Syntacticness, Cycle-Syntacticness and Shallow Theories

Hubert Comon, Marianne Haberstrau and Jean-Pierre Jouannaud*†
CNRS and LRI,
Bat. 490, Université de Paris Sud
91405 ORSAY cedex, France
E-mail {comon,haberstr,jouannaud}@lri.lri.fr

December 21, 1993

Abstract

Solving equations in the free algebra $T(F,X)$ (i.e. *unification*) uses the two rules: $f(\vec{s}) = f(\vec{t}) \rightarrow \vec{s} = \vec{t}$ (decomposition) and $s[x] = x \rightarrow \perp$ (occur-check). These two rules are not correct in quotients of $T(F,X)$ by a finitely generated congruence $=_E$. Following C. Kirchner, we first define classes of equational theories (called *syntactic* and *cycle syntactic* respectively) for which it is possible to derive some rules replacing the two above ones. Then, we show that these abstract classes are relevant: all *shallow theories*, i.e. theories which can be generated by equations in which variables occur at depth at most one, are both syntactic and cycle syntactic. Moreover, the new set of unification rules is terminating, which proves that unification is decidable and finitary in shallow theories. We give still further extensions. If the set of equivalence classes is infinite, a problem which turns out to be decidable in shallow theories, then shallow theories fulfill Colmerauer's *independence of disequations principle* (a conjunction of n disequations is solvable iff each disequation alone is solvable). This allows us to derive quantifier-elimination rules. It turns out that these rules do terminate for shallow theories, hence the first-order theory of the quotient algebra $T(F)/_{=_E}$ is decidable when F is finite and E is shallow. This extends Mal'cev results on the classes of (permutative) locally-free algebras that are completely axiomatizable [25].

1 Introduction

There is a well-known class of equational theories for which all interesting problems, e.g. the word problem [26, 21, 29], the unification problem [22], the disunification problem [4], are decidable: the class of ground theories. Even if ground theories are very important for practice, many interesting problems do not fall into this category. On the other hand, there are some axioms, such as commutativity, or more generally Malcev's permutative axioms [25], for which all the investigated above problems are also known to be decidable. The fundamental characterization of these axioms is that all their variables occur at depth at most one. We call *shallow* an equational theory which can be presented by finitely many such axioms. Our goal in this paper is to show that the class of shallow theories generalizes the class of ground theories with respect to the decidability of the above problems.

An interesting characteristic of the decidability results for ground theories is their uniformity: they take for argument a ground theory and an instance of the problem, and give the answer as result. There are very few classes of theories that enjoy a uniform unification or disunification algorithm: theories presented by

*also, Groupe FORMEL, INRIA, Batiment 11, Domaine de Voluceau - Rocquencourt, BP 105, 78150 Le Chesnay,France, email: jouannaud@margaux.inria.fr

†This research has been partly supported by *Greco de Programmation du CNRS* and *ESPRIT working group COMPASS.*

a convergent set of rewrite rules [11, 12], and *syntactic* theories for which a proof between any two terms can be performed by using at most one *non-collapsing* top equality step [18, 4]. For a syntactic theory, the decomposition schema of an equation $f(\vec{s}) = g(\vec{t})$ depends only upon the axioms of the form $f(\vec{l}) = g(\vec{r})$. Commutativity is syntactic. Surprisingly, the theory of associativity and commutativity is also syntactic [19, 27]. Both techniques allow to enumerate the solutions of a unification or disunification problem, but, in general, they do not allow to compute a finite representation of them. This enumeration does actually terminate for ground theories. And, indeed, both decidability results by Kozen [22] and Comon [4] are based on this underlying idea that ground theories are syntactic.

Generalizing these results to shallow theories turned out to be more difficult than expected. In particular, we found the need to generalize the idea of a syntactic theory in two different, yet related ways. First, we allow *collapse axioms* of the form $l[x] = x$. As a consequence, there are new decomposition schemes, since the variable x can be instantiated by terms headed by arbitrary function symbols, hence allowing to move from the head of l to an arbitrary symbol. Second, we allow, unlike Kirchner, cyclic equations (or cycles) such as $s[x] = x$ to have solutions. As a consequence, the usual occur-check rule becomes unsound in an arbitrary shallow theory: the cycle $x = f(x)$ has the solution $\{x \mapsto a\}$ in the theory generated by the (ground) axiom $a = f(a)$. To solve such cycles, we introduce the new key concept of a *cycle-syntactic* theory, which plays for cycles a role similar to syntacticness for non-cyclic equations. In a cycle-syntactic theory, any cycle $t[u] = u$ enjoys a proof which uses a *collapsing* axiom on the path from the root of t to the root of u in order to shrink t onto u. Collapsing axioms are defined as those axioms that decrease the size for all instantiations, such as $x + x \to x$. Solving cycles is then performed by non-deterministically guessing a collapsing rule along that path.

Another difficulty arose when dealing with negation: the so-called "lemma of independence of disequations"[5] is unsound if the ground term algebra is partitioned into finitely many congruence classes. On one hand, the latter case can be trivially solved. On the other, we must decide whether a shallow theory has finitely or infinitely many congruence classes in order to apply the appropriate technique.

Unfortunately, shallow presentations are not syntactic in general, nor cycle-syntactic. The main technique used throughout this paper is the computation by ordered completion techniques of conservative extensions of the starting shallow presentation which are respectively, ground convergent, syntactic and cycle-syntactic. In all cases, the property that variables occur at depth at most one appears to be crucial.

Shallow theories thus emerge as a fundamental non-trivial, union-closed subclass of equational theories for which all important questions are decidable: the word problem is decidable (section 4), the finiteness of the number of equivalence classes is decidable (section 10), unification is finitary (section 8), disunification is decidable (section 9), equational problems are decidable (section 9), hence the corresponding quotient algebras are completely axiomatizable.

2 Basic Definitions

Our notations and definitions are consistent with [10]. We recall the most important ones for our purposes. $T(F, X)$ is the algebra of terms over a finite set F of function symbols and a set X of variable symbols.

$T_1(F, X)$ is the subset of terms of depth at most 1. The root of $s = f(s_1, \ldots, s_n)$ is f, its size is $|s| = 1 + |s_1| + \cdots + |s_n|$. The root of a variable x is x itself and its size is one. A position in a term is a sequence of natural numbers used to select a subterm. Positions are ordered by the prefix ordering. The empty position is denoted Λ and corresponds to the whole term. The subterm of s at position p is written $s|_p$ and $s[u]_p$ indicates that the subterm of s at position p has been replaced by u. $s(p)$ is the root of the term $s|_p$. Two terms s and t are literally similar, written $s \simeq t$ if s can be obtained from t by an injection from X to X. We also use $\equiv$ for the syntactic identity between terms.
Equations are unordered pairs of terms. If E is a set of equations, we write

$$s \xleftrightarrow[E]{p} t \text{ iff } \exists l = r \in E,\ \sigma \in \Sigma,\ s|_p \equiv l\sigma \text{ and } t \equiv s[r\sigma]_p$$

We may make the rule and substitution precise as in $s \xleftrightarrow[l\sigma = r\sigma]{p} t$, or instead drop irrelevant information as in $s \xleftrightarrow{p} t$. We may also use predicate on positions instead of the positions themselves to indicate any position satisfying the predicate, as in $s \xleftrightarrow{\neq\Lambda} t$ (standing for some position $p \neq \Lambda$) $\xleftrightarrow{(>p)^*}$ (standing for $\xleftrightarrow{p \cdot p_1} \ldots \xleftrightarrow{p \cdot p_n}$ for some $p_1, \ldots, p_n$ distinct from Λ).
The reflexive transitive closure of a relation R is denoted by R^*, its reflexive closure by $R^=$ and its transitive closure by R^+. We also write, e.g. $\xleftrightarrow{(\neq\Lambda)^*}$, instead of $(\xleftrightarrow{\neq\Lambda})^*$.
Homomorphisms from $T(F, X)$ into $T(F)$ (resp. endomorphisms of $T(F, X)$) are called *ground substitutions* or *ground assignments* (resp. *substitutions*). Thanks to the universal property of $T(F, X)$, the set of ground substitutions (resp. the set of substitutions) is identified to the set $T(F)^X$ of mappings from X into $T(F)$ (resp. the set of mappings $T(F, X)^X$ of mappings from X into $T(F, X)$). The notation $\{x_1 \mapsto t_1;\ \ldots; x_n \mapsto t_n\}$ where $x_1, \ldots, x_n$ are distinct elements of X and $t_1, \ldots, t_n \in T(F, X)$ stands for the substitution mapping each x_i to t_i and the other elements of X to themselves.
Given a set E of equations and a first order formula ϕ whose atoms are equations, the set $[\![\phi]\!]_E$ of *E-solutions* of ϕ (resp. *solutions in $T(F, X)/_{=_E}$*) is the set of homomorphisms from $T(F, X)$ into $T(F)$ (resp. endomorphisms of $T(F, X)$) defined by:

$$\begin{array}{rcl}
[\![\top]\!]_E & = & T(F)^X \\
[\![s = t]\!]_E & = & \{\sigma \in T(F)^X \mid s\sigma =_E t\sigma\} \\
[\![\phi \wedge \psi]\!]_E & = & [\![\phi]\!]_E \cap [\![\psi]\!]_E \\
[\![\exists x :\ \phi]\!]_E & = & \bigcup_{t \in T(F)} [\![\phi\{x \mapsto t\}]\!]_E \\
[\![\neg\phi]\!]_E & = & T(F)^X \setminus [\![\phi]\!]_E
\end{array}$$

The other definitions follow from the above ones.
By notation, x, y, z (with possible subscripts) are variables, f, g, h are non-constant function symbols, a, b are constants and $s, t, u, v, w, l, r, g, d$ are terms, $\gamma, \sigma, \theta, \tau$ are substitutions and ϕ, ψ are formulae.

3 Syntactic Theories

In the following E is a finite set of equations.

Syntacticness is a property of equational proofs which has been introduced first by C. Kirchner [18] in the case of "collapse-free" theories. This property states that, for every equation such that $E \models s = t$, there is a proof of $s = t$ with at most one step at the root position. More formally:

Definition 3.1 *A* syntactic *presentation is a finite set E of equational axioms satisfying*

$$\underset{E}{\overset{*}{\longleftrightarrow}} \subseteq \underset{E}{\overset{(>\Lambda)*}{\longleftrightarrow}} \underset{E}{\overset{=}{\longleftrightarrow}} \underset{E}{\overset{(>\Lambda)*}{\longleftrightarrow}}$$

A theory is syntactic *if it can be axiomatized by a syntactic presentation.*

Commutativity is syntactic. Associativity is syntactic. Associativity and commutativity is a syntactic theory although the usual presentation is not syntactic [19]. Left + Right distributivity is not a syntactic theory [19]. We will also see in section 4 that all "shallow" theories are syntactic.

Syntactic theories enjoy an important property: equations between non-variable terms can be decomposed according to the rules in the presentation [18]. Such rules are called mutations. The idea is that if two terms s and t are unifiable by σ, then $s\sigma = t\sigma$ enjoys a syntactic proof. If there is no proof step at the top, then the proof breaks down into smaller subproofs between the subterms of $s\sigma$ and $t\sigma$. This case is part of the rule **Mutate 2** below. Otherwise, some equation $l = r$ is used at the top, and the proof breaks down into smaller proofs between the subterms of $s\sigma$ and $l\sigma$, and between the subterms of $r\sigma$ and $t\sigma$. This happens in both **Mutate 1** and **Mutate 2**. The same idea is used by Kozen for unifying modulo a ground theory. Kirchner [18] assumes the axioms in the syntactic presentation to be non-collapsing, allowing for simpler rules than ours. We allow here any kind of axioms (hence collapse axioms) as in [15]. This makes the expression of the rules more complex, since we must consider the case where l or r is a variable, hence the immediate subterms of $l\sigma$ are no more immediate subterms of l instantiated by σ.
Accordingly, we get the set of mutation rules of figure 1.

Lemma 3.2 *If $\widehat{E}$ is a finite syntactic set of equations defining the same equational theory as E, then* **Mutate 1** *and* **Mutate 2** *preserve the set of solutions in $T(F,X)/_{=_E}$.*

Proof
Straightforward adaptation of Kirchner's proof [18].

4 Shallow Theories are Syntactic

Definition 4.1 *A term t is shallow if all its variables occur at depth at most one. An equational axiom $l = r$ is* shallow *if both l and r are shallow. A* shallow presentation *is a finite set of shallow axioms. A theory is* shallow *if it can be axiomatized by a shallow presentation.*

For example, commutativity, identity and idempotency are shallow axioms, but associativity is not. Note that shallow theories are closed under union.
For every shallow presentation E, we are going to construct an equivalent syntactic shallow presentation $\widehat{E}$.

Mutate 1

$$
\begin{aligned}
f(\vec{s}) = g(\vec{t}) \quad \Longrightarrow \quad & \bigvee_{f(\vec{l})=g(\vec{r})\in\widehat{E}} \exists Var(\vec{l},\vec{r}),\ \big(\bigwedge_{i=1}^{n} s_i = l_i \bigwedge_{i=1}^{m} r_i = t_i\big) \\
& \vee \bigvee_{f(\vec{l})=x\in\widehat{E}} \exists Var(\vec{l}),\ \big(\bigwedge_{i=1}^{n} s_i = l_i \wedge x = g(\vec{t})\big) \\
& \vee \bigvee_{g(\vec{r})=x\in\widehat{E}} \exists Var(\vec{r}),\ \big(\bigwedge_{i=1}^{m} r_i = t_i \wedge x = f(\vec{s})\big)
\end{aligned}
$$

If $f \neq g$.

