Equational Logic

From now on First-order Logic is considered with equality. In this chapter, I investigate properties of a set of unit equations. For a set of unit equations I write E.

Full first-order clauses with equality are studied in the chapter on first-order superposition with equality. I recall certain definitions from Section 1.6 and Chapter 3.

The main reasoning problem considered in this chapter is given a set of unit equations E and an additional equation $s \approx t$, does $E \models s \approx t$ hold?

As usual, all variables are implicitely universally quantified. The idea is to turn the equations E into a convergent term rewrite system (TRS) R such that the above problem can be solved by checking identity of the respective normal forms: $s \downarrow_R = t \downarrow_R$.

Showing $E \models s \approx t$ is as difficult as proving validity of any first-order formula, see the section on complexity.

4.0.1 Definition (Equivalence Relation, Congruence Relation)

An *equivalence* relation $\sim$ on a term set $T(\Sigma, \mathcal{X})$ is a reflexive, transitive, symmetric binary relation on $T(\Sigma, \mathcal{X})$ such that if $s \sim t$ then $\mathrm{sort}(s) = \mathrm{sort}(t)$.
Two terms s and t are called *equivalent*, if $s \sim t$.
An equivalence $\sim$ is called a *congruence* if $s \sim t$ implies $u[s] \sim u[t]$, for all terms $s, t, u \in T(\Sigma, \mathcal{X})$. Given a term $t \in T(\Sigma, \mathcal{X})$, the set of all terms equivalent to t is called the *equivalence class of t by* $\sim$, denoted by $[t]_\sim := \{t' \in T(\Sigma, \mathcal{X}) \mid t' \sim t\}$.

If the matter of discussion does not depend on a particular equivalence relation or it is unambiguously known from the context, $[t]$ is used instead of $[t]_\sim$. The above definition is equivalent to Definition 3.2.3.

The set of all equivalence classes in $T(\Sigma, \mathcal{X})$ defined by the equivalence relation is called a *quotient by* $\sim$, denoted by $T(\Sigma, \mathcal{X})|_\sim := \{[t] \mid t \in T(\Sigma, \mathcal{X})\}$. Let E be a set of equations then $\sim_E$ denotes the smallest congruence relation “containing” E, that is, $(l \approx r) \in E$ implies $l \sim_E r$. The equivalence class $[t]_{\sim_E}$ of a term t by the equivalence (congruence) $\sim_E$ is usually denoted, for short, by $[t]_E$. Likewise, $T(\Sigma, \mathcal{X})|_E$ is used for the quotient $T(\Sigma, \mathcal{X})|_{\sim_E}$ of $T(\Sigma, \mathcal{X})$ by the equivalence (congruence) $\sim_E$.

4.1.1 Definition (Rewrite Rule, Term Rewrite System)

A *rewrite rule* is an equation $l \approx r$ between two terms l and r so that l is not a variable and $vars(l) \supseteq vars(r)$. A *term rewrite system R*, or a TRS for short, is a set of rewrite rules.

4.1.2 Definition (Rewrite Relation)

Let E be a set of (implicitly universally quantified) equations, i.e., unit clauses containing exactly one positive equation. The *rewrite relation* $\rightarrow_E \subseteq T(\Sigma, \mathcal{X}) \times T(\Sigma, \mathcal{X})$ is defined by

$$\begin{array}{lll} s \rightarrow_E t & \text{iff} & \text{there exist } (l \approx r) \in E, p \in pos(s), \\ & & \text{and matcher } \sigma, \text{ so that } s|_p = l\sigma \text{ and } t = s[r\sigma]_p. \end{array}$$

Note that in particular for any equation $l \approx r \in E$ it holds $l \to_E r$, so the equation can also be written $l \to r \in E$.

Often $s = t\downarrow_R$ is written to denote that s is a normal form of t with respect to the rewrite relation $\to_R$. Notions $\to_R^0, \to_R^+, \to_R^*, \leftrightarrow_R^*$, etc. are defined accordingly, see Section 1.6.

An instance of the left-hand side of an equation is called a *redex* (reducible expression). *Contracting* a redex means replacing it with the corresponding instance of the right-hand side of the rule.

