instance to arrive at a valid derivation. While parameters are always local to a subderivation, we consider existential variables to be global in a derivation.[4]

The second concept we need is that of a *substitution* for existential variables. We use a new notation, because this form of substitution is quite different from substitutions for bound variables x or parameters a.

$$\textit{Substitutions} \quad \theta \quad ::= \quad \cdot \mid \theta, X \mapsto t$$

We require that all variables X defined by a substitution are distinct. We write $\mathrm{dom}(\theta)$ for the variables defined by a substitution and $\mathrm{cod}(\theta)$ for all the variables occuring in the terms t. For a ground substitution $\mathrm{cod}(\theta)$ is empty. For the technical development it is convenient to assume that the domain and co-domain of a substitution share no variables. This rules out "circular" substitutions such as $X \mapsto f(X)$ and it also disallows identity substitutions $X \mapsto X$. The latter restriction can be dropped, but it does no harm and is closer to the implementation. As for contexts, we consider the order of the definitions in a substitution to be irrelevant.

We write $t[\theta]$, $A[\theta]$, and $\Gamma[\theta]$ for the application of a substitution to a term, proposition, or context. This is defined to be the identity on existential variables which are not explicitly defined in the substitution.

We also need an operation of composition, written as $\theta_1 \circ \theta_2$ with the property that $t[\theta_1 \circ \theta_2] = (t[\theta_1])[\theta_2]$ and similarly for propositions and contexts. Composition is defined by

$$\begin{aligned} (\cdot) \circ \theta_2 &= \theta_2 \\ (\theta_1, X \mapsto t) \circ \theta_2 &= (\theta_1 \circ \theta_2), X \mapsto t[\theta_2] \end{aligned}$$

In order for composition to be well-defined and have the desired properties we require that $\mathrm{dom}(\theta_1)$, $\mathrm{dom}(\theta_2)$ and $\mathrm{cod}(\theta_2)$ are disjoint, but of course variables in the co-domain of θ_1 can be defined by θ_2.

Now we introduce the judgment which implicitly defines an algorithm for unification. We write

$$\models F \;/\; \theta \quad \theta \text{ is a most general unifier for } F.$$

The intent (to be proven later) is that if $\models F \;/\; \theta$ then $\models F[\theta]$, which means that θ is a *unifier* for F. Moreover, we show that whenever $\models F[\theta']$ then there exists a substitution θ'' such that $\theta' = \theta \circ \theta''$, which means that θ is a *most general unifier* for F (any unifier is an instance of θ).

Conjunction and Truth. Algorithmically, we impose a left-to-right order on the solution of F_1 and F_2, but this just fixes a don't care non-deterministic choice.

$$\frac{\models F_1 \;/\; \theta_1 \qquad \models F_2[\theta_1] \;/\; \theta_2}{\models F_1 \wedge F_2 \;/\; \theta_1 \circ \theta_2}\wedge I \qquad \frac{}{\models \top \;/\; \cdot}\top I$$

After all the rules have been shown, it will be easy to see that the side conditions on composition are satisfied and $\theta_1 \circ \theta_2$ is well-defined.

[4] *[example]*

Existential Quantification. Existential variables are introduced for existential quantifiers. They must be "new" (even though the judgment is not parametric). Because of the way existential variables are global to a derivation, this freshness requirement is a global requirement: in a complete derivation, the existential variables chosen for all existential quantifiers must be distinct. To be completely formal about this condition would require to thread a list of existential variables or a counter through a derivation. We will dispense with this complication here. We define $(\theta', X \mapsto t) - X = \theta'$ and $\theta - X = \theta$ if X is not in the domain of θ.

$$\frac{\models [X/x]F \;/\; \theta \quad X \text{ globally new}}{\models \exists x.\, F \;/\; \theta - X}\,\exists\mathrm{I}$$

Despite the strong freshness requirement on X, the derivation of the premise is not parametric in X. That is, we cannot substitute an arbitrary term t for X in a derivation of the premiss and obtain a valid derivation, since the vr, rv, vv$\neq$, and vv$=$ rules below require one or both sides of the equation to be an existential variable. Substituting for such a variable invalidates the application of these rules. Moreover X can still appear in the co-domain of θ in the generated substitution.