Mutate 2

$$
\begin{aligned}
f(\vec{s}) = f(\vec{t}) \quad \Longrightarrow \quad & \big(\bigwedge_{i=1}^{n} s_i = t_i\big) \\
& \vee \bigvee_{f(\vec{l})=f(\vec{r})\in\widehat{E}} \exists Var(\vec{l},\vec{r}),\ \big(\bigwedge_{i=1}^{n}(s_i = l_i \wedge r_i = t_i)\big) \\
& \vee \bigvee_{f(\vec{l})=x\in\widehat{E}} \exists Var(\vec{l}),\ \big(\bigwedge_{i=1}^{n} s_i = l_i \wedge x = f(\vec{t})\big) \\
& \vee \bigvee_{f(\vec{l})=x\in\widehat{E}} \exists Var(\vec{l}),\ \big(\bigwedge_{i=1}^{n} t_i = l_i \wedge x = f(\vec{s})\big)
\end{aligned}
$$

Figure 1: Mutation and Decomposition Rules

4.1 Completion of E

We first construct an alphabet C' and a conservative extension of the theory E, by abstracting ground terms by new constant symbols taken from C'. The following rule operates on pairs (C', E'), initialized by $(\emptyset, E)$, of an alphabet of constant symbols disjoint from C and a set of equations E':

$$\begin{array}{lcl} (C', E' \cup \{l[u]_i = r\}) & \rightsquigarrow & (C' \cup \{a\}, E' \cup \{l[a]_i = r, u = a\}) \\ & & \textbf{if } |i| = 1, u \in T(F \cup C') \setminus (C \cup C') \textit{ and } a \notin F \cup C' \end{array}$$

Lemma 4.2 *$\rightsquigarrow$ is terminating and any irreducible form E' of E by $\rightsquigarrow$ satisfies the following properties:*

- $\forall s \in T(F \cup C', X)$, $\exists t \in T(F, X)$, $s =_{E'} t$
- $\forall s, t \in T(F, X)$, $s =_{E'} t \Rightarrow s =_E t$
- *For every $s = t \in E'$, s and t have a maximal depth one*

Proof

Straightforward. □

From now on, we assume that E has been completed using the above rule and therefore that every equation in E has a depth smaller or equal to one.

We are now going to complete the set of equations obtained from E by adding new equational consequences in order to obtain a syntactic shallow presentation of E. This completion process is similar to unfailing completion (see e.g the survey [10]), but more equational consequences are actually added. For example, if $l \to r$ and $g \to d$ are two rules whose left hand sides do not overlap, we may still add an equational consequence $l\sigma = g\sigma$, if σ is a unifier of r and d: we add "shortcuts" in the equational proofs which would need two steps otherwise. (See also example 4.4.) Here are the inference rules:

$$(1) \quad \frac{g = d \quad l = r}{d\sigma = r\sigma} \quad \text{If } l, g \notin X \text{ and } \sigma = \mathrm{mgu}(l, g)$$

$$(2) \quad \frac{x = d \quad y = r}{d = r\{y \mapsto x\}} \quad \text{If } y \in X \text{ and } x \in C \cup X$$

$$(3) \quad \frac{g[a] = d \quad a = b}{g[b] = d} \quad \text{If } a, b \in C$$

Let $\widehat{E}$ be the closure of E with respect to the three previous rules.

Lemma 4.3 *$\widehat{E}$ is finite and, for every equation $s = t \in \widehat{E}$, the depth of s and t is at most one.*

Proof

Termination results from the finiteness of $\widehat{E}$ which follows from the finiteness of

$$\{f(t_1, \ldots, t_n)/\simeq \ \mid t_1, \ldots, t_n \in C \cup X, f \in F \cup C\}$$

For the second part of the claim, we need to show that if the premisses of a rule are equations whose members have a depth smaller or equal to one, then the conclusion of the rule satisfies the same property. This is straightforward for the rules (2) and (3). For (1), note that σ is a substitution of the form

$$\{x_1 \mapsto z_1; \ldots; x_n \mapsto z_n\}$$

where $z_1, \ldots, z_n$ are variables or constants. $\square$

Example 4.4 Let us start with the set of equations expressing the commutativity, idempotency and neutral element for a binary operation $*$:

$$E = \{x * x = x;\ x * y = y * x;\ x * 1 = x\}.$$

The completion process leads to

$$\widehat{E} = E \cup \{1 * x = x;\ 1 * 1 = 1;\ 1 * x = x * 1;\ 1 * x = x * x;\ x * 1 = x * x\}$$

(the trivial equations $s = s$ have been deleted).

It is also possible to rule out some cases of inconsistent theories (we will see later in lemma 4.14 that these are actually all cases of inconsistent theories): if a deduced equation $x = s$ is such that x is a variable which does not occur in s, then $=_E$ is inconsistent[1].

> We assume in the following that there is no equation $x = s \in \widehat{E}$ such that x is a variable which does not occur in s.

4.2 Properties of $\widehat{E}$

Let $\geq_{lpo}$ be the lexicographic path ordering (see, e.g. [10]) extending any total precedence over F such that any constant symbol is smaller than any non-constant symbol. Let $\rightarrow$ be the ordered rewriting relation generated by $\widehat{E}$ according to $\geq_{lpo}$. Let us recall that this relation is defined on ground terms by:

$$s \xrightarrow[l\sigma \rightarrow r\sigma]{p} t \text{ if } \begin{cases} l = r \in \widehat{E} \\ s|_p \equiv l\sigma \\ t \equiv s[r\sigma]_p \\ l\sigma >_{lpo} r\sigma \end{cases}$$

Lemma 4.5 *If* $s \xrightarrow[l\sigma \rightarrow r\sigma]{p} t$ *then* $l \notin X$.

<u>Proof</u>
This is a consequence of the assumption that $\widehat{E}$ does not contain any equation $x = s$ with $x \notin Var(s)$ and of the subterm property of $\geq_{lpo}$. $\square$

As indicated before, the above process performs the ordered completion of E:

[1]Inconsistent theories do have obviously all decidability properties. Hence they are out of interest.

Lemma 4.6 *Let u, v, w be ground terms, then*

$$v \leftarrow u \rightarrow w \;\Rightarrow\; v \overset{*}{\rightarrow} \;\overset{*}{\leftarrow} w$$

Hence $\widehat{E}$ is ground confluent.

Proof

We consider the possible critical overlaps between equations of $\widehat{E}$. We may assume they do not share any variables (they may be renamed apart). Assume that $l = r, g = d \in \widehat{E}$ and that $l\sigma >_{lpo} r\sigma, g\sigma >_{lpo} d\sigma, l|_p\sigma \equiv g\sigma$. By lemma 4.5, $l, g \notin X$. There are two cases:

If $p = \Lambda$, then σ is a unifier of l and g and therefore $\sigma = \rho\theta$ where ρ is a most general unifier of l and g. Then, by closure property of $\widehat{E}$ with respect to rule (1), $r\rho = d\rho \in \widehat{E}$. Hence, we have the proof $v \underset{r\rho\theta = d\rho\theta}{\longleftrightarrow} w$.

If $p \neq \Lambda$, then the only possible case is $g\sigma \in C$, because l and g have a depth smaller or equal to one. Since $g \notin X$, $g \in C$. Then $d \in C$ because constant symbols are smaller than non-constant symbols and d cannot be a variable. Therefore, $l[d]_p = r \in \widehat{E}$ by the closure property of $\widehat{E}$ with respect to rule (3). Hence, we have the proof $v \underset{l[d]\sigma = r\sigma}{\longleftrightarrow} w$.

□

Corollary 4.7 *The word problem is decidable in shallow theories.*

We now study commutation properties involving a proof step at the top :

Lemma 4.8 *Let v, w be ground terms.*

$$v \overset{\Lambda}{\longrightarrow} \;\overset{\neq\Lambda}{\longrightarrow} w \;\Rightarrow\; v \xrightarrow{(\neq\Lambda)^*} \;\overset{\Lambda}{\longrightarrow}\; \xleftarrow{(\neq\Lambda)^*} w.$$

Proof

Assume that $v \xrightarrow[l\sigma \to r\sigma]{\Lambda} u \xrightarrow[g\theta \to d\theta]{p} w$ with $p \neq \Lambda$. There are three cases :

- If $r|_p \in C$, then g and d are constants and therefore, $v \xrightarrow[l\sigma \to r[d]\sigma]{\Lambda} w$ by closure property of $\widehat{E}$ with respect to rule (3).
- If p is below a variable x occurring in r and not occurring in l, then $v \xrightarrow[l\tau \to r\tau]{\Lambda} u\{x\sigma \mapsto x\tau\} \xleftarrow[d\theta \leftarrow g\theta]{(\neq\Lambda)^*} w$ where τ is the substitution equal to θ everywhere except on x for which $x\tau \equiv x\sigma[d\theta]_q$. We use moreover the notation $\{x\sigma \mapsto x\tau\}$ (as for substitutions) to indicate that, for all occurrences of x, $x\sigma$ is replaced with $x\tau$. We keep this notation in the next lemma.
- If p is below a variable x of r that occurs in l, then $v \xrightarrow[g\theta \to d\theta]{(\neq\Lambda)^*} v\{x\sigma \mapsto x\tau\} \xrightarrow[l\tau \to r\tau]{\Lambda} u\{x\sigma \mapsto x\tau\} \xleftarrow[d\theta \leftarrow g\theta]{\neq\Lambda} w$ where τ is the substitution equal to θ everywhere except on x for which $x\tau \equiv x\sigma[d\theta]_q$.

  □

Lemma 4.9 *Let v, w be ground terms.*

$$v \xrightarrow{\Lambda} \xrightarrow{\Lambda} w \Rightarrow v \xrightarrow{(\neq\Lambda)^*} \xrightarrow{\Lambda} w.$$

Proof

Assume that $v \xrightarrow[l\sigma\to r\sigma]{\Lambda} u \xrightarrow[g\theta\to d\theta]{\Lambda} w$. Then $r\sigma \equiv g\theta$. There are two cases :

If $r \notin X$**,** then $v \xrightarrow[l\tau\rho\to d\tau\rho]{\Lambda} w$ by closure property of $\widehat{E}$ with respect to rule (1) ($\tau = mgu(r, g)$).

If $r \in X$**,** then x occurs in l since we assume that for every generated equation $x = s$, x occurs in s. Then $v \xrightarrow[g\theta\to d\theta]{(\neq\Lambda)^*} l\{x\sigma \mapsto d\theta\} \xrightarrow[l\tau\to g\tau]{\Lambda} w$ where τ is the substitution equal to σ everywhere except on x for which $x\tau \equiv d\theta$. □

Corollary 4.10 *Let $u \in T(F)$. Then u has a unique irreducible form v (w.r.t. $\widehat{E}$) and there is a reduction $u \xrightarrow{(\neq\Lambda)^*} \xrightarrow{\Lambda^=} v$.*

Proof

By lemma 4.6, the normal form of u is unique. By lemma 4.8, top reductions can be pushed to the right. By lemma 4.9, the sequence of top reductions can be replaced with a single top reduction. □

Lemma 4.11 *Let v, w be ground terms.*

$$v \xrightarrow{\Lambda} \xleftarrow{\Lambda} w \Rightarrow v \xleftrightarrow{(\neq\Lambda)^*} \xleftrightarrow{\Lambda} \xleftrightarrow{(\neq\Lambda)^*} w.$$

The directions of the rewrites in the above proof depend on v, w, hence the use of the double arrow.

Proof

Let $v \xrightarrow[l\sigma\to r\sigma]{\Lambda} u \xleftarrow[d\theta\leftarrow g\theta]{\Lambda} w$. Then $r\sigma \equiv u \equiv d\theta$ with $l\sigma >_{lpo} r\sigma$ and $g\theta >_{lpo} d\theta$. There are three different cases (the others follow by symmetry) :

- If neither r nor d are variables, then, by closure property of $\widehat{E}$ with respect to rule (1), $l\phi = g\phi \in \widehat{E}$ where ϕ is a most general unifier of r and d. This means that $v \xleftrightarrow{\Lambda} w$.
- If r and d are both variables, we use a similar argument based on the closure by rule (2) instead of the closure by rule (1).
- If $r \in X$ and $d \notin X$, then the two proof steps commute, yielding a proof

  $$v \xleftarrow[d\theta\leftarrow g\theta]{(\neq\Lambda)^*} l\sigma\{r\sigma \leftarrow g\theta\} \xrightarrow[l\tau\to r\tau]{\Lambda} w$$

  where τ is identical to σ, but on r where $r\tau \equiv r\theta$.