A term rewrite system R is called *convergent* if the rewrite relation $\rightarrow_R$ is confluent and terminating. A set of equations E or a TRS R is terminating if the rewrite relation $\rightarrow_E$ or $\rightarrow_R$ has this property. Furthermore, if E is terminating then it is a TRS.

A rewrite system is called *right-reduced* if for all rewrite rules $l \rightarrow r$ in R, the term r is irreducible by R. A rewrite system R is called *left-reduced* if for all rewrite rules $l \rightarrow r$ in R, the term l is irreducible by $R \setminus \{l \rightarrow r\}$. A rewrite system is called *reduced* if it is left- and right-reduced.

4.1.3 Lemma (Left-Reduced TRS)

Left-reduced terminating rewrite systems are convergent. Convergent rewrite systems define unique normal forms.

4.1.4 Lemma (TRS Termination)

A rewrite system R terminates iff there exists a reduction ordering $\succ$ so that $l \succ r$, for each rule $l \to r$ in R.

Let E be a set of universally quantified equations. A model $\mathcal{A}$ of E is also called an *E-algebra*. If $E \models \forall \vec{x}(s \approx t)$, i.e., $\forall \vec{x}(s \approx t)$ is valid in all E-algebras, this is also denoted with $s \approx_E t$. The goal is to use the rewrite relation $\rightarrow_E$ to express the semantic consequence relation syntactically: $s \approx_E t$ if and only if $s \leftrightarrow_E^* t$.

Let E be a set of (well-sorted) equations over $T(\Sigma, \mathcal{X})$ where all variables are implicitly universally quantified. The following inference system allows to derive consequences of E:

Reflexivity $E \Rightarrow_{\mathsf{E}} E \cup \{t \approx t\}$

Symmetry $E \uplus \{t \approx t'\} \Rightarrow_{\mathsf{E}} E \cup \{t \approx t'\} \cup \{t' \approx t\}$

Transitivity $E \uplus \{t \approx t', t' \approx t''\} \Rightarrow_{\mathsf{E}}$
$E \cup \{t \approx t', t' \approx t''\} \cup \{t \approx t''\}$

Congruence $E \uplus \{t_1 \approx t'_1, \ldots, t_n \approx t'_n\} \;\Rightarrow_{\mathsf{E}}$
$E \cup \{t_1 \approx t'_1, \ldots, t_n \approx t'_n\} \cup \{f(t_1, \ldots, t_n) \approx f(t'_1, \ldots, t'_n)\}$
for any function $f : \mathrm{sort}(t_1) \times \ldots \times \mathrm{sort}(t_n) \to S$ for some S

Instance $E \uplus \{t \approx t'\} \;\Rightarrow_{\mathsf{E}}\; E \cup \{t \approx t'\} \cup \{t\sigma \approx t'\sigma\}$
for any well-sorted substitution σ

4.1.5 Lemma (Equivalence of $\leftrightarrow_E^*$ and $\Rightarrow_E^*$)

The following properties are equivalent:

1. $s \leftrightarrow_E^* t$
2. $E \Rightarrow_E^* s \approx t$ is derivable.

where $E \Rightarrow_E^* s \approx t$ is an abbreviation for $E \Rightarrow_E^* E'$ and $s \approx t \in E'$.

4.1.6 Corollary (Convergence of E)

If a set of equations E is convergent then $s \approx_E t$ if and only if $s \leftrightarrow^* t$ if and only if $s \downarrow_E = t \downarrow_E$.

4.1.7 Corollary (Decidability of $\approx_E$)

If a set of equations E is finite and convergent then $\approx_E$ is decidable.

The above Lemma 4.1.5 shows equivalence of the syntactically defined relations $\leftrightarrow_E^*$ and $Rightarrow_E^*$. What is missing, in analogy to Herbrand's theorem for first-order logic without equality Theorem 3.5.5, is a semantic characterization of the relations by a particular algebra.

