Journal of Symbolic Computation 37 (2004) 1–33

Journal of Symbolic Computation

www.elsevier.com/locate/jsc

Superposition with completely built-in Abelian groups

Guillem Godoy, Robert Nieuwenhuis*

Department of LSI, Technical University of Catalonia, Jordi Girona 1, 08034 Barcelona, Spain

Received 14 August 2001; accepted 17 December 2002

Abstract

A new technique is presented for superposition with first-order clauses with built-in Abelian groups (AG). Compared with previous approaches, it is simpler, and AG-unification is used instead of the computationally more expensive unification modulo associativity and commutativity. Furthermore, no inferences with the AG axioms or abstraction rules are needed; in this sense this is the first approach where AG is completely built in.

1. Introduction

It is crucial for the performance of a deduction system that it incorporates specialized techniques to work efficiently with certain theories, since a naive handling of their axioms leads to an explosion of the search space. Perhaps the most important example of this is *paramodulation*, an inference rule specialized to equality in the context of resolution-based systems. Essentially, paramodulation builds the congruence axioms inside the inference system.

Another well-investigated line of research concerns building-in equational theories inside paramodulation and resolution-based systems. Some axioms generate many slightly different permuted versions of clauses, and for efficiency reasons it is many times better to treat all these clauses together as a single one representing the whole class, i.e. to work with a *built-in* equational theory E, and performing deduction with specialized E-matching and E-unification algorithms.

Early results on paramodulation *modulo* E were given by Plotkin (1972), Slagle (1974) and Lankford and Ballantyne (1977) and *extended* E*-rewriting* was defined by Peterson and Stickel (1981). Special attention has always been devoted to the case where E

* Corresponding author.
E-mail addresses: ggodoy@lsi.upc.es (G. Godoy), roberto@lsi.upc.es (R. Nieuwenhuis).

0747-7171/04/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0747-7171(03)00070-1

includes axioms of associativity and commutativity (AC), which occur very frequently in practical applications, and are well suited for being built in due to their permutative nature. Note that in general there is no unique most general E-unifier for a given E-unification problem, and that new variables may appear: for example, if f is an AC-symbol, then $f(x,a)$ and $f(y,b)$ have the two AC-unifiers $\sigma_1 = \{x \mapsto b, y \mapsto a\}$ and $\sigma_2 = \{x \mapsto f(b,z), y \mapsto f(a,z)\}$.

Resolution modulo E is relatively simple: there exist general completeness results for resolution with constraints, which essentially say that completeness is preserved when unification is replaced by E-unification. The reason is that resolution inferences, which take place at the atomic level, do not interfere with the built-in equational theories, which affect only the term level, and hence *lifting* can still be done (see Nieuwenhuis and Rubio, 2001). Unfortunately, for paramodulation this is far from true, and for each built-in theory special inference rules have to be designed and their completeness proved.

Paramodulation with built-in Abelian groups (AG) has been investigated by many authors: Chenadec (1986), Zhang (1993), Marché (1994, 1996), Ganzinger and Waldmann (1996), Waldmann (1998, 1999), Stuber (1998). This is not surprising since AG are of course ubiquitous in many applications of (semi-)automated reasoning. But building-in AG is also attractive for at least two more reasons.

On the one hand, due to the fact that Diophantine equation solving is easier in the integers than in the natural numbers, AG unification is easier than AC and AC1 (i.e. Abelian monoid) unification. If all free symbols are constants, then there is one single most general AG unifier and the decision problem is polynomial, whereas for AC and AC1 the decision problems are NP-complete, and for AC one may need to consider exponentially many unifiers. Although with arbitrary free symbols the decision problem is NP-complete in all three cases, AG unification behaves better in practice. Also the number of unifiers to be considered is usually much smaller and not doubly exponential as for AC (see Baader and Siekmann, 1993; Baader and Snyder, 2001 for surveys on these results).

Another aspect that makes building-in AG attractive is called *symmetrization* (e.g. in Chenadec, 1986): modulo AG $(+,-,0)$, every ground equation can be written as $u + \cdots + u \simeq t$, where the summand u is greater (w.r.t. the given term ordering $\succ$) than the summands in t. As we will see, this allows one to restrict inferences to this maximal summand and to avoid the prolific inferences with extended equations that appear in the AC case.

Symmetrization is also exploited in Marché's framework for Knuth–Bendix completion of unit equations with built-in theories (ranging from AC to commutative rings) Marché (1994, 1996). His completion procedure decides the ground word problem modulo AG by building a finite convergent rewrite system. However, his procedure is not refutation complete for equations with variables: in many cases it fails since it cannot handle symmetrization at the non-ground level.

Full first-order clauses are considered in Ganzinger and Waldmann (1996) and Waldmann (1997), where symmetrization is also central. This work focusses not on AG, but on the more general theory of cancellative Abelian monoids. It applies AC1 unification and *abstraction* rules, which, roughly, turn clauses like $C \vee f(s) \simeq t$ into $C \vee x \not\simeq s \vee f(x) \simeq t$, where x is a new variable; this of course increases the number of possible inferences on f. By specializing to torsion-free divisible AG, AC-unification and inferences into variables can be avoided, but abstraction remains necessary Waldmann (1998, 1999).

In Stuber's work on paramodulation for AG represented as integer modules (Stuber, 1998), symmetrization is again crucial, but AG unification is not applied. Instead, AC unification is used, and hence paramodulation inferences with the AG axioms on the remaining clauses are needed. For example, refuting a clause like $f(-b + x + a) \not\simeq f(0)$ requires inferences with the AG axioms, instead of directly finding the contradictory instance $b - a$ for x by AG-unification. Technically, even for the ground case, his inference rules and proofs are rather involved. In Stuber's Ph.D. Thesis (1999), proofs for the ground case are given in a uniform framework for AG and several other commutative theories.