□

This shows in particular that $\widehat{E}$ is syntactic:

Lemma 4.12 *For every terms $s, t \in T(F, X)$ such that $s =_E t$, there is a proof in $\widehat{E}$:*

$$s \xleftrightarrow{(\neq\Lambda)^*} \xleftrightarrow{(\Lambda)^=} \xleftrightarrow{(\neq\Lambda)^*} t$$

Proof

First, we may assume that s, t are ground terms since they involve only finitely many symbols and adding a finite set of free constants to F does not change $\widehat{E}$. The syntactic proof between s and t is obtained in two steps. Using first lemma 4.6 and corollary 4.10 twice, there is a proof of $s = t$ which has the form:

$$\xrightarrow{(\neq\Lambda)^*} \xrightarrow{(\Lambda)^=} \xleftarrow{(\Lambda)^=} \xleftarrow{(\neq\Lambda)^*}$$

Using then lemma 4.11 results in a proof with at most one step at the top. □

Corollary 4.13 *Shallow theories are syntactic.*

As a simple consequence of lemma 4.12, the computation of $\widehat{E}$ decides whether $=_E$ is consistent or not:

Lemma 4.14 *E is inconsistent iff an inconsistency $x = s$ where $x \notin Var(s)$ has been found during the computation of $\widehat{E}$.*

Proof

By contradiction, assume that $\widehat{E}$ has been computed without generating an inconsistent equation. Since E is inconsistent, $x =_E y$, for any two variables x, y. Consider a proof of $x = y$ in $\widehat{E}$ which involves at most one step at the root. Since the first and the last step are necessary performed at the root, they have to be identical: $x \underset{\widehat{E}}{\longleftrightarrow} y$. Contradiction. □

As a consequence, it is undecidable whether a finite set of equations defines a shallow theory.

5 Cycle-syntacticness

Definition 5.1 *Given an equational theory E, a* cycle *for E is a proof between some term $t[u]_p$ and some of its strict subterms u. A* cyclic equation *is any equation of the form $t[x]_p = x$, where $p \neq \Lambda$ and $x \in X$.*

In his paper introducing syntacticness, Kirchner considers only those syntactic theories that do not admit cycles, hence cyclic equations cannot have solutions [18]. Shallow theories, however, may admit cycles, hence cyclic equations may have solutions. For example, if $E = \{f(0) = 0\}$, $\{x \mapsto 0\}$ is a solution of $f(x) = x$. To solve this problem, we will take our inspiration in the notion of a syntactic theory by allowing for a *single* paramodulation step on $t[x]$ in order to break the cycle. But this condition alone does not ensure termination of paramodulations. For example, to solve $f(x) = x$ with $E = \{f(x) = g(x)\}$ requires a paramodulation on $f(x)$ yielding $g(x) = x$ which in turn gives $f(x) = x$. In order to force termination, we will use an ordered version of paramodulation, according to some reduction ordering total on ground terms. For sake of completeness, however, the set of equations will have to be completed beforehand. This completion process is the subject of section 6.1.

Using a single paramodulation step, possibly not at the top, will raise a second completeness problem. A new property of proofs will be needed, playing for cycles the role played by syntacticness for the other kind of proofs. A first important remark before to characterize such a property is the following: cycles are solved thanks to the use of particular equations, those able to shrink a term so as to yield one of its subterms. We call such equations *collapsing rules.*

Definition 5.2 *A rule $l \to r$ is* collapsing *if, for every substitution σ, $|l\sigma| > |r\sigma|$. An equation $l = r$ or a rule $l \to r$ is* size-preserving *if for every substitution σ $|l\sigma| = |r\sigma|$*

This definition of a collapsing rule is more general than the classical one. Note that a rule is collapsing iff the multiset of variables occurring in its right hand side is included in the multiset of variables occurring in its left hand side and the size of the right hand side is smaller than the size of the left-hand side. The rule $f(x) \to g(a)$ is not collapsing, but so is $f(x) \to a$.

Collapsing rules are defined without referring to any particular ordering such as the ordering used in section 4. When using an ordering, we will have to make sure that it is compatible with the notion of collapsing rules. Lemma 6.5 shows that this is the case for the ordering we will use in the next section.

Usual collapsing rules ($s[u] \to u$ where u is a strict subterm of s), as well as non-regular equations ($s = t$ with $Var(s) \neq Var(t)$) are known to play a key role for breaking cycles [1]. What is new here is that we use rules instead of equations, and have a more general form of collapsing rules. We are now ready for defining the new property of proofs:

Definition 5.3 *A finite set of equational axioms E is* cycle-syntactic *if any cycle between $t[u]$ and u has a proof of the form*

$$t[u]_{p.q} \xleftrightarrow{(>p)^*} \xrightarrow[l\sigma \to r\sigma]{p} \xleftrightarrow{*} u$$

where $q \neq \Lambda$ and $l \to r$ is a collapsing rule of E. A theory E is cycle-syntactic if it has a cycle-syntactic presentation.

Note the dissymmetry introduced between the left and right hand sides of the proof between $t[u]$ and u, reflecting the dissymmetry between $t[u]$ and u themselves. It will permit to decompose the left part of the proof into smaller subproofs in a way similar to **Mutate**. The condition that the left proof-steps are below the collapsing step is essential for that. Here is an example:

Let $g >_F f$ and $E = \{f(a) = g(a);\ h(x, a) = a\}$. Consider the cyclic equation $f(h(x, h(y, z))) = x$ whose one solution can be obtained from the cycle

$$f(h(g(a), h(a, a))) \to f(h(g(a), a)) \to f(a) \xleftarrow{\Lambda} g(a).$$

This cycle does not suggest us an ordered paramodulation at the top, however, since the last step does not go the right direction nor use a collapsing rule. It suggests instead a paramodulation at position 1 using the collapsing rule $h(z', a) \to a$. This yields the new equational problem

$$\exists z'\ f(a) = x \wedge z' = x \wedge h(y, z) = a.$$

Cycle

$$t[x]_p = x \quad \Longrightarrow \quad \bigvee_{\substack{q < p \\ f(\vec{l}) \to r \in E_c}} \exists \vec{y},\ \Big(t[r]_q = x \wedge \bigwedge_{i=1}^{n} l_i = v_i\Big)$$

If $t|_q \equiv f(\vec{v})$ and $\vec{y} = \mathcal{V}ar(\vec{l})$.

Figure 2: Cycle rule

Since the cycle corresponding to the cyclic equation is of course not known beforehand, we will use an oriented paramodulation in $t[x]$ at any non-variable position. This is the desired rule which corresponds to **Mutate**.

Let E be a cycle-syntactic presentation whose set of collapsing rules is E_c. The rule for breaking cycles is displayed in figure 2.

Lemma 5.4 *If E is cycle-syntactic, then* **Cycle** *preserves the set of solutions in $T(F, X)/_{=_E}$.*

Proof

Obviously, every solution of the right hand side is a solution of the left hand side. Assume now that σ is a solution of the left hand side. Since x is a strict subterm of t, $t\sigma \overset{*}{\longleftrightarrow} x\sigma$ is a cycle. But E is a cycle-syntactic presentation. Therefore, it admits a proof $t\sigma \overset{(>q)^*}{\longleftrightarrow} t\sigma[f(\vec{l})\sigma]_q \xrightarrow[f(\vec{l})\sigma \to r\sigma]{q} t\sigma[r\sigma]_q \overset{*}{\longleftrightarrow} x\sigma$ where $f(\vec{l}) \to r$ is a collapsing rule, and $q < p$. Hence $t|_q \equiv f(\vec{v})$ and $v_i\sigma \overset{*}{\longleftrightarrow} l_i\sigma$. But $t[r]_q\sigma \equiv t\sigma[r\sigma]_q \overset{*}{\longleftrightarrow} x\sigma$. Hence σ is a solution of the right hand side. $\Box$

6 Shallow Theories are Cycle-Syntactic

We are now going to show that shallow theories are syntactic, using a method similar to the one we used in section 4.

6.1 Completing the shallow presentation

E is assumed to be consistent and to contain only equations $s = t$ where s, t have a depth smaller or equal to one. We use a linear ordering $>_F$ on F in which arity(f) > arity(g) implies $f >_F g$, and consider the lexicographic path ordering $>_{lpo}$ extending the linear ordering on the set of function symbols.

Our goal here is to construct a new set of equations E^* from E such that E^* is cycle-syntactic and defines the same equational theory. The construction is given as a set of inference rules applying on quadruples $(E, S, F, >_F)$, where S collects all equations generated so far and is initialized with E. There are two kinds of rules: one deduction rule **Deduce** will add new equational consequences from E; three rewrite rules **Simplify, Eliminate** and **Divide** will add new equations and remove some previously obtained ones. This situation is similar to completion. The resulting output may of course depend on a particular sequence of application of these rules. As we will see, the process terminates for any strategy, resulting in a finite set E^*. We denote by $(E_n, S_n, F_n, >_{F_n})$ the quadruple obtained at step n. C_n is the set of constants of F_n.

$$\textbf{Deduce} \quad \frac{(E \cup \{f(\vec{l}) = r, f(\vec{g}) = d\}, S, F, >_F)}{(E \cup \{f(\vec{l}) = r, f(\vec{g}) = d, r\sigma = d\sigma\}, S \cup \{r\sigma = d\sigma\}, F, >_F)}$$

if $arity(f) > 0$,
either $r \in Var(\vec{l})$ or else $r = h(\vec{r})$ and $arity(f) \geq_N arity(h)$,
either $d \in Var(\vec{g})$ or else $d = k(\vec{d})$ and $arity(f) \geq_N arity(k)$,
$r\sigma = d\sigma$ is not an instance of an equation in S.

where $\sigma = \text{mgu}(\vec{l}, \vec{g})$

$$\textbf{Simplify} \quad \frac{(E \cup \{s = t[u]\}, S, F, >_F)}{(E \cup \{s = t[v]\}, S \cup \{s = t[v]\}, F, >_F)}$$

If $t[u] \underset{u \to v}{\longrightarrow} t[v]$, u, v are constants and $u = v \in E$

$$\textbf{Eliminate} \quad \frac{(E \cup \{s = s\}, S, F, >_F)}{(E, S, F, >_F)}$$

$$\textbf{Divide} \quad \frac{(E \cup \{f(\vec{s}) = g(\vec{t})\}, S, F, >_F)}{(E \cup \{f(\vec{s}) \to h(\vec{x}), g(\vec{t}) \to h(\vec{x})\}, S \cup \{f(\vec{s}) \to h(\vec{x}), g(\vec{t}) \to h(\vec{x})\}, F \cup \{h\}, >_F \oplus h)}$$

where $\vec{x} = Var(\vec{s}) \cap Var(\vec{t})$

if $arity(f) > |\vec{x}|$ and $arity(g) > |\vec{x}|$

where $>_F \oplus h$ is any ordering on $F \cup \{h\}$ extending $>_F$, satisfying the arity condition, and such that any added constant is bigger than all previously existing ones.

Among the four deduction rules, the last three are simplification rules while the first is a generation rule. As usual in a completion process, the resulting set of equations is obtained by collecting the persisting equations obtained at each step of the process when using a fair strategy. We denote by $(E_n, S_n, F_n, >_{F_n})$ the quadruple obtained at the nth inference step, and $(E^*, S^*, F^*, >_{F^*})$ its limit. Since the sequence $(E_n)_{n \in N}$ is not monotonic with respect to set inclusion, E^* is the set of persisting equations:

$$E^* = \bigcup_{i=0}^{\infty} \bigcap_{j=i}^{\infty} E_j.$$

We first prove that **Deduce** and **Simplify** terminate. Then we prove that the whole completion process terminates.

Lemma 6.1 **Deduce** *and* **Simplify** *terminate for any presentation E whose axioms have a depth at most one.*

Proof

Since the set of equations that can possibly be inferred is finite (using the same argument as in the proof of lemma 4.3), we simply need to prove that a loop cannot occur. This is guaranteed by the condition that $\{r\sigma = d\sigma\}$ does not belong to S; this means that an equation is deduced only if it had never been deduced before. $\square$

Lemma 6.2 *Let $f(\vec{s}) = t$ be an equation in E_n for some n. Then, $f(\vec{s}) \in T_1((F_n \setminus C_n) \cup C_0, X) \cup C_n$.*

Proof

First, by a straightforward induction on n, if $s = t \in E_n$, then $s, t \in T_1(F_n, X)$. Now, again by induction and from the definition of $>_F \oplus h$ among constants, there is no occurrence of $a \in C_n \setminus C_0$ as a strict subterm of a member of E_n. $\square$

Lemma 6.3 *Starting with $(E, E, F, >_F)$, the four above rules terminate for any presentation E whose axioms have a depth at most one, resulting in a finite presentation E^* of the theory.*

Proof

Let $T_1(F_n, X)$ be the set of terms in $T(F_n, X)$ whose depth is at most 1. We now interpret a set of equations E_n by the following multiset of natural numbers:

$$\{\min(\text{arity}(f), \text{arity}(g)) | f(\vec{s}), g(\vec{t}) \in (T_1((F_n \setminus C_n) \cup C_0, X) \cup C_n)/\simeq,\ f(\vec{s}) = g(\vec{t}) \in E_n \cup \overline{S_n}\}$$

where $\overline{S_n}$ is the complement of S_n. Note that the union $E_n \cup \overline{S_n}$ is actually a disjoint union since E_n is always a subset of S_n. In the above definition, variables are considered as constants, hence a rule $l \to x$ is interpreted as 0.

This interpretation is unchanged when using **Deduce** since generated terms are in $(T_1((F_n \setminus C_n) \cup C_0, X) \cup C_n)$ by lemma 6.2, and previously deduced equations cannot be deduced anymore by definition of **Deduce**. Hence, **Deduce** moves an equation from $\overline{S_n}$ into E_n.

The interpretation either decreases or remains constant when using **Simplify**. Indeed, the equation which is removed does belong to E_n and the equation which is added either belongs to $\overline{S_n}$ (in which case the interpretation is decreasing) or belongs to S_n. In the latter case, $E_n = \{s = t[u]\} \cup E$, $S_n = \{s = t[u], s = t[v]\} \cup S$, $E_{n+1} = \{s = t[v]\} \cup E$ and $S_{n+1} = S_n$. Hence, $E_{n+1} \cup \overline{S_{n+1}} = (E_n \cup \overline{S_n} \cup \{s = t[v]\}) \setminus \{s = t[u]\}$. Then, the interpretation remains unchanged since the arities are the same.

