## CSC\_51051\_EP: First-order logic

Samuel Mimram

2025

École polytechnique

# Part I

# First-order logic

Recall that in the first part of the course we have defined a natural deduction calculus for propositional logic.

We are ultimately going to extend this calculus for dependent types (the core of Agda).

For now, we begin by looking at first order logic (which should be familiar from INF412).

This is a first kind of dependency where types can depend on terms.

There are two syntactic classes in first-order logic:

- terms construct elements of the model,
- formulas express logical properties,

and predicate allow constructing formulas form terms.

For instance, we can express associativity of a multiplication (e.g. in a group) with



A signature  $\Sigma$  is a set of *function symbols* together with an *arity* for each symbol.

For instance, the signature for groups is

$$\Sigma = \{m: 2, e: 0, i: 1\}$$

We write  $\mathcal{T}_{\Sigma}$  for the **terms** generated by the signature, e.g. m(m(e, x), m(x, i(y)))

which is what we would usually write

 $(1 \times x) \times (x \times y^{-1})$ 

#### Terms

We suppose fixed an infinite countable set of variables:

 $\mathcal{X} = \{x, y, z, \ldots\}$ 

The set  $\mathcal{T}_{\Sigma}$  of **terms** over  $\Sigma$  is the smallest set such that

• every variable is a term:

 $\mathcal{X}\subseteq \mathcal{T}_{\Sigma}$ 

• terms are closed under the operations:

if  $f \in \Sigma$  is a function symbol of arity *n* and  $t_1, \ldots, t_n$  are terms then

 $f(t_1,\ldots,t_n)$ 

is also a term (this is a *formal* application).

In short, terms are generated by the grammar:  $t ::= x | f(t_1, \ldots, t_n)$ 



The " $\lambda$ -terms" are the terms over the signature

$$\Sigma = \{ \mathsf{app} : 2, \lambda_x : 1 \mid x \in Var \}$$

not really: we don't take  $\alpha$ -conversion in account.

For instance,

 $\lambda_x(\operatorname{app}(x,\lambda_y(y)))$ 

which is simply another notation for

 $\lambda x.x(\lambda y.y)$ 

We write FV(t) for the set of variables occurring in a term (no variable is bound).

For instance,

$$\mathsf{FV}(m(m(e,x),m(x,i(y)))) = \{x,y\}$$

Formally, this is defined by induction by

$$FV(x) = \{x\}$$
$$FV(f(t_1,...,t_n)) = \bigcup_{i=1}^n FV(t_i)$$

A term *t* is closed when  $FV(t) = \emptyset$ .

For instance, natural numbers can be defined as the set of closed terms over

 $\Sigma = \{Z: 0, S: 1\}$ 

Namely, the closed terms are

 $Z() \qquad S(Z()) \qquad S(S(Z()) \qquad \dots$ 

(a non-closed term is S(S(x)))

We suppose fixed set  $\mathcal{P}$  of **predicates** together with an arity.

For instance,

$$\mathcal{P} = \{=: 2, \mathsf{even}: 1, \ldots\}$$

The set of formulas (or propositions) is generated by

 $A ::= P(t_1, \ldots, t_n) \mid A \Rightarrow B \mid A \land B \mid \top \mid A \lor B \mid \bot \mid \neg A \mid \exists x.A \mid \forall x.A$ 

where *P* is a predicate of arity *n*, the  $t_i$  are terms and  $x \in \mathcal{X}$  is a variable.

For instance, in the signature of groups

$$\Sigma = \{m: 2, e: 0, i: 1\}$$
  $\mathcal{P} = \{=: 2, \ldots\}$ 

we have the formula

$$\forall x. \forall y. \forall z. \quad m(m(x, y), z) = m(x, m(y, z)) \land m(e, x) = x \land m(x, e) = x$$

With

$$\mathcal{P} = \{D: 1, \ldots\}$$

the drinker formula is

 $\exists x.(D(x) \Rightarrow \forall y.D(y))$ 

#### $\alpha$ -equivalence

In a formula of the form  $\exists x.A$  or  $\forall x.A$ , the variable x is **bound** in A (a variable which is not bound is **free**).

As usual, we consider formulas up to renaming of bound variables.

Formally, we define set FV(A) of free variables of A by

 $FV(P(t_1, ..., t_n)) = FV(t_1) \cup ... \cup FV(t_n)$   $FV(A \Rightarrow B) = FV(A \times B) = FV(A + B) = FV(A) \cup FV(B)$   $FV(\top) = FV(\bot) = \emptyset$   $FV(\neg A) = FV(A)$   $FV(\forall x.A) = FV(\exists x.A) = FV(A) \setminus \{x\}$ 

#### Given a formula A, a term t and a variable x, we write

## A[t/x]

for the formula A where all the free occurrences of x have been substituted by t

$$A = (\exists y.x + x = y) \lor (\exists x.x = y)$$
$$A[z + z/x] = (\exists y.(z + z) + (z + z) = y) \lor (\exists x.x = y)$$
$$A[y + y/x] = (\exists z.(y + y) + (y + y) = z) \lor (\exists x.x = y)$$
$$A[y + y/x] \neq (\exists y.(y + y) + (y + y) = y) \lor (\exists x.x = y)$$

As usual, we might have to rename variables to avoid captures!

Rules

The rules of (intuitionistic) logic are the usual ones:

$$\overline{\Gamma, A, \Gamma' \vdash A}^{(ax)}$$

$$\frac{\Gamma \vdash A \Rightarrow B \qquad \Gamma \vdash A}{\Gamma \vdash B} (\Rightarrow_{E})$$

$$\frac{\Gamma \vdash A \land B}{\Gamma \vdash A} (\land_{E}^{I}) \qquad \frac{\Gamma \vdash A \land B}{\Gamma \vdash B} (\land_{E}^{r})$$

$$\frac{\Gamma \vdash A \lor B \qquad \Gamma, A \vdash C \qquad \Gamma, B \vdash C}{\Gamma \vdash C} (\lor_{E})$$

$$\frac{\Gamma \vdash A \lor B \qquad \Gamma, A \vdash C \qquad \Gamma, B \vdash C}{\Gamma \vdash A} (\downarrow_{E})$$

$$\frac{\Gamma \vdash \neg A \qquad \Gamma \vdash A}{\Gamma \vdash \bot} (\neg_{E})$$

$$\frac{\Gamma, A \vdash B}{\Gamma \vdash A \Rightarrow B} (\Rightarrow_{1})$$

$$\frac{\Gamma \vdash A \qquad \Gamma \vdash B}{\Gamma \vdash A \land B} (\land_{1})$$

$$\frac{\Gamma \vdash A}{-A \lor B} (\lor_{1}^{l}) \qquad \frac{\Gamma \vdash B}{\Gamma \vdash A \lor B} (\lor_{1}^{r})$$

$$\frac{\overline{\Gamma \vdash T}}{\Gamma \vdash \neg A} (\neg_{1})$$

ГΙ

14

Rules

Together with four new rules:

$$\frac{\Gamma \vdash \forall x.A}{\Gamma \vdash A[t/x]} (\forall_{\mathsf{E}}) \qquad \qquad \frac{\Gamma \vdash A}{\Gamma \vdash \forall x.A} (\forall_{\mathsf{I}})$$
$$\frac{\Gamma \vdash \exists x.A \quad \Gamma, A \vdash B}{\Gamma \vdash B} (\exists_{\mathsf{E}}) \qquad \qquad \frac{\Gamma \vdash A[t/x]}{\Gamma \vdash \exists x.A} (\exists_{\mathsf{I}})$$

These rules are subject to the following (important!) side conditions:

- in  $(\forall_I)$ , we suppose  $x \notin FV(\Gamma)$ ,
- in  $(\exists_{\mathsf{E}})$ , we suppose  $x \notin \mathsf{FV}(\Gamma) \cup \mathsf{FV}(B)$ .