Here we apply a variant of Bachmair and Ganzinger's model generation technique (Bachmair and Ganzinger, 1994), where the model is defined by rewriting, modulo AC of $+$, with the well-known convergent rewrite system R_{AG} for AG, plus a set of ground rewrite rules R that consists of symmetrized rules $nu \to t$ (here nu denotes $u + \cdots + u$ where u occurs n times) and their *inverse* version $-u \to (n-1)u - t$. Hence $\succ$ has to be an AC-compatible reduction ordering orienting these rules, which can be fulfilled by simple general-purpose orderings like RPO (this was already mentioned by Marché). This gives relatively simple completeness proofs for full first-order ground clauses. From our results it is easy to obtain a decision procedure for the satisfiability of arbitrary sets of ground clauses modulo AG.

For completely building-in AG at the non-ground level, and hence avoiding all inferences with the AG axioms by applying AG-unification, the main problem is: how to lift, to inferences on non-ground clauses C, the rewrite steps with $R \cup R_{\mathrm{AG}}$ on ground instances $C\sigma$? The steps with R indeed become inferences, but for the steps with R_{AG} this is precisely what we want to avoid. The key ideas to our solution are roughly as follows. We keep non-ground clauses C fully simplified w.r.t. R_{AG} (which is a cheap and useful simplification anyway). Furthermore, in the completeness proofs we consider instances with reduced[1] substitutions σ (extending some ideas from the *basic* superposition approach of Nieuwenhuis and Rubio, 1995; Bachmair et al., 1995). Some steps with R_{AG} may then still be needed in $C\sigma$ at the frontier between C and σ. But a careful analysis of these steps reveals that they can be covered by considering inferences with AG-unification on adequate subterms.

Our AG-superposition inference rules have strong ordering restrictions implying that inferences only need to involve the maximal summands of the clause. This generalizes standard superposition: summands play the role of terms.

Due to the simplicity and restrictiveness of our inference system, its compatibility with redundancy notions and constraints, and the fact that standard term orderings like RPO can be used, we believe that our techniques will become the method of choice for practice. On the theoretical side, we expect that our techniques and results will also lead to logic-based decidability and complexity results, along the lines of, e.g. Basin and Ganzinger (1996),

[1] In the preliminary version of this work, (Godoy and Nieuwenhuis, 2000), we used a different notion of irreducibility. In this article the definitions are more intuitive and we obtain shorter and simpler proofs.

Nieuwenhuis (1996, 1998), Ganzinger et al. (1999), Ganzinger and de Nivelle (1999) and Waldmann (1999).

This article is structured as follows. After the basic notions and notation given in Section 2, in Section 3 we introduce our techniques for the simple case of ground Horn clauses, and show that this can be used for deciding the satisfiability of set of general ground clauses modulo AG. Sections 4–6 are the core of this article. There, the ideas of the ground case are extended to Horn clauses with variables. This is again extended to general clauses with variables in Section 7. Finally, in Section 8 we give conclusions and mention some optimizations and other ideas for further work.

2. Basic notions

We use the standard notation and terminology for terms and rewriting of Dershowitz and Jouannaud (1990), for inference rules, clauses and equality Herbrand models of Bachmair and Ganzinger (1994) and Nieuwenhuis and Rubio (2001), and for constraints and constrained clauses of Nieuwenhuis and Rubio (1995, 2001).

Furthermore, we use the following terminology for positions p and q in a term t: we say that p is (strictly) *below* q if q is a (proper) prefix of p, and then q is (strictly) *above* p. Similarly, p is *beside* q (or *disjoint* with q) if no one is a prefix of the other. We also say that p is *below* a function symbol f in t if $t|_q$ is headed by f for some q above p, and then p is *immediately below* f if p is $q \cdot i$ for some natural number i.

The rewrite system R_{AG} consists of the following five rules:

$$
\begin{aligned}
&x + 0 \to x\\
&-x + x \to 0\\
&-(-x) \to x\\
&-0 \to 0\\
&-(x + y) \to (-x) + (-y).
\end{aligned}
$$

By AG we denote the set of seven equations consisting of these five rules (seen as equations) plus AC, the AC axioms for $+$. By $=_{\mathrm{AC}}$ and $=_{\mathrm{AG}}$ we denote the corresponding congruences on terms. In this article, rewriting with a set of rules R is always considered *modulo* AC, that is, when writing $\to_R$, we mean the relation $=_{\mathrm{AC}} \to_R =_{\mathrm{AC}}$. We denote by $nf_R(t)$ the normal form of a term t by rewriting with R, and instead of writing $nf_{R_{\mathrm{AG}}}(t)$ we sometimes write AG-$nf(t)$. By *free* function symbols we mean symbols different from $+$, $-$ and 0.

We sometimes write terms with $+$ in infix notation, without parenthesis. For example, $+(a, +(+(b, c), d))$ is written $a + b + c + d$. But we remark that this is only done at the notation level (and terms are not considered to be in flattened form as in other approaches, but this is not relevant here since we work with the rewrite relation $=_{\mathrm{AC}} \to_R =_{\mathrm{AC}}$, i.e. before each rewrite step we can apply AC-steps on the whole term, not only on the subterm that is rewritten). A *summand* is a term u headed by a free symbol. We write nu as a shorthand for the expression $u + \cdots + u$ where u occurs n times, and $-nu$ as a shorthand for $n(-u)$, and $a - b$ as a shorthand for $a + (-b)$.

An AG-*context position* in a term s is either λ or a position $p \cdot i$ such that the topmost symbol of $s|_p$ is neither $+$ nor $-$. An AG-*context* in a term s is any occurrence of a subterm of s at an AG-context position.

In this article, we assume that $\succ$ is a well-founded strict ordering on ground terms satisfying:

1. $\succ$ is AC-*compatible*, that is, $s' =_{\mathrm{AC}} s \succ t =_{\mathrm{AC}} t'$ implies $s' \succ t'$.
2. $\succ$ is total up to $=_{\mathrm{AC}}$ on the set of ground terms, that is, for all ground terms s and t, we have $s \succ t$ or $t \succ s$ or $s =_{\mathrm{AC}} t$.
3. $\succ$ orients all rules of R_{AG}, that is, $l\sigma \succ r\sigma$ for every rule $l \to r$ of R_{AG} and all grounding substitutions σ.
4. $\succ$ is monotonic on ground terms, that is, for all ground terms s, t and u, we have $u[s]_p \succ u[t]_p$ whenever $s \succ t$.