On the other hand, **Eliminate** and **Divide** decrease strictly the interpretation. This is straightforward for **Eliminate**. Concerning **Divide**, assume that $E_{n+1} = E_n \setminus \{f(\vec{s}) = g(\vec{t})\} \cup \{f(\vec{s}) \to h(\vec{x});\ g(\vec{t}) \to h(\vec{x})\}$. Then $a(h) < a(f)$ and $a(h) < a(g)$ and any equation $s = t \in E_{n+1} \cup \overline{S_{n+1}}$ which is not in $E_n \cup \overline{S_n}$ and such that $s, t \in T_1((F_{n+1} \setminus C_{n+1}) \cup C_0, X) \cup C_{n+1}$ has a member whose top symbol is h (symbols of C_{n+1} are not allowed strictly inside the terms). Hence the interpretation of $s = t$ is smaller than or equal to $a(h)$ which

is in turn strictly smaller than $\min(a(f), a(g))$ which is the interpretation of the removed equation. This means, that the interpretation decreased, by definition of the multiset extension of an ordering.
Now termination is obtained by applying lemma 6.1. □

The termination proof was complicated by **Simplify**. It would be simpler if the equation $s = t[v]$ were not added to E when it already belongs to S. This slight change, however, would on the other hand complicate the correctness proof of **Simplify** (currently the correctness of all rules is trivial).

Example 6.4 Considering again example 4.4, The above completion process computes a set

$$E^* = E \cup \{1 * x \to x\}$$

since only one deduction step is possible. If we add the equation $x * x = 1 * x$ to E (and this is indeed an equational consequence of E), then the **Divide** rule applies and then **Deduce**, leading to

$$E \cup \{1 * x \to x;\ x * x \to f(x);\ 1 * x \to f(x);\ f(x) \to x\}$$

6.2 Properties of E^*

The above set of inference rules has been designed so as to obtain a conservative extension of the starting theory made of size-preserving equations or rules and collapsing rules:

Lemma 6.5 *Any equation $u = v \in E^*$ is of one of the following forms:*

1. *$u \equiv f(\vec{s}), v \equiv f(\vec{t})$: Then $f(\vec{s}) = f(\vec{t})$ is a (size-preserving) permutative equation in the sense of Malcev, that is, $\vec{s}$ and $\vec{t}$ are different permutations of the same set of variables.*
2. *$u \equiv f(\vec{s}), v \equiv g(\vec{t}), arity(f) = arity(g)$ and $f >_{F^*} g$: Then $\vec{s}$ and $\vec{t}$ are both permutations of the same (possibly empty) set of variables, hence $f(\vec{s}) \to g(\vec{t})$ is a size-preserving rule.*
3. *$u \equiv f(\vec{s}), v \equiv g(\vec{t}), arity(f) > arity(g)$, hence $f >_{F^*} g$: Then $\vec{t}$ is a (linear) subset of $Var(\vec{s})$, hence $f(\vec{s}) \to g(\vec{t})$ is a collapsing rule.*
4. *$u = f(\vec{s}), v \in Var(\vec{s})$: Then $u \to v$ is a collapsing rule.*

<u>Proof</u>
In all cases, we denote by $\vec{x}$ the set $Var(\vec{s}) \cap Var(\vec{t})$. We distinguish four possible cases. In the first three, neither u nor v are variables, hence $u \equiv f(\vec{s})$, and $v \equiv g(\vec{t})$.

- $f = g$: Then $\vec{s}$ and $\vec{t}$ must be the set $\vec{x}$ itself (up to permutation), otherwise, **Divide** would apply.
- $f >_{F^*} g$ and arity(f) = arity(g): then $\vec{s}$ and $\vec{t}$ must be the set $\vec{x}$ itself (up to permutation), otherwise **Divide** would apply. Hence $f(\vec{s}) >_{lpo} g(\vec{t})$ and $f(\vec{s}) \to g(\vec{t})$ is a size-preserving rule.
- $f >_{F^*} g$ and arity$(f) >$ arity(g): $\vec{t}$ is a set of (distinct) variables and $\vec{t} \subset \vec{s}$ since **Divide** cannot be applied. Now, $f >_{F^*} g$ implies that the rule can be oriented into $f(\vec{s}) \to g(\vec{t})$ which is a collapsing rule because arity$(f) >$ arity(g).

- $v \in X$. $u \not\equiv v$, otherwise **Eliminate** would apply. Then $x \in Var(u)$, otherwise the theory E would be inconsistent, a case that has been ruled out already. Hence, $u \rightarrow v$ is a collapsing rule.□

Lemma 6.6 *E^* is ground convergent.*

Proof

Let us consider the possible critical overlaps between equations of E^*. We may assume any two equations do not share any variables (one may be renamed). Assume that $l = r, g = d \in E^*$ and that $l\sigma >_{lpo} r\sigma, g\sigma >_{lpo} d\sigma, l|_p\sigma \equiv g\sigma$ and that σ itself is irreducible. We may assume that $g \notin C$, otherwise $l = r$ can be simplified using $g = d$. Then $p = \Lambda$ and $l \notin C$. Hence σ is a unifier of l and g and therefore $\sigma = \rho\theta$ where ρ is a most general unifier of l and g. Furthermore, by lemma 6.5, arity$(l(\Lambda)) \geq_N$ arity$(r(\Lambda))$ and arity$(l(\Lambda)) \geq_N$ arity$(d(\Lambda))$. Then, by closure property of E^* with respect to **Deduce** we know that either $r\rho = d\rho \in E^*$ or else one of the equations $l = r$ and $g = d$ can be simplified. This means that all critical overlaps can be simplified: E^* is ground convergent. □

Lemma 6.7 *Assume $t[u]_p \overset{*}{\underset{E}{\longleftrightarrow}} u$ where $t, u \in T(F)$, u is irreducible and $p \neq \Lambda$. Then, there is a position $q < p$ such that $t[u]_p \xrightarrow[E^*]{(>q)^*} \xrightarrow[s \rightarrow t]{q} \xrightarrow[E^*]{*} u$ where $s \rightarrow t$ is a collapsing rule.*

Proof

From lemma 6.6, there is a rewrite proof of $t[u]_p =_{E^*} u$. Since u is irreducible, this is a rewriting from $t[u]$ to u and there is no rewrite step inside u or at position p.

We choose to normalize $t[u]$ bottom up, by applying first the collapsing rules, then the other rules that do not change the size of terms. Among all such normalizations, we choose one of minimal length. Since E^* is shallow, and because of the properties stated in lemma 6.5, the normalization process is non-increasing with respect to the prefix ordering on positions, that is applying a rule at position q does not create new redexes at a position qq' for a non-empty q'.

Since all rules are size preserving except the collapsing ones, there must exist such a rule applied at a strict prefix q of p. Consider the first such application in the rewriting from $t[u]$ to u. We must show that the rule applied just before, if any, is applied at a strict suffix of q.

Assume by contradiction that a size-preserving rule or equation is applied before the collapsing one at position q, that is

$$t[u] \rightarrow^* w \xrightarrow[g\sigma \rightarrow d\sigma]{q} v \xrightarrow[f(\vec{l})\sigma \rightarrow r\sigma]{q} v' \rightarrow^* u$$

where $g = d$ is size-preserving, $f(\vec{l}) \rightarrow r$ is collapsing, and both equations are supposed to have their variables renamed apart so as to use the same substitution σ. Then $v|_q \equiv f(\vec{l\sigma}) \equiv d\sigma$, hence d is headed by f, and since $g = d$ is size-preserving, then the head of g has the same arity as f. On the other hand, since $f(\vec{l}) \rightarrow r$ is collapsing, r is either a variable of $\vec{l}$ or is headed by a function symbol whose arity is not greater than the arity of f.

Deduce applies, yielding the equation $g\theta = r\theta$ such that $\sigma = \theta\rho$ for some ρ. $g\theta$ cannot be simplified since the constants occurring in $g\theta$ occur either in g or in r which cannot be simplified by definition of E^*. Now, either $g\theta = r\theta \in E^*$ (in which case, we let $r' \equiv r\theta$ by convention) or else, there is a rule $g\theta \to r' \in E^*$ which is obtained by dividing and simplifying $g\theta = r\theta$. In any case, $|g\theta| > |r\theta| \geq_N |r'|$ which means, by lemma 6.5 that $g\theta >_{lpo} r'$ and $g\theta \to r'$ is a collapsing rule. This rule applies to w at position q, yielding a new smaller proof following our strategy, which is a contradiction. □

This shows in particular the cycle-syntacticness of E^*.

Corollary 6.8 *For every term $t \in T(F, X)$ and strict subterm u of t such that $t[u] =_E u$, there is a proof*

$$t[u]_{q.q'} \xleftrightarrow{(>q)^*} \xrightarrow[l\sigma \to r\sigma]{q} \xleftrightarrow{*} u$$

where $l \to r$ is a collapsing rule of E^*.

Proof

First, we may consider t as ground. Indeed, adding the variables of t as new free constants in F does not change E^*. Let $t[u]_p = u$ be a ground cycle ($p = q.q'$). Let $u\!\downarrow$ be the normal form of u. Then $t[u\!\downarrow]_p \xleftrightarrow{(\geq p)^*} t[u]_p \xleftrightarrow{*} u \xleftrightarrow{*} u\!\downarrow$. The result follows from lemma 6.7. □

As a consequence, we may use the rule **Cycle** choosing for E_c the (finite) set of collapsing rules in E^*.

7 Unification in Syntactic and Cycle-Syntactic Theories

In addition to the rules **Mutate 1**, **Mutate 2** and **Cycle** we use the classical unification rules; the whole set of rules consists of all rules displayed in figures 1, 2 and 3.

Lemma 7.1 *Assume that $\widehat{E}, E^*$ and E define the same equational theory and that $\widehat{E}$ is syntactic and E^* is cycle-syntactic. Let E_c be the set of collapsing rules in E^*. Then all rules given in figures 1, 2 and 3 are sound. (I.e. they preserve the set of solutions).*

Formulas are supposed to be kept in disjunctive forms.
Solved forms are tree-solved forms in [16]:

Definition 7.2 *A* solved form *is either $\top$, $\bot$ or a finite disjunction of formulae of the form*

$$\exists \vec{z}\; x_1 = t_1 \wedge \ldots \wedge x_n = t_n$$

where $x_1, \ldots, x_n$ are distinct free variables occurring only once.

The completeness result below follows from a detailed inspection of the left hand sides of the rules displayed in figures 1, 2 and 3.

Theorem 7.3 *Every irreducible form (w.r.t. the rules of figures 1, 2 and 3) of a conjunction of equations is a solved form.*

Existential Quantifiers Elimination

$$\exists x : x = s \wedge P \implies P\{x \mapsto s\}$$

If x does not occur in s

Variable Elimination

$$x = s \wedge P \implies x = s \wedge P\{x \mapsto s\}$$

If $x \notin Var(s)$, $x \in Var(P)$ and if s is a variable, then $s \in Var(P)$

Delete

$$s = t \implies \top$$

If $s =_E t$

Figure 3: Elimination rules

8 Unification in Shallow Theories

The question now is whether solved forms can be computed in finite time. This is not the case in general since unification in syntactic theories may be undecidable [20]. We can therefore seek for subclasses of syntactic theories for which our set of rules terminates for every control, or may terminate for some control. Shallow theories are an example of syntactic theories in which the unification procedure of the previous section terminates for any application of the rules. In addition, they enjoy an important property: existential quantifiers are not necessary as they can be eliminated as soon as they are introduced.

Lemma 8.1 *If ϕ does not contain any existential quantifier and $\phi \implies \psi$, then all normal forms of ψ w.r.t. the Existential Quantifiers Elimination Rule are quantifier-free.*

Proof

This is a consequence of lemmas 4.3 and 6.2. $\square$

Again we can apply the previous results to shallow theories. Let E^* and $\widehat{E}$ be as defined in sections 6 and 4 respectively. Then

Theorem 8.2 *Assuming that the existential quantifiers elimination rules is applied eagerly, the rules of figures 1 and 3 are terminating when E is shallow. Moreover, irreducible formulae are quantifier-free solved forms.*

Proof

We interpret the disjunctions of conjunctions of equations in an ordered set whose ordering is well-founded and such that any rule entails a strict decreasingness of the interpretation.

We use the following interpretation:

- If $s = t$ is an equation, $I(s = t) = |s| + |t|$.
- $I_1(s_1 = t_1 \wedge \ldots \wedge s_n = t_n)$ is the pair $(N, \{I(s_1 = t_1), \ldots, I(s_n = t_n)\})$ where N is the number of *unsolved* variables in the conjunction (let us recall that a variable x is solved in a conjunction of equations if it occurs only once in the conjunction and if it is a member of an equation). The second component is the multiset of interpretations of each equation. Such pairs are ordered using the lexicographic extension of the usual ordering on natural numbers on one hand, and the multiset extension of the above ordering on the other hand.
- $I_2(c_1 \vee \ldots \vee c_n)$ where $c_1, \ldots, c_n$ are conjunctions of equations is the multiset $\{I_1(c_1), \ldots, I_1(c_n)\}$. Such multisets are ordered using the multiset extension of the above ordering.

We now need to check the decreasingness on the irreducible forms of the formulas w.r.t. the Existential Quantifiers Elimination Rule (the termination of this latter rule alone being straightforward).

Variable Elimination This rule strictly decreases the number of unsolved variables in some conjunction.

