Decidability and Complexity Analysis by Basic Paramodulation*

Robert Nieuwenhuis
Technical University of Catalonia
Jordi Girona 1, 08034 Barcelona, Spain
E-mail: roberto@lsi.upc.es

*An extended abstract of an early version of this paper was published in proc. LICS'96. The author is partially supported by the EU HCM network *Console* and the EU Esprit WG CCL-II.

Running head:

Decidability by Basic Paramodulation

Please send proofs to:

Robert Nieuwenhuis
Technical University of Catalonia
Dept. LSI, Campus Nord C6
Jordi Girona 1, 08034 Barcelona, Spain
E-mail: roberto@lsi.upc.es

Abstract

It is shown that for sets of Horn clauses saturated under *basic paramodulation*, the word and unifiability problems are in NP, and the number of minimal unifiers is simply exponential (i). For Horn sets saturated wrt. a special ordering under the more restrictive inference rule of *basic superposition*, the word and unifiability problems are still decidable and unification is finitary (ii). These two results are applied to the following languages.

For *shallow* presentations (equations with variables at depth at most one) we show that the closure under paramodulation can be computed in polynomial time. Applying result (i), it follows that shallow unifiability is in NP, which is optimal since unifiability in ground theories is already NP-hard. The shallow word problem is even shown to be polynomial.

Generalizing shallow theories to the Horn case, we obtain (two versions of) a language we call *Catalog*, a natural extension of Datalog to include functions and equality. The closure under paramodulation is finite for Catalog sets, hence (i) still applies. For Catalog sets S the decidability of the full first-order theory of $\mathcal{T}(\mathcal{F})/_{=_S}$ is shown as well.

Finally we define *standard theories*, which include and significantly extend *shallow* theories. Standard presentations can be finitely closed under superposition and result (ii) applies, thus obtaining a new fundamental class with decidable word and unifiability problems and where unification is finitary.

1 Introduction

Word and unification problems (entailment and answer computation) in equational Horn theories have significant applications in logic in computer science: (constraint) logic and functional programming, automated deduction, knowledge-based systems, computational linguistics, etc. Hence the decidability and computational feasibility of these problems for diverse classes of formulae have been extensively studied.

Some of these classes are defined by a combination of semantic and syntactic restrictions on the language. For example, for the class of ground equations where some symbols are associative and commutative (AC) a finite convergent rewrite system can always be computed [Narendran and Rusinowitch, 1991, Marché, 1991] by which the word problem is decidable. In the same class, the unification problem is also decidable [Narendran and Rusinowitch, 1993] (see also [Marche, 1994] for decidability of word problems in ground presentations *modulo* several other theories different from AC).

Some well-known classes of first-order formulae have been proved decidable by means of model-theoretic methods, and satisfiability of their corresponding clause sets can also be decided by means of more practical methods like resolution and paramodulation. This is the case e.g. for the Ackermann class [Fermüller and Salzer, 1993] or the *monadic* class, which is moreover equivalent to a type of *set constraints* [Bachmair et al., 1993a, Bachmair et al., 1993b].

An important purely syntactically defined class is the Datalog language of *flat* Horn clauses without equality, which is well-known from deductive databases.

Another such class are the equational *shallow theories*, the ones axiomatized by equations where all variables are *shallow*, i.e. they appear only at depth at most one in each side of the equation. This is a fundamental class with decidable word and unification problems and even a decidable first-order theory [Comon et al., 1994]. These results on shallow theories subsume much earlier ones on classes like *ground, permutative, compact* or *quasi-free* theories. They were obtained, following [Kirchner, 1986], by transforming shallow presentations into equivalent *(cycle-)syntactic* ones, for which complete and terminating unification rules were derived (cf. also the techniques for dealing with shallow equations and rewrite rules given in [Christian, 1992, Domenjoud, 1993]).

Here we proceed in a completely different way, namely by a careful termination analysis of *basic* paramodulation [Nieuwenhuis and Rubio, 1995, Bachmair et al., 1995], for Horn clauses with equality. Basicness means in this context that no inferences are needed on certain *blocked* terms of the clauses, typically the terms created in unifiers of previous inferences, as in *basic narrowing*.

Completeness results have been given for several basic strategies. The main idea behind these results is roughly to consider a particular rewrite strategy at the ground level, which is lifted to the non-ground level. For example in innermost rewrite strategies a rewrite step at a certain position is performed only if all its subterms are irreducible; when lifting such a rewrite step to an inference, one can hence impose irreducibility restrictions on these subterms and block them for inferences.

In basic strategies, a trade-off exists between the number of positions that can be blocked and the ordering restrictions of the inference rules. Let us consider for instance an (in this case purely equational) paramodulation inference:

$$\frac{l \simeq r \qquad s \simeq t}{(s[r]_p \simeq t)\sigma} \quad \text{where } \sigma \text{ is the mgu of } l \text{ and } s|_p.$$

Parameterized by an ordering on terms $\succ$, such an inference is only needed if it has some ground instance θ compatible with σ such that $l\theta \succ r\theta$, i.e. paramodulation is done only with maximal sides of equations. Completeness is also preserved if one only paramodulates *on* maximal sides, i.e. $s\theta \succ t\theta$ is required as well; in that case the inference rule is called basic *superposition*. But if paramodulation steps are also performed on non-maximal sides, then, roughly speaking, all inferences on r have already been considered, and the term r in the conclusion can be blocked for further inferences; in the latter case the inference rule is called basic *paramodulation*.

Basic paramodulation and superposition techniques turn out to be a surprisingly powerful uniform method for obtaining complexity and decidability results for several classes of union-closed languages. We prove the following results:

Section 3: for sets of Horn clauses saturated under basic *paramodulation* (i), the word and unifiability problems are in NP, and the number of minimal unifiers is simply exponential. For Horn sets S saturated under the more restrictive inference rule of basic *superposition* (ii), the word and unifiability problems are still decidable and unification is finitary, provided S fulfils some natural syntactic and ordering requirements.

Section 4 is on shallow theories and their extension to the Horn case. We obtain (two versions of) a language we call *Catalog*, a natural extension of Datalog to include functions and equality. The closure under paramodulation is finite for Catalog sets, hence (i) applies. Since for shallow sets this closure is even polynomial, shallow unifiability is in NP, which is optimal: unifiability in ground theories is already NP-hard. We even go beyond: the shallow word problem is polynomial and for Catalog sets S we prove decidability of the full first-order theory of $\mathcal{T}(\mathcal{F})/_{=_S}$.

In section 5 we define *linear standard theories*, generalized in section 6 to full *standard theories*, which include shallow theories and in fact significantly extend them since in several common situations non-shallow variables are allowed. Standard presentations can be finitely closed under superposition fulfilling the requirements of result (ii), which hence applies.

2 Basic Paramodulation and other Basic Notions

Our notations and definitions on equations, orderings and rewriting will be consistent with [Dershowitz and Jouannaud, 1990]; the ones on ordered deduction with first-order equality clauses are the ones of [Bachmair and Ganzinger, 1994].

A variable x (or a constant a) *occurs at depth* 0 in the term x (resp. a) and occurs at depth $k+1$ in a term $f(t_1, \ldots, t_n)$ if it occurs at depth k in t_i for some $i \in 1 \ldots n$. A variable is *shallow* in a term s if it occurs only at depth 0 or 1 in s. Non-shallow variables, the ones occurring at depth two or more, are sometimes called *deep* variables. A term (equation) is shallow if all its variables are shallow. The *depth* of a term s is defined $depth(s) = max\{k \mid$ *some variable or constant occurs at depth* k *in* $s\}$. The *top symbol* $top(f(\ldots))$ of a term $f(\ldots)$ is f. A term s is *linear* if each variable in $\mathcal{V}ars(s)$ occurs only once in s. An equation $s \simeq t$ is linear if s and t are linear terms and is *collapsing* if at least one of s or t is a variable. A variable is *linear* in a term s if it occurs only once in s. Two terms *share* a variable x if x occurs in both terms. A linear equation is a *permuter* if it is of the form $f(x_1, \ldots, x_n) \simeq f(x_{\pi(1)} \ldots x_{\pi(n)})$ where π is a permutation and all x_i are variables for $i \in 1 \ldots n$.