#### For instance,



The side conditions

$$\frac{\Gamma \vdash A}{\Gamma \vdash \forall x.A} (\forall_{\mathsf{I}})$$

with  $x \notin FV(\Gamma)$ 

avoid clearly problematic proofs:

$$\frac{\overline{A(x) \vdash A(x)}}{A(x) \vdash \forall x.A(x)} \stackrel{(ax)}{(\forall_{1})} \\ \frac{\overline{A(x) \vdash \forall x.A(x)}}{\vdash A(x) \Rightarrow \forall x.A(x)} \stackrel{(\Rightarrow_{1})}{(\Rightarrow_{1})} \\ \frac{\overline{\vdash \forall x.(A(x) \Rightarrow \forall x.A(x))}}{\vdash A(t) \Rightarrow \forall x.A(x)} \stackrel{(\forall_{E})}{(\forall_{E})}$$

We have  $x \in FV(A(x))!$ 

The rules are not completely satisfactory, for instance we can prove:

$$\frac{\overline{\forall x.A \vdash \forall x.A}}{\forall x.A \vdash A} \stackrel{(ax)}{(\forall_{\mathsf{E}})} \\ \frac{\overline{\forall x.A \vdash A}}{\forall x.A \vdash \exists x.A} \stackrel{(\exists_{\mathsf{I}})}{(\exists_{\mathsf{I}})} \\ \overline{+ (\forall x.A) \Rightarrow \exists x.A} \stackrel{(ax)}{(ax)}$$

(this can be fixed but we will simply ignore it)

As for predicate logic, when a proof contains cuts, we can eliminate those.

A cut means that

- you use an introduction rule,
- followed by an elimination rule for the introduced connective.

We still have the cut elimination property: the two new cases are



**Theorem** If  $\Gamma \vdash A$  is provable then it admits a cut-free proof. As before, we have

Proposition

A cut-free intuitionistic proof of  $\vdash$  A necessarily ends with an introduction rule.

In particular,

**Theorem (Witness property)**  $If \vdash \exists x.A$  is provable intuitionistically then there exists t such that A[t/x] is provable. **Proof.** 

The proof can be supposed to be cut-free and then necessarily ends on  $\frac{\vdash A[t/x]}{\vdash \exists x.A}$  ( $\exists_1$ ).

We can easily extend the Curry-Howard correspondence to first-order logic.

The expressions e corresponding to programs are

$$e ::= \ldots \mid \lambda x.e \mid et \mid \langle t, e 
angle \mid ext{let} \langle x, y 
angle = e ext{ in } e'$$

and typing rules are

$$\frac{\Gamma \vdash e : A}{\Gamma \vdash \lambda x. e : \forall x. A} (\forall_{\mathsf{I}}) \qquad \qquad \frac{\Gamma \vdash e : A[t/x]}{\Gamma \vdash \langle t, e \rangle : \exists x. A} (\exists_{\mathsf{I}})$$

and

$$\frac{\Gamma \vdash e : \forall x.A}{\Gamma \vdash e t : A[t/x]} (\forall_{\mathsf{E}})$$

$$\frac{\Gamma \vdash e : \exists x.A \qquad \Gamma, y : A \vdash e' : B}{\Gamma \vdash \operatorname{let} \langle x, y \rangle = e \operatorname{in} e' : B} (\exists_{\mathsf{E}})$$

Reduction rules correspond to cut-elimination:

$$\frac{\pi}{\frac{\Gamma \vdash e : A[t/x]}{\Gamma \vdash \langle t, e \rangle : \exists x.A}} (\exists_{I}) \qquad \frac{\pi'}{\frac{\Gamma, y : A \vdash e' : B}{\Gamma \vdash e' : B}} (\exists_{E}) \qquad \rightsquigarrow \qquad \frac{\pi'[t/x][\pi/A]}{\Gamma \vdash e'[t/x, e/y] : B}$$
  
i.e.  
$$let \langle x, y \rangle = \langle t, e \rangle in e' \qquad \longrightarrow_{\beta} \qquad e'[t/x, e/y]$$

#### **Curry-Howard**

Note that there are two abstractions:

- $\lambda x^A \cdot e : A \Rightarrow B$
- $\lambda x.e: \forall x.A$

Similarly, there are two pairs:

- $\langle e, e' \rangle : A \wedge B$
- $\langle t, e \rangle : \exists x.A$

They behave similarly but they are not the same (we should write them differently).

They will become the same in dependent types!

#### **Curry-Howard**

For instance, recall our proof of  $(\forall x. \neg A) \Rightarrow \neg(\exists x. A)$ :



The corresponding term is

 $\lambda f.\lambda e. \mathtt{let} \langle x, a \rangle = e \operatorname{in} f x a \qquad : \qquad (\forall x. (A \Rightarrow \bot)) \Rightarrow (\exists x. A) \Rightarrow \bot$ 

As before, the rules we have presented implement intuitionistic first-order logic.

Classical first order logic can be obtained by adding the usual axioms, e.g.

#### $\neg\neg A \Rightarrow A$

$$\vdash \neg(\forall x.\neg A(x)) \Rightarrow \exists x.A(x)$$

$$\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\vdash \neg (\forall x. \neg A(x)) \Rightarrow \exists x. A(x)} (\Rightarrow_1)$$

$$\frac{\neg \forall x. \neg A(x) \vdash \neg \neg \exists x. A(x)}{\neg \forall x. \neg A(x) \vdash \exists x. A(x)} (\neg \neg_{\mathsf{E}}) \\
+ \neg (\forall x. \neg A(x)) \Rightarrow \exists x. A(x) (\Rightarrow_{\mathsf{I}})$$

$$\frac{\neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \bot}{\neg \forall x. \neg A(x) \vdash \neg \neg \exists x. A(x)} (\neg_{1}) \\ \hline \neg \forall x. \neg A(x) \vdash \neg \neg \exists x. A(x) \\ \hline \neg \forall x. \neg A(x) \vdash \exists x. A(x) \\ \vdash \neg (\forall x. \neg A(x)) \Rightarrow \exists x. A(x) \\ \hline (\Rightarrow_{1}) \\ \hline \end{pmatrix}$$

$$\begin{array}{c} \begin{array}{c} & \neg \forall x. \neg A(x), \neg \exists x. A(x), A(x_0) \vdash \bot & (\neg_1) \\ & \neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \neg A(x_0) & (\forall_1) \\ & & \neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \forall x. \neg A(x) & (\neg_E) \\ & & \neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \bot & (\neg_1) \\ & & \neg \forall x. \neg A(x) \vdash \neg \neg \exists x. A(x) & (\neg \neg_E) \\ & & \neg \forall x. \neg A(x) \vdash \exists x. A(x) & (\Rightarrow_1) \\ & & \vdash \neg (\forall x. \neg A(x)) \Rightarrow \exists x. A(x) & (\Rightarrow_1) \end{array}$$
# Classical first-order logic