Mutate (1,2) The normal form of the formula does not contain any new (existentially quantified) variable by lemma 8.1. Therefore, the number of unsolved variables can only decrease. Let $s = t$ be the equation on which a **Mutate** rule is applied. If both sides are constants, then the right hand side of the **Mutate** rule is either $\top$ or $\perp$ (depending on whether $s = t \in \widehat{E}$ or not). We now eliminate this case, and assume that (say) $|s| > 1$ and $|s| \geq |t|$. Each new equation has one of the following forms

- $s_i = t_j$ (corresponding to $l_i \equiv r_j \in X$, leading to $s_i = t_j$ by the **EQE** rule)
- $s_i = s_j$ (corresponding to $l_i \equiv l_j \in X$)
- $t_i = t_j$ (corresponding to $r_i \equiv r_j \in X$)
- $s_i = c$ where $c \in C$ (corresponding to $l_i \in C$)
- $c = t_j$ where $c \in C$ (corresponding to $r_j \in C$)
- $s_i = t$ (corresponding to $l_i \equiv x$)
- $s = t_j$ (corresponding to $r_i \equiv x$)

There is no other situation since every l_i, r_i is either a variable or a constant. It is now easy to check that, for all i, j, $|s_i| + |t_j| < |s| + |t|$, $|s_i| + |s_j| < |s| + |t|$, $|t_i| + |t_j| < |s| + |t|$, $|s_i| + |c| < |s| + |t|$, $c + |t_j| < |s| + |t|$, $|s_i| + |t| < |s| + |t|$, $|s| + |t_j| < |s| + |t|$. Hence I_2 is decreasing, by definition of the multiset extension of an ordering.

Cycle By lemma 6.5, for every $f(\vec{l}) \to s \in E_c^*$, s is linear. On the other hand, every variable $x \in Var(s)$ occurs in $\vec{l}$. Therefore, each equation produced by the **Cycle** rule has one of the following forms:

- $t[c]_q = x$ (corresponding to $s \in C$)
- $t[v_i]_q = x$ (corresponding to $s \equiv l_i \in X$)

- $t[g(w_1, \ldots, w_k)]_q = x$ where $\{w_1, \ldots, w_k\} \subset \{v_1, \ldots, v_n\} \cup C$ and $k < n$ (corresponding to $s \equiv g(y_1, \ldots, y_k)$)
- $c = v_i$ (corresponding to $l_i \in C$)
- $v_i = v_j$ (corresponding to $l_i \equiv l_j \in X$)

In all cases, the sum of the sizes of the members of the equation is strictly smaller than $1 + |t[x]_p| = 1 + |t[f(\vec{v})]_q|$. Then, by definition of the multiset extensions of the ordering, I_2 is indeed decreasing.

□

Corollary 8.3 *Unification is decidable and finitary in shallow theories.*

Proof

Note that $=_E$ is decidable in shallow theories. Hence, the applicability of the rules is decidable. Now, the result follows from theorems 8.2 and 7.3. □

Example 8.4 Let us consider again example 4.4. Assume that we want to solve the equation $x*y = x*(x*y)$. Using **Mutate 2** combined with the existential quantifier elimination rule, and removing trivial equations, we get the problem:

$$\begin{array}{llllll} & y = x*y & \vee & (x=1 \wedge x=y \wedge x=x*y) & \vee & (x=1 \wedge x=y \wedge x*y=1) \\ \vee & (x=y \wedge x=x*y) & \vee & (y=1 \wedge x=x*y) & \vee & (x=1 \wedge x=y \wedge 1=x*y) \\ \vee & (x=y \wedge x=1 \wedge x=x*y) & \vee & (x=y \wedge x*y=1) & \vee & (x=y \wedge x=x*(x*y)) \\ \vee & (y=1 \wedge x=x*(x*y)) & \vee & (x=1 \wedge y=x*(x*y)) & \vee & (x=x*y \wedge x=x*y) \\ \vee & (x*y=1 \wedge x=x*y) & \vee & x=1 & & \end{array}$$

Now, we use the following transformations:

$$\begin{array}{rll} y = x*y & \Longrightarrow_{\text{Cycle}} & x=y \vee x=1 \vee (y=1 \wedge x=y) \\ x = x*y & \Longrightarrow_{\text{Cycle}} & x=y \vee y=1 \vee (x=1 \wedge x=y) \\ x = x*(x*y) & \Longrightarrow^*_{\text{Cycle}} & (x=1 \wedge x=x*y) \vee x*y=1 \vee x=x*y \\ & \Longrightarrow^*_{\text{Cycle}} & (x=1 \wedge x=y) \vee (x=1 \wedge y=1 \wedge x=y) \vee x*y=1 \\ & & \vee x=y \vee x=1 \vee (y=1 \wedge x=y) \\ x*y=1 & \Longrightarrow_{\text{Mutate}} & x=1 \wedge y=1 \end{array}$$

Now, eliminating redundant parts (i.e. using the normalization rule $P \vee (P \wedge Q) \mapsto P$), we obtain a disjunction of three tree solved forms:

$$x = y \vee x = 1 \vee y = 1$$

Mutations and Cycle rules are no more than particular narrowings, in which the syntacticness and cycle-syntacticness properties of the theory is used. For example, syntacticness allows to consider only the root position, hence reducing the non-determinism of narrowing. Therefore, our method must be more efficient than arbitrary basic narrowing (although, as pointed out by a referee, more elaborated narrowing strategies might be more efficient. Also, the comparison is not really accurate since the set E^* can be larger than the

set of permutations of the rules of a convergent system generating the same equational theory). As it was recently pointed out to us by Domenjoud and Rusinowitch (personal communications), the non-permutative equations of E^* are indeed rewrite rules (w.r.t. the ordering used) which can be shown to be convergent modulo the permutative axioms in E^*. Since these rules are of depth one, basic narrowing terminates as recently shown in [2]. This yields of course an alternative decidability proof of unification in shallow theories with the above mentioned drawbacks. Moreover, this would not allow to proceed with the next section.
Our rules are based on Variable Elimination instead of Merge, which is somewhat against our usual philosophy. Indeed, using Merge allows sharing hence avoiding exponential growth of terms in the unification process [16]. The problem is that Merge and Mutation together may not terminate when the presentation contains collapse axioms. Recovering termination is probably possible, but would require the use of a complicated control. The former version of this paper which appeared in the proceedings of LICS 92, indeed used Merge together with a control expressed by a new predicate $\stackrel{\neq\Lambda}{=}$. This solution however does not work since $x \stackrel{\neq\Lambda}{=} s \wedge x = t$ does not imply $t \stackrel{\neq\Lambda}{=} s$ as implicitly assumed in the paper, but $t = s$ which would then cause non-termination.

9 Disunification in Shallow Theories

An *equational formula* is a first order formula whose atoms are equations. Our goal is now to lift the previous results to equational formulae, thus deriving complete axiomatizations for a class of quotient algebras (as done in [6]).
In order to make the section easier to read, we divided it in several parts: we first set up precisely the problem and show the difficulties through examples. Then we give the rules and prove their correctness in section 9.2 and their termination and completeness in section 9.3. Finally we conclude in section 9.4

9.1 The setting

First, we may restrict our attention to the Σ_2 fragment: it is sufficient to show on one hand how to reduce any formula $\exists^*\forall^*\phi$ where ϕ is quantifier-free in an equivalent formula $\exists^*\psi$ where ψ is quantifier-free, and, on the other hand, to show how to solve any purely existential formula. That is why we assume now that equational formulae are (disjunction of) formulae which have the following form:

$$\exists\vec{x}, \forall\vec{y} : d_1 \wedge \ldots \wedge d_n$$

where $d_1, \ldots, d_n$ are disjunctions of atomic formulae (either equations $s = t$ or *disequations* $s \neq t$). Note that this assumes that formulae are kept in a conjunctive, prenex form.
Of course, we may use the techniques of the previous sections and simplify equations using **Mutate** rules and **Cycle** rules. Since these rules preserve the set of solutions, we may also use their negation for simplifying disequations. However, there are additional difficulties arising with the quantifiers. Let us point out some of them. First, we should be able to solve the formula $\forall x : x = t_1 \vee \ldots \vee x = t_n$ where $t_1, \ldots, t_n$ are ground terms. Solving such a formula amounts to deciding whether $T(F)/_{=_E}$ is a subset of $\{\overline{t_1}, \ldots, \overline{t_n}\}$ (where $\overline{s}$ is the class of s modulo $=_E$). Actually, if $T(F)/_{=_E}$ is finite and if $=_E$ is decidable, then there is a

straightforward decision procedure for the first order theory of $T(F)/_{=_E}$, which is completely different from the procedure of [7, 4]. Now, assume that $T(F)/_{=_E}$ is infinite (we keep this assumption until the end of this section). We need more generally to solve formulae $\forall \vec{y} : y_1 = t_1 \vee \ldots \vee y_n = t_n$ where $y_1, \ldots, y_n \in \vec{y}$. In the free algebra, it is possible to show that such formulae cannot have a solution. The proof relies on an *independence of disequations lemma* [23], which is also used for solving e.g. conjunction of disequations. Roughly, such a lemma states that a conjunction of disequations has a solution iff each disequation has a solution. This result fails in the case of quotient algebras, even when $T(F)/_{=_E}$ is infinite. For example, let $F = \{0, s, e\}$ where 0 is a constant and s, e are unary. Let $E = \{e(0) = 0;\ e(s(s(x))) = e(x);\ e(s(0)) = s(0)\}$. Then, it is easy to see that $\{s^n(0) \mid n \in \mathbf{N}\}$ is a set of representatives of $T(F)/_{=_E}$ and $e(x) \neq 0 \wedge e(x) \neq s(0)$ has no solution whereas each disequation has a solution.

Again, we aim at giving general techniques for solving such formulae. For, we will use the notion of *compact equational theories* introduced in [6]. In this class of theories, there is indeed a kind of "independence of disequations". We use the definitions given in [6]:

Definition 9.1 *The disequations $t_1 \neq u_1, \ldots t_n \neq u_n$ are* independent *(w.r.t. E) if*

- *either there is an index i such that $t_i =_E u_i$*
- *or else $t_1 \neq u_1 \wedge \ldots \wedge t_n \neq u_n$ has at least one solution in $T(F)/_{=_E}$.*

Definition 9.2 *A* unification problem *is a purely existential equational formula which does not involve negation (i.e. no disequation).*

A unification problem is *solved* if it is of the form given in definition 7.2.

Given two equational formulae ϕ, ψ, we write $\phi \approx_E \psi$ if ϕ and ψ have the same set of E-solutions.

Definition 9.3 *A finite set of equations is* finitary *if there is a terminating algorithm which transforms any unification problem $\mathcal{P}$ into a finite disjunction of completely solved unification problems $\mathcal{SF}(\mathcal{P})$ such that $\mathcal{P} \approx_E \bigvee_{\mathcal{E} \in \mathcal{SF}(\mathcal{P})} \mathcal{E}$.*

Definition 9.4 *A set of equations E is* compact *if it is finitary and $=_E$ is decidable and any n disequations are independent (w.r.t. E).*

Examples of compact theories are given in [6]. But it is worth to note that

Proposition 9.5 *If E is shallow and if $T(F)/_{=_E}$ is infinite then E is compact.*

<u>Proof</u>

The proof is similar to the proof given in [6] for quasi-free theories.

It relies on the following criterion:

Lemma 9.6 ([6]) *If*

- *E is finitary,*

- $T(F)/_{=_E}$ *is infinite,*
- $=_E$ *is decidable,*
- *for every finite set of new constants* C', *for every equation* $s = t$ *such that*
  - $s, t \in T(F \cup C', \{x\})$
  - $s \neq_E t$,

  the set $\{\overline{u} \mid u \in T(F \cup C'),\ s\{x \mapsto u\} =_E t\{x \mapsto u\}\}$ *is finite.*

then E *is compact.*

Now, consider an equation $s = t$ such that $s, t \in T(F \cup C', \{x\})$ and $s \neq_E t$. If we apply the unification rules, as in section 8 (i.e. the rules of figures 1 and 3 with an eager application of quantifier elimination), then, by theorem 8.2, we get an equivalent finite disjunction of formulae of the form $x = t$ where $t \in T(F \cup C')$. Hence $\{\overline{u} \mid u \in T(F \cup C'),\ s\{x \mapsto u\} =_E t\{x \mapsto u\}\}$ is finite. □

9.2 Transformation rules

In compact theories, it is possible to derive the quantifier elimination rules of figure 4:

Lemma 9.7 *If* E *is compact, then the rules of figure 4 preserve the set of solutions.*

Proof

All rules, except the third one, are consequences of the equality axioms. Their correctness does not depend on any particular interpretation. The correctness of the third rule is a direct consequence of the independence of disequations w.r.t. E. □

Moreover, we use the *explosion rule* which expresses the *domain closure axiom* (we know that this rule is necessary, even in the case of a free algebra over a finite alphabet [25, 24]) and the negation of mutation and cycle rules.

Now, we could develop, as in previous sections, a theory of disunification in "syntactic, cycle-syntactic and compact theories". However, the mutation rules add innermost existential quantifiers when they are applied to equations and they add innermost universal quantifiers when they are applied to disequations. This means that the Σ_2 fragment is not closed under the application of this rule. Actually, the mutation rules act, in this case, in the opposite way of quantifier elimination. As we have seen in section 8, these additional quantifiers can be removed eagerly in the case of shallow theories. For these reasons, we will only consider disunification in shallow theories.