Predicate and Function Constants. An equation between the same function constant applied to arguments is decomposed into equations between the arguments. Unification fails if different function symbols are compared, but this is only indirectly reflected by an absence of an appropriate rule. Failure can also be explicitly incorporated in the algorithm (see Exercise **??**).

$$\frac{\models t_1 \doteq s_1 \wedge \cdots \wedge t_n \doteq s_n \;/\; \theta}{\models p(t_1, \ldots, t_n) \doteq p(s_1, \ldots, s_n) \;/\; \theta}\,\mathrm{pp} \qquad \frac{\models t_1 \doteq s_1 \wedge \cdots \wedge t_n \doteq s_n \;/\; \theta}{\models f(t_1, \ldots, t_n) \doteq f(s_1, \ldots, s_n) \;/\; \theta}\,\mathrm{rr}$$

These rules violate orthogonality by relying on conjunction in the premises for the sake of conciseness of the presentation. We could avoid this by introducing a separate judgment for the unification of lists of terms. When f or p have no arguments, the empty conjunction in the premise should be read as $\top$.

Existential Variables. There are four rules for variables. We write r for terms of the form $f(t_1, \ldots, t_n)$. Existential variables always range over terms (and not propositions), so we do not need rules for equations of the form $X \doteq P$ or $P \doteq X$.

$$\frac{X \text{ not in } r}{\models X \doteq r \;/\; (X \mapsto r)}\,\mathrm{vr} \qquad \frac{X \text{ not in } r}{\models r \doteq X \;/\; (X \mapsto r)}\,\mathrm{rv}$$

These two rules come with the proviso that the existential variable X does not occur in the term r. This is necessary to ensure that the substitution $X \mapsto r$ is indeed a unifier. Otherwise unification fails and we can recognize formulas such as $\exists x.\, x \doteq f(x)$ as false. This leaves equations of the form $X \doteq Y$ between two existential variables.

$$\frac{Y \neq X}{\models X \doteq Y \,/\, (X \mapsto Y)}\,\mathrm{vv}{\neq} \qquad \frac{}{\models X \doteq X \,/\, \cdot}\,\mathrm{vv}{=}$$

We now explore the soundness and completeness of these rules, and then analyze the rules as the basis of an algorithm. In the statement of the properties below we take some care so that for a judgment $\models F$, F never contains free existential variables which are seen only as part of the algorithm, not the definition of the logic. We write σ for ground substitutions. We say σ *grounds* a formula F if $F[\sigma]$ contains no existential variables. We assume that there is a closed term to be substituted for each variable. This assumption is necessary, for example, to prove that $\exists x.\, x \doteq x$ is valid.

Theorem 4.12 (Soundness of Unification)
If $\models F \,/\, \theta$ *then for any substitution* σ *which grounds* $F[\theta]$ *we have* $\models (F[\theta])[\sigma]$.

Proof: By induction on the structure of the derivation $\mathcal{U}$ for $\models F \,/\, \theta$.

Case:

$$\mathcal{U} = \frac{}{\models \top \,/\, \cdot}\,\top\mathrm{I}$$

$\top[\cdot][\sigma] = \top$	By definition of substitution
$\models \top$	By rule $\top$I

Case:

$$\mathcal{U} = \frac{\begin{array}{c}\mathcal{U}_1 \\ \models F_1 \,/\, \theta_1\end{array} \qquad \begin{array}{c}\mathcal{U}_2 \\ \models F_2[\theta_1] \,/\, \theta_2\end{array}}{\models F_1 \wedge F_2 \,/\, \theta_1 \circ \theta_2}\,\wedge\mathrm{I}$$

σ grounds $(F_1 \wedge F_2)[\theta_1 \circ \theta_2]$	Assumption
$\theta_2 \circ \sigma$ grounds $F_1[\theta_1]$	By properties of substitution
$\models (F_1[\theta_1])[\theta_2 \circ \sigma]$	By i.h. on $\mathcal{U}_1$
σ grounds $(F_2[\theta_1])[\theta_2]$	By properties of substitution
$\models (F_2[\theta_1][\theta_2])[\sigma]$	By i.h. on $\mathcal{U}_2$
$\models (F_1 \wedge F_2)[\theta_1 \circ \theta_2][\sigma]$	By rule $\wedge$I and properties of substitution

Case:

$$\mathcal{U} = \frac{\begin{array}{c}\mathcal{U}_1 \\ \models [X/x]F_1 \,/\, \theta'\end{array}}{\models \exists x.\, F_1 \,/\, \theta' - X}\,\exists\mathrm{I}$$

In this case we distinguish subcases, depending on whether X is in the domain of θ'.