The inference rules for *basic* paramodulation and superposition are expressed here by means of *equality constrained clauses* [Nieuwenhuis and Rubio, 1995]. A clause is a disjunction of equations and negated equations; non-equality atoms A are expressed by equations $A \simeq \mathit{true}$. An equality constraint is a conjunction of equalities $s = t$, where s and t are terms in $\mathcal{T}(\mathcal{F}, \mathcal{X})$. The constraints are interpreted in $\mathcal{T}(\mathcal{F})$ and $=$ is interpreted as syntactic equality. In other words, a constraint is *satisfied* by a ground substitution σ if $s\sigma$ and $t\sigma$ are the same element in $\mathcal{T}(\mathcal{F})$ for all $s = t$ in the conjunction, i.e., if σ is a (simultaneous) ground unifier of all $s = t$. The semantics of a constrained clause $C \,[\![T]\!]$ is the set of all its *ground instances*: ground clauses $C\sigma$ such that σ satifies T. Hence a clause with an unsatisfiable constraint is a tautology.

The inference rules below are parameterized by a simplification ordering $\succ$ on ground terms, which, unless stated otherwise, will be be the *recursive path ordering* (RPO) (see e.g. [Dershowitz, 1987] for details and properties)[1]. Assume given a *precedence* $\succ_p$ on the function symbols, let $==$ be the congruence relation of equality up to permutations of the direct subterms of any symbol and let $\succeq$ stand for $\succ \vee ==$. Then the RPO $\succ$ is defined: $s \succ t$ iff (i) s is $f(s_1, \ldots, s_n)$ and $\exists i \in \{1, \ldots, n\}$, $s_i \succeq t$ or (ii) s is $f(s_1, \ldots, s_n)$, t is $g(t_1, \ldots, t_m)$ and $f \succ_p g$ and $s \succ t_i$ for all $i = 1, \ldots, m$ or (iii) s is $f(s_1, \ldots, s_n)$, t is $f(t_1, \ldots, t_n)$ and $\{s_1, \ldots, s_n\} \gg \{t_1, \ldots, t_m\}$ where $\gg$ is the multiset extension of $\succ$.

The inference rules of *paramodulation* and *equality resolution* are, respectively:

$$\frac{C' \vee s \simeq t \,[\![T']\!] \qquad C \,[\![T]\!]}{C' \vee C[t]_p \,[\![T' \wedge T \wedge C|_p = s]\!]} \qquad\qquad \frac{C \vee s \not\simeq t \,[\![T]\!]}{C \,[\![T \wedge s = t]\!]}$$

In paramodulation, the term $C|_p$ is never a variable. The equality constraints express in a natural way the *basicness restriction*: no inferences take place on subterms introduced by unifiers of previous inferences. In this paper, assume that in every clause C having at least one negative literal, one such negative literal is *selected* [Bachmair et al., 1995], i.e. all inferences with C as a premise are paramodulations and equality resolutions on this literal. This implies for the Horn case that in all paramodulation inferences the left premise is a positive unit clause (hence the part C' of the left premise is empty). Furthermore, depending on the ordering restrictions with $\succ$, we obtain *basic paramodulation* or *basic superposition*. Paramodulations with $s \simeq t$ on a clause C are always restricted to the cases where $s\sigma \succ t\sigma$ for some σ satisfying the constraint of the conclusion. If only this restriction is imposed, then the inference rule is called *basic paramodulation*. In this case, the basicness restriction can be imposed also on the subterm t in the conclusion $C' \vee C[t]_p \,[\![T' \wedge T \wedge C|_p = s]\!]$ and hence the conclusion can be replaced by $C' \vee C[x]_p \,[\![T' \wedge T \wedge C|_p = s \wedge x = t]\!]$ where x is a new variable.

Now suppose wlog. that $C|_p$ is a position in the term s' of a (possibly negated) equation $s' \simeq t'$ in the right premise C. Then if in addition also $s'\sigma \succ t'\sigma$ is required we obtain *basic superposition*, in which the term t cannot be moved into the constraint.

From [Nieuwenhuis and Rubio, 1995, Bachmair et al., 1995] we have the following results. Basic paramodulation (superposition), together with equality resolution is refutationally complete for Horn clauses (for the non-Horn case a factoring rule is also needed). It is compatible with selection of negative literals and *redundancy* criteria including e.g. simplification by rewriting with unconstrained equations or deletion of tautologies. *Saturation* is the process of closure under the inference rules up to redundancy. From

[1] Although RPO is not total on ground terms, it can be extended to a total ordering, which suffices for the completeness results of paramodulation and superposition.

[Nieuwenhuis, 1995] we have completeness results for answer computation: e.g. if E is a set of Horn clauses, the saturation of $E \cup \{\neg s_1 \simeq t_1 \vee \ldots \vee \neg s_n \simeq t_n\}$ will produce an empty clause for every (simultaneous, irreducible) E-unifier σ s.t. $s_i\sigma =_E t_i\sigma$ for all i, and σ can be recovered from the accumulated instantiations of the variables of $s \simeq t$. These results are also compatible with *propagation* [Kirchner et al., 1990] of (parts of) a constraint to the clause part: for example, $f(x) \simeq a \,[\![x = g(b)]\!]$ propagates into $f(g(y)) \simeq a \,[\![y = b]\!]$. Eager propagation of the whole constraint gives us normal ordered paramodulation and superposition. A set is saturated under basic paramodulation (resp. superposition) whenever it is saturated under normal paramodulation (resp. superposition).

3 Deciding Unifiability by Basic Inference Rules

In this section we observe that for some kinds of sets of clauses closed under the basic inference rules, unification (answer computation) by again applying these inference rules can be proved terminating.

Theorem 3.1 Let S be a finite set of Horn clauses that is saturated under basic paramodulation with selection.

Then the word and unifiability problems in S are in NP, and unification (answer computation) is finitary: the number of minimal unifiers is simply exponential.

Proof: Let ϕ be a formula of the form $s_1 \simeq t_1 \wedge \ldots \wedge s_n \simeq t_n$ where $\mathcal{V}ars(\phi) = \vec{x}$. Let C be the clause $\neg s_1 \simeq t_1 \vee \ldots \vee \neg s_n \simeq t_n$. Then $S \models \exists \vec{x}\phi$ iff $S \cup \{C \,[\![true]\!]\}$ is inconsistent. Note that the word problem is a particular case in which C is Skolemized (ground). We proceed by basic paramodulation. S is closed (up to redundant inferences) under paramodulation, hence only steps on descendants $C' \,[\![T']\!]$ of $C \,[\![true]\!]$ are performed. Since all non-unit clauses have a selected negative literal, the only applicable clauses of S are positive unit clauses. Clearly, from each $C' \,[\![T']\!]$ we get only finitely many new $C_i \,[\![T_i]\!]$ and in C_i the number of non-variable positions is strictly smaller than in C' (recall that in basic *paramodulation* the term inserted in the conclusion is replaced by a variable). We obtain a so-called narrowing tree whose depth is at most the number of nonvariable positions in ϕ and with a branching factor polynomial in the number of such positions in ϕ and the number of unit clauses in S. This shows membership in NP of unifiability: simply guess the right path in the tree. By computing this whole tree, *all* (irreducible, i.e. minimal) unifiers (answers) are computed [Nieuwenhuis, 1995] (and can be extracted by recording the instantiations of the variables, as in logic programming), which implies that the number of unifiers is simply exponential. □

Later on, we wil show that for example shallow theories can be closed in polynomial time under paramodulation. The result of membership in NP of shallow unifiability is indeed optimal. For example, the commutativity axiom (C) is shallow and the C-unifiability problem is already NP-hard. Also ground theories are shallow, and the following is well-known:

Property 3.2 The unifiability problem in ground equational theories is NP-hard.