A typical formula which is only provable classically:

$$\frac{\neg \forall x. \neg A(x), \neg \exists x. A(x), A(x_0) \vdash A(x_0)}{\neg \forall x. \neg A(x), \neg \exists x. A(x), A(x_0) \vdash \exists x. A(x)} (\exists_1) \\
\frac{\neg \forall x. \neg A(x), \neg \exists x. A(x), A(x_0) \vdash \bot}{\neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \neg A(x_0)} (\neg_1) \\
\frac{\neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \neg A(x_0)}{\neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \forall x. \neg A(x)} (\neg_1) \\
\frac{\neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \bot}{\neg \forall x. \neg A(x) \vdash \neg \exists x. A(x)} (\neg_1) \\
\frac{\neg \forall x. \neg A(x) \vdash \neg \exists x. A(x)}{\neg \forall x. \neg A(x) \vdash \exists x. A(x)} (\neg_1) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg \neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg \neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash \exists x. A(x)}{\neg z} (\neg z) \\
\frac{\neg \forall x. \neg A(x) \vdash z}{\neg z} (\neg z) \\
\frac{\neg z}{\neg z} \\
\frac{\neg z} (\neg z) \\
\frac{\neg z} (\neg z) \\
\frac$$

# Classical first-order logic

A typical formula which is only provable classically:

$$\frac{\neg \forall x. \neg A(x), \neg \exists x. A(x), A(x_0) \vdash A(x_0)}{\neg \forall x. \neg A(x), \neg \exists x. A(x), A(x_0) \vdash \exists x. A(x)} (\exists_{1}) \\ \neg \forall x. \neg A(x), \neg \exists x. A(x), A(x_0) \vdash \exists x. A(x)} (\neg_{E}) \\ \neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \neg A(x_0) \\ \neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \forall x. \neg A(x)} (\forall_{1}) \\ \neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \forall x. \neg A(x)} (\neg_{E}) \\ \neg \forall x. \neg A(x), \neg \exists x. A(x) \vdash \bot} (\neg_{1}) \\ \neg \forall x. \neg A(x) \vdash \neg \neg \exists x. A(x)} (\neg_{1}) \\ \neg \forall x. \neg A(x) \vdash \forall x. A(x)} (\neg_{1}) \\ \neg \forall x. \neg A(x) \vdash \forall x. A(x)} (\neg_{1}) \\ \neg \forall x. \neg A(x) \vdash \forall x. A(x)} (\neg_{1}) \\ \neg \forall x. \neg A(x) \vdash \forall x. A(x)} (\neg_{1}) \\ \neg \forall x. \neg A(x) \vdash \forall x. A(x)} (\neg_{1}) \\ \neg \forall x. \neg A(x) \vdash \forall x. A(x)} (\neg_{1}) \neg \forall x. A(x)} (\neg_{1}) \neg \forall x. A(x)} (\neg_{1}) \neg \forall x. A(x) \vdash \forall x. A(x)} (\neg_{1}) \neg \forall x. A(x) \neg \forall x. A(x) \neg \forall x. A(x)} (\neg_{1}) \neg \forall x. A(x)} (\neg_{1}) \neg \forall x. A(x) \neg \forall x. A(x) \neg \forall x. A(x)} (\neg_{1}) \neg \forall x. A(x) \neg \forall$$

# Classical first-order logic

A typical formula which is only provable classically:

We see that we cannot expect the witness property in classical first-order logic!

# The Drinker formula

Another one is the Drinker formula  $A = \exists x.(D(x) \Rightarrow (\forall y.D(y)))$ :  $\frac{\frac{1}{\dots,\neg D(y)\vdash \neg D(y)} (\mathsf{ax})}{\neg A, D(x), \neg D(y), D(y)\vdash \bot} (\mathsf{ax}) (\mathsf{ax})$  $(\neg_{\mathsf{F}})$  $-(\perp_E)$  $\neg A, D(x), \neg D(y), D(y) \vdash \forall y. D(y)$ — (⇒<sub>I</sub>)  $\neg A, D(x), \neg D(y) \vdash D(y) \Rightarrow (\forall y. D(y))$ — (∃<sub>I</sub>)  $\neg A, D(x), \neg D(y) \vdash \exists x. (D(x) \Rightarrow (\forall y. D(y)))$ —— (¬<sub>F</sub>)  $\neg A, D(x), \neg D(y) \vdash \bot$ - (¬ı)  $\neg A, D(x) \vdash \neg \neg D(y)$ - (¬¬<sub>E</sub>)  $\neg A. D(x) \vdash D(y)$ (∀ı)  $\neg A, D(x) \vdash \forall y. D(y)$ - (⇒ı)  $\neg A \vdash D(x) \Rightarrow \forall v.D(v)$ – (∃ı)  $\neg A \vdash \exists x.(D(x) \Rightarrow \forall y.D(y))$ (raa)  $\vdash A$ 

28

# Part II

# Theories

# A theory consists of

- a signature Σ,
- a set  $\mathcal{P}$  of predicates,
- a set  $\mathcal{T}$  of axioms (formulas that are supposed to be true).

### Groups

For instance, the theory of groups has

• signature

 $\Sigma = \{m: 2, e: 0, i: 1\}$ 

• predicates

$$\mathcal{P} = \{=:2\}$$

• axioms:

$$\forall x.\forall x'.\forall y.\forall y'.(x = x') \Rightarrow (y = y') \Rightarrow (m(x, y) = m(x', y'))$$
  
$$\forall x.\forall x'.(x = x') \Rightarrow (i(x) = i(x'))$$

(we generally omit congruence axioms in the following)

A formula is provable in a theory  $\mathcal{T}$  when it can be proved in using the rules of logic extended with one rule  $\frac{1}{\vdash A} \text{ (axiom)}$ 

for each  $A \in \mathcal{T}$ .

Let us quickly recall (see INF412) that a model of a theory consists of

- a set *M*,
- a function  $\llbracket f \rrbracket : M^n \to M$  for every function symbol f of arity n,
- a relation  $\llbracket P \rrbracket \subseteq M^n$  for every relation symbol R of arity n,

such that the interpretation of every axiom is true.

For instance, a model of a the theory of groups is... a group!

The following (not very difficult) theorem ensures that our syntax is alright:

**Theorem (Correctness)** If a formula can be derived in a theory then it is true in every model of the theory. Correctness

In the theory of groups, one can show

$$\forall x.\forall y.(m(x,y)=e) \Rightarrow (y=i(x))$$

In traditional notation:

 $x \times y = 1$  $x^{-1} \times (x \times y) = x^{-1} \times 1$  $x^{-1} \times (x \times y) = x^{-1}$  $(x^{-1} \times x) \times y = x^{-1}$  $1 \times y = x^{-1}$  $y = x^{-1}$ 

By correctness, it holds in every group!

In the theory of groups, the following formula cannot be shown:

 $\forall x.\forall y.m(x,y) = m(y,x)$ 

Namely, any non-commutative group (look in a math book to have a concrete example) is a model of the theory and does not satisfy the above formula and we would have a contradiction by correctness.

A theory is **consistent** if it does not allow to prove  $\perp$ .