One way to build such an ordering $\succ$ is to simply use the recursive path ordering (RPO) Dershowitz (1982), applied to the terms to be compared in *flattened* form w.r.t. $+$. This flattening consists of removing all operators $+$ that are immediately below another $+$. For example, $+(a, +(f(+(a, +(b, c))), c))$ becomes $+(a, f(+(a, b, c)), c)$, which can also be written $a + f(a + b + c) + c$. Note that in the flattened form of a term t, denoted by $\mathrm{flat}(t)$, different occurrences of $+$ can have different arities (but all greater than 1).

Lemma 2. *Let $\succ$ be defined by: $s \succ t$ if $\mathrm{flat}(s) \succ_{\mathrm{rpo}} \mathrm{flat}(t)$, where $\succ_{\mathrm{rpo}}$ is an RPO with a total precedence $\succ_{\mathcal{F}}$ such that $f \succ_{\mathcal{F}} - \succ_{\mathcal{F}} + \succ_{\mathcal{F}} 0$ for all free symbols f and where all symbols have a lexicographic status, except $+$, whose status is multiset. Then $\succ$ fulfills the aforementioned requirements.*

Definition 3. A ground equation $nu \simeq n_1v_1 + \cdots + n_kv_k$ in normal form w.r.t. R_{AG} is said to be in *reductive form* if $n > 0$, the n_i are non-zero integers, and u and the v_i are summands with $u \succ v_i$. The (logically equivalent w.r.t. AG-models) *inverse reductive form* of this equation is $-u \simeq (n-1)u - n_1v_1 - \cdots - n_kv_k$.

For every equation $s \simeq t$, its reductive form can be obtained by normalizing $s + (-t) \simeq 0$ w.r.t. R_{AG} into $n_1u_1 + \cdots + n_ku_k \simeq 0$ where, say, u_1 is the maximal summand, and then, if n_1 is positive, the reductive form is $n_1u_1 \simeq -n_2u_2 - \cdots - n_ku_k$; otherwise, it is $-n_1u_1 \simeq n_2u_2 + \cdots + n_ku_k$. Note that the unary minus operator is overloaded in our notation since it is also applied to coefficients (but remember that coefficients are not part of our logical language but just a shorthand in our notation).

Example 4. If $a \succ b \succ c$ then the equation $(-a) + c + 0 + (-(-c)) + (-b) \simeq (-c) + a + b + 0$ is equivalent to $(-a) + (-a) + c + c + c + (-b) + (-b) \simeq 0$, written shortly $-2a + 3c - 2b \simeq 0$, and becomes in reductive form $2a \simeq 3c - 2b$, and in inverse reductive form $-a \simeq a - 3c + 2b$. □

Example 5. Equations in reductive form can be adequately used as terminating rewrite rules. Assume we have $a \succ b \succ c$ and the equation (in reductive form) $3a \simeq -b + c$. It can be applied either as it is, or in its inverse form $-a \simeq 2a + b - c$.

For example, $4a$ is AG-equivalent by this equation to $-2a - 2b + 2c$. Let us prove it by rewriting both terms into their respective normal forms. On the one hand, by simply applying the equation to three of its four a's, $4a$ rewrites into the normal form $a - b + c$. On the other hand, by applying the inverse form, $-2a - 2b + 2c$ rewrites into $-a - 2b + 2c + 2a + b - c$ which simplifies with R_{AG} into $a - b + c$.

Note that normal forms w.r.t. both ways of rewriting with such equations $nu \simeq v$ will always have a positive number of u's between 0 and $n - 1$, and that the inverse kind of steps is not needed if $n = 1$. The two ground inference rules of AG-superposition that are given below, in fact, correspond to these two ways of rewriting. □

3. Ground Horn case

Here we first introduce part of our techniques on the simple subcase of ground Horn clauses. We assume all equations in clauses to be eagerly maintained in reductive form, and moreover we assume negative equations $0 \not\simeq 0$ to be removed eagerly from all clauses.

Definition 6. The inference rules for ground AG-superposition are as follows:

$$\text{direct AG-superposition: } \frac{C \vee nu \simeq r \qquad D[nu]_p}{C \vee D[r]_p}$$

$$\text{inverse AG-superposition: } \frac{C \vee nu \simeq r \qquad D[-u]_p}{C \vee D[(n-1)u - r]_p} \qquad \text{if } n > 1$$

where $D|_p$ denotes a subterm of D modulo AC, that is, each $D'|_q$ is such a subterm if $D =_{\mathrm{AC}} D'$.

The ordering restrictions of AG-superposition are such that inferences are needed only if they take place *with* the strictly maximal summand and *on* a maximal summand (that is strictly maximal if it occurs in a positive equation), that is, denoting by $s \succ C$ the fact that $s \succ t$ for every summand t occurring in C, these inferences are needed only if:

1. $u \succ C$ (and remind that, by expression in reductive form, also $u \succ r$)
2. $s \succ D'$ whenever D is $D' \vee ms \simeq t$ (in reductive form) with $D|_p$ in ms
3. $s \succeq D'$ whenever D is $D' \vee ms \not\simeq t$ (in reductive form) with $D|_p$ in ms.

Note that hence inverse AG-superposition is needed only on proper subterms of summands s since in an (in)equation in reductive form the term $-u$ cannot occur elsewhere.

3.1. Completeness for the ground Horn case

We now use multiset extensions for lifting the ordering $\succ$ on terms to orderings on ground equations (in reductive form) and clauses in the usual way.