Proof: Reduce for example 3-sat: $(x_1 \vee \overline{x_2} \vee x_3) \wedge \ldots$ is expressed by an equation $and(or(x_1, not(x_2), x_3), and(\ldots)\ldots) \simeq t$, to be unified in the theory defined by the ground

equations
$not(t) \simeq f, \quad not(f) \simeq t,$
$and(f,f) \simeq f, \ldots, and(t,t) \simeq t,$
$or(f,f,f) \simeq f, \ldots, or(t,t,t) \simeq t.$ □

Definition 3.3 Let $\succ$ be an RPO ordering based on a total precedence $\succ_p$. Let σ_0 be the substitution mapping all variables to the smallest constant wrt. $\succ$. The ordering $\succ_u$ (the "u" stands for unification) is defined by $s \succ_u t$ iff $s\sigma_0 \succ t\sigma_0$. It is well-founded and monotonic, but not stable under substitutions. Its multiset extension $\gg_u$ is well-founded on clauses (seen as multisets of terms).

Theorem 3.4 Let S be a finite set of Horn clauses that is closed under basic superposition with selection and where all positive unit clauses are unconstrained permuters or equations of the form $s \simeq t \; [\![T]\!]$ where $top(s) = f$ and $f(x \ldots x) \succ_u t$ for a variable x and $s\sigma \succ t\sigma$ for all its ground instances $s\sigma \simeq t\sigma$. Then the word and unifiability problems in S are decidable, and unification is finitary.

Proof: Similar to the previous theorem, except that we apply basic superposition. From each $C' \; [\![T']\!]$, again only finitely many new descendants $C_i \; [\![T_i]\!]$ are obtained with $C' \gg_u C_i$, except in the case of innermost superpositions with permuters, which do not modify C'. But only finitely many of such permuting steps can be done consecutively without cycling. Hence, if we avoid cycling, termination follows by König's lemma. □

4 Shallow theories and Catalog

Datalog is a well-known language of Horn clauses in which all literals are flat, i.e. its atoms are predicates applied to variables or constants. Here we extend this language in a natural way to include equalities and function symbols. We first revisit shallow *equations*:

Theorem 4.1 Every finite set E of shallow equations can be closed under paramodulation in polynomial time.

Proof: Wlog. assume E is a set of *flat* equations: at depth one there are only variables and constants: otherwise, replace non-constant ground arguments t everywhere with a new constant a and add a new shallow equation $a \simeq t$ (iterate this if the arguments of t are again non-constant). This process generates a conservative extension [Comon et al., 1994]. It only adds a linear number of new constants and equations.

Now close E under paramodulation wrt. an RPO $\succ$ compatible with arities. This means that constants are small in the precedence, and hence all non-topmost inferences are replacements of constants by constants. The language of flat equations is closed under paramodulation: if σ is the most general unifier of two non-variable flat terms, then $x\sigma$ is either a variable or a constant for all x in the domain of σ. No unifications take place between a variable and another term, because no maximal side of an equation is a variable (except for equations $x \simeq t$ where $x \notin \mathcal{V}ars(t)$, but then the closure can be stopped with the —inconsistent— outcome $x = y$, where x and y are distinct variables). This language of shallow equations has a polynomial cardinality for a given signature, which means that only polynomially many inferences are considered in the closure. □

Definition 4.2 A *shallow atom* is an atomic formula of the form $P(t_1, \ldots, t_n)$ where P is a non-equality predicate symbol and where each t_i for $i \in 1 \ldots m$ is a ground term or a variable.
A Horn clause of the form $\neg A_1 \vee \ldots \vee \neg A_n \vee A_0$ where $n \geq 0$, is a *Catalog-1 clause* iff all A_i for $i \in 1 \ldots n$ are shallow atoms and A_0 is a shallow equation or a shallow atom. It is a *Catalog-2 clause* iff each A_i for $i \in 0 \ldots n$, is either a shallow atom or a non-collapsing shallow equation.

Note that in the case of Catalog-2 clauses, the non-collapsing requirement (i.e., no side of an equation is a variable) is essential, since otherwise arbitrary equations (for which obviously all these problems are undecidable) could be expressed. For example, an equation like $f(g(x), g(x)) \simeq h(x)$ could be expressed by the logically equivalent Horn clause $\neg y \simeq g(x) \vee f(y, y) \simeq h(x)$.

Lemma 4.3 Let S be a finite set of only Catalog-1 clauses or only Catalog-2 clauses. Then S can be finitely saturated under paramodulation.

Proof: Wlog. as above assume S is a set of *flat* Catalog-1 (or -2) clauses. Shallow atoms A are treated as equations $A \simeq \mathit{true}$ where *true* is a new constant. The languages of flat Catalog-1 (or -2) clauses are closed under paramodulation and equality resolution steps[2]. Finiteness of the closure follows from finiteness of the number of different flat clauses that can appear (due to the unit strategy the number of literals per clause does not increase). □

Theorem 4.4 Let S be a finite set of only Catalog-1 clauses or only Catalog-2 clauses. Let E be a set of shallow equations.
(i) The word problem in E is polynomial, E-unifiability is in NP, and the number of minimal E-unifiers is simply exponential.
(ii) The word and unifiability problems in S are decidable and unification (answer computation) is finitary.
(iii) The first-order theory of $\mathcal{T}(\mathcal{F})/_{=_S}$ is decidable.

Proof: For deciding $E \models s \simeq t$ in polynomial time, let $\overline{s} \simeq \overline{t}$ be the ground Skolemization of $s \simeq t$. Let E' be the (polynomial size) closure under paramodulation of the shallow set $E \cup \{\ \overline{s} \simeq a,\ \overline{t} \simeq b\ \}$ where a and b are new constants. The refutation by basic paramodulation (like in Theorem 3.1) of $E' \cup \{a \not\simeq b\}$ is linear in E'.

For (i) and (ii), E and S can be saturated as above; then apply Theorem 3.1. For (iii), let E be the set of unit clauses of the saturation. It is not difficult to prove that $\mathcal{T}(\mathcal{F})/_{=_S}$ and $\mathcal{T}(\mathcal{F})/_{=_E}$ are the same algebra. Since E is a set of shallow equations, from [Comon et al., 1994] we have that the first-order theory of $\mathcal{T}(\mathcal{F})/_{=_E}$ is decidable. □

Let us conclude this section with a comment on the complexity of deduction and unification in Catalog. Deduction in the well-known language of Datalog is said to be polynomial. But this is only true for a *constant* set of non-unit deduction rules (in databases they are considered part of the —fixed— intensional database). If we consider the word problem for flat Horn clauses without equality, in which we have as instance a Datalog program P and a ground query atom Q and the question "does Q follow from P?", then we can easily encode 3-sat problems:

[2]Note that collapsing negative literals $s \simeq x$ are not created. Equality resolution would instantiate x by s and create non-shallow literals if x occurs elsewhere at depth 1.

Property 4.5 The word problem for flat Horn clauses without equality is NP-hard.

Proof: Encoding 3-sat, using four predicates P_0, P_1, P_2, P_3, where each P_i is used for a 3-sat clause with i negative atoms. These predicates are defined by the unit clauses

$P_0(t,x,y)$ $P_0(x,t,y)$ $P_0(x,y,t)$
$P_1(t,x,y)$ $P_1(x,t,y)$ $P_1(x,y,f)$
$P_2(t,x,y)$ $P_2(x,f,y)$ $P_2(x,y,f)$
$P_3(f,x,y)$ $P_3(x,f,y)$ $P_3(x,y,f)$

Clearly, a 3-sat problem like

$(x_1 \vee x_2 \vee \overline{x_3}) \wedge (\overline{x_4} \vee \overline{x_5} \vee \overline{x_6}) \wedge \ldots$

is satisfiable iff the atom Sat follows from the Datalog program consisting of the unit clauses for the P_i plus the flat Horn clause:

$\neg P_1(x_1,x_2,x_3) \vee \neg P_3(x_4,x_5,x_6) \vee \ldots \vee Sat$ □

Since Catalog includes Datalog, deduction in Catalog can be polynomial only with the same assumptions as for databases. Indeed, in that case for Catalog we obtain the same results as in the previous theorem for shallow equations.