The idea is to "lift" universally quantified variables, until a rule of figure 4 can be applied. "Lifting" y consists in transforming the formula into an equivalent one in which y occurs at positions of smaller size (see also the example below). Iterating such liftings will end eventually in situations where y only occurs as a member of equations or disequations. Then the rules of figure 4 can be applied.

$$\forall y: P \implies P$$

If $y \notin Var(P)$.

$$\forall y : P \wedge (y \neq s \vee d) \implies \forall y : P \wedge d\{y \mapsto s\}$$

If y does not occur in s.

$$\forall \vec{y} : P \wedge (y_1 = t_1 \vee \ldots \vee y_n = t_n \vee d) \implies \forall \vec{y} : P \wedge d$$

If 1. $y_1, \ldots, y_n \in \vec{y}$
2. d is a disjunction of equations and disequations without any universally quantified variable
3. For all i, $y_i \notin Var(t_i)$

$$\exists x_1, \ldots, x_n : x_1 \neq t_1 \wedge \ldots \wedge x_n \neq t_n \wedge P \implies P$$

If 1. $x_1, \ldots, x_n \notin Var(P)$
2. For all i, $x_i \neq_E t_i$

Figure 4: Quantifier Elimination Rules

To this end, we use mutation rules which decompose the equations and disequations until (for all equations and disequations), a variable occurs on one side of the equation (resp. disequation). Then, if a universally quantified variable y occurs in an equation $x = t[y]$ (resp. a disequation $x \neq t[y]$), we "explode" x, i.e. we replace x by all possible constructions $f(\vec{z_f})$, which leads hopefully to a lifting of y. There is however a difficulty with the mutation rules: they may decompose only one side of the atomic formula; this may cause some trouble when the side which is not decomposed is the only one which contains universally quantified variables. Let us give an example.

Example 9.8

Let $E = \{f(x) = x\}$ (over the alphabet $F = \{f, g, a\}$) and consider the problem $\forall y : x \neq g(y)$. In order to eliminate the quantifier, we have to lift y until it is a member of an equation or a disequation. We use the domain closure axiom which states that $\exists z : x = a \vee x = f(z) \vee z = g(z)$. The problem becomes

$$(\forall y : x = a \wedge a \neq g(y)) \vee (\exists z, \forall y : x = g(z) \wedge g(z) \neq g(y)) \vee (\exists z, \forall y : x = f(z) \wedge f(z) \neq g(y))$$

Using mutation rules and quantifier elimination rules, it is easy to see that the first disjunct reduces to $x = a$ and the second reduces to $\perp$. But what about the last one? If we use the mutation rule as displayed up to now, we would get

$$\exists z, \forall y : x = f(z) \wedge z \neq g(y)$$

which is a renaming of the original problem.

In order to solve this problem, we first observe that, after an explosion, we need not consider the "one side decomposition rule". In the example above, the last disjunct needs not be considered since all we will get are solutions which are solutions of the first disjunct. This is actually a general rule: we will show that the general *explosion rule* of figure 5 does preserve the set of solutions. The transformation Tr defined on figure 5 is a bit complicated. It only explains that the replacement of x with $f(\vec{z_f})$, if it leads to a mutation somewhere in the formula, this mutation should not involve the one-side decomposition of $f(\vec{z_f})$ (which is the non-termination case of example 9.8).

This might be sufficient for termination, but would lead to very complicated proofs. We prefer to give some more control on the mutations. To this end, we introduce a new predicate symbol $\stackrel{P}{=}$, which has the same interpretation as $=$, but is used operationally in another way: it is a "mark" of the atomic formula. This mark means "Priority" which implies that decompositions and mutations will be performed first on marked equations or disequations. We introduce these symbols in order to make the termination proof easier. In our termination proof, one major component will be the size of the positions of universally quantified variables. This measure is decreasing each time a universally quantified variable is lifted. Applying a mutation may, as we have seen, only decompose one member of an (dis)equation. Assume, as in example 9.8, that the member which is decomposed is precisely the one which does not contain any universally quantified variable. Then our main measure is constant. However, this main component may have been duplicated while putting the formula in conjunctive normal form. The mutation rule itself may also produce more than one such equation out of a single one. In other words, if we have an equation $s = t[y]$, after a mutation, we may get several equations of the form $s' = t[y]$ which will have the same interpretation according to the main measure we consider (the size of the positions of y). This is a problem for termination, as long as s' is not a variable. Imposing some priority on the mutations and decompositions of equations of the form $s' = t[y]$ ensures precisely that, eventually, either y is lifted or s' is a variable.

The whole set of rules ($\mathcal{R}$) consists of the rules displayed on figures 4, 5, 6 and 7. In these rules, $\widehat{E}$ is supposed to be a syntactic presentation defining the same equational theory as E. Similarly, E_c is assumed to be the set of collapsing rules occurring in a cycle-syntactic presentation E^* of the same theory as E. Finally, all rules of $\mathcal{R}$ assume (as before) the symmetry of $=$ and the associativity-commutativity of $\vee$ and $\wedge$.

Let us first establish the correctness of the rules.

Lemma 9.9 *All rules in $\mathcal{R}$ do preserve the set of solutions when E is shallow and $T(F)/_{=_E}$ is infinite.*

Proof

This is a consequence of lemmas 9.7, 7.1 for all rules, except the explosion rule. Indeed, for example, mutations are obtained from the previous mutation rules by applying eagerly an existential quantifier elimination. The additional symbol $\stackrel{P}{=}$ does not change anything for the correctness since it has the same interpretation as $=$: $s = t$ has the same set of solutions as $\stackrel{P}{=}$. Moreover, the rules are (by assumption) the same for $=$ and for $\stackrel{P}{=}$.

Explosion

$$\exists \vec{x}, \forall \vec{y} : P \quad \Longrightarrow \quad \bigvee_{f \in F} \exists \vec{x}, \vec{z_f}, \forall \vec{y} :\ Tr(P, x, f(\vec{z_f}))$$

Where $Tr(P, x, f(\vec{z_f}))$ is defined as follows:

$$\begin{array}{rcll}
Tr(d_1 \wedge \ldots \wedge d_n, x, f(\vec{z_f})) & \stackrel{\text{def}}{=} & Tr(d_1, x, f(\vec{z_f})) \wedge \ldots \wedge Tr(d_n, x, f(\vec{z_f})) & \\
Tr(e_1 \vee \ldots \vee e_m, x, f(\vec{z_f})) & \stackrel{\text{def}}{=} & Tr(e_1, x, f(\vec{z_f})) \vee \ldots \vee Tr(e_m, x, f(\vec{z_f})) & \\
Tr(e, x, f(\vec{z_f})) & \stackrel{\text{def}}{=} & e\{x \mapsto f(\vec{z_f})\} & \text{If } x \text{ is not a member of } e \\
 & & & \text{Or if } e \text{ does not contain a universally quantified variable}
\end{array}$$

$$\begin{array}{rcl}
Tr(x = g(\vec{t}), x, f(\vec{z_f})) & \stackrel{\text{def}}{=} & \displaystyle\bigvee_{f(\vec{l}) = g(\vec{r}) \in \widehat{E} \cup R(F)} \Big(\bigwedge_{\substack{l_i \equiv r_j \\ l_i \text{variable}}} z_i = t_j \bigwedge_{r_i \text{ constant}} r_i = t_i \bigwedge_{\substack{r_i \equiv r_j \\ r_i \text{ variable} \\ i \neq j}} t_i = t_j \Big) \\
 & & \displaystyle\bigvee_{g(\vec{l}) = x' \in \widehat{E}} \Big(\bigwedge_{\substack{l_i \equiv l_j \\ l_i \text{ variable} \\ i \neq j}} t_i = t_j \bigwedge_{l_i \text{ constant}} l_i = t_i \bigwedge_{l_i \equiv x'} t_i = f(\vec{z_f}) \Big))
\end{array}$$

If $\vec{t}$ contains a universally quantified variable

$$\begin{array}{rcl}
Tr(x \neq g(\vec{t}), x, f(\vec{z_f})) & \stackrel{\text{def}}{=} & \displaystyle\bigwedge_{f(\vec{l}) = g(\vec{r}) \in \widehat{E} \cup R(F)} \Big(\bigvee_{\substack{l_i \equiv r_j \\ l_i \text{variable}}} z_i \neq t_j \bigvee_{r_i \text{ constant}} r_i \neq t_i \bigvee_{\substack{r_i \equiv r_j \\ r_i \text{ variable} \\ i \neq j}} t_i \neq t_j \Big) \\
 & & \displaystyle\bigwedge_{g(\vec{l}) z = x' \in \widehat{E}} \Big(\bigvee_{\substack{l_i \equiv l_j \\ l_i \text{ variable} \\ i \neq j}} t_i \neq t_j \bigvee_{l_i \text{ constant}} l_i \neq t_i \bigvee_{l_i \equiv x'} t_i \neq f(\vec{z_f}) \Big))
\end{array}$$

If $\vec{t}$ contains a universally quantified variable

If

1. there is an equation $x = g(\vec{t})$ or a disequation $x \neq g(\vec{t})$ in P such that $\vec{t}$ contains an occurrence of a universally bound variable.
2. x is not universally bound and $\vec{z} \cap (Var(P) \cup \vec{y}) = \emptyset$
3. no other rule can be applied
4. $R(F)$ is the set of functional reflexive axioms $f(\vec{x}) = f(\vec{x})$, for $f \in F$

Figure 5: Explosion rules

Mutate 1

$$f(\vec{s}) = g(\vec{t}) \quad\Longrightarrow\quad \bigvee_{f(\vec{l})=g(\vec{r})\in\widehat{E}\cup R(F)} \Big(\bigwedge_{\substack{l_i\equiv l_j\\ l_i \text{ variable}\\ i\neq j}} s_i = s_j \bigwedge_{\substack{r_i\equiv r_j\\ r_i\text{ variable}\\ i\neq j}} t_i=t_j \bigwedge_{\substack{l_i\equiv r_j\\ l_i\text{ variable}}} s_i = t_j$$

$$\bigwedge_{\mathrm{l_i \ constant}} l_i = s_i \bigwedge_{\mathrm{r_i \ constant}} r_i = t_i\Big)$$

$$\vee \bigvee_{f(\vec{l})=x\in\widehat{E}} \Big(\bigwedge_{\substack{l_i\equiv l_j\\ l_i \text{ variable}\\ i\neq j}} s_i = s_j \bigwedge_{\mathrm{l_i \ constant}} l_i = s_i \bigwedge_{\mathrm{l_i \equiv x}} s_i \overset{P?}{=} g(\vec{t})\Big)$$

$$\vee \bigvee_{g(\vec{r})=x\in\widehat{E}} \Big(\bigwedge_{\substack{r_i\equiv r_j\\ r_i \text{ variable}\\ i\neq j}} t_i = t_j \bigwedge_{\mathrm{r_i \ constant}} r_i = t_i \bigwedge_{\mathrm{r_i \equiv x}} t_i = f(\vec{s})\Big)$$

Mutate 2

$$f(\vec{s}) \neq g(\vec{t}) \quad\Longrightarrow\quad \bigwedge_{f(\vec{l})=g(\vec{r})\in\widehat{E}\cup R(F)} \Big(\bigvee_{\substack{l_i\equiv l_j\\ l_i \text{ variable}\\ i\neq j}} s_i \neq s_j \bigvee_{\substack{r_i\equiv r_j\\ r_i\text{ variable}\\ i\neq j}} t_i \neq t_j \bigvee_{\substack{l_i\equiv r_j\\ l_i\text{ variable}}} s_i \neq t_j$$

$$\bigvee_{\mathrm{l_i \ constant}} l_i \neq s_i \bigvee_{\mathrm{r_i \ constant}} r_i \neq t_i\Big)$$

$$\wedge \bigwedge_{f(\vec{l})=x\in\widehat{E}} \Big(\bigvee_{\substack{l_i\equiv l_j\\ l_i \text{ variable}\\ i\neq j}} s_i \neq s_j \bigvee_{\mathrm{l_i \ constant}} l_i \neq s_i \bigvee_{\mathrm{l_i \equiv x}} s_i \overset{P?}{\neq} g(\vec{t})\Big)$$

$$\wedge \bigwedge_{g(\vec{r})=x\in\widehat{E}} \Big(\bigvee_{\substack{r_i\equiv r_j\\ r_i \text{ variable}\\ i\neq j}} t_i \neq t_j \bigvee_{\mathrm{r_i \ constant}} r_i \neq t_i \bigvee_{\mathrm{r_i \equiv x}} t_i \neq f(\vec{s})\Big)$$

These rules assume that $g(\vec{t})$ contains at least one occurrence of a universally quantified variable. The symbol $\overset{P?}{=}$ (resp. $\overset{P?}{\neq}$) stands for $=$ (resp. $\neq$) if $f(\vec{s})$ contains at least one occurrence of a universally quantified variable, and for $\overset{P}{=}$ (resp. $\overset{P}{\neq}$) otherwise.