**Theorem** *A theory with a model is consistent.* 

Proof.

Suppose that  $\perp$  can be proved in the theory. By correctness, this means that  $\perp$  is true in the model, but it is not by definition of a model, contradiction.

## Presburger arithmetic

The **Presburger arithmetic** is the theory over  $\Sigma = \{0: 0, S: 1, +: 2\} \qquad \qquad \mathcal{P} = \{=: 2\}$ 

with axioms

$$\forall x.0 = S(x) \Rightarrow \bot$$
  
$$\forall x.\forall y.S(x) = S(y) \Rightarrow x = y$$
  
$$\forall x.0 + x = x$$
  
$$\forall x.\forall y.S(x) + y = S(x + y)$$

together with, for every formula A(x),

 $A(0) \Rightarrow (\forall x.A(x) \Rightarrow A(S(x))) \Rightarrow \forall x.A(x)$ 

lt is

- decidable (in  $O(2^{2^{cn}})$ )
- coherent

and the expected formulas are derivable, e.g.

$$\forall x.x + 0 = x$$
  
$$\forall x.\forall y.\forall z.(x + y) + z = x + (y + z)$$
  
$$\forall x.\forall y.x + y = y + x$$

It admits  $(\mathbb{N}, 0, \_+1, +)$  as a model but this is not the only model.

For instance,  $\forall x.x + 0 = x$  can be proved by recurrence on x.

Consider the formula A(x) being x + 0 = x. We have

- A(0): 0+0=0.
- Suppose A(x), we have A(S(x)), namely

S(x) + 0 = S(x + 0) = S(x)

The **Peano arithmetic** is the theory over  $\Sigma = \{0: 0, S: 1, +: 2, \times: 2\} \qquad \qquad \mathcal{P} = \{=: 2\}$ 

with the previous axioms together with

 $\forall x.0 \times x = 0 \\ \forall x.\forall y.S(x) \times y = y + (x \times y)$ 

lt is

- undecidable (Hilbert problem #2),
- coherent.

For coherence, we can use the fact that  $(\mathbb{N}, 0, \_+1, +, \times)$  is a model and use correctness.

This is however unsatisfactory because it lives in ZFC.

Gentzen introduced cut-elimination in 1936 to show the consistency of PA. This requires a transfinite induction up to the ordinal  $\varepsilon_0$ ( $\varepsilon_0 = \omega^{\varepsilon_0}$ , finite rooted non-planar trees). By Gödel's second incompleteness theorem, we need more than usual recurrence!

# Part III

# Set theory

An important first order theory is set theory, which axiomatizes sets.

This means that every term of the theory should be interpreted as a set.

It was initiated in 1870's by Cantor and Dedekind.

We will need a superficial understanding of it.

We consider the signature

$$\Sigma = \{\} \qquad \qquad \mathcal{P} = \{\in : 2\}$$

together with an axiom for every formula A(x):

 $\exists y.\forall x.x \in y \Leftrightarrow A(x)$ 

the unrestricted comprehension scheme, which states the existence of

 $y = \{x \mid A(x)\}$ 

From this, we can define all the usual operations. For instance,

• the empty set:

$$\emptyset = \{x \mid \bot\}$$

• union of x and y:

 $x \cup y = \{z \mid z \in x \lor z \in y\}$ 

• ...

All good!

There is a "slight" problem: theory is inconsistent, Russell found a paradox in 1901.

Consider A(x) being  $\neg (x \in x)$ .

There exists  $y = \{x \mid A(x)\}$  such that  $\forall x.x \in y \Leftrightarrow \neg(x \in x)$ .

In particular, for x being y, we have  $y \in y \Leftrightarrow \neg(y \in y)$ .

Therefore,

- if  $y \in y$ , we have  $\neg(y \in y)$  and thus  $\bot$ , i.e. we have  $\neg(y \in y)$ ,
- if  $\neg(y \in y)$ , we have  $y \in y$  and thus  $\bot$ , i.e. we have  $\neg \neg(y \in y)$ .

Can we have an intuitionistic proof? This is an intuitionistic proof!

# Naive set theory

The core of this proof consists in showing (see TD) that

 $(A \Leftrightarrow \neg A) \Rightarrow \bot$ 

or equivalently

$$(A \Rightarrow \neg A) \Rightarrow (\neg A \Rightarrow A) \Rightarrow \bot$$

and conclude using  $A = y \in y$ .

Interestingly, the corresponding  $\lambda$ -term is

 $\lambda f.\lambda g.f(g(\lambda a.f a a))(g(\lambda a.f a a))$ 

If we set f and g to be the identity, we recover the looping term

 $\Omega = (\lambda a.a a)(\lambda a.a a)$ 

We can think of a set t as a predicate, i.e. t(u) is true when  $u \in t$ .

We are thus tempted to adopt the following notations:

instead of	we write
$u \in t$	t u
$\{x \mid t\}$	$\lambda x.t$

Namely,

 $u \in \{x \mid t(x)\} \quad \Leftrightarrow \quad (\lambda x.t)u \quad \Leftrightarrow \quad t(u)$ 

This was in fact Church initial intuition behind  $\lambda$ -calculus!

# Recovering the fixpoint combinator

Russell's set  $r = \{x \mid \neg(x \in x)\}$  can be written as

 $r = \lambda x. \neg (xx)$ 

The fact that  $r \in r \Leftrightarrow \neg (r \in r)$  translates as

 $rr = _{\beta} \neg (rr)$ 

In other words rr is a fixpoint for  $\neg$ !

Generalizing this to any f instead of  $\neg$ , we recover Church's fixpoint combinator:

 $r = \lambda x.f(xx)$   $Y = \lambda f.rr$ 

In order to avoid this paradox, Zermelo and Fraenkel proposed a new axiomatization of **set theory**, roughly between 1900 and 1925, which is now the "standard".

The idea in order to avoid paradoxes is that some collections are "too big" to be sets. In particular,

there is no set of all sets.

We now list the six axioms (some variants have different axioms, but are equivalent) on

 $\mathcal{P} = \{=: 2, \in : 2\}$ 

#### Axiom of extensionality

Two sets with the same elements are equal:

 $\forall x.\forall y.((\forall z.z \in x \Leftrightarrow z \in y) \Rightarrow x = y)$ 

In usual notation,

 $x \subseteq y \land y \subseteq x \Rightarrow x = y$ 

#### Axiom of union

The union of a family of sets exists:

$$\forall x. \exists y. \forall z. (z \in y \Leftrightarrow \exists t. (t \in x \land z \in t))$$

In usual notation, we can construct

$$y = \bigcup_{t \in x} t = \{z \mid z \in t, t \in x\}$$

Axiom of powerset

There is a set of subsets of a set:

 $\forall x.\exists y.\forall z.(z \in y \Leftrightarrow (\forall t.t \in z \Rightarrow t \in x))$ 

In usual notation,

$$\forall x.\exists y.\forall z.(z \in y \Leftrightarrow z \subseteq x)$$

i.e. we can construct

 $y = \mathcal{P}(x) = \{z \mid z \subseteq x\}$ 

### Axiom schema of replacement

The image of a set under a partial function is a set:

 $(\forall x.\forall y.\forall y'.(A(x,y) \land A(x,y') \Rightarrow y = y')) \Rightarrow \forall t.\exists u.\forall y.(y \in u \Leftrightarrow \exists x.(x \in t \land A(x,y)))$ 

This means that given a relation A(x, y) encoding a partial function we can define the set u of images of t under the partial function:

 $u = \{y \mid \exists x \in t.A(x,y)\}$ 

This is also why we restrict to *functional* relations: otherwise, for a given x,  $\{y \mid A(x, y)\}$  could be "too big" to be a set.