5 Linear Standard Theories

Before studying in the next section full standard theories, in order to give the reader a better understanding of our proof techniques, we first introduce them for the following simpler linear sublanguage (recall that a term is linear if no variable occurs more than once in it):

Definition 5.1 An equation $s \simeq t$ is *linear standard* iff s and t are linear terms sharing only shallow variables. A *linear standard theory* is a theory axiomatisable by a set of linear standard equations.

Our aim will be, for every linear standard presentation E, to show that it can be finitely saturated under superposition. For this purpose, an interesting property (straightforward to show) is that the language of linear standard equations is closed under superposition. Our termination proof is based on the fact that only finitely many different equations $s \simeq t$ exist (up to variable renamings) if s and t belong to $\mathcal{T}(\mathcal{F},\mathcal{X})$ for some finite $\mathcal{F}$ and $depth(s) < k$ and $depth(t) < k$ for some k. Unfortunately, although inferences at the top position between linear standard equations are non-depth increasing, this is not the case for inferences at other positions.

Example 5.2 From the linear standard equations $f(g(x,y)) \simeq a$ and $g(x,y) \simeq h(x,f(f(z)))$, by one paramodulation step on $g(x,y)$, under any ordering where a is small, we obtain $f(h(x,f(f(z)))) \simeq a$ which has a greater depth than the premises. □

A solution to the problem appearing in the previous example is to eagerly *split* each equation:

Definition 5.3 Let $s \simeq t$ be a non-collapsing linear standard equation which is not already of the form $s \simeq h(x_1,\ldots,x_n)$ where $\{x_1,\ldots,x_n\}$ is a subset of the set of (shallow)

variables of s. Then $s \simeq t$ is *split* by replacing it by two new ones $s \simeq h(x_1, \ldots, x_n)$ and $t \simeq h(x_1, \ldots, x_n)$, where h is a new function symbol and $\{x_1, \ldots, x_n\}$ is the set of (shallow) variables shared by s and t.

Lemma 5.4 For all sets of equations E over $\mathcal{T}(\mathcal{F}, \mathcal{X})$, the theory of the set E' over $\mathcal{T}(\mathcal{F} \cup \{h\}, \mathcal{X})$, obtained from E by splitting some equation in E is a conservative extension, i.e. for all terms s and t in $\mathcal{T}(\mathcal{F}, \mathcal{X})$, $E \models s \simeq t$ iff $E' \models s \simeq t$.

Moreover all ground terms $h(t_1, \ldots, t_n)$ are congruent under E' with some term in $\mathcal{T}(\mathcal{F})$, hence the initial models $\mathcal{T}(\mathcal{F})/=_E$ and $\mathcal{T}(\mathcal{F} \cup \{h\})/=_{E'}$ are isomorphic.

Note that by splitting an equation $s \simeq t$ the new symbol h introduced has a strictly smaller arity than each of the top symbols of s and t. In the following, we will eagerly split all equations (the initial ones and the ones generated by inferences) for which this is possible (recall that splitting is only defined for equations which are not already of the required form). In the context of saturation, splitting a conclusion $s \simeq t$ of an inference into $s \simeq h(\vec{x})$ and $t \simeq h(\vec{x})$ preserves completeness since these two new equations make $s \simeq t$ redundant if the symbol h is small in the ordering. It is not difficult to see that, for equations of this form, superposition inferences are indeed non depth increasing:

Lemma 5.5 Let $s \simeq t$ be a linear standard equation where t is either a variable occurring in s or a term $h(x_1, \ldots, x_n)$ where all x_i are variables occurring in s and where $top(s) \succ_p h$. In every inference among linear standard equations by superposition wrt. $\succ$ with left premise $s \simeq t$, the depth of the conclusion, which is again a linear standard equation, is smaller or equal than the depth of the premises.

Proof: Case analysis using the fact that if s and t are linear terms and $\sigma = mgu(s, t)$ then $depth(s\sigma) = depth(t\sigma) = max(depth(s), depth(t))$. □

However, splitting only postpones the termination proof, because if the number of new symbols introduced by splittings is infinite, the bounded depth argument does not suffice for proving termination of the closure process. A very simple case of such a bad situation is described in the following example:

Example 5.6 Assume we have the equations:

1. $f(x, g(a)) \simeq g(x)$
2. $f(x, y) \simeq h(y)$

By one inference at the top position between 1. and 2. we obtain $g(x) \simeq h(g(a))$, which has to be split into 3. $g(x) \simeq c_1$ and 4. $h(g(a)) \simeq c_1$ for some new constant symbol c_1.

On the other hand, from 1. and 3. we get a new version of 1., namely $f(x, c_1) \simeq g(x)$, which with 2. produces $g(x) \simeq h(c_1)$, which has to be split into 5. $g(x) \simeq c_2$ and 6. $h(c_1) \simeq c_2$ for some new symbol c_2.

It is clear that 5. can now again play the role of equation 3., etc, leading to non-termination, in spite of the fact that the depth of the equations does not increase. □

In the previous example, it is clear that in fact all the infinitely many constants $c_1, c_2, \ldots$ are equivalent (in fact, they are all equivalent to $g(x)$). Following this intuition, by a careful analysis of standard equations an interesting property (with a clearly *basic* flavour) can be shown: let e and e' be two equations and let E and E' be the sets of all

descendants by zero or more inferences at non-topmost positions on e and e' respectively. Then, out of all possible equations that can be obtained by superposing at the topmost position equations of E on equations of E', *only one* of them is needed.

Example 5.7 Let e be a (split linear standard) equation $f(\ldots, t, \ldots) \simeq r$. If e_1 is a new equation obtained by an inference below the subterm t in e then, due to the variable restrictions of the language, e_1 only differs from e in the part below t, and hence e_1 is of the form $f(\ldots, s, \ldots) \simeq r$.

Now let e' be another equation $f(\ldots, t', \ldots) \simeq r'$ by which an topmost inference with e is possible. If in e' the argument t' is not a variable occurring in r' then the conclusion of the topmost inference with e is identical to the one with e_1 (if the latter one exists).

Otherwise e' is of the form $f(\ldots, x, \ldots) \simeq h(\ldots, x, \ldots)$; then the conclusion with e is $h(\ldots, t, \ldots) \simeq r\sigma$ and the one with e' is $h(\ldots, t', \ldots) \simeq r\sigma$, which can also be obtained from the latter by the same inference on t. □

Definition 5.8 Let e_1 and e_2 be linear standard equations of the form $f(s_1, \ldots, s_n) \simeq s$ and $f(t_1, \ldots, t_n) \simeq t$ where s is a variable x or of the form $g(\vec{x})$, and t is a variable y or of the form $h(\vec{y})$, such that $\vec{x} \subseteq \{s_1, \ldots, s_n\}$ and $\vec{y} \subseteq \{t_1, \ldots, t_n\}$, and let σ be the substitution with domain $\vec{x} \cup \vec{y}$ defined by $s_i\sigma = t_i$ and $t_i\sigma = t_i$ if $s_i \in \vec{x}$, and $t_i\sigma = s_i$ otherwise. Then $s\sigma \simeq t\sigma$ is called the *needed* conclusion of e_1 and e_2.

Note that the needed conclusion of e_1 and e_2 is equal up to variable renamings to the needed conclusion of e_2 and e_1, and if there is any topmost inference possible between e_1 and e_2, then its conclusion is exactly the needed conclusion.

Lemma 5.9 Let E be a set of split linear standard equations, and let e_1 and e_2 be two equations with a needed conclusion e. Let D_1 and D_2 be the sets of all descendants by zero or more inferences at non-topmost positions with equations of E on e_1 and e_2 respectively, and let D be the set of all conclusions that can be obtained by superposing at the topmost position an equation of D_1 on an equation of D_2.

If $D \neq \emptyset$ then $e \in D$. Furthermore, all equations in D can be obtained by inferences at non-topmost positions with equations of E on e.

Proof: Let e_1 and e_2 be as in the previous definition. All equations in D_1 will have the form $f(s'_1, \ldots, s'_n) \simeq s$ and the ones in D_2 will be of the form $f(t'_1, \ldots, t'_n) \simeq t$.