Definition 7. Let C be a ground clause, and let $\mathrm{emul}(s \simeq t)$ be $\{s, t\}$ if $s \simeq t$ is a positive equation in C, and $\{s, s, t, t\}$ if it is negative. Then we define the ordering $\succ_e$ on (occurrences of) ground equations in a clause by $e \succ_e e'$ if $\mathrm{emul}(e) \succ_{\mathrm{mul}} \mathrm{emul}(e')$. Similarly, $\succ_c$ on ground clauses is defined $C \succ_c D$ if $\mathrm{mse}(C)\ (\succ_{\mathrm{mul}})_{\mathrm{mul}}\ \mathrm{mse}(D)$, where $\mathrm{mse}(C)$ is the multiset of all $\mathrm{emul}(e)$ for occurrences e of equations in C.

Lemma 8. *Let C and D be ground clauses. If D is the reductive form of C then $C \succeq_c D$.*

Proof. Let u be the maximal summand of an equation $s \simeq t$ occurring (positively or negatively) in C. If u does not occur in the reductive form of $s \simeq t$, i.e. it has been cancelled out, then the reductive form is smaller. Otherwise the reductive form of $s \simeq t$ is of the form $nu \simeq r$ where $u \succ r$. If $-u$ occurs in $s \simeq t$ then again $nu \simeq r$ is smaller. Otherwise $s \simeq t$ is of the form $nu + s' \simeq t$ and $nu \simeq r$ is smaller (if s' is non-empty) or equal (if s' is empty). $\square$

We now show how to construct a model for sets S of ground Horn clauses closed under ground AG-superposition and where $\square \notin S$ (note that this implies the refutation completeness of ground AG-superposition). As usual (see Bachmair and Ganzinger, 1994), in order to construct the model we will generate a set of rewrite rules R_S by induction on $\succ_c$. But here the model will contain as well the rules of R_{AG}, and, as said, all rules will be applied modulo AC.

Definition 9. Let S be a set of ground Horn clauses in reductive form, and let C be a clause in S of the form $C \vee nu \simeq r$. Then C *generates* the rule $nu \to r$ if the following three conditions are satisfied:

1. $(R_C \cup \mathrm{AG})^* \not\models C$
2. $u \succ r$ and $u \succ C$
3. nu is irreducible by R_C

where R_C is the set of rules generated by clauses of S smaller than C w.r.t. $\succ_c$. Furthermore, if C generates $nu \to r$ with $n > 1$, in addition C generates its inverse form $-u \to (n-1)u - r$. The set of all rules generated by clauses in S is denoted by R_S.

We now state an essential result: $R_S \cup R_{\mathrm{AG}}$ is convergent modulo AC.

Lemma 10. *Let S be a set of ground Horn clauses in reductive form. $R_S \cup R_{\mathrm{AG}}$ is terminating and confluent modulo AC on ground terms.*

Proof. All rules in $R_S \cup R_{\mathrm{AG}}$ are oriented w.r.t. $\succ$, and hence $R_S \cup R_{\mathrm{AG}}$ is terminating for rewriting modulo AC, since $\succ$ is AC-compatible, well founded, and monotonic on ground terms. Confluence is a consequence of the following facts. By construction of R_S, for all ground rules $l \to r$ in R_S, the term l is irreducible by the ground rules in $R_S \backslash \{l \to r\}$. Furthermore, R_{AG} is well known to be confluent. Finally, the (extended) critical pairs between R_{AG} and R_S are easily shown to be joinable. This is straightforward but long, so we omit this part here (similar results are given in the literature, e.g. in Chenadec, 1986; Marché, 1996, but we have not found the exact result needed here). $\square$

Theorem 11. *AG-superposition is refutation complete for ground Horn clauses.*

Proof. Let S be a set of ground Horn clauses (whose equations are in reductive form) such that S is closed under AG-superposition and $\square \notin S$. We prove that then S is satisfiable by exhibiting an AG-model I for S, where I is the equality Herbrand interpretation defined as the congruence on ground terms generated by $R_S \cup \mathrm{AG}$. Note that, since $R_S \cup R_{\mathrm{AG}}$ is terminating and confluent, $I \models s \simeq t$ if, and only if, $s \to^*_{R_S \cup R_{\mathrm{AG}}} \leftarrow^*_{R_S \cup R_{\mathrm{AG}}} t$. We proceed by induction on $\succ_c$, that is, we derive a contradiction from the existence of a minimal (w.r.t. $\succ_c$) clause D (in reductive form) of S such that $I \not\models D$.

Let s be the maximal summand in D. Then D is either of the form $D' \vee ms \simeq t$ with $s \succ D'$ (a), or else it is $D' \vee ms \not\simeq t$ with $s \succeq D'$ (b). We first show that in both cases ms is reducible by R_S.

(a) Since $I \not\models D$, it has generated no rule of R_S. According to Definition 9, this can only be because ms is reducible by R_D. (b) Since $I \not\models D$, we have $I \models ms \simeq t$. Therefore ms and t are joinable by $R_S \cup R_{\mathrm{AG}}$, and since $ms \succ t$, the maximal side ms, which is in normal form w.r.t. R_{AG}, has to be reducible by R_S. The rule reducing ms has been generated by a clause of the form $C \vee nu \simeq v$, and there exists an inference by (direct or inverse) AG-superposition

$$\frac{C \vee l \simeq r \qquad D[l]_p}{C \vee D[r]_p}$$

where $I \not\models C \vee D[r]_p$ and D is larger w.r.t. $\succ_c$ than $C \vee D[r]_p$, and therefore, by Lemma 8, also larger than the reductive form of $C \vee D[r]_p$, contradicting the minimality of D. $\square$

3.2. *Selecting negative literals*

It is easy to see that our inference rules remain complete with *selection* of negative literals (see, e.g. Bachmair and Ganzinger, 1994), where it is assumed that in each clause with a non-empty antecedent one of its negative equations has been *selected*. In the Horn case this leads to positive unit strategies (and in the non-Horn case to positive strategies): all left premises of AG-superpositions are positive unit clauses, and the only inferences involving non-unit clauses are AG-superpositions on the selected negative equation. The following result is a simple modification of the previous one; it is immediate if we define R_S such that only unit clauses generate rules:

Theorem 12. *AG-superposition with selection is refutation complete for ground Horn clauses.*

3.3. *Deciding the satisfiability of sets of ground clauses*

From our results it is not difficult to obtain a decision procedure for the satisfiability of arbitrary sets of ground clauses modulo AG.