We assume exactly the same rules with $f(\vec{s}) \overset{P}{=} g(\vec{t})$ instead of $f(\vec{s}) = g(\vec{t})$ in the left sides (resp. $f(\vec{s}) \overset{P}{\neq} g(\vec{t})$ instead of $f(\vec{s}) \neq g(\vec{t})$)

Figure 6: Mutations rules for disunification in shallow theories

Cycle 1

$$t[x]_p = x \quad \Longrightarrow \quad \bigvee_{\substack{q<p \\ f(\vec{l})\to r \in E_c}} \Big((t[r]_q = x)\{l_{i_1} \mapsto v_{i_1};\ldots;l_{i_k} \mapsto v_{i_k}\} \bigwedge_{\substack{l_i \equiv l_j \\ l_i \text{ variable} \\ i \neq j}} v_i = v_j \bigwedge_{\text{l}_\text{i} \text{ constant}} l_i = v_i\Big)$$

Cycle 2

$$t[x]_p \neq x \quad \Longrightarrow \quad \bigwedge_{\substack{q<p \\ f(\vec{l})\to r \in E_c}} \Big((t[r]_q \neq x)\{l_{i_1} \mapsto v_{i_1};\ldots;l_{i_k} \mapsto v_{i_k}\} \bigvee_{\substack{l_i \equiv l_j \\ l_i \text{ variable} \\ i \neq j}} v_i \neq v_j \bigvee_{\text{l}_\text{i} \text{ constant}} l_i \neq v_i\Big)$$

If

1. x is universally bound.
2. $t|_q \equiv f(\vec{v})$
3. $\{l_{i_1},\ldots,l_{i_k}\} = \mathcal{V}ar(\vec{l})$

$\stackrel{P}{=}$-elimination

$$x \stackrel{P}{=} s \quad \Longrightarrow \quad x = s$$
$$x \stackrel{P}{\neq} s \quad \Longrightarrow \quad x \neq s$$

If x is a variable.

Delete

$$s = t \quad \Longrightarrow \quad \top$$
$$s \neq t \quad \Longrightarrow \quad \bot$$

If $s =_E t$.

Figure 7: Disunification rules: solving cycles

Now let us consider the explosion rule. The only difficulty is to explain why we can omit the disjunction $\bigvee_{f(\vec{l})=x} \ldots$. Obviously, all solutions of the right hand side of the rule are also solution of the left hand side. Let now γ be a solution of the left hand side. This means that, for all assignments δ to the variables in $\vec{y}$, $\gamma\delta \models P \wedge (d \vee x = g(\vec{t}))$. Let u be an irreducible form of $x\gamma$ w.r.t. $\widehat{E}$ (let us recall that $\widehat{E}$ is ground convergent by lemma 4.10). By lemma 4.10 again, for every v such that $v \xleftrightarrow[\widehat{E}]{*} u$, there is a proof $u \xleftarrow[\widehat{E}]{\Lambda} \xleftarrow[\widehat{E}]{(\neq\Lambda)^*} v$ or else $u \xleftarrow[\widehat{E}]{(\neq\Lambda)^*} v$. Let f be the top symbol of u: $u \equiv f(\vec{z_f})\lambda$. Then, it is sufficient to show that, each time $\gamma\delta$ is a solution of an atomic formula ϕ occurring in the left hand side of the rule, $\gamma\delta\sigma$ is a solution of the formula by which ϕ has been replaced in the right hand side. This is straightforward when ϕ is replaced with $\phi\{x \mapsto f(\vec{z_f})\}$. Let us now consider the interesting parts of the right hand side. $x\gamma \xleftrightarrow{*} g(\vec{t})\gamma\delta$ iff $x\gamma \xleftrightarrow{*} u \xleftarrow{\Lambda^=} \xleftarrow{(\neq\Lambda)^*} g(\vec{t})\gamma\delta$. The rules applied at position Λ are either non-collapsing or else of the form $g(\vec{l}) \to x$, applied from right to left. In the first case, $\gamma\delta\lambda$ satisfies the first disjunct $(\bigwedge z_i = t_j \bigwedge r_i = t_i \bigwedge l_i = t_j)$ and in the second case $\gamma\delta\lambda$ satisfies the second disjunct $(\bigwedge t_i = t_j \bigwedge l_i = t_i \bigwedge t_i = f(\vec{z_f}))$ of the right hand side of the rule. $\square$

9.3 Termination

Now, the main result is the termination of $\mathcal{R}$:

Theorem 9.10 *If mutations are applied first to equations $s \stackrel{P}{=} t$ and disequations $s \stackrel{P}{\neq} t$ as long as possible and the other rules are applied only on problems without these marked predicates, then any reduction w.r.t. $\mathcal{R}$ is terminating.*

Proof

Let first $\mathcal{R}_0$ be the system consisting of the rules whose left hand side is an annotated atom $s \stackrel{P}{=} t$ or $s \stackrel{P}{\neq} t$ only. $\mathcal{R}_0$ is terminating. For, consider the multiset, for all equations $s \stackrel{P}{=} t$ and disequations $s \stackrel{P}{\neq} t$ occurring in the formula of the sum $|s| + |t|$. This multiset is strictly decreasing (w.r.t. the multiset extension of the ordering on natural numbers) by any application of these rules. Applying $\mathcal{R}_0$ eventually results in a formula which does not contain the symbol $\stackrel{P}{=}$.

Now, we interpret the formulae in a well-founded domain in such a way that the interpretation is decreasing by any application of a rule: the interpretation of ϕ is the interpretation of any normal form of ϕ w.r.t. $\mathcal{R}_0$. Such normal forms (which are supposed to be disjunctions of formulae $\exists\vec{x}, \forall\vec{y} : P$ where P is quantifier-free; this assumes some normalization at each reduction step) are interpreted as follows:

- $I(\bigvee_{i\in I} \exists\vec{x_i}, \forall\vec{y_i} : P_i)$ where P is quantifier-free is the multiset of interpretations of $\forall\vec{y_i} : P_i$.
- $I(\forall\vec{y} : d_1 \wedge \ldots \wedge d_n)$ where each d_i is a disjunction of atomic formulae, is the multiset of interpretations of formulae $\forall\vec{y} : d_i$
- $I(\forall\vec{y} : e_1 \vee \ldots \vee e_n)$ where each e_i is either an equation $s_i = t_i$ or a disequation $s_i \neq t_i$ is a pair consisting of

- the number $|Var(e_1, \ldots, e_n) \cap \vec{y}|$ of universally quantified variables occurring in the disjunction
- the multiset of interpretations of formulae $\forall \vec{y} : \; s_i = t_i$

- $I(\forall \vec{y} : s = t) = I(\forall \vec{y} : s \neq t)$ is a triple which consists of
  - the sum of sizes of all positions of a variable $y \in \vec{y}$ in s or t:

$$I(\forall \vec{y} : s = t) = \sum_{s|_p \in \vec{y}} |p| + \sum_{t|_q \in \vec{y}} |q|$$

  - the number 0 if s or t (or both) is a variable, or if $Var(s,t) \cap \vec{y} = \emptyset$ and 1 otherwise.
  - the number 0 if $s = t$ is not an equation of the form $y = t[y]$ where $y \in \vec{y}$, and $|t[y]|$ otherwise.

We use the multiset and lexicographic extensions in the usual way for the definition of the ordering on the interpretation. Let us show that the interpretation of any formula is decreased by application of any rule of $\mathcal{R}$. First note some kind of "monotonicity" of the interpretation: assume that an atomic formula a is replaced with any Boolean combination of atomic formulae which are strictly smaller in the interpretation:

$$\exists \vec{x}, \forall \vec{y} : P \wedge (a \vee d) \Longrightarrow \bigvee_{i \in I} \exists \vec{x_i}, \forall \vec{y} : P \wedge (b \vee d)$$

The right side is then normalized into $\bigvee_{i \in I} \phi_i$ where each ϕ_i is in prenex conjunctive normal form. Altogether, we replace a formula ϕ with a formula ψ by this process. If $Var(b \vee d) \cap \vec{y} \subseteq Var(a \vee d) \cap \vec{y}$, then $I(\psi) < I(\phi)$, by definition of the multiset extension of an ordering.[2] This shows, that, for the rules of the above form (or which are obtained by iterating some rules of the above form) we have only to check the decreasingness locally. Let us now consider each family of rules:

Quantifier Eliminations The decreasingness of the interpretation is straightforward: one universally quantified variable is eliminated in some disjunction for the two first rules.

Mutations For any term s occurring in a formula ϕ, let $|s|_U$ be the sum of the sizes of positions of universally quantified variables in s. For all atomic formulae $s = t$ or $s \neq t$ in the right side of the mutation rules whose left hand side is $u = v$ (or $u \neq v$), $|s|_U + |t|_U < |u|_U + |v|_U$. (This is because we only apply mutations to equations $s = t$ or disequations $s \neq t$ such that s or t contains a universally quantified variable). If the mutation introduces an equation $s \overset{P}{=} t$ (or a disequation $s \overset{P}{\neq} t$), then $|s|_U + |t|_U = |u|_U + |v|_U$. Hence, in irreducible formulae w.r.t. $\mathcal{R}_0$, we find only atoms $s = t$ where $|s|_U + |t|_U < |u|_U + |v|_U$ or else atoms $x = u$ or $x = v$ where x is a variable. In this latter case, $I(x = u) = (N, 0, k) < I(u = v) = (M, 1, 0)$ since $N \leq M$. In all cases, after applying a mutation rule, all atomic formulae in the irreducible form w.r.t. $\mathcal{R}_0$ of the right hand side are strictly smaller than the left hand side. This means that the interpretation is indeed decreasing, by definition of the multiset extensions.

[2] This can be shown by induction on the number of nested conjunctions. Assume for example that b is a conjunction of disjunctions: $b = b_1 \wedge \ldots \wedge b_n$, normalizing the formula leads to $\bigvee_{i \in I} \exists \vec{x_i}, \forall \vec{y} : P \wedge (b_1 \vee d) \wedge \ldots \wedge (b_n \vee d)$. And $I(P \wedge (a \vee d)) = I(P) \cup \{(k, I(\forall \vec{y} : d) \cup \{I(\forall \vec{y} : a)\})\}$ whereas $I(P \wedge (b_1 \vee d) \wedge \ldots \wedge (b_n \vee d)) = I(P) \cup \{(k_1, I(\forall \vec{y} : d) \cup \{I(\forall \vec{y} : b_1)\}), \ldots (k_n, I(\forall \vec{y} : d) \cup \{I(\forall \vec{y} : b_n)\})\}$. On the other hand, by hypothesis, $k_i \leq k$ and $I(\forall \vec{y} : b_i) < I(\forall \vec{y} : a)$, hence, by definition of the multiset extension of an ordering, I is strictly decreasing.

Explosion As above, it is sufficient to show that, for every atomic formula ψ, ψ is replaced either with an atomic formula ψ' such that $I(\psi') \leq I(\psi)$ or else ψ is replaced with a Boolean combination of atomic formulae ψ_i such that, for all i, $I(\psi_i) < I(\psi)$, and this last situation occurs for at least one atomic formula.

According to the definition of Tr, for every atomic formula ψ, there are 3 cases

1. ψ does not contain any universally quantified variable. In this case, ψ is simply replaced with $\psi\{x \mapsto f(\vec{z_f})\}$ which has the same interpretation as ψ. (I.e. $(0,0,0)$)
2. ψ does contain a universally quantified variable, but x is not a member of ψ. Then ψ has the form $z = u$ or $z \neq u$ where z is an existentially quantified variable or a free variable. (This is so because explosion is applied only if no other rule can be applied). Then ψ is replaced with $\psi\{x \mapsto f(\vec{z_f})\}$ and $I(\psi) = (N,0,0) = I(\psi\{x \mapsto f(\vec{z_f})\})$.
3. ψ has x as a member and $I(\psi) = (N,0,0)$. Then, according to the definition of Tr, a lot of new atomic formulae are introduced. Let us check that their interpretation is indeed smaller than $I(\psi)$:
   - $I(z_i = t_j) = (\sum_{t_j|_p \in \vec{y}} |p|, 0, 0) < (\sum_{g(\vec{t})|_q \in \vec{y}} |q|, 0, 0)$ since $g(\vec{t})$ is assumed to contain at least one occurrence of a universally quantified variable. Similarly, $I(r_i = t_i), I(l_i = t_i), I(z_i \neq t_j), I(r_i \neq t_i), I(l_i \neq t_i)$ are strictly smaller than $I(\psi)$.
   - $I(t_i = t_j) = I(t_i \neq t_j) = (\sum_{t_i|_p \in \vec{y}} |p| + \sum_{t_j|_q \in \vec{y}} |q|, \epsilon, k)$ ($\epsilon \in \{0,1\}$). The fist component of the pair is strictly smaller than the first component of the interpretation of ψ since $i \neq j$.
   - $I(t_i = f(\vec{z_f})) = I(t_i \neq f(\vec{z_f})) = (|t_i|_U, 1, 0) < I(\psi)$ since $|g(\vec{t})|_U > |t_i|_U$ and the comparison of the pair is lexicographic.

Cycles As for the previous rules, all atomic formulae $v_i = v_j$, $l_i = v_j$, $v_i \neq v_j$, $l_i \neq v_j$ are strictly smaller than the left hand side. It only remains to consider the atomic formula $(t[r]_q = x)\{l_{i_1} \mapsto v_{i_1}; \ldots; l_{i_k} \mapsto v_{i_k}\}$ (resp. its negation). According to lemma 6.5, r is a linear term when $f(\vec{l}) \to r \in E_c$. This shows that $|t[r]_q\{l_{i_1} \mapsto v_{i_1}; \ldots; l_{i_k} \mapsto v_{i_k}\}|_U \leq |t[x]_p|_U$. But it can be equal. In case of equality, $I(t[x]_p, = x) = (N, 0, |t[x]_p|)$ and $I((t[r]_q = x)\{l_{i_1} \mapsto v_{i_1}; \ldots; l_{i_k} \mapsto v_{i_k}\}) = (N, 0, |(t[r]_q = x)\{l_{i_1} \mapsto v_{i_1}; \ldots; l_{i_k} \mapsto v_{i_k}\}|)$ which is strictly smaller, thanks to the properties of collapsing rules.

