$$(S-RCDWIDTH) \quad \{l_1 : T_1\}_{i \in I..n+k} < \{l_1 : T_1\}_{i \in I..n}$$

$$(S-RCDDEPTH) \quad \frac{\text{for each } i \quad s_i < T_i}{\{l_1 : s_1\}_{i \in I..n} < \{l_1 : T_1\}_{i \in I..n}}$$

$$(S-RCDPERM) \quad \frac{\{k_j : s_j\}_{j \in I..n} < \{l_1 : T_1\}_{i \in I..n}}{\{k_j : s_j\}_{j \in I..n} \text{ is a permutation of } \{l_1 : T_1\}_{i \in I..n}}$$

Subtype relation

$$(S-ARROW) \quad \frac{T_1 < S_1 \quad S_2 < T_2}{S_1 \rightarrow S_2 < T_1 \rightarrow T_2}$$

$$(S-TOP) \quad S < \text{Top}$$

Even simpler subtype relation

$$\begin{array}{l}
 \text{(S-REFL)} \quad S < S \\
 \text{(S-TRANS)} \quad \frac{S < U \quad U < T}{S < T} \\
 \text{(S-RCD)} \quad \frac{\{k_j : S_j\}_{j \in I \dots n} \subseteq \{k_j\}_{j \in I \dots m} \quad k_j = T_i \text{ implies } S_j < T_i}{\{k_j : S_j\}_{j \in I \dots m} < \{T_i : T_i\}_{i \in I \dots n}} \\
 \text{(S-ARROW)} \quad \frac{T_1 < S_1 \quad S_2 < T_2 \quad S_1 \rightarrow S_2 < T_1 \rightarrow T_2}{S < \text{Top}} \\
 \text{(S-TOP)} \quad S < \text{Top}
 \end{array}$$

Step 3: Get rid of transitivity

Observation: S-TRANS is unnecessary.
 Lemma: If $S < T$ can be derived, then there is a derivation that does not use S-TRANS.

Simpler subtype relation

$$\begin{array}{l}
 \text{(S-REFL)} \quad S < S \\
 \text{(S-TRANS)} \quad \frac{S < U \quad U < T}{S < T} \\
 \text{(S-RCD)} \quad \frac{\{k_j : S_j\}_{j \in I \dots n} \subseteq \{k_j\}_{j \in I \dots m} \quad k_j = T_i \text{ implies } S_j < T_i}{\{k_j : S_j\}_{j \in I \dots m} < \{T_i : T_i\}_{i \in I \dots n}} \\
 \text{(S-ARROW)} \quad \frac{T_1 < S_1 \quad S_2 < T_2 \quad S_1 \rightarrow S_2 < T_1 \rightarrow T_2}{S < \text{Top}} \\
 \text{(S-TOP)} \quad S < \text{Top}
 \end{array}$$

Step 2: Get rid of reflexivity

Observation: S-REFL is unnecessary.
 Lemma: $S < S$ can be derived for every type without using S-REFL.

Subtyping Algorithm (pseudo-code)

The algorithmic rules can be translated directly into code:

```

subtyping(S, T) = if T = Top, then true
                else if S = S1 → S2 and T = T1 → T2
                    then subtyping(T1, S1) ∨ subtyping(S2, T2)
                else if S = {kj : Sj}i ∈ I..m and T = {li : Ti}i ∈ I..n
                    then {li ∈ I..n} ⊆ {kj}i ∈ I..m
                ∨ for all i there is some j ∈ I..m with kj = li
                else false.
    
```

“Algorithmic” subtyping relation

(SA-Top) $\vdash S <: \text{Top}$

(SA-ARROW) $\frac{\vdash T_1 <: S_1 \quad \vdash S_2 <: T_2}{\vdash S_1 \rightarrow S_2 <: T_1 \rightarrow T_2}$

(SA-RCD) $\frac{\vdash \{k_j : S_j\}_{j \in I..m} <: \{l_i : T_i\}_{i \in I..n}}{\text{for each } k_j = l_i, \vdash S_j <: T_i}$

Decision Procedures

A **decision procedure** for a relation $R \subseteq U$ is a total function p from U to $\{\text{true}, \text{false}\}$ such that $p(u) = \text{true}$ iff $u \in R$.