### Axiom of infinity

There exists a set with infinitely many elements:

 $\exists x. \emptyset \in x \land \forall y. y \in x \Rightarrow S(y) \in x$ 

with  $S(y) = y \cup \{y\}$ .

In particular, we can define  $\mathbb{N}$  as the intersection of all sets containing  $\emptyset$  and closed under *S*, with the von Newman coding of natural numbers:

 $0 = \{\} \qquad 1 = 0 \cup \{0\} = \{\{\}\} \qquad 2 = 1 \cup \{1\} = \{\{\}, \{\{\}\}\}\}$ 

etc.

### Axiom of foundation

Every non-empty set contains a member which is disjoint from the whole set:

$$\forall x.(\exists t.t \in x) \Rightarrow \exists y.(y \in x \land \neg \exists z.(z \in y \land z \in x))$$

or, in usual notation,

$$\forall x. x \neq \emptyset \Rightarrow \exists y \in x. y \cap x = \emptyset$$

Together with other axioms, this is equivalent to the principle of  $\in$ -induction:

 $(\forall x.(\forall y.y \in x \Rightarrow A(y)) \Rightarrow A(x)) \Rightarrow \forall x.A(x)$ 

*Optionally*, one can add the

### Axiom of choice

Given a collection of non-empty sets, one can construct a **choice function** which chooses an element in each of the sets:

$$\forall x. \emptyset \notin x \Rightarrow \exists (f: x \to \cup x). \forall y \in x. f(y) \in y$$

This is very natural at first.

From a constructivist point of view, the axiom of choice is difficult to accept though. It chooses for us elements in sets, but we do not know how exactly.

In particular, given a non-empty set, one can choose an element of it

 $x \neq \emptyset \Rightarrow \exists y.y \in x$ 

which is difficult to accept constructively:

 $\neg \neg \exists y. y \in x \Rightarrow \exists y. y \in x$ 

There are various formulations of the axiom of choice:

• for every set of non-empty sets there is a choice function

 $\forall x. \emptyset \notin x \Rightarrow \exists (f: x \to \cup x). \forall y \in x. f(y) \in y$ 

- every surjective function admits a section,
- every set can be well-ordered,
- . . .
A surjective function

 $f: A \rightarrow B$ 

is "the same" as a collection of subsets of B (forming a partition):

 $(f^{-1}(y))_{y\in B}$ 

A choice function for this collection is a function. We have the following alternative formulation of the axiom of choice:

 $\begin{array}{c} g:B\rightarrow A\\ every \ surjective \ function \ admits \ a \ section. \end{array}$ 

which is a *section* of f.

Note that the naive translation of the axiom of choice in constructive type theory is provable:

AC : {A B : Set} (f : A  $\rightarrow$  B)  $\rightarrow$ ((y : B)  $\rightarrow \Sigma$  A ( $\lambda$  x  $\rightarrow$  f x  $\equiv$  y))  $\rightarrow$  $\Sigma$  (B  $\rightarrow$  A) ( $\lambda$  g  $\rightarrow$  (y : B)  $\rightarrow$  f (g y)  $\equiv$  y) AC f s = ( $\lambda$  y  $\rightarrow$  proj<sub>1</sub> (s y)) , ( $\lambda$  y  $\rightarrow$  proj<sub>2</sub> (s y))

The axiom of choice would be more like a doubly negated version of that:

postulate CAC : {A B : Set} (f : A 
$$\rightarrow$$
 B)  $\rightarrow$   
 $\neg \neg$  ((y : B)  $\rightarrow \Sigma$  A ( $\lambda \ge y$ ))  $\rightarrow$   
 $\neg \neg$  ( $\Sigma$  (B  $\rightarrow$  A) ( $\lambda \ge y$  (y : B)  $\rightarrow$  f (g y)  $\equiv$  y))

although we would rather use an "intuitionistic version of double negation".

So, people started to investigate the status of AC with respect to ZF.

In 1938, Gödel, showed that AC is *consistent* with ZF by constructing a model of ZF+AC inside a model of ZF.

In 1963, Cohen showed that ZF+ $\neg$ AC is *consistent* by constructing a model using forcing.

AC is thus *independent* of ZF: we can add it or not.

If we are hardcore constructivists, we want to use intuitionistic logic, we will see that in this case we also have to give up (some variants of) AC.

The theory ZF is usually taken to be in classical logic, but it makes sense in intuitionistic logic (it is then called IZF).

In this world, things do not behave as nicely as usual, but we are constructive!

Let's investigate this.

# IZF: deciding membership

#### Given a proposition A not involving y, consider

 $x = \{y \in \mathbb{N} \mid A\}$ 

We have

 $(0 \in x) \Leftrightarrow A$ 

**Lemma** In IZF, the formula

 $\forall y.\forall x.(y \in x) \lor \neg (y \in x)$ 

is equivalent to the excluded middle.

Proof.

The right-to-left implication is immediate. For the other implication, given a formula A, take y = 0 and  $x = \{y \in \mathbb{N} \mid A\}$ : the formula is then equivalent to  $A \lor \neg A$ .

What is the intuition behind that?

Take

 $h = \{m \in \mathbb{N} \mid \text{the Turing machine } m \text{ is halting}\}$ 

then, if we had a proof of

 $\forall x \in \mathbb{N}. (x \in h) \lor \neg (x \in h)$ 

by Curry-Howard, we would have a function which to every program x indicates whether it is halting or not, which is impossible.

# IZF: deciding emptyness

#### Given a proposition A not involving y, consider

 $x = \{y \in \mathbb{N} \mid A\}$ 

We have

 $(x = \emptyset) \Leftrightarrow \neg A$ 

**Lemma** In IZF, the formula

 $\forall x.(x = \emptyset) \lor \neg (x = \emptyset)$ 

is equivalent to having

 $\neg A \lor \neg \neg A$ 

for every formula A, which does not hold in intuitionistic logic.

**Lemma** In IZF, the formula

$$\forall x.\forall y.(x=y) \lor \neg (x=y)$$

does not hold.

Proof.

We have seen that it does not hold in the particular case where  $y = \emptyset$ .

# IZF: deciding equality

This does not mean that we cannot decide equality for any set!

For instance, we define the booleans  $\mathbb B$  as

 $\mathbb{B} = \{ x \mid x = 0 \lor x = 1 \} \qquad \qquad 0 = \emptyset \qquad \qquad 1 = \{ \emptyset \}$ 

We have  $0 \neq 1$ : namely 0 = 1 would imply that  $\emptyset \in 0$  since  $\emptyset \in 1$ , thus  $\bot$ .

Every element  $x \in \mathbb{B}$  is either 0 or 1 (by definition), so that we can decide equality:

**Lemma** The formula  $\forall x \in \mathbb{B} . \forall y \in \mathbb{B} . (x = y) \lor (x \neq y)$  holds.

The same would hold for the set  $\mathbb{N}$  of natural numbers.