4.1.8 Definition (Quotient Algebra)

For sets of unit equations this is a *quotient algebra*: Let X be a set of variables. For $t \in T(\Sigma, \mathcal{X})$ let $[t] = \{t' \in T(\Sigma, \mathcal{X})) \mid E \Rightarrow_{\mathrm{E}}^* t \approx t'\}$ be the *congruence class* of t. Define a Σ-algebra $\mathcal{I}_E$, called the *quotient algebra*, technically $T(\Sigma, \mathcal{X})/E$, as follows: $S^{\mathcal{I}_E} = \{[t] \mid t \in T_S(\Sigma, \mathcal{X})\}$ for all sorts S and $f^{\mathcal{I}_E}([t_1], \ldots, [t_n]) = [f(t_1, \ldots, t_n)]$ for $f : \mathrm{sort}(t_1) \times \ldots \times \mathrm{sort}(t_n) \to T \in \Omega$ for some sort T.

4.1.9 Lemma ($\mathcal{I}_E$ is an E-algebra)

$\mathcal{I}_E = T(\Sigma, \mathcal{X})/E$ is an E-algebra.

4.1.10 Lemma ($\Rightarrow_E$ is complete)

Let $\mathcal{X}$ be a countably infinite set of variables; let $s, t \in T_S(\Sigma, \mathcal{X})$. If $\mathcal{I}_E \models \forall \vec{x}(s \approx t)$, then $E \Rightarrow_E^* s \approx t$ is derivable.

4.1.11 Theorem (Birkhoff's Theorem)

Let $\mathcal{X}$ be a countably infinite set of variables, let E be a set of (universally quantified) equations. Then the following properties are equivalent for all $s, t \in T_S(\Sigma, \mathcal{X})$:

1. $s \leftrightarrow_E^* t$.
2. $E \Rightarrow_E^* s \approx t$ is derivable.
3. $s \approx_E t$, i.e., $E \models \forall \vec{x}(s \approx t)$.
4. $\mathcal{I}_E \models \forall \vec{x}(s \approx t)$.

By Theorem 4.1.11 the semantics of E and $\leftrightarrow_E^*$ conincide. In order to decide $\leftrightarrow_E^*$ we need to turn $\rightarrow_E^*$ in a confluent and terminating relation.

If $\leftrightarrow_E^*$ is terminating then confluence is equivalent to local confluence, see Newman's Lemma, Lemma 1.6.6. Local confluence is the following problem for TRS: if $t_1 \ {}_E\!\leftarrow t_0 \rightarrow_E t_2$, does there exist a term s so that $t_1 \rightarrow_E^* s \ {}_E^*\!\leftarrow t_2$?

If the two rewrite steps happen in different subtrees (disjoint redexes) then a repitition of the respective other step yields the common term s.

If the two rewrite steps happen below each other (overlap at or below a variable position) again a repetition of the respective other step yields the common term s.

If the left-hand sides of the two rules overlap at a non-variable position there is no ovious way to generate s.

More technically two rewrite rules $l_1 \rightarrow r_1$ and $l_2 \rightarrow r_2$ overlap if there exist some non-variable subterm $l_1|_p$ such that l_2 and $l_1|_p$ have a common instance $(l_1|_p)\sigma_1 = l_2\sigma_2$. If the two rewrite rules do not have common variables, then only a single substitution is necessary, the mgu σ of $(l_1|_p)$ and l_2.

4.2.1 Definition (Critical Pair)

Let $l_i \to r_i$ ($i = 1, 2$) be two rewrite rules in a TRS R whithout common variables, i.e., $\mathsf{vars}(l_1) \cap \mathsf{vars}(l_2) = \emptyset$. Let $p \in \mathsf{pos}(l_1)$ be a position so that $l_1|_p$ is not a variable and σ is an mgu of $l_1|_p$ and l_2. Then $r_1\sigma \leftarrow l_1\sigma \to (l_1\sigma)[r_2\sigma]_p$.

$\langle r_1\sigma, (l_1\sigma)[r_2\sigma]_p\rangle$ is called a *critical pair* of R.

The critical pair is *joinable* (or: converges), if $r_1\sigma \downarrow_R (l_1\sigma)[r_2\sigma]_p$.

4.2.2 Theorem ("Critical Pair Theorem")

A TRS R is locally confluent iff all its critical pairs are joinable.