Soundness and completeness

Theorem: $S <: T$ iff $\vdash S <: T$.

Proof: ...

Terminology:

- ◆ The algorithmic presentation of subtyping is **sound** with respect to the original if $\vdash S <: T$ implies $S <: T$.
(Everything validated by the algorithm is actually true.)
- ◆ The algorithmic presentation of subtyping is **complete** with respect to the original if $S <: T$ implies $\vdash S <: T$.
(Everything true is validated by the algorithm.)

Decision Procedures

A **decision procedure** for a relation $R \subseteq U$ is a total function p from U to $\{true, false\}$ such that $p(u) = true$ iff $u \in R$.

Is our *subtype* function a decision procedure?

have

1. if $subtype(S, T) = true$, then $\vdash S <: T$

(hence, by soundness of the algorithmic rules, $S <: T$)

2. if $subtype(S, T) = false$, then not $\vdash S <: T$

(hence, by completeness of the algorithmic rules, not $S <: T$)

Is our *subtype* function a decision procedure?

Since *subtype* is just an implementation of the algorithmic subtyping rules, we

Q: What's missing?

(hence, by completeness of the algorithmic rules, not $S <: T$)

2. if $subtype(S, T) = false$, then not $\vdash S <: T$

(hence, by soundness of the algorithmic rules, $S <: T$)

1. if $subtype(S, T) = true$, then $\vdash S <: T$

have

Since *subtype* is just an implementation of the algorithmic subtyping rules, we

Is our *subtype* function a decision procedure?

A **decision procedure** for a relation $R \subseteq U$ is a total function p from U to $\{true, false\}$ such that $p(u) = true$ iff $u \in R$.

Decision Procedures

Decision Procedures

A **decision procedure** for a relation $R \subseteq U$ is a total function p from U to $\{true, false\}$ such that $p(u) = true$ iff $u \in R$.

Is our *subtype* function a decision procedure?

Since *subtype* is just an implementation of the algorithmic subtyping rules, we

have

1. if $subtype(S, T) = true$, then $\vdash S <: T$

(hence, by soundness of the algorithmic rules, $S <: T$)

2. if $subtype(S, T) = false$, then not $\vdash S <: T$

(hence, by completeness of the algorithmic rules, not $S <: T$)

Is our *subtype* function a decision procedure?

Since *subtype* is just an implementation of the algorithmic subtyping rules, we

have

1. if $subtype(S, T) = true$, then $\vdash S <: T$

(hence, by soundness of the algorithmic rules, $S <: T$)

2. if $subtype(S, T) = false$, then not $\vdash S <: T$

(hence, by completeness of the algorithmic rules, not $S <: T$)

Q: What's missing?

A: How do we know that *subtype* is a **total** function?

Where is this rule really needed?

$$\frac{\Gamma \vdash t : T}{\Gamma \vdash t : S \quad S <: T} \text{ (T-SUB)}$$

For the typing relation, we have just one problematic rule to deal with: substitution.

Issue

Where is this rule really needed?

For applications. E.g., the term $(\lambda r:\{x:\text{Nat}\}. r.x) \{x=0, y=1\}$

is not typable without using substitution.

$$\frac{\Gamma \vdash t : T}{\Gamma \vdash t : S \quad S <: T} \text{ (T-SUB)}$$

For the typing relation, we have just one problematic rule to deal with: substitution.

Issue

A: How do we know that *subtype* is a **total** function? Prove it!

Q: What's missing?

- 1. if $\text{subtype}(S, T) = \text{true}$, then $S <: T$
 - 2. if $\text{subtype}(S, T) = \text{false}$, then not $S <: T$
- (hence, by soundness of the algorithmic rules, $S <: T$)
- (hence, by completeness of the algorithmic rules, not $S <: T$)

Since *subtype* is just an implementation of the algorithmic subtyping rules, we have

Is our *subtype* function a decision procedure?

A **decision procedure** for a relation $R \subseteq U$ is a total function p from U to $\{\text{true}, \text{false}\}$ such that $p(u) = \text{true}$ iff $u \in R$.