If no topmost inference is possible between e_1 and e_2 this is because for some i the non-variable linear terms t_i and s_i are not unifiable. If some inference between descendants of e_1 and e_2 is possible, then this is because the corresponding s'_i and t'_i have become unifiable. By performing non-topmost inferences only on such positions i, we can clearly obtain the needed conclusion, since we leave unchanged the terms s_j and t_j, where $s_j \in \vec{x}$ or $t_j \in \vec{y}$ which determine how the conclusion is going to be.

If some topmost inference with a substitution σ' between descendants of e_1 and e_2 produces a conclusion e' with $e' \neq e$, e.g. because $x\sigma' = s'_i \neq s_i$ for some $x \in \vec{x}$, then, since s'_i has been obtained from s_i by non-topmost steps on e_1, we can apply the same steps to obtain e' from e, since non-variable arguments only appear in the left hand sides after splitting of e. □

Definition 5.10 An *almost permuter* is a linear equation of the form $f(x_1, \ldots, x_n) \simeq g(x_{\pi(1)} \ldots x_{\pi(n)})$ where π is a permutation, all x_i are variables and where $f \neq g$.

We now show that sets of linear standard equations can be finitely closed under unconstrained superposition (and hence under basic superposition). The closure process is parameterized by a recursive path ordering (RPO) $\succ$ on terms where the precedence $\succ_p$ is compatible with arities, i.e. if $arity(f) > arity(g)$ then $f \succ_p g$. We always eagerly split initial equations and conclusions of inferences. Note that this implies that all equations in E are orientable as rewrite rules wrt. $\succ$ except permuters. For efficiency and simplicity reasons, almost permuters are supposed to be applied as simplification rules to eagerly eliminate by rewriting everywhere all occurrences of symbols like the f of the previous definition. After this, no inferences with it will exist. Note that a particular case of this is rewriting with equations between constants.

Definition 5.11 Let e be a (split) linear standard equation.
If e is a permuter $f(x_1, \ldots, x_n) \simeq f(x_{\pi(1)} \ldots x_{\pi(n)})$ we define $degree(e) = n$. Otherwise e is of the form $s \simeq t$ with $s \succ t$ and if t is a constant then $degree(e) = 0$ else $degree(e) = arity(top(s))$.

For non-empty sets of standard equations E we define $degree(E) = max\{degree(e) \mid e \in E\}$.

Theorem 5.12 Every set E of linear standard equations can be finitely closed under superposition.

Proof: The initial splitting of such a set E is clearly finite, so we can assume E to be closed under splitting. In the closure of E we have the following case analysis of every possible kind of inferences:

Non-topmost inferences:

1. A superposition of some other equation on a position inside the linear term s of an equation $f(\ldots, s, \ldots) \simeq t$, where t is a variable or a term $h(\vec{x})$: in this case we obtain a conclusion $f(\ldots, s', \ldots) \simeq t$, i.e. the only change is that s is replaced by some other linear term s' with smaller or equal depth. No splitting of the conclusion is needed.

Topmost inferences:

1. A superposition at the topmost position between two collapsing equations $f(\ldots, x, \ldots) \simeq x$ and $f(\ldots, y, \ldots) \simeq y$
   (a) If x and y coincide at the same argument position of f, this produces as conclusion a tautology $x \simeq x$.
   (b) Else, if for one of x or y, the corresponding argument term is a variable z, an inconsistency of the form $z \simeq t$, where $z \notin \mathcal{V}ars(t)$, is obtained.
   (c) Otherwise, an equation $s \simeq t$ is obtained where s and t are linear non-variable terms not sharing any variables, which will be split into $s \simeq a$ and $t \simeq a$ where a is a new constant symbol.

2. A superposition at the topmost position between a collapsing equation $f(\ldots, x, \ldots) \simeq x$ and a non-collapsing equation $f(\ldots, s, \ldots) \simeq h(y_1, \ldots, y_n)$ where s is at the same argument position of f as x.

   (a) If s is one of the variables y_i, a new collapsing equation $h(\ldots, x, \ldots) \simeq x$ is obtained, for which splitting is not needed, and whose depth is smaller or equal to the one of the first premise.

   (b) If s is a variable z, but not one of the variables y_i, an inconsistency of the form $z \simeq t$, where $z \notin \mathcal{V}ars(t)$, is obtained.

   (c) Otherwise, s is a linear non-variable term, and an equation $s \simeq t$ is obtained, where s and t are linear terms not sharing any variables, which will be split into $s \simeq a$ and $t \simeq a$ where a is a new constant symbol.

3. A superposition at the topmost position between two non-collapsing equations $f(t_1, \ldots, t_n) \simeq h(x_1, \ldots, x_p)$ and $f(s_1, \ldots, s_n) \simeq h'(y_1, \ldots, y_q)$ of degree p and q respectively, with $p \geq q$, and where for each x_i and y_j we have $x_i \in \{t_1, \ldots, t_n\}$ and $y_j \in \{s_1, \ldots, s_n\}$.

   The conclusion will be a non-collapsing equation $h(x_1, \ldots, x_p)\sigma \simeq h'(y_1, \ldots, y_q)\sigma$.

   (a) If every variable y_i appears at the same argument position in $\{s_1, \ldots, s_n\}$ as some x_i in $\{t_1, \ldots, t_n\}$ then the conclusion needs no splitting.

   In particular, this case applies if at least one of the right hand sides (which must be $h'(y_1, \ldots, y_q)$ since we suppose $p \geq q$) is a constant. Then also the conclusion will have the constant h' as a constant side.

   This case also applies if at least one of the premises is a permuter (which must be $f(t_1, \ldots, t_n) \simeq h(x_1, \ldots, x_p)$ since we suppose $p \geq q$).

   In this case, the conclusion is an (almost) permuter if $p = q$; it has degree q if $p > q$.

   (b) Otherwise the conclusion needs splitting and both resulting equations will be of degree strictly smaller than q.

We now prove, by induction on $degree(E)$, that the number of new symbols introduced in the following closure process of E is finite, and that the closure process of E is finite as well:

If $degree(E) = 0$ then there are only equations with constant right hand sides. This language of degree 0 equations is closed under superposition inferences and no new symbols will be ever introduced. Since the depth of equations is not increased, the topmost closure of E is finite.

For the induction step, let $degree(E) = n$ for the initial set E, and let E^n denote the set of all equations of degree n of the successive sets E of the closure process. To clarify the proof, we distinguish the following phases in the closure process (although in practice any order between the inferences is allowed):

1. The subset of permuters of E^n can be independently closed under inferences. This is a finite process where no new symbols are introduced. Note that after this no new permuters of E^n can appear since such permuters can only be generated by inferences between equations of degree greater than n, which do not exist.

2. After this, the topmost inferences between permuter and non-permuter equations of E^n can be computed, which is again a finite process where no new symbols are introduced. Let k be the number of non-permuter equations in E^n after this step.

3. There are at most k^2 topmost *needed* inferences between non-permuter equations of E^n: by Lemma 5.9, only such needed inferences between equations in E^n are necessary and there are at most k^2 of them. In this phase at most k^2 new symbols are introduced, all of them appearing in equations of degree smaller than n.

4. By the induction hypothesis, the closure of the set of equations $E \setminus E^n$ (which has degree smaller than n) produces only finitely many new symbols and is finite.

5. The only remaining steps are non-topmost inferences with or on equations in E^n. But these steps introduce no new function symbols and are non-depth increasing, hence there can only be finitely many of them, which completes the proof.

□

6 Standard Theories

Definition 6.1 A *standard signature* $\mathcal{F}$ is a signature where every function symbol f with arity n in $\mathcal{F}$ has an associated set of *shallow* positions $sh(f)$ and a set of *linear* positions $lin(f)$, such that $sh(f) \cap lin(f) = \emptyset$ and $sh(f) \cup lin(f) = \{1 \ldots n\}$.

A depth-one argument s_i of $f(s_1, \ldots, s_n)$ is a *linear position argument* of it if $i \in lin(f)$ and a *shallow position* argument if $i \in sh(f)$.