For the Horn inference system with selection, each inference of $l \simeq r$ on a clause D produces a smaller clause D'. Furthermore, D is a logical consequence (modulo AG) of the smaller clauses $l \simeq r$ and D', i.e. D has become *redundant* in the sense of Bachmair and Ganzinger (1994). In our procedure such redundant clauses can be removed without loss of completeness (redundant clauses never generate any rules, and in the proof of the completeness theorem, they are never the smallest counter example; see, e.g. Bachmair and Ganzinger, 1994 for details). Hence, if after each inference the maximal premise D is removed, the procedure remains complete, and at each inference the clause set decreases w.r.t. the multiset extension of the ordering and hence the process terminates, thus deciding satisfiability.

A decision procedure for the satisfiability of sets of arbitrary ground clauses modulo AG can be obtained by first transforming into Horn clauses (where $S \cup C \vee A_1 \vee \cdots \vee A_n$

is split into the disjunction of sets S_i of the form $S \cup C \vee A_i$; then S is satisfiable if some of the S_i is).

Theorem 13. *AG-superposition with selection decides the satisfiability of sets of ground clauses modulo AG.*

4. Inference rules for clauses with variables

In this section, we adapt the inference system in order to deal with equality constrained clauses with variables, where constraints are conjunctions of equalities $s = t$. As usual, the semantics of a constrained clause $C \mid T$ is the set of its ground instances, that is, the ground instances $C\sigma$ such that $T\sigma$ evaluates to true if $=$ is interpreted as $=_{\mathrm{AG}}$. Then σ is called a solution for T. The empty clause with a constraint T is hence a contradiction, denoted simply by $\square$, if, and only if, $T\theta$ is true for some ground θ.

Very roughly, the following is needed for lifting our completeness results from the ground case to equality constrained clauses with variables. If for clauses $C_1 \mid T_1$ and $C_2 \mid T_2$ there is an inference between ground instances

$$\frac{C_1\sigma \qquad C_2\sigma}{D}$$

then there exists an inference by the non-ground version of the inference rules

$$\frac{C_1 \mid T_1 \qquad C_2 \mid T_2}{D' \mid T}$$

such that D is a ground instance of $D' \mid T$.

As we will see in Section 5, for completeness it suffices to be able to do this only for instances with σ of C_1 and C_2 that are, in some technically rather involved sense, *irreducible* w.r.t. R_S, where R_S is the set of rules generated in a way similar to the previous section (but now by ground instances of clauses).

Definition 14. An equation $s = t$ is in *one-sided form* if it is of the form $e \simeq 0$ where e is in normal form w.r.t. R_{AG}.

Note that each equation has two (AG-equivalent) one-sided forms: for example, $x + y - z \simeq 0$ is equivalent to $-x - y + z \simeq 0$. In the following, we assume that all equations in clauses are kept in one-sided form. Unless explicitly stated otherwise, it does not matter which one of the two. Furthermore, for all substitutions σ, we assume w.l.o.g. that $x\sigma$ is in normal form w.r.t. R_{AG} for all x.

In order to define the non-ground inference rules, we now analyze for each inference rule how their premises have to be expressed. For simplicity, we omit the constraints, since they do not matter at this point; let us only remark that the amount of possible inferences can be further restricted in many different ways by checking their compatibility with the constraints.

4.1. Left premises of direct AG-superposition

Intuitively, our aim is the following. Let C be a clause with a positive equation $e \simeq 0$, and assume a ground instance $C\sigma$ of it generates a rule $nu \to r$ with $n > 0$, and $C\sigma$ is the left premise of an AG-superposition. Then for the non-ground case we have to be able to express $e \simeq 0$ as $s \simeq t$ such that the terms $s\sigma$ and $t\sigma$ have, respectively, nu and r as normal forms w.r.t. R_{AG}, and then perform the inference with AG-unification between s and the corresponding subterm of the right premise. *Orienting* $e \simeq 0$ as $s \simeq t$ in this way may require to *split* the variables of e into two parts:

Example 15. Consider the clauses $a + 2x \simeq b$ and $f(4a) \not\simeq f(a + b - 2c)$, where $a \succ b \succ c$. Assume that, for the instance where $x \mapsto a + c$, the equation $a + 2x \simeq b$ generates the rule $3a \to b - 2c$. Then there exists a ground inference

$$\frac{3a \to b - 2c \qquad f(4a) \not\simeq f(a + b - 2c)}{f(a + b - 2c) \not\simeq f(a + b - 2c)}$$

applied to three of the a's in $f(4a)$, where the conclusion in reductive form becomes $0 \not\simeq 0$ and hence the empty clause.

To cover this inference at the non-ground level, x has to be split into y (which, roughly, will contain the maximal summands in $x\sigma$) and z (for the remaining summands). Hence $a + 2x \simeq b$ can be *oriented* as $a + 2y \simeq b - 2z$. Then there is a non-ground inference

$$\frac{a + 2y \simeq b - 2z \qquad f(4a) \not\simeq f(a + b - 2c)}{f(a + b - 2z) \not\simeq f(a + b - 2c)}$$

unifying $a + 2y$ with three of the a's in $f(4a)$. AG-unifying both sides of the conclusion (which will be another inference rule; see below) detects the instance where z is c; the corresponding instance has a reductive form $0 \not\simeq 0$ and hence the contradiction is found. $\square$

Definition 16. Let e be a term of the form $n_1 s_1 + \cdots + n_p s_p + m_1 x_1 + \cdots + m_q x_q$ where the s_i are non-variable summands, the x_i are variables, and the n_i and m_i are non-zero integers. By splitting each x_i into two new variables y_i and z_i, and splitting the summands into two disjoint sets, the equation $e \simeq 0$ can be written as an equivalent equation $s \simeq t$ of the form

$$n_1 s_1 + \cdots + n_k s_k + m_1 y_1 + \cdots + m_q y_q \simeq -n_{k+1} s_{k+1} \\ - \cdots - n_p s_p - m_1 z_1 - \cdots - m_q z_q.$$

In the following, we call each such an equation $s \simeq t$ an *orientation* for $e \simeq 0$ and we call the corresponding constraint τ of the form

$$x_1 = y_1 + z_1 \wedge \cdots \wedge x_q = y_q + z_q$$

the *splitting constraint* for this orientation.