Decision Procedures

Metatheory of Typing

$$\frac{\frac{\vdots}{\Gamma, x:s_1 \vdash s_2 : T_2} \quad \frac{\vdots}{s_2 <: T_2}}{\Gamma, x:s_1 \vdash s_2 : T_2} \text{(T-Sub)}}{\Gamma \vdash \lambda x:s_1.s_2 : s_1 \rightarrow T_2} \text{(T-Abs)}$$

Example

$$\frac{\frac{\vdots}{\Gamma, x:s_1 \vdash s_2 : s_2} \quad \frac{\vdots}{s_2 <: T_2}}{\Gamma, x:s_1 \vdash s_2 : T_2} \text{(S-REPL)} \quad \frac{\vdots}{s_1 <: s_1} \quad \frac{\vdots}{s_1 \rightarrow s_2 <: s_1 \rightarrow T_2}}{\Gamma \vdash \lambda x:s_1.s_2 : s_1 \rightarrow T_2} \text{(T-Sub)}}{\Gamma \vdash \lambda x:s_1.s_2 : s_1 \rightarrow T_2} \text{(S-ARROW)}$$

becomes

$$\frac{\frac{\vdots}{\Gamma, x:s_1 \vdash s_2 : s_2} \quad \frac{\vdots}{s_2 <: T_2}}{\Gamma, x:s_1 \vdash s_2 : T_2} \text{(T-Sub)}}{\Gamma \vdash \lambda x:s_1.s_2 : T_2} \text{(T-Abs)}$$

Example

Where is this rule really needed?
 For applications. E.g., the term
 $(\lambda r:\{x:\text{Nat}\}. r.x) \{x=0, y=1\}$
 is not typable without using subsupmption.
 Where else??

$$\frac{\Gamma \vdash t : T}{\Gamma \vdash t : S \quad S <: T} \text{(T-Sub)}$$

Issue

For the typing relation, we have just one problematic rule to deal with:
 subsupmption.

Where is this rule really needed?
 For applications. E.g., the term
 $(\lambda r:\{x:\text{Nat}\}. r.x) \{x=0, y=1\}$
 is not typable without using subsupmption.
 Where else??
Nowhere else! Uses of subsupmption to help typecheck applications are the only interesting ones.

$$\frac{\Gamma \vdash t : T}{\Gamma \vdash t : S \quad S <: T} \text{(T-Sub)}$$

Issue

For the typing relation, we have just one problematic rule to deal with:
 subsupmption.

$$\begin{array}{c}
 \vdots \\
 \vdots \\
 \vdots \\
 \hline
 \Gamma s : s \quad \Gamma s : u \\
 \text{(T-Sub)} \quad \text{(T-Sub)} \\
 \vdots \\
 \vdots \\
 \vdots \\
 \hline
 \Gamma s : t \\
 \text{(T-Sub)}
 \end{array}$$

Example

$$\begin{array}{c}
 \vdots \\
 \vdots \\
 \vdots \\
 \hline
 \Gamma s : s \quad \Gamma s : u \\
 \text{(T-Sub)} \quad \text{(T-Sub)} \\
 \vdots \\
 \vdots \\
 \vdots \\
 \hline
 \Gamma s : t \\
 \text{(T-Sub)}
 \end{array}$$

Example

becomes

$$\begin{array}{c}
 \vdots \\
 \vdots \\
 \vdots \\
 \hline
 \Gamma s : s \quad \Gamma s : u \\
 \text{(T-Sub)} \quad \text{(T-Sub)} \\
 \vdots \\
 \vdots \\
 \vdots \\
 \hline
 \Gamma s : t \\
 \text{(T-Sub)}
 \end{array}$$

$$\begin{array}{c}
 \vdots \\
 \vdots \\
 \vdots \\
 \hline
 \Gamma s : s \quad \Gamma s : u \\
 \text{(T-Sub)} \quad \text{(T-Sub)} \\
 \vdots \\
 \vdots \\
 \vdots \\
 \hline
 \Gamma s : t \\
 \text{(S-TRANS)}
 \end{array}$$