Definition 6.2 A term s is is a *standard* term iff it is a variable or a term of the form $f(s_1, \ldots, s_n)$ where if $i \in sh(f)$ then s_i is a variable or a ground term and if $i \in lin(f)$ then all variables in s_i are linear in s.

Note that acording to the previous definition, all ground terms are standard, but not all linear terms are, because at shallow positions no terms with deep variables are allowed. Furthermore, the only non-linear variables of a standard term s are shallow variables at shallow positions.

Definition 6.3 An equation $s \simeq t$ is *standard* iff
(i) s is linear and t is ground or
(ii) s is a standard term $f(\ldots, g(t), \ldots)$ and t is a variable or
(iii) s and t are standard terms sharing only shallow variables and no variable x is both a shallow position argument and a linear position argument in $s \simeq t$.
A *standard presentation* is a set of standard equations and a *standard theory* is a theory axiomatizable by a standard presentation. □

Note that for standard equations of type (i) s is not required to be a standard term. The standard equations of type (ii) are the only ones in which both sides of the equation share a non-shallow variable. Note however that this variable has strong restrictions: it must appear at depth two below a unary symbol.

In what follows we will consider that all terms are built over a standard signature $\mathcal{F}$ (in which sufficiently many new symbols for splitting are available). Let us also remark that for monadic symbols wlog. we can choose its unique position to be linear.

Clearly, shallow presentations are standard: just define $lin(f) = \emptyset$ for all function symbols f; on the other hand, linear standard presentations are also standard: take $sh(f) = \emptyset$ for all f. The union of two standard presentations over the same standard signature is again standard.

Example 6.4 $f(x, x, g(y)) \simeq h(x, g(z))$, $succ(pred(x)) \simeq x$ and $not(not(x)) \simeq x$ are standard equations which are not shallow nor linear standard. □

Example 6.5 The presentation

$\{ \ f(g(y), x) \simeq h(x), \ \ f(x, x) \simeq g(x) \ \}$

is *not* standard because the first position of f cannot be linear and shallow at the same time. It is clear why this kind of equation cannot be allowed: by one superposition inference we obtain $g(g(x)) \simeq h(g(x))$ sharing non-shallow variables; and indeed if $g \succ_p h$ in the closure under superposition the infinite set of equations schematized by $g(h^n(g(x))) \simeq h^{n+1}(g(x))$ for $n \geq 0$ is generated. □

For standard theories, we will proceed as we did for linear standard ones. Our aim will be to show that every standard presentation E can be finitely closed under superposition satisfying the conditions of Theorem 3.4, and hence the word problem in E is decidable and unification in E is decidable and finitary. In this case, inferences are still non-depth increasing, if splitting is done eagerly. For standard equations splitting has to be defined more precisely than was done for linear standard ones:

Definition 6.6 A non-collapsing standard equation $s \simeq t$ which is not already of the form $s \simeq h(x_1, \ldots, x_n)$ where $\{x_1, \ldots, x_n\}$ is a subset of the set of (shallow) variables of s, is *split* by replacing it by two new ones $s \simeq h(x_1, \ldots, x_n)$ and $t \simeq h(x_1, \ldots, x_n)$, where h is a new function symbol and $\{x_1, \ldots, x_n\}$ is the set of (shallow) variables shared by s and t.

The new symbol h will have linear and shallow positions such that a variable x will be at a linear position of h iff it was at linear positions in $s \simeq t$. □

The following three lemmas are straightforward from the definition of the language of standard equations, and they provide the reader with most of the intuition for the reasons behind the finiteness of the closure under superposition for standard equations.

Lemma 6.7 Let $f(\ldots, g(x), \ldots) = x$ be a split type (ii) standard equation and let t be a non-variable term where $\sigma = mgu(s, t)$. If t is linear or standard, then $x\sigma$ is linear and $depth(x\sigma) \leq max\{depth(s), depth(t)\}$. □

Lemma 6.8 Let $s = h(x_1, \ldots, x_n)$ be a split type (iii) standard equation and let t be a non-variable term where $\sigma = mgu(s, t)$.

- If t is linear, then $h(x_1\sigma, \ldots, x_n\sigma)$ is linear.
- If t is standard, then $h(x_1\sigma, \ldots, x_n\sigma)$ is standard.

- If t is linear or standard, $depth(h(x_1\sigma, \ldots, x_n\sigma)) \leq max\{depth(s), depth(t)\}$ and $x_i\sigma$ is linear for all i in $1 \ldots n$. □

Lemma 6.9 Let $s = x$ be a split type (iii) standard equation and let t be a non-variable term where $\sigma = mgu(s, t)$. If t is linear or standard, then $x\sigma$ is linear and $depth(x\sigma) \leq max\{depth(s), depth(t)\}$. □

Due to the richer language (e.g. the collapsing equations of case (ii) of the definition) the case analysis of all possible inferences is more tedious than in the previous section, although all cases are very simple. Below we analyse all cases according to the type of standard equation (i),(ii) or (iii) of Definition 6.3 in all their possible combinations. For some combinations, again several subcases appear, as for example, the five subcases for topmost inferences between type (ii) and type (iii) equations. It is clear that some cases could have been merged; however, we prefer the current presentation in which by construction no cases can be overlooked.

In order to reduce somewhat the length of the case analysis, we first provide some remarks and intuitive conventions:

- Due to eager splitting, in all equations $s \simeq t$ of type (i) the ground term t is in fact a constant, which below will be denoted with $a, b, c, \ldots$
- Variables will be denoted with $x, y, z, \ldots$
- In some cases an inconsistency of the form $s = x$ is obtained where x is a variable not occurring in s. In such a case obviously the closure process can be stopped with outcome $x = y$.
- In all cases, when it is not explicitly mentioned, no splitting is required. Note that splitting creating a new non-constant symbol is needed only in a topmost superposition of two non-collapsing type (iii) equations.
- In all cases, it is easy to check that the conclusions obtained have depth smaller than or equal to the maximal depth of any of the premises.
- In many cases an equation $s = t$ is obtained where s and t do not share any variables. In all these cases, the following (omitted) analysis applies: If one or both sides is a variable, this is an inconsistency. Otherwise, if neither side is a constant, splitting into $s = a$ and $t = a$ is required, introducing a new constant symbol a. If e.g. s is linear, $s = a$ is a type (i) equation. If it is standard, the result is a type (iii) equation.

Non-topmost inferences:

1. (i) on (i):
   $s = a$ on $t[s'] = b$. Due to linearity of $t[s']$, unifying s and s' does not affect the variables of $t[\cdot]$ and hence the result is $t[a] = b$.
2. (i) on (ii):
   $g(s) = a$ on $f(\ldots, g(x), \ldots) = x$. Since x occurs only once in $f(\ldots, g(x), \ldots)$, unifying s and x does not affect the context $f(\ldots, [\cdot], \ldots)$ and the result is $f(\ldots, a, \ldots) = s$, where the left and right hand side do not share any variables.

3. (i) on (iii):

   (a) $s = a$ on $f(\ldots, t[s'], \ldots) = h(x_1, \ldots, x_n)$. Since the variables of s' occur only once in $f(\ldots, t[s'], \ldots) = h(x_1, \ldots, x_n)$, the result is $f(\ldots, t[a], \ldots) = h(x_1, \ldots, x_n)$ of type (iii).

   (b) $s = a$ on $f(\ldots, t[s'], \ldots) = x$. Like in the previous case, the result is $f(\ldots, t[a], \ldots) = x$ of type (iii).

4. (ii) on (i): $f(\ldots, g(x), \ldots) = x$ on $s[f(\ldots, t, \ldots)] = a$. Since x and the variables of s do not occur elsewhere in any of the premises, the result is $s[t'] = a$ of type (i), where t' is some linear subterm of t (or t' is x, if t is a variable).

5. (ii) on (ii):

   (a) $f(\ldots, g(x), \ldots) = x$ on $h(\ldots, s[f(\ldots, t, \ldots)], \ldots, g(y), \ldots) = y$. Since x and y do not occur elsewhere, the result is $h(\ldots, s[t'], \ldots, g(y), \ldots) = y$ of type (ii), where t' is some linear subterm of t like in the previous case.