Other rules The other rules obviously do not increase the interpretation. They are terminating on their own since they remove parts of the formula.

$\square$.

$\mathcal{R}$ is also complete:

Lemma 9.11 *Equational formulae that are irreducible w.r.t. the rules of $\mathcal{R}$ are purely existential.*

Proof

We only have to check that each formula which contains an occurrence of a universally quantified variable is reducible by $\mathcal{R}$. Let y a universally quantified variable

- If y occurs in an equation or a disequation between two non-variable terms, then a mutation rule applies.
- If y occurs in an equation or a disequation whose one member is a universally quantified variable, then either a universal quantifier elimination rules can be applied or else a **Cycle** or a **Delete** rule.

Now, if all universally quantified variables occur in equations $x = t[y]_p$ or in disequations $x \neq t[y]_p$ with $p \neq \Lambda$, **Explosion** can be applied. $\square$

9.4 Decidability of the first-order theory of $T(F)/_{=_E}$

Now, we are left to solve purely existential formulae. For, we can use a last result of [6]:

Theorem 9.12 ([6]) *If E is compact, then the existential fragment of the theory of $T(F)/_{=_E}$ is decidable.*

Putting together all previous results, we get the following:

Corollary 9.13 *If $=_E$ is a shallow equational theory then the first order theory of $T(F)/_{=_E}$ is decidable. Actually, there is a (recursive) complete axiomatization of this algebra that can directly be borrowed to the set of rules .*

<u>Proof</u>
If $T(F)/_{=_E}$ is infinite (which can be decided according to the results of the next section), then this is a consequence of theorem 9.10, lemma 9.11, proposition 9.5 and theorem 9.12. If $T(F)/_{=_E}$ is finite, the first-order theory of $T(F)/_{=_E}$ is equivalent to quantified Boolean formulae (hence it is PSPACE-complete). $\square$

Actually, inspecting the rules that are used in the transformation of purely universally quantified sentences which do not involve negation, we may derive the following result:

Corollary 9.14 *When $T(F)/_{=_E}$ is infinite and E is a shallow presentation, then the equational theory is ω-complete. (I.e. the inductive theory equals the equational theory).*

10 Finiteness of $T(F)/_{=_E}$

In this section, we show that the finiteness of $T(F)/_{=_E}$ is decidable when E is shallow. The infiniteness of the quotient is a property which has been actually used in the previous section. The proof is based on a pumping technique.
Let E be a shallow presentation. $T(F)/_{=_E}$ is finite iff $T(F^*)/_{=_{E^*}}$ is finite. Therefore, we may assume in this section the properties of E^* which are described in section 6. For example, by lemma 6.6, E^* is ground convergent, according to the ordering $\geq$ on terms, extending the precedence $>_{F^*}$ on F^*.

Lemma 10.1 *Let f be the greatest symbol of F^* and $a_1, \ldots, a_n$ be constants. If $f(a_1, \ldots, a_n)$ is irreducible and t is any irreducible term of depth at least 1, then $f(a_1, \ldots, a_{n-1}, t)$ is irreducible.*

Proof

The only possible reduction is at the root position. Assume that $f(a_1, \ldots, a_{n-1}, t)$ is reducible. Then there is an equation $f(s_1, \ldots, s_n) = g(t_1, \ldots, t_m) \in E^*$ and a substitution σ such that

1. $s_n\sigma \equiv t$
2. for all $i \leq n-1$, $s_i\sigma \equiv a_i$
3. $f(a_1, \ldots, a_{n-1}, t) > g(t_1, \ldots, t_m)\sigma$

The first condition implies that s_n is a variable (because E^* is shallow) and that it occurs only once in $f(s_1, \ldots, s_n)$. Now, $f(a_1, \ldots, a_n)$ being irreducible, we must have $f(s_1, \ldots, s_n)\theta = f(a_1, \ldots, a_n) \leq g(t_1, \ldots, t_m)\theta$ where θ is identical to σ but on s_n where $s_n\theta = a_n$. Which means first that $f(s_1, \ldots, s_n) = g(t_1, \ldots, t_m)$ cannot be oriented into a rule. Then, thanks to lemma 6.5 and by maximality of f, we must have $f = g$ and $\vec{s} = \vec{t}$ up to permutation. In particular, s_n is one of the t_i's and $t \equiv t_i\sigma$. By condition 3 and because the status of f is lexicographic, we must have $i = n$, i.e. $t_n \equiv t$. Moreover, for every $i \leq n-1$, $a_i \geq t_i\sigma$ by the condition 3 and $a_i \leq t_i\sigma$ because $f(a_1, \ldots, a_n) \leq f(t_1, \ldots, t_n)\sigma$. This proves that $t_i\sigma \equiv a_i$ for all $i \leq n-1$. We get a contradiction since we proved that $f(s_1, \ldots, s_n)\sigma$ and $g(t_1, \ldots, t_m)\sigma$ have to be identical. □

Lemma 10.2 *$T(F^*)/_{=_{E^*}}$ is infinite iff there is an irreducible term of the form $g(a_1, \ldots, a_m)$ where $a_1, \ldots, a_m$ are constants and g is not a constant.*

Proof

The only if part of this lemma is straightforward since F^* is finite. Now, assume that there is an irreducible term $g(a_1, \ldots, a_m)$. Let F_0 be the set of function symbols that are smaller or equal to g and let E_0 be the subset of E^* which contains all equations whose functions symbols belong to F_0. F_0, E_0 satisfy the same properties as E^*, F^* i.e. lemmas 6.2, 6.5, 6.6 still hold true if we replace E^*, F^* with E_0, F_0 (if the reader is not convinced, then replace F_0, E_0 with F_0^*, E_0^* obtained by the completion of section 6; we may keep g maximal since every symbol introduced by **Divide** has a strictly smaller arity. The maximality of g is all we need). Hence, by lemma 10.1, $g(a_1, \ldots, a_{m-1}, t)$ is irreducible for all irreducible t. In particular, we can construct by induction the infinite sequence of irreducible terms: $t_0 = g(a_1, \ldots, a_m)$, $t_{n+1} = g(a_1, \ldots, a_{n-1}, t_n)$. Now, any two terms in $T(F_0)$ that are irreducible by E_0 are also irreducible by E^* because any equation in $E^* \setminus E_0$ must contain a symbol $f >_{F^*} g$. By ground convergence of E^*, this means that $T(F^*)/_{=_{E^*}}$ is infinite. □

As an easy consequence of this lemma, we get:

Theorem 10.3 *$T(F)/_{=_E}$ is finite iff the only irreducible terms in $T(F^*)$ are constants. Moreover, there is a terminating algorithm which decides whether $T(F)/_{=_E}$ is finite or not and, in case of finiteness, gives explicitly a finite set of representatives.*

11 Conclusion

This paper is a start of a systematic study of shallow presentations. Besides proving new decidability and complexity results, we are interested into two kinds of questions:

We think that the notion of a shallow theory can be extended so as to encompass many more practical cases, by allowing non-overlapping contexts on top of shallow axioms. Finite automata techniques would then be used in order to deterministically get rid of contexts before to use the rules described here.

Considering ground rules rather than equations, all interesting questions are again decidable, e.g. termination [14], confluence [8, 28], and reachability [9]. Moreover, a set of ground equations can always be transformed into a finite convergent set of ground rules in polynomial time [13, 30]. Unfortunately, shallow rewriting does not enjoy all properties of ground rewriting: termination is indeed decidable, but the first order theory of the embedding relation on ground terms is known to be undecidable [31], although embedding can be expressed as the reachability property of the rewrite system defined by finitely many shallow rewrite rules of the form $f(\ldots, x_i, \ldots) \to x_i$. On the other hand, it would be very interesting to solve the decidability of the existential fragment of the first order theory of embedding, since this would open the way to the decidability of the same fragment for the recursive path ordering[3], which is known for a total precedence only [3, 17].

References

[1] A. Boudet, J.-P. Jouannaud, and M. Schmidt-Schauß. Unification in Boolean rings and Abelian groups. *Journal of Symbolic Computation*, 8:449–477, 1989.

[2] J. Christian. Some termination criteria for narrowing and e-narrowing. In D. Kapur, editor, *Proc. 11th Int. Conf. on Automated Deduction, Saratoga Springs, NY, LNCS 607*, pages 582–588, 1992.

[3] H. Comon. Solving symbolic ordering constraints. *International Journal of Foundations of Computer Science*, 1(4):387–411, 1990.

[4] H. Comon. Complete axiomatizations of some quotient term algebras. In *Proc. 18th Int. Coll. on Automata, Languages and Programming, Madrid, LNCS 510*, July 1991.

[5] H. Comon. Disunification: a survey. In J.-L. Lassez and G. Plotkin, editors, *Computational Logic: Essays in Honor of Alan Robinson*. MIT Press, 1991.

[6] H. Comon. Complete axiomatizations of some quotient term algebras. *Theoretical Comput. Sci.*, 118(2), Sept. 1993.

[7] H. Comon and P. Lescanne. Equational problems and disunification. *Journal of Symbolic Computation*, 7:371–425, 1989.

[8] M. Dauchet, T. Heuillard, P. Lescanne, and S. Tison. The confluence of ground term rewriting systems is decidable. In *Proc. 3rd IEEE Symp. Logic in Computer Science, Edinburgh*, 1988.

[3]The recursive path ordering with an empty precedence is the embedding.

[9] M. Dauchet and S. Tison. The theory of ground rewrite systems is decidable. Research Report IT 182, Laboratoire d'Informatique Fondamentale de Lille, Université des Sciences et Techniques de Lille Flandres Artois, France, Mar. 1990. Also in Proc. 5th IEEE Symp. Logic in Computer Science, Philadelphia.

[10] N. Dershowitz and J.-P. Jouannaud. Rewrite systems. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 243–309. North-Holland, 1990.

[11] M. Fay. First-order unification in an equational theory. In *Proc. 4th Workshop on Automated Deduction, Austin, Texas*, pages 161–167, 1979.

[12] M. Fernández. Narrowing based procedures for equational disunification. *Applicable Algebra in Engineering Communication and Computing*, 3:1–26, 1992.

[13] J. Gallier, P. Narendran, D. Plaisted, S. Raatz, and W. Snyder. An algorithm for finding canonical sets of ground rewrite rules in polynomial time. Draft, Jan. 1988.

[14] G. Huet and D. S. Lankford. On the uniform halting problem for term rewriting systems. Research Report 283, INRIA, Mar. 1978.

[15] J.-P. Jouannaud. Syntactic theories. In B. Rovan, editor, *Proc. 15th Mathematical Foundations of Computer Science, Banskà Bystrica*, 1990. LNCS 452.

[16] J.-P. Jouannaud and C. Kirchner. Solving equations in abstract algebras: A rule-based survey of unification. In J.-L. Lassez and G. Plotkin, editors, *Computational Logic: Essays in Honor of Alan Robinson*. MIT-Press, 1991.

[17] J.-P. Jouannaud and M. Okada. Satisfiability of systems of ordinal notations with the subterm property is decidable. In *Proc. 18th Int. Coll. on Automata, Languages and Programming, Madrid, LNCS 510*, 1991.

[18] C. Kirchner. Computing unification algorithms. In *Proc. 1st IEEE Symp. Logic in Computer Science, Cambridge, Mass.*, pages 206–216, 1986.

[19] C. Kirchner and F. Klay. Syntactic theories and unification. In *Proc. 5th IEEE Symp. Logic in Computer Science, Philadelphia*, June 1990.

[20] F. Klay. Undecidable properties of syntactic theories. In *Proc. 4th Rewriting Techniques and Applications, Como, LNCS 488*, 1991.

[21] D. Kozen. Complexity of finitely presented algebras. In *Proc. 9th STOC*, pages 164–177, 1977.

[22] D. Kozen. Positive first order logic is NP-complete. *IBM Journal of Res. Developp.*, 25(4):327–332, July 1981.

[23] J.-L. Lassez, M. J. Maher, and K. G. Marriot. Unification revisited. In *Proc. Workshop on Found. of Logic and Functional Programming, Trento, LNCS 306*. Springer-Verlag, Dec. 1986.

[24] M. J. Maher. Complete axiomatizations of the algebras of finite, rational and infinite trees. In *Proc. 3rd IEEE Symp. Logic in Computer Science, Edinburgh*, pages 348–357, July 1988.

[25] A. Mal'cev. On the elementary theories of locally free algebras. *Soviet Math. Doklady*, 1961.

[26] G. Nelson and D. C. Oppen. Fast decision procedures based on congruence closure. *J. ACM*, 27:356–364, 1980.

[27] T. Nipkow. Proof transformations for equational theories. In *Proc. 5th IEEE Symp. Logic in Computer Science, Philadelphia*, June 1990.

[28] M. Oyamaguchi. The church-rosser property for ground term rewriting systems is decidable. *Theoretical Comput. Sci.*, 49:43–79, 1987.

[29] R. E. Shostak. An efficient decision procedure for arithmetic with function symbols. *J. of the Association for Computing Machinery*, 26(2):351–360, Apr. 1979.

[30] W. Snyder. Efficient ground completion: A fast algorithm for generating reduced ground rewriting systems from a set of ground equations. Boston University, Nov. 1989.

[31] R. Treinen. A new method for undecidability proofs of first order theories. Tech. Report A-09/90, Universität des Saarladandes, Saarbrücken, May 1990.