# IZF: deciding equality

However, given a proposition A(b), we cannot perform case analysis and say that

 $x = \{b \in \mathbb{B} \mid A(b)\}$ 

is either

 $\emptyset$  {0} {1} {0,1} (thus  $A = \emptyset$  or not).

This is because an element of  $\mathbf{x}$  is more than a boolean: it is

• a boolean *b*,

• together with a proof that A(b) holds,

and we would have to reason by case analysis on all proofs of A(b) which we can't.

#### **Baby Diaconescu**

**Theorem** *IZF+AC implies NNE.* 

**Proof.** Given a formula A, suppose  $\neg \neg A$  and consider the set

 $x = \{y \in \mathbb{N} \mid A\}$ 

We have  $\neg(x = \emptyset)$ . Namely, if  $x = \emptyset$ , then  $\neg A$ , and therefore  $\bot$  (by  $\neg \neg A$ ). By AC, there is an element  $y \in x$  and thus A.

We have used the following formulation of AC: for every family of non-empty sets there is a choice function

whereas we would rather have:

for every family of sets with an element there is a choice function

The two are classically equivalent, but not intuitionistically so.

## Diaconescu's theorem

**Theorem** *IZF+AC implies LEM.* 

#### Proof.

Fix an arbitrary proposition A: we are going to show  $A \lor \neg A$ . Consider the sets  $x = \{b \in \mathbb{B} \mid (b = 0) \lor A\}$  and  $y = \{b \in \mathbb{B} \mid (b = 1) \lor A\}$ . They have an element since  $0 \in x$  and  $1 \in y$ . By AC, there is a function  $f : \{x, y\} \to \mathbb{B}$  such that  $f(x) \in x$  and  $f(y) \in y$ . Now, f(x) and f(y) are booleans so we can reason by case analysis on those:

- if f(x) = 1 then  $1 \in x$ , thus  $(1 = 0) \lor A$  holds, thus A holds,
- if f(y) = 0 then  $0 \in y$ , thus  $(0 = 1) \lor A$  holds, thus A holds,
- if f(x) = 0 ≠ 1 = f(y), then x ≠ y (they would have the same image under f otherwise), if we suppose A then x = y = B, thus ⊥, therefore ¬A holds.

Therefore  $A \lor \neg A$ .

## Diaconescu's theorem

We have defined

 $x = \{b \in \mathbb{B} \mid (b = 0) \lor A\}$   $y = \{b \in \mathbb{B} \mid (b = 1) \lor A\}$ 

Note that even though  $0 \in x$  and  $1 \in y$ , we cannot say that

$$egin{array}{ll} f:\{A,B\} o \mathbb{B}\ &x\mapsto 0\ &y\mapsto 1 \end{array}$$

is a choice function because it would not be well defined in the case x = y.

If x = y, then  $f : x \mapsto 0$  is suitable.

But in order to use those facts to define a function, we would need to decide the equality between x and y in the first place!

Constructive mathematics are hard.

Hilbert would say

Taking the principle of excluded middle from the mathematician would be the same, say, as proscribing the telescope to the astronomer or to the boxer the use of his fists. To prohibit existence statements and the principle of excluded middle is tantamount to relinquishing the science of mathematics altogether.

Without LEM (and thus AC), we have to give away very useful results and tools (we are going to see some of those).

But they are satisfactory: when something is true, we know why.

Let's see what life in IZF looks like.

# Subsets of finite sets

A set *A* is **finite** when there exists  $n \in \mathbb{N}$  and an enumeration  $(a_i)_{1 \leq i \leq n}$  of its elements:  $A = \{a_i \mid 1 \leq i \leq n\}$ 

**Theorem** *LEM is equivalent to "every subset of a finite set is finite".* 

In the same vein, the axiom of choice is also equivalent to the **trichotomy** principle: two given sets either have the same cardinality or one has smaller cardinality than the other.

Some models of IZF satisfy the following properties:

- There is a set that can be partitioned into strictly more equivalence classes than the original set has elements, and a function whose domain is strictly smaller than its range. In fact, this is the case in all known models.
- There is an infinite set of real numbers without a countably infinite subset.
- The real numbers are a countable union of countable sets. This does not imply that the real numbers are countable: we need AC to show that a countable union of countable sets is itself countable requires the Axiom of countable choice.

Another striking property:

**Theorem** AC is equivalent to "every vector space has a basis".

In fact, we know models where

- there is a vector space with no basis,
- there is a vector space with two basis of different cardinalities.

Life without choice is difficult, but this does not necessarily means that AC is the right way to go. In fact it is suspiciously powerful.

Although ZFC is equiconsistent with ZF, we can think of it as the rule

# $\Gamma \vdash A$

which would clearly be too powerful.

Two sets *A* and *B* of points in  $\mathbb{R}^3$  are **congruent** if one can be obtained from the other by an isometry (i.e. by using translations, rotations and reflections).

**Theorem (Banach-Tarski)** Given two bounded subsets of  $\mathbb{R}^3$  of non-empty interior, there are partitions

 $A = A_1 \uplus \ldots \uplus A_n \qquad \qquad B = B_1 \uplus \ldots \uplus B_n$ 

such that each  $A_i$  is congruent to  $B_i$ . **Proof:** AC + many other things.



A last remark is that, even if we are not interested in constructivity, it is not entirely fair to think of the proofs in IZF as a subset of the proofs in ZF.

Sometimes, less is more.

The notion of infinitesimal  $\varepsilon$  is difficult to properly define in mathematics.

Typically, one would like to be able to write something like

 $f(a+\varepsilon)=f(a)+f'(a)\varepsilon$ 

for an arbitrary infinitesimal  $\varepsilon$ .

The idea is that the derivative of f is the first-order variation of f around a point.



Consider  $f(x) = x^2$ . We want to compute f'(a).

Suppose given  $\varepsilon$  which is *very* small:  $\varepsilon^2 = 0$ . We have

$$f(a + \varepsilon) = (a + \varepsilon)^{2}$$
$$= a^{2} + 2a\varepsilon + \varepsilon^{2}$$
$$= a^{2} + 2a\varepsilon$$

Therefore, we can define

f'(a) = 2a

to be the linear part.

#### Infinitesimals

This suggests that we define the set of infinitesimals as

 $D = \{ \varepsilon \in \mathbb{R} \mid \varepsilon^2 = 0 \}$ 

and postulate the principle of microaffineness:

Axiom (Kock-Lawvere) Every function  $f : D \to \mathbb{R}$  is of the form

 $f(\varepsilon) = a + b\varepsilon$ 

for some unique a and b (with a being necessarily f(0)).

We can the define f'(x) by writing

 $f(x+\varepsilon)=f(x)+b\varepsilon$ 

and defining f'(x) = b.

We can compute the usual law for deriving products:

$$(f \times g)(a + \varepsilon) = f(a + \varepsilon) \times g(a + \varepsilon)$$
  
=  $(f(a) + f'(a)\varepsilon) \times (g(a) + g'(a)\varepsilon)$   
=  $f(a)g(a) + (f'(a)g(a) + f(a)g'(a))\varepsilon + f'(a)g'(a)\varepsilon^{2}$   
=  $f(a)g(a) + (f'(a)g(a) + f(a)g'(a))\varepsilon$ 

Therefore,

$$(f \times g)'(a) = f'(a) \times g(a) + f(a) \times g'(a)$$

Similarly, we can compute the chain rule by

$$egin{array}{lll} (g\circ f)(a+arepsilon)&=g(f(a)+f'(a)arepsilon)\ &=g(f(a))+g'(f(a)) imes f'(a)arepsilon \end{array}$$

since  $f'(a)\varepsilon \in D$ .