   (b) $g(h(y)) = y$ on $f(\ldots, g(x), \ldots) = x$. Since x does not occur elsewhere in the second premise, the result is the type (iii) equation $f(\ldots, y, \ldots) = h(y)$. Note that y is on a linear position in the left hand side, and that, since h is unary, it does not mind whether its unique argument position is linear or shallow.

6. (ii) on (iii): $f(\ldots, g(x), \ldots) = x$ on $g(\ldots, s[f(\ldots, t, \ldots)], \ldots) = r$. Since x and the variables of s do not occur elsewhere in the premises, the result is a new type (iii) equation $g(\ldots, s[t'], \ldots) = r$ where t' is some linear subterm of t like in the previous two cases.

7. (iii) on (i):

   (a) $t = h(x_1, \ldots, x_n)$ on $s[t'] = a$. The result is a type (i) equation $s[h(x_1\sigma, \ldots, x_n\sigma)] = a$, since by Lemma 6.8 $h(x_1\sigma, \ldots, x_n\sigma)$ is linear.

   (b) $t = x$ on $s[t'] = a$. The result is a type (i) equation $s[x\sigma] = a$, since by Lemma 6.9 $x\sigma$ is linear.

8. (iii) on (ii):

   (a) $g(y) = h(y)$ on $f(\ldots, g(x), \ldots) = x$. The result is a new type (ii) equation $f(\ldots, h(x), \ldots) = x$. In fact, this inference corresponds to a simplification step with the almost permuter $g(x) = h(x)$.

   (b) $g(x) = x$ on $f(\ldots, g(x), \ldots) = x$. The result is a new type (iii) equation $f(\ldots, x, \ldots) = x$. In fact, this inference corresponds to a simplification step with $g(x) = x$.

   (c) $t = h(x_1, \ldots, x_n)$ on $f(\ldots, s[t'], \ldots, g(x), \ldots) = x$. The result is a new type (ii) equation $f(\ldots, s[h(x_1\sigma, \ldots, x_n\sigma)], \ldots, g(x), \ldots) = x$.

   (d) $t = y$ on $f(\ldots, s[t'], \ldots, g(x), \ldots) = x$. The result is a new type (ii) equation $f(\ldots, s[x\sigma], \ldots, g(x), \ldots) = x$.

9. (iii) on (iii):

(a) $t = h(x_1, \ldots, x_n)$ on $f(\ldots, s[t'], \ldots) = r$. The result is a new type (iii) equation $f(\ldots, s[h(x_1\sigma, \ldots, x_n\sigma)], \ldots) = r$.

(b) $t = x$ on $f(\ldots, s[t'], \ldots) = r$. The result is a new type (iii) equation $f(\ldots, s[x\sigma], \ldots) = r$.

Topmost inferences:

1. (i) on (i):
   $s = a$ on $t = b$. The result is an equation between constants $a = b$.

2. (i) on (ii):
   $s = a$ on $f(\ldots, g(x), \ldots) = x$. The result is a new type (i) equation of the form $s|_p = a$ where p is the position of x in the term $f(\ldots, g(x), \ldots)$.

3. (i) on (iii):

   (a) $s = a$ on $t = h(x_1, \ldots, x_n)$. The result is a new type (i) equation of the form $h(x1\sigma, \ldots, x_n\sigma) = a$.

   (b) $s = a$ on $t = x$. The result is a new type (i) equation of the form $x\sigma = a$.

4. (ii) on (ii):

   (a) $f(\ldots, g(x), \ldots, s, \ldots) = x$ on $f(\ldots, t, \ldots, h(y)\ldots) = y$. Since $g(x)$ and $h(y)$ are on linear positions in both premises, the variables of the terms $g(x)$, $h(y)$, s and t do not occur elsewhere in the premises. Hence the result is an equality $t' = s'$ between two linear terms not sharing any variables.

   (b) $f(\ldots, g(x), \ldots) = x$ on $f(\ldots, g(y), \ldots) = y$. The result is a tautology $x = x$.

5. (ii) on (iii):

   (a) $f(\ldots, g(x), \ldots) = x$ on $f(\ldots, x_i, \ldots) = h(x_1, \ldots, x_n)$. The result is a new type (ii) equation $h(x_1\sigma, \ldots, x_{i-1}\sigma, g(x), x_{i+1}\sigma, \ldots, x_n\sigma) = x$. Note that x_i does not occur elsewhere in the second premise since here (and in the following four cases) $g(x)$ must be at a linear argument position of f.

   (b) $f(\ldots, g(x), \ldots) = x$ on $f(\ldots, y, \ldots) = y$. The result is a new type (iii) equation $g(x) = x$.

   (c) $f(\ldots, g(x), \ldots) = x$ on $f(\ldots, g(s), \ldots) = h(x_1, \ldots, x_n)$. The result is an equation $h(x_1\sigma, \ldots, x_n\sigma) = s$ where both sides share no variables and where s is linear and by Lemma 6.8 $h(x_1\sigma, \ldots, x_n\sigma)$ is standard.

   (d) $f(\ldots, g(x), \ldots) = x$ on $f(\ldots, g(s), \ldots) = y$. The result is an equation $y\sigma = s$ where both sides share no variables, s is linear, and by Lemma 6.9 $y\sigma$ is linear as well.

   (e) $f(\ldots, g(x), \ldots) = x$ on $f(\ldots, y, \ldots) = r$ where y is a variable not in r. The result is an inconsistency $r\sigma = x$ where x does not occur in $r\sigma$.

6. (iii) on (iii):

(a) $t = h(x_1, \ldots, x_n)$ on $s = g(y_1, \ldots, y_m)$.
The result is $h(x_1\sigma, \ldots, x_n\sigma) = g(y_1\sigma, \ldots, y_m\sigma)$, which are both standard terms by Lemma 6.8. The result may need splitting into $h(x_1\sigma, \ldots, x_n\sigma) = h'(z_1, \ldots, z_k)$ and $g(y_1\sigma, \ldots, y_m\sigma) = h'(z_1, \ldots, z_k)$ where $m > k$ and $n > k$.

(b) $f(t_1, \ldots, t_k) = h(x_1, \ldots, x_n)$ on $f(s_1, \ldots, s_k) = y$, where the variable y occurs once or more as some s_i in $f(s_1, \ldots, s_k)$. The result is $h(x_1\sigma, \ldots, x_n\sigma) = y\sigma$. Since both left hand sides are standard terms, $y\sigma$ can be:

i. One of the variables among the $x_j\sigma$. Then the result is a new type (iii) equation.

ii. A variable not in $\{x_1\sigma, \ldots, x_n\sigma\}$. Then the result is an inconsistency.

iii. A non-variable linear term sharing no variables with $h(x_1\sigma, \ldots, x_n\sigma)$.

(c) $f(t_1, \ldots, t_k) = x$ on $f(s_1, \ldots, s_k) = y$, where the variable x occurs once or more as some t_i in $f(t_1, \ldots, t_k)$ and the same happens with y in $f(s_1, \ldots, s_k)$. The result is $x\sigma = y\sigma$, where both sides are linear.

i. If t_i is x and s_i is y for some i the result is a tautology. The same thing happens if x and y get unified through a chain of shallow variables, like when unifying $f(x, z, z)$ with $f(z', z', y)$.

ii. Otherwise the result is an equation $x\sigma = y\sigma$ where both sides share no variables.

As we did in the previous section for the linear subcase, sets of standard equations can be finitely closed under unconstrained superposition (and hence under basic superposition). Again the ordering $\succ$ applied is an RPO compatible with arities, i.e. if $arity(f) > arity(g)$ then $f \succ_p g$. Since initial equations and conclusions of inferences are eagerly split, all equations in E are orientable as rewrite rules wrt. $\succ$ except permuters. The "almost" permuters are again applied as simplification rules.

At this point, it is clear that the proof of the following theorem is analogous to that one of Theorem 5.12 of the previous section, applying the same notion of *degree* of an equation and the same results for *needed* conclusions of a topmost inference between two equations.