Subcase: $\theta' = \theta, X \mapsto t$. Recall that we assume that there are ground terms.

σ grounds $(\exists x.\ F_1)[\theta]$ — Assumption
$\sigma \circ \sigma'$ grounds $([X/x]F_1)[\theta, X \mapsto t]$
for any appropriate σ' which grounds $t[\sigma]$
$\models ([X/x]F_1)[\theta, X \mapsto t][\sigma \circ \sigma']$ — By i.h. on $\mathcal{U}_\infty$
$\models [t[\sigma \circ \sigma']/x](F_1[\theta][\sigma])$ — By properties of substitution
$\models (\exists x.\ F_1)[\theta][\sigma]$ — By rule $\exists$I and properties of substitution

Subcase: $\theta' = \theta$ and contains no binding for X. Then we proceed as in the previous subcase, using $\sigma' = X \mapsto t'$ for some arbitrary ground term t'.

Case:

$$\mathcal{U} = \frac{X \text{ not in } r}{\models X \doteq r \ / \ (X \mapsto r)}\text{vr}$$

σ grounds $(X \doteq r)[X \mapsto r]$ — Assumption
$(X \doteq r)[X \mapsto r] = (r \doteq r)$ — By precondition for rule
$\models (X \doteq r)[X \mapsto r][\sigma]$ — By rule $\doteq$I

Case: Rule rv for $r \doteq X$ is symmetric.

Case: Rule vv$\neq$ for $X \doteq X$ is similar.

Case: Rule vv= for $X \doteq X$ and $X \doteq Y$ for $X \neq Y$ is also similar.

Case:

$$\mathcal{U} = \frac{\begin{array}{c}\mathcal{U}_1 \\ \models s_1 = t_1 \wedge \ldots \wedge s_n = t_n \ / \ \theta\end{array}}{\models f(s_1, \ldots, s_n) = f(t_1, \ldots, t_n) \ / \ \theta}\text{rr}$$

σ grounds $s_1[\theta], \ldots, s_n[\theta]$ and $t_1[\theta], \ldots, t_n[\theta]$ — From assumption
$\models (s_1 \doteq t_1 \wedge \ldots \wedge s_n \doteq t_n)[\theta][\sigma]$ — By i.h. on $\mathcal{U}_1$
$s_1[\theta][\sigma] = t_1[\theta][\sigma], \ldots, s_n[\theta][\sigma] = t_n[\theta][\sigma]$ — By inversion
$\models f(s_1, \ldots, s_n) \doteq f(t_1, \ldots, t_n)[\theta][\sigma]$ — By rule $\doteq$I

□

For the completeness theorem, it is convenient to introduce an intermediate system in which equality is structural, rather than allowing general equality of the form $t \doteq t$ as axioms. This is because there is a mismatch between the algorithm and the specification in that the algorithm analyzes and equality layer-by-layer, while the specification proves $\models t \doteq t$ all at once. We write $\models^{-} F$ for a system which contains all the rules in $\models F$, except that $\doteq$I is replaced by the following family of rules (one for each constant, function symbol, or predicate symbol h).

$$\frac{\models^{=} s_1 \doteq t_1 \wedge \ldots \wedge s_n \doteq t_n}{\models^{=} h(s_1, \ldots, s_n) \doteq h(t_1, \ldots, t_n)} \doteq\mathrm{I}_h$$

Again, as in the case of the unification algorighm, we take the empty conjunction (for $n = 0$) to be $\top$. In the extension with parameters, we would have a separate rule for equality between parameters.

Lemma 4.13 (Structural Equality)
$\models F$ if and only if $\models^{=} F$.

Proof: See Exercise 4.3. □

Using this property, we can show completeness for structural equality and easily lift it to the general system. In this completeness lemma we need to show that the solution found by the unification algorithm is minimally committed, that is, any actual ground solution is an instance of the solution found by the algorithm.