It is not difficult to see that this notion of orientation fulfills what we wanted: if $e\sigma \simeq 0$ generates a rule $nu \to r$ then indeed for some orientation $s \simeq t$ of $e \simeq 0$ and some extension of σ in order to include the y_i and z_i, the terms $s\sigma$ and $t\sigma$ have, respectively, nu and r as normal forms w.r.t. R_{AG}. This we will see in detail in the completeness proofs.

Of course, the fewer orientations have to be considered for a given equation $e \simeq 0$, the fewer inferences will be performed, which is better for efficiency in practice. Indeed, a little more careful analysis reveals that a large number of optimizations are possible. In Section 8 we will mention some of them. It is also important for efficiency to exploit the unifiability and ordering restrictions as the strongest possible filters to avoid redundant inferences with such orientations $s \simeq t$. For example, apart from the unification restrictions of the inference itself, where s is unified with a subterm of the right premise, in the above orientation we can add $s_1 = \cdots = s_k$ to the constraint; in particular, this means, e.g. that if e is $f(\cdots) + g(\cdots) + \cdots$, then no orientation $s \simeq t$ is needed where both summands headed with f and g are in the left hand side s. In Section 8 the problem of checking the ordering restrictions is addressed.

Note that this notion of orientation does not depend on which one of $e \simeq 0$ or $-e \simeq 0$ we consider as the one-sided form, and that the non-deterministic aspect of orientation is the guess of a subset $s_1 \ldots s_k$ of the (non-variable) summands (where the guess is constrained by the requirement that all of them are AG-unifiable and by the requirements on $\succ$).

4.2. *Left premises of inverse AG-superposition*

Example 17. Consider $a \succ b \succ c$ and the clauses $f(-a+b+c) \not\simeq f(a-c)$ and $2x \simeq b$. With the instance $x \mapsto a-c$, the second equation becomes $2a \simeq b + 2c$. At the ground level, there exists an inference with inverse AG-superposition which produces $f(a-c) \not\simeq f(a-c)$. At the non-ground level, x is split into $y+z$, and the inference is performed with $-y \simeq y + 2z - b$, and we obtain $f(a+c+2z) \not\simeq f(a-c)$. From this, by AG-unification the instance $z \mapsto -c$ is found and the empty clause is obtained. $\square$

Definition 18. Let e (or $-e$) be a term of the form $x_1 + \cdots + x_n + v$, where v contains only negative variables and (positive or negative) summands, and let e' be e but where every occurrence of x_i at top-level position has been replaced by $y_i - z_i$, where y_i and z_i are new variables. The splitting constraint τ is $x_1 = y_1 - z_1 \wedge \cdots \wedge x_n = y_n - z_n$. Hence e' is of the form $y_1 - z_1 + \cdots + y_n - z_n + v$.

Then, if e' is of the form $s + e''$ where s is a positive summand, then $-s \simeq e''$ is an *inverse orientation* for $e \simeq 0$ with splitting constraint τ.

Furthermore, if e' is of the form $w + e''$ where w is a variable (i.e. w is some y_i), then $-w_1 \simeq w_2 + e''$ is an *inverse orientation* for $e \simeq 0$ with splitting constraint $\tau \wedge w = w_1 + w_2$.

Finally, if e' is of the form $-w + e''$ where w is a variable, but none of the z_i, then $w_1 \simeq -w_2 + e''$ is an *inverse orientation* for $e \simeq 0$ with splitting constraint $\tau \wedge w = w_1 + w_2$.

The splitting of the variable w in the second case of inverse orientation is the one illustrated by the previous example. Example 24 shows the necessity of the splittings of the constraint τ.

4.3. *Right premises for direct AG-superposition*

Example 19. Consider $a \succ b \succ c$, the left premise $3a \simeq b$, and the right premise $f(2x, x) \not\simeq f(a+b+2c, 2a+c)$. With the instance $\{x \mapsto 2a+c\}$, the right premise is $f(4a + 2c, 2a + c) \not\simeq f(a + b + 2c, 2a + c)$ which gives in one ground inference $f(a + b + 2c, 2a + c) \not\simeq f(a + b + 2c, 2a + c)$, which in reductive form is $0 \not\simeq 0$.

Now the question is: how can we, at the non-ground level, perform the inference into the term $2x$? (which is the term t in the definition below). By splitting x into only the two variables y and z, one gets $f(y + y + z + z, y + z) \not\simeq f(a + b + 2c, 2a + c)$, for which the ground inference cannot be lifted: it is impossible to split $y + y + z + z$ into $t_1 + t_2$ such that $t_1\sigma$ is $3a$, and $t_2\sigma$ is $a + 2c$ for some σ.

As we will see, by splitting x into three variables y, y', and z, lifting is always possible. In our example, then one gets $f(2y + 2y' + 2z, y + y' + z) \not\simeq f(a + b + 2c, 2a + c)$, where $2y + 2y' + 2z$ is split into $2y + y'$ and $y' + 2z$ (these are the terms t_1 and t_2 in the definition below). Then an AG-unifier of $3a$ and $2y + y'$ instantiates y and y' with a, and the conclusion of the non-ground inference is $f(a+b+2z, 2a+z) \not\simeq f(a+b+2c, 2a+c)$, which by one more AG-unification, where z is instantiated with c, becomes $0 \not\simeq 0$. $\square$

Definition 20. Let t be a non-variable subterm of e in a literal $e \simeq 0$ or $e \not\simeq 0$ where t is not immediately below an AG-symbol and the head symbol of t is free or $+$. W.l.o.g., let t be of the form

$$n_1 s_1 + \cdots + n_p s_p + m_1 x_1 + \cdots + m_q x_q + t'$$

where all s_i are summands, all x_i variables, all n_i and m_i are positive coefficients, and t' contains only negative summands and variables.