Therefore,

 $(g \circ f)'(a) = g'(f(a)) \times f'(a)$ 

# A bug?

Oh wait, there is a "slight" problem: our axiom

Axiom Every function  $f : D \to \mathbb{R}$  is of the form  $f(\varepsilon) = a + b\varepsilon$  for some unique a and b.

is clearly wrong.

Namely, we have  $D = \{0\}$  and thus

 $f(a + \varepsilon) = f(a) + b\varepsilon$   $f(a + \varepsilon) = f(a) + c\varepsilon$ 

for any **b** and **c** since  $\varepsilon = 0$ .

Also, our axiom implies that every function is differentiable and we know that's not the way things are.

# Synthetic differential geometry

Why is it the case that  $D = \{\varepsilon \mid \varepsilon^2 = 0\} = \{0\}$ ?

Well, it's obvious: take any  $\varepsilon \in D$ , if  $\varepsilon \neq 0$  then  $\varepsilon^2 > 0$  and therefore  $\varepsilon \notin D$ .

Oh wait, we have used classical logic. Namely, in order to make this reasoning we have implicitly used

- nne: we have shown that  $\varepsilon$  cannot be non-zero, and deduced that  $\varepsilon = 0$ , or
- lem: either  $\varepsilon = 0$  or  $\varepsilon \neq 0$ , and in the second case we have a contradiction.

In intuitionistic logic, all we can show is that

 $\forall \varepsilon \in D. \neg \neg (\varepsilon = 0)$ 

In this sense, an infinitesimal is "almost 0".

Is it a field? No: since  $\varepsilon^2 = 0$ , we would have

 $\varepsilon = \varepsilon^2 / \varepsilon = 0 / \varepsilon = 0$ 

However, we still have the fact that  $x \neq 0$  implies that x is invertible, and this is not a problem because we cannot show  $\varepsilon \neq 0$ , since we have  $\neg \neg (\varepsilon = 0)$ .

# Synthetic differential geometry

The axiom implies that every function is differentiable.

But this is not the case, for instance,



which cannot be defined by

$$f(x) = egin{cases} -1 & ext{if } x < 0 \ 1 & ext{otherwise} \end{cases}$$

# The field of **synthetic differential geometry** studies differential geometry in this way, by considering infinitesimals in intuitionistic logic.

What we did not show here is that we actually have a model satisfying the axioms...

# Part IV

# Unification

#### Unification

Suppose fixed a signature  $\Sigma$ , generating a set  $\mathcal{T}$  of terms.

We are going to solve systems of equations over  $\mathcal{T}$ .

For instance, with

$$\Sigma = \{f : 1, g : 2, a : 0\}$$

consider the system of equations

 $\begin{cases} f(x) \neq f(f(a)) \\ g(x,x) \neq g(x,y) \end{cases}$ 

What are the solutions?

The unification algorithm solves such equational problems.

In particular, we have seen that we could express typing à la Curry in this way.

A substitution  $\sigma$  is a function which to some variables associates a term.

We write  $dom(\sigma)$  for its domain, i.e. the variables on which it is defined.

Given a term t with  $FV(t) \subseteq dom(\sigma)$ , we write  $\sigma(t)$  for the term t where each variable x has been replaced by  $\sigma(x)$ .

# Systems of equations

A system of equations *E* is a set

$$E = \{t_1 \neq u_1, \ldots, t_n \neq u_n\}$$

of pairs of terms.

A substitution is a **solution** (or **unifier**) of a system of equations E, if for every equation  $t_i \neq u_i$ , we have  $FV(t_i) \cup FV(u_i) \subseteq dom(\sigma)$  and  $\sigma(t_i) = \sigma(u_i)$ 

A solution of our example  $\{f(x) \neq f(f(a)), g(x, x) \neq g(x, y)\}$  is

$$\sigma(x) = f(a)$$
  $\sigma(y) = f(a)$ 

We first have to explain what we mean by "the solution":

- $f(x) \neq f(a)$  has one solution:  $x \mapsto a$ ,
- $x \neq f(y)$  has many solutions:  $[y \mapsto a, x \mapsto f(a)], [y \mapsto f(a), x \mapsto f(f(a))],$  etc.
- $f(x) \neq g(y, z)$  has no solution,
- $x \neq f(x)$  has no solution.
We are going to describe an algorithm which transforms equation systems akin to gaussian elimination:

$$\begin{cases} x + 3y = 0 \\ 2x + 8y = 2z \end{cases} \xrightarrow{\sim} \begin{cases} x + 3y = 0 \\ 2y = 2z \end{cases} \xrightarrow{\sim} \begin{cases} x + 3y = 0 \\ y = z \end{cases} \xrightarrow{\sim} \begin{cases} x + 3z = 0 \\ y = z \end{cases}$$
$$\xrightarrow{\sim} \begin{cases} x = -3z \\ y = z \end{cases}$$

# Solved form

### An equation system E is in solved form if is of the form

 $E = \{x_1 \neq t_1, \ldots, x_n \neq t_n\}$ 

where

- $x_i \neq x_j$  for  $i \neq j$ ,
- $x_i \notin FV(t_j)$  for every *i* and *j*.

The strategy is thus to transform our equation system

$$E \rightsquigarrow E_1 \rightsquigarrow E_2 \rightsquigarrow \ldots \rightsquigarrow E_n$$

in such a way that

- $E_i$  and  $E_{i+1}$  have the same solutions,
- $E_n$  is in solved form.

# Unification

The unification algorithm consists in applying the following transformations.

• Delete:

 $\{t \neq t\} \uplus E \quad \rightsquigarrow \quad E$ 

• Decompose:

 $\{f(t_1,\ldots,t_n) \neq f(u_1,\ldots,u_n)\} \uplus E \quad \rightsquigarrow \quad \{t_1 \neq u_1,\ldots,t_n \neq u_n\} \uplus E$ 

• **Orient**: when *t* is not a variable

 $\{t \neq x\} \uplus E \quad \rightsquigarrow \quad \{x \neq t\} \uplus E$ 

• Eliminate: if  $x \in FV(E) \setminus FV(t)$ ,

 $\{x \neq t\} \uplus E \quad \rightsquigarrow \quad \{x \neq t\} \uplus \sigma(E)$ <sup>99</sup>

In addition, the algorithm will fail on the following cases:

• Clash: for  $f \neq g$ ,

$$\{f(t_1,\ldots,t_n) \neq g(u_1,\ldots,u_m)\} \uplus E \quad \rightsquigarrow \quad fail$$

• Occurs-check: for  $x \in FV(t)$ ,

$$\{x \neq t\} \uplus E \quad \rightsquigarrow \quad fail$$

An example

Let's execute our algorithm on

$$\{x \neq f(a), g(x, x) \neq g(x, y)\}$$

$$\Rightarrow \{x \neq f(a), g(f(a), f(a)) \neq g(f(a), y)\}$$
 by Eliminate
$$\Rightarrow \{x \neq f(a), f(a) \neq f(a), f(a) \neq y\}$$
 by Decompose
$$\Rightarrow \{x \neq f(a), f(a) \neq y\}$$
 by Delete
$$\Rightarrow \{x \neq f(a), y \neq f(a)\}$$
 by Orient

A solution is thus

$$\sigma(x) = f(a)$$
  $\sigma(y) = f(a)$ 

The unification algorithm consists, starting from E, in

- applying a transformation when one applies,
- when no transformation applies anymore and we have not failed, we return the corresponding substitution.