Lemma 4.14 (Completeness Lemma for Unification)
If $\models^{=} F[\sigma]$ for a substitution σ which grounds F, then $\models F \;/\; \theta$ for some θ and $\sigma = \theta \circ \sigma'$ for some σ'.

Proof: By induction on the structure of derivation $\mathcal{E}$ of $\models^{=} F[\sigma]$. We show some of the cases.

Case:

$$\mathcal{E} = \frac{}{\models^{=} \top[\sigma]} \top\mathrm{I}$$

$\models \top \;/\; \cdot$	By rule $\top$I
$\sigma = \cdot \circ \sigma$	By definition of composition
$\theta = \cdot$ and $\sigma' = \sigma$ satisfy the claim	

Case:

$$\mathcal{E} = \frac{\begin{array}{c}\mathcal{E}_1\\ \models^{=} F_1[\sigma]\end{array} \qquad \begin{array}{c}\mathcal{E}_2\\ \models^{=} F_2[\sigma]\end{array}}{\models^{=} (F_1 \wedge F_2)[\sigma]} \wedge\mathrm{I}$$

$\models F_1 \;/\; \theta_1$ and $\sigma = \theta_1 \circ \sigma_1$ for some θ_1 and σ_1	By i.h. on $\mathcal{E}_1$
$F_2[\sigma] = (F_2[\theta_1])[\sigma_1]$ where σ_1 grounds $F_2[\theta_1]$	By props. of substitution
$\models F_2[\theta_1] \;/\; \theta_2$ and $\sigma_1 = \theta_2 \circ \sigma_2$ for some θ_2 and σ_2	By i.h. on $\mathcal{E}_2$
$\models (F_1 \wedge F_1) \;/\; \theta_1 \circ \theta_2$	By rule $\wedge$I
$\sigma = \theta_1 \circ \sigma_1 = \theta_1 \circ (\theta_2 \circ \sigma_2) = (\theta_1 \circ \theta_2) \circ \sigma_2$	By above
$\theta = \theta_1 \circ \theta_2$ and $\sigma' = \sigma_2$ satisfy the claim	

Case:

$$\mathcal{E} = \cfrac{\begin{matrix}\mathcal{E}_1 \\ \models^{-} [t/x]F_1[\sigma]\end{matrix}}{\models^{-} (\exists x.\ f_1)[\sigma]}\ \exists\mathrm{I}$$

$[t/x]F_1[\sigma] = ([X/x]F_1)[\sigma, X \mapsto t]$ for a fresh X By definition of substitution
$\models [X/x]F_1 \ / \ \theta_1$ and $\sigma, X \mapsto t = \theta_1 \circ \sigma_1$
for some θ_1 and σ_1 By i.h. on $\mathcal{E}_1$
$\models \exists x.\ F_1 \ / \ \theta_1 - X$ By rule $\exists$I

Now we distinguish two subcases: X is in the domain of θ_1 and X is not.

Subcase: $\theta_1 = \theta_1', X \mapsto t'$.

$\sigma, X \mapsto t = (\theta_1', X \mapsto t') \circ \sigma_1$
$= (\theta_1' \circ \sigma_1), X \mapsto t'[\sigma_1]$ By definition of substitution
$\theta = \theta_1'$ and $\sigma' = \sigma_1$ satisfy the claim

Subcase: X is not in the domain of θ_1.

$\sigma, X \mapsto t = \theta_1 \circ \sigma_1$ By above
$\sigma_1 = \sigma_1', X \mapsto t$ Since $X \notin \mathrm{dom}(\theta_1)$
$\theta = \theta_1$ and σ_1' satisfy the claim

Case:

$$\mathcal{E} = \cfrac{\begin{matrix}\mathcal{E}_1 \\ \models^{-} s_1' \doteq t_1' \wedge \ldots \wedge s_n' \doteq t_n'\end{matrix}}{\models^{-} (s \doteq t)[\sigma]}\ \doteq\mathrm{I}_f$$

where $s[\sigma] = f(s_1', \ldots, s_n')$ and $t[\sigma] = f(t_1', \ldots, t_n')$. Here we distinguish several subcases, depending on the structure of s and t.

Subcase: $s = t = X$ for some existential variable X.