Then $t_1 + t_2$ is a *splitting* for t if t_1 is a term whose head symbol is free or $+$ of the form

$$k_1 s_1 + \cdots + k_p s_p + m_1 y_1 + \cdots + m_q y_q + l_1 y'_1 + \cdots + l_q y'_q$$

where $0 \le k_i \le n_i$ and $0 \le l_i < m_i$, and t_2 is

$$(n_1 - k_1)s_1 + \cdots + (n_p - k_p)s_p + m_1 z_1 + \cdots + m_q z_q + l'_1 y'_1 + \cdots + l'_q y'_q + t'$$

where l'_i is 0 if l_i is 0 (i.e. then x_i is split only into two parts y_i and z_i), and l'_i is $m_i - l_i$ otherwise. Again we denote by τ the corresponding splitting constraint.

As before, other restrictions apply; for example it is also not necessary to consider t_1 of the form $y_i + y'_i$ (i.e. if m_i is 1).

4.4. *Right premises for inverse AG-superposition*

Definition 21. Let t be a non-variable subterm of e in a literal $e \simeq 0$ or $e \not\simeq 0$ where t is not immediately below an AG-symbol.

If t is of the form $-s + t'$, where s is a summand, then $t_1 + t_2$ is an *inverse splitting* for t with empty splitting constraint τ if t_1 is $-s$ and t_2 is t'.

If t is of the form $-x + t'$, where x is a variable, then $t_1 + t_2$ is an *inverse splitting* for t if t_1 is $-y$ and t_2 is $-z + t'$, and the splitting constraint τ is $x = y + z$.

4.5. *AG-superposition rules*

Based on the notions of orientations and splittings defined in the previous subsections, we are now ready to define the inference system for Horn clauses with variables.

Definition 22. In the left premise $C \vee l \simeq r$ of the direct AG-superposition rule below, it is assumed that the actual clause is $C \vee e \simeq 0$ and that $l \simeq r$ is an orientation of $e \simeq 0$. Similarly, in the right premise, $D[t_1 + t_2]_p$ denotes that $D|_p$ is a non-variable term t that is not immediately below an AG symbol, with a splitting $t_1 + t_2$. In the same way, for the inverse AG-superposition rule, they denote inverse orientations and splittings. In all cases, τ is the conjunction of the splitting constraints of the two premises. The inference system $\mathcal{H}$ consists of the following three[2] rules for constrained clauses:

direct AG-superposition:

$$\frac{C \vee l \simeq r \mid T \qquad D[t_1 + t_2]_p \mid T'}{C \vee D[r + t_2]_p \mid T \wedge T' \wedge l = t_1 \wedge \tau}$$

inverse AG-superposition:

$$\frac{C \vee l \simeq r \mid T \qquad D[t_1 + t_2]_p \mid T'}{C \vee D[r + t_2]_p \mid T \wedge T' \wedge l = t_1 \wedge \tau}$$

AG-zero-instance:

$$\frac{C \vee e \not\simeq 0 \mid T}{C \mid T \wedge e = 0}.$$

The ordering restrictions of the superposition rules are the ones corresponding to the ground rules. More precisely, a direct (or inverse) superposition with premises $C_1 \mid T_1$ and $C_2 \mid T_2$ and conclusion $D \mid T$ is needed if, for some solution θ of T, there is a ground direct (resp. inverse) inference between the reductive forms of $C_1\theta$ and $C_2\theta$, and with conclusion $D\theta$. The AG-zero-instance rules can be restricted to maximal equations of the clause.

In the following sections, we will prove the refutation completeness of this inference system. But let us first illustrate some of the limitations and technical difficulties when dealing with constrained clauses, by means of an example taken from Nieuwenhuis and Rubio (2001). Note that in such examples where only free symbols occur, AG-superposition boils down to normal superposition.

Example 23. Consider the unsatisfiable clause set, with the ordering as in Lemma 2 based on $f \succ_{\mathcal{F}} a \succ_{\mathcal{F}} b \succ_{\mathcal{F}} c$:

1. $a \simeq b$
2. $f(x) \simeq c \mid x = a$
3. $f(b) \not\simeq c$.

No inferences that are compatible with the constraint of the second clause can be made (a superposition inference between 2 and 3 leads to a clause with an unsatisfiable constraint $x = a \wedge b = x$). This incompleteness is due to the fact that the usual lifting arguments for superposition (see Nieuwenhuis and Rubio, 2001) do not work here, since they are based on the existence of *all* ground instances of the clauses; in this case, it requires an instance $f(b) \simeq c$ of clause 2, which does not exist. This example also shows that one cannot

[2] For explanation purposes we prefer to keep the direct and inverse versions of the superposition rules, like in the ground case, in spite of the fact that the two rules itself are written identically here.

deal with arbitrary initial constraints. For constrained clauses, the alternative technique for lifting is based on the notion of irreducible instances (Nieuwenhuis and Rubio, 2001). In this article we extend this idea of irreducible substitution. It becomes technically more complex due to the built-in properties of AG (Example 29 gives an idea of it). $\square$

Example 24. In this example it is shown how the inference system performs and also the need of the splitting of variables in the right premise of inverse AG-superposition is illustrated. Consider the clause $f(x) \not\simeq f(-a) \vee x + 3a \simeq 0$. With the instance $\{x \mapsto -a\}$, the negative equation in reductive form is $0 \not\simeq 0$. The positive equation is $2a \simeq 0$, which may generate the two rules $2a \to 0$ and $-a \to a$. If one wants to refute $f(-3a) \not\simeq f(a)$, then the inverse rule has to be used. Indeed, with $-a \to a$, the term $f(-3a)$ rewrites into $f(-2a + a)$, which is $f(-a)$, which rewrites into $f(a)$.