Theorem 6.10 Every standard presentation E can be finitely closed under superposition satisfying the conditions of Theorem 3.4. Hence the word problem in E is decidable and unification in E is decidable and finitary.

7 Conclusions and Further Work

By applying new proof techniques based on basic paramodulation and superposition, which are more powerful than the previously existing ones, larger, union-closed, purely syntactically defined classes of formulae have been identified for which the word and unification problems are decidable.

Furthermore, membership in NP of the word and unification problems has been shown for all theories for which a presentation as a saturated Horn set under basic paramodulation is provided. For shallow theories such a presentation is computable in polynomial time and the word problem is even polynomial.

It is not clear whether the results of this paper will have direct applications in practical areas such as functional programming, deductive databases, program verification, hardware verification, or natural language modeling. In practice it may be more likely to use basic paramodulation and superposition directly in such applications. None the less, the results seem significant in providing evidence that basic paramodulation and superposition are powerful — they provide the intellectual foundation needed for a variety of decidability and complexity results.

Several interesting open problems appear as a consequence of this work. A first problem (which we have started working on) is decidability of the full first-order theory of standard presentations. On the other hand, it would be interesting to more precisely determine the complexity of the word and unification problems in standard theories.

Regarding the identification of more general classes, undecidability seems to be close:
(i) For equations $s \simeq t$ where only t is required to be shallow, these problems are undecidable even when s has depth at most two and there are only unary function symbols (and hence at most one occurence of a variable per term): word equations $ab = c$ can be written as $a(b(x)) \simeq c(x)$, and it is undecidable whether an equality $a = b$ follows from a set of word equations $a_1 \ldots a_n = b_1 \ldots b_m$ even when $n \leq 2$ and $m = 1$: replace too long words abw by cw and add a word equation $ab = c$ where c is a new symbol[3].
(ii) For collapsing equations $s \simeq x$ where x is a variable, even when s has depth at most three and there are only unary function symbols (and hence at most one occurence of a variable per term) $a(b(x)) \simeq c(x)$ can be encoded by $d(a(b(x)) \simeq x$ and $c(d(y)) \simeq y$ with a new symbol d.
(iii) With (non-linear) collapsing presentations E of depth two, $a(b(x)) \simeq c(x)$ can be encoded by $g(b(x), y, a(y)) \simeq x$ and $c(g(x, x, y)) \simeq y$.
(iv) The word problem is undecidable in theories presented by equations with one ground side [Oyamaguchi, 1990].

However, in several other interesting directions there is still room. The theorems of section 3 might be applicable to other languages. Standard theories include equations $s \simeq t$ where s is linear and t is ground. The linearity requirement on s cannot be dropped (see (iv) above), but can it be weakened in some sense? How much progress can be made towards the whole class $\{s \simeq t \mid s \textit{ and } t \textit{ share only shallow variables}\}$, which is also known to be undecidable by (iv)? The only standard equations where s and t share non-shallow variables are the collapsing equations of case (ii) of the definition; can they be generalized? Is it necessary that the deep variable occurs only at depth two and below a unary symbol? Finally, it should also be possible to define an extension to Horn clauses of standard theories in the same way as shallow equations extend to Catalog.

References

[Bachmair and Ganzinger, 1994] Bachmair, L. and Ganzinger, H. (1994). Rewrite-based equational theorem proving with selection and simplification. *Journal of Logic and Computation*, 4(3):217–247.

[3] This idea subsumes the undecidability proof (reducing Turing machines) given in [Fermüller and Salzer, 1993] for the more general class $\mathcal{T}^=$ in which both sides of equations may have depth two.

[Bachmair et al., 1995] Bachmair, L., Ganzinger, H., Lynch, C., and Snyder, W. (1995). Basic paramodulation. *Information and Computation*, 121(2):172–192.

[Bachmair et al., 1993a] Bachmair, L., Ganzinger, H., and Waldmann, U. (1993a). Set constraints are the monadic class. In *Eighth Annual IEEE Symposium on Logic in Computer Science*, pages 75–83, Montreal, canada. IEEE Computer Society Press.

[Bachmair et al., 1993b] Bachmair, L., Ganzinger, H., and Waldmann, U. (1993b). Superposition with simplification as a decision procedure for the monadic class with equality. In *3rd Kurt Gödel Colloquium: Computational Logic and Proof Theory*, LNCS 713, pages 83–96. SpringerVerlag.

[Christian, 1992] Christian, J. (1992). Some termination criteria for narrowing and E-narrowing. In Kapur, D., editor, *11th International Conference on Automated Deduction*, LNAI 607, pages 582–588, Saratoga Springs, New York, USA. Springer-Verlag.

[Comon et al., 1994] Comon, H., Haberstrau, M., and Jouannaud, J.-P. (1994). Syntacticness, Cycle-Syntacticness and Shallow Theories. *Information and Computation*, 111(1):154–191.

[Dershowitz, 1987] Dershowitz, N. (1987). Termination of rewriting. *Journal of Symbolic Computation*, 3:69–116.

[Dershowitz and Jouannaud, 1990] Dershowitz, N. and Jouannaud, J.-P. (1990). Rewrite systems. In van Leeuwen, J., editor, *Handbook of Theoretical Computer Science*, volume B: Formal Models and Semantics, chapter 6, pages 244–320. Elsevier Science Publishers B.V., Amsterdam, New York, Oxford, Tokyo.

[Domenjoud, 1993] Domenjoud, E. (1993). Shallow AC Theories. In *Abstracts of the CCL Workshop*, La Escala, Spain. Also INRIA Nancy (France) research report.

[Fermüller and Salzer, 1993] Fermüller, C. and Salzer, G. (1993). Ordered Paramodulation and Resolution as a Decision Procedure. In Voronkov, A., editor, *Logic Programming and Automated Reasoning, Int. Conf.*, LNAI 698, pages 122–133. Springer-Verlag.

[Kirchner, 1986] Kirchner, C. (1986). Computing unification algorithms. In *1st IEEE Symposium on Logic in Computer Science*, pages 206–216, Cambridge, Mass, USA.

[Kirchner et al., 1990] Kirchner, C., Kirchner, H., and Rusinowitch, M. (1990). Deduction with symbolic constraints. *Revue Française d'Intelligence Artificielle*, 4(3):9–52.

[Marché, 1991] Marché, C. (1991). On ground AC-completion. In Book, R. V., editor, *Rewriting Techniques and Applications, 4th International Conference*, LNCS 488, pages 411–422, Como, Italy. Springer-Verlag.

[Marche, 1994] Marche, C. (1994). Normalised Rewriting and Normalised Completion. In *Ninth Annual IEEE Symposium on Logic in Computer Science*, Paris, France. IEEE Computer Society Press, Los Alamitos, CA, USA.

[Narendran and Rusinowitch, 1991] Narendran, P. and Rusinowitch, M. (1991). Any ground associative commutative theory has a finite canonical system. In *Fourth int. conf. on Rewriting Techniques and Applications*, LNCS 488, pages 423–434, Como, Italy. Springer-Verlag.

[Narendran and Rusinowitch, 1993] Narendran, P. and Rusinowitch, M. (1993). The Unifiability Problem in Ground AC Theories. In *Eighth Annual IEEE Symposium on Logic in Computer Science*, pages 364–370, Montreal, canada. IEEE Computer Society Press.

[Nieuwenhuis, 1995] Nieuwenhuis, R. (1995). On Narrowing, Refutation Proofs and Constraints. In Hsiang, J., editor, *6th International Conference on Rewriting Techniques and Applications*, LNCS 914, pages 56–70, Kaiserslautern, Germany. Springer-Verlag.

[Nieuwenhuis and Rubio, 1995] Nieuwenhuis, R. and Rubio, A. (1995). Theorem Proving with Ordering and Equality Constrained Clauses. *J. of Symbolic Computation*, 19(4):321–351.

[Oyamaguchi, 1990] Oyamaguchi, M. (1990). On the word problem for right-ground term-rewriting systems. *The Transactions of the IEICE*, 73(5):718–723.