Note that

- we have to show that our transformations preserve the solutions
- we have to show that if the equation system has a solution then we have a normal form when we stop (without failing)
- the algorithm is not deterministic so that it is not clear that it will always give the same solution
- an equation system might have multiple solutions so what is the status of the solution provided by unification?
- we have to show that the algorithm terminates

**Lemma** If  $E \rightsquigarrow E'$  then E and E' admit the same solutions.

**Proof.** By inspection of the rules.

- Delete:  $\{t \neq t\} \uplus E \rightsquigarrow E$
- Decompose:  $\{f(t_1,\ldots,t_n) \neq f(u_1,\ldots,u_n)\} \uplus E \rightsquigarrow \{t_1 \neq u_1,\ldots,t_n \neq u_n\} \uplus E$
- Orient:  $\{t \neq x\} \uplus E \rightsquigarrow \{x \neq t\} \uplus E$  when t is not a variable
- Eliminate:  $\{x \neq t\} \uplus E \rightsquigarrow \{x \neq t\} \uplus \sigma(E) \text{ if } x \in FV(E) \setminus FV(t)$

**Lemma** If  $E \nleftrightarrow$  then E is in solved form.

**Proof.** If E is not in solved form then either

- it contains some  $f(t_1, \ldots, t_n) \neq u$  with either
  - u = x: orient, or
  - $u = f(t_1, \ldots, t_n)$ : delete, or
  - $u = f(u_1, \ldots, u_n)$ : decompose, or
  - $u = g(u_1, \ldots, u_m)$ : clash, or
- there is some  $x_i \neq t_i$  with  $x_i \in FV(t_j)$ : eliminate / occurs-check.

# Termination

How do we show the termination of the algorithm?

We could associate a natural number  $n^E$  to each system of equations such that

$$E \rightsquigarrow E'$$
 implies  $n^E > n^{E'}$ 

It turns out that it will be more convenient to associate a triple  $(n_1^E, n_2^E, n_3^3)$  of natural numbers such that

 $E \rightsquigarrow E'$  implies  $(n_1^E, n_2^E, n_3^E) > (n_1^{E'}, n_2^{E'}, n_3^{E'})$ where  $(m_1, m_2, m_3) > (n_1, n_2, n_3)$  when  $m_1 > n_1$  or  $m_1 = n_1$  and  $m_2 > n_2$  or  $m_1 = n_1$  and  $m_2 = n_2$  and  $m_3 > n_3$ (lexicographic order).

106

# The lexicographic order

We order the set  $\mathbb{N} \times \mathbb{N}$  by  $(m_1, m_2) > (n_1, n_2)$  whenever

```
(a) m_1 > n_1, or
(b) m_1 = n_1 + n_2
```

(b)  $m_1 = n_1$  and  $m_2 > n_2$ .

A decreasing sequence is

 $(5,10) > (4,8) > (3,18) > (3,15) > (2,40) > \dots$ 

#### Proposition

There is no infinite decreasing sequence in  $\mathbb{N} \times \mathbb{N}$ .

#### Proof.

In the sequence there is an infinite number of (a) or of (b).

- (a) This is impossible because the first  $\mathbb{N}$  is noetherian.
- (b) After the last (a), there is an infinite sequence of (b) which is impossible because the second N is noetherian.

We need some terminology.

A variable is **solved** in E when it occurs exactly once, as a left member of an equation.

The size of a term is the number of function symbols occurring in it:

|f(g(a(),x),y)|=3

The size of an equation system is the sum of the size of the terms in a left or right member of an equation.

# Termination

# Theorem

The unification algorithm is terminating for all inputs.

**Proof.** To every equation system *E*, we associate:

- $n_1$ : the number of variables in E which are not solved,
- $n_2$ : the size of E,
- $n_3$ : the number of equations of the form  $t \neq x$  in E.

We then have

 $\begin{array}{ccc} n_1 & n_2 & n_3 \\ \text{Delete} & \geqslant & > \\ \text{Decompose} & \geqslant & > \\ \text{Orient} & \geqslant & = & > \\ \text{Eliminate} & > \end{array}$ 

Recall that  $\sigma$  is a partial function from variables to terms.

By convention, we can suppose that it is a total function by declaring that

 $\sigma(x) = x$ 

for  $x \notin \operatorname{dom}(\sigma)$ .

Given two substitution  $\sigma$  and  $\tau$ , we have a **composite** substitution defined by

 $(\tau \circ \sigma)(x) = \tau(\sigma(x))$ 

A substitution  $\sigma$  is a  $\mathbf{renaming}$  when

- $\sigma(x)$  is a variable for every variable x,
- $\sigma(x) = \sigma(y)$  implies x = y.

In particular, we have the identity substitution id defined by

id(x) = x

It can be noticed that if  $\sigma$  is a solution of E and  $\tau$  is an arbitrary substitution then  $\tau \circ \sigma$  is also a solution of E.

We say that  $\sigma$  is more general than  $\sigma'$ 

 $\sigma \preceq \sigma'$ 

when there exists  $\tau$  such that  $\sigma' = \tau \circ \sigma$ .

# **Ordering unifiers**

**Lemma** The relation  $\leq$  is a preorder.

**Proof.** We have

- *Reflexivity*. We have  $\sigma \preceq \sigma$  since  $\sigma = id \circ \sigma$ .
- Transitivity. Suppose  $\sigma \preceq \sigma' \preceq \sigma''$ . We have

$$\sigma' = \tau \circ \sigma \qquad \text{and} \qquad \sigma'' = \tau' \circ \sigma'$$

therefore

$$\sigma'' = \tau' \circ \sigma' = \tau' \circ (\tau \circ \sigma) = (\tau' \circ \tau) \circ \sigma$$

ls <u>≺</u> antisymmetric? No.

Take  $\sigma(x) = y$  and  $\sigma(y) = x$ . We have

- id  $\leq \sigma$  since  $\sigma = \sigma \circ id$ ,
- $\sigma \preceq \text{id since id} = \sigma \circ \sigma$ .

#### Lemma

Two substitutions  $\sigma$  and  $\tau$  are such that  $\sigma \leq \tau$  and  $\tau \leq \sigma$  if and only if they differ by renamings only.

## A most general unifier for *E* is a solution which is minimal wrt $\leq$ .

**Theorem** *The unification algorithm computes a most general unifier.* 

# A last example

For instance,

 $\{g(x,y) \neq g(y,f(z))\} \quad \rightsquigarrow \quad \{x \neq y, y \neq f(z)\} \quad \rightsquigarrow \quad \{x \neq f(z), y \neq f(z)\}$ 

The most general unifier is thus

 $\sigma(x) = f(z)$   $\sigma(y) = f(z)$ 

Another most general unifier is thus

$$\sigma(x) = f(x)$$
  $\sigma(y) = f(x)$ 

A non-minimal solution is

 $\sigma(x) = f(f(a)) \qquad \qquad \sigma(y) = f(f(a))$