$\models X \doteq X \ / \ \cdot$ By rule vv=
$\sigma = \cdot \circ \sigma$ By definition of substitition
$\theta = \cdot$ and $\sigma' = \sigma$ satisfy the claim

Subcase: $s = X$ and $t = Y$ for $X \neq Y$.

$X[\sigma] = Y[\sigma]$ By Lemma 4.13
$\sigma = \sigma_1, X \mapsto X[\sigma], Y \mapsto Y[\sigma]$ By definition of substitution
$\models X \doteq Y \ / \ X \mapsto Y$ By rule vv$\neq$
$\sigma = (X \mapsto Y) \circ (\sigma_1, Y \mapsto Y[\sigma])$ By definition of composition
$\theta = X \mapsto Y$ and $\sigma' = \sigma_1, Y \mapsto Y[\sigma]$ satisfy claim

Subcase: $s = X$ and $r = f(t_1, \ldots, t_n)$.

$X[\sigma] = r[\sigma]$ By Lemma 4.13
X not in r From above
$\models X \doteq r \;/\; X \mapsto r$ By rule vr
$\sigma = \sigma_1, X \mapsto r[\sigma_1]$ By properties of substitution
$\sigma = (X \mapsto r) \circ \sigma_1$ By definition of composition
$\theta = X \mapsto r$ and $\sigma = \sigma_1$ satisfy the claim

Subcase: $s = f(s_1, \ldots, s_n)$ and $r = Y$. Symmetric to previous case.

Subcase: $s = f(s_1, \ldots, s_n)$ and $t = f(t_1, \ldots, t_n)$.

$\models s_1 \doteq t_1 \wedge \ldots \wedge s_n \doteq t_n \;/\; \theta$ and
$\sigma = \theta \circ \sigma'$ for some θ and σ' By i.h. on $\mathcal{E}_1$
$\models f(s_1, \ldots, s_n) \doteq f(t_1, \ldots, t_n) \;/\; \theta$ By rule rr
θ and σ' satisfy the claim

□

Theorem 4.15 (Completeness of Unification)
If $\models F$ (where F contains no existential variables) then $\models F \;/\; \cdot$.

Proof: By Lemma 4.13 we can reduce the problem to structural equality. On this, we apply Lemma 4.14 using the empty substitution for σ. □

Soundness and completeness of the deductive system for unification are not enough to yield all the desired properties of an implementation for unification. We also need to verify that, when given an F it always terminates, either failing or returning a substitution θ. By the properties above we know that failure will mean that there is no unifier, and that θ is indeed a most general unifier.

Termination is not a trivial property of the system for unification. This is because in the rule for conjunction of $F_1 \wedge F_2$ we apply unification to $F_2[\theta_1]$ where θ_1 is a unifier for F_1. In general, $F_2[\theta_1]$ could be a much bigger than F_2, so a simple termination argument based on the structure of the formula F will fail. If no special measures are taken, an implementation would be exponential, since the textual size of the unifier of two terms may be exponential in the size of the inputs. Consider, for example,

$$\exists x_1.\, \exists x_2.\, \exists x_2.\, f(x_1, x_2, x_3) \doteq f(g(x_2, x_2), g(x_3, x_3), g(x_4, x_4))$$

In practice, this rather unfortunate complexity bound does not seem to be much of a problem. Most implementations are straightforward, since the size of the terms in realistic theorem proving problems tends to remain relatively small. Moreover, the practically expensive part of unification (the occurs-check in the vr and rv rules) can largely be compiled away and does not need to be carried out very often.

For the proof of termination of unification, see Exercise 4.4.

4.4 Exercises

Exercise 4.1 Give an alternative proof of the inversion properties (Theorem 4.1) which does not use induction, but instead relies on admissibility of cut in the sequent calculus (Theorem 3.11).

Exercise 4.2 Formulate one or several cut rules directly on inversion sequents as presented in Section 4.1 and prove that they are admissible. Does this simplify the development of the completeness result for inversion proofs? Show how admissibility might be used, or illustrate why it is not much help.

Exercise 4.3 Prove Lemma 4.13.

Exercise 4.4 Prove that the rules for unification read as an algorithm for computing θ from F always terminate, either with failure to construct a derivation (in which case there is no unifier) or with a θ (in which case θ is a most general unifier).