Now we want to perform, at the non-ground level, the ground refutation corresponding to these two rewrite steps. Assume that, at the non-ground level, we consider the orientation $-a \simeq 2a + x$, i.e. without the additional splitting of x as explained in Definition 18. Then, by the corresponding inverse AG-superposition inference we obtain $f(x) \not\simeq f(-a) \vee f(x) \not\simeq f(a)$. If one adds constraints forcing a to be the maximal summand in the clause $f(x) \not\simeq f(-a) \vee x + 3a \simeq 0$ and such constraints are inherited, then no substitution different from $\{x \mapsto -a\}$ is possible (such constraints can be handled with the methods presented in Godoy and Nieuwenhuis, 2001). Now, one would want to do a new inference on x, but in $\mathcal{H}$ no inferences below variables are computed. So this shows the need of a splitting of x into $y - z$ in an inverse AG-superposition inference.

Indeed, if we do this additional splitting, the orientation becomes $-a \simeq 2a + y - z$. Then the instance under consideration is extended such that $\{y \mapsto 0, z \mapsto a\}$, and the obtained clause is $f(x) \not\simeq f(-a) \vee f(y - z) \not\simeq f(a)$, with the splitting constraint $x = y - z$. Now, it is possible to do the second inverse AG-superposition inference (the one corresponding to the second rewrite step with $-a \to a$). Applying $-a$ on $-z$, one obtains $f(x) \not\simeq f(-a) \vee f(x') \not\simeq f(-a) \vee f(y + 2a + y' - z') \not\simeq f(a)$ (here, the x of the left premise is renamed into x') with the splitting constraint $x' = y' - z'$, and extending the substitution $\{y' \mapsto 0, z' \mapsto a\}$. With this substitution, all these equations are of the form $0 \simeq 0$, and three AG-zero-instance inferences give us the desired refutation. $\square$

5. Completeness for a simple subcase

For explanation purposes, in this section we consider the simpler subcase where all free symbols are constants. Hence this is assumed in all results of this section. It is interesting to observe that in this subcase the inference rule of inverse AG-superposition is not needed.

As said before, we will deal with instances with ground substitutions σ of clauses C that are in some sense irreducible with respect to R_S, where R_S is the set of rules generated in a way similar to how it was done for the ground case in the previous section.

Example 25. Let s be a term and σ a substitution, both in normal form w.r.t. R_{AG}. Then still $s\sigma$ needs not be in normal form w.r.t. R_{AG}.

For example, if s is $-x + y + a$, $x\sigma$ is $a + b$, and $y\sigma$ is b, then $-x\sigma$ is AG-equal to $-a + (-b)$ and $s\sigma$ in AG-normal form is 0. $\square$

Example 26. The problems illustrated in Example 23 still occur in this simple case where all free symbols are constants. Again with the ordering $a \succ b \succ c$, consider

1. $a \simeq b$
2. $b + x \simeq c \mid x = a$
3. $2b \not\simeq c$.

No inferences are possible on this unsatisfiable set. □

Definition 27. Let C be a clause, let t be a term, let σ be a substitution in AG-normal form, and let R be a ground TRS.

The pair (t, σ) is *irreducible* w.r.t. R if for all variables x occurring in t, the term $x\sigma$ is irreducible w.r.t. $R^{\preceq -u}$ where u is the maximal (w.r.t. $\succ$) summand of AG-$nf(t\sigma)$.

The pair (C, σ) is irreducible w.r.t. R if (e, σ) is irreducible w.r.t. R for all equations $e \simeq 0$ of C.

Note that the notion of irreducibility for (C, σ) does not depend on which one-sided form $e \simeq 0$ is considered.

We now adapt the notion of rule generation to the non-ground case. Instead of having the rules generated by ground clauses in reductive form, now the rules are generated by the reductive forms of instances $C\sigma$ of clauses $C \mid T$ of S, where (C, σ) is irreducible:

Definition 28. Let S be a set of constrained Horn clauses, let $C \mid T$ be a clause in S with a ground instance $C\sigma$, and let G be the (ground) reductive form of $C\sigma$, where G is of the form $G' \vee nu \simeq r$. Then G *generates* the rule $nu \rightarrow r$ if the following four conditions are satisfied:

1. $(R_G \cup AG)^* \not\models G$
2. $u \succ r$ and $u \succ G'$
3. nu is irreducible by R_G
4. (C, σ) is irreducible w.r.t. R_G

where R_G is the set of rules generated by reductive forms of instances of clauses of S that are smaller than G w.r.t. $\succ_c$. Furthermore, for each generated rule $nu \rightarrow r$ with $n > 1$, in addition the rule $-u \rightarrow (n-1)u - r$ is generated. The set of all rules generated by clauses in S is denoted by R_S.

In the remainder of this section R_S always denotes the ground TRS generated for a given S as in the previous definition.

Example 29. This example illustrates how the application of generated rules correspond to inferences at the non-ground level. It also shows why the irreducibility notion is more complicated than the standard one of superposition with constraints of Bachmair et al. (1995) and Nieuwenhuis and Rubio (1995), where, roughly speaking, one simply imposes that for every variable x the term $x\sigma$ has to be irreducible w.r.t. the rewrite system R.

Consider the equation $e \simeq 0$ of the form $2x - 2a - 2b + c \simeq 0$ where $a \succ b \succ c$, and the substitution σ such that $x\sigma$ is $a + b$. We have that $e\sigma \simeq 0$ is $2a + 2b - 2a - 2b + c \simeq 0$, and its reductive form is $c \simeq 0$. The corresponding orientation at the non-ground level is $c \simeq 2a + 2b - 2x$. Due to this instance the rule $c \rightarrow 0$ may be generated. Later on, the rule

$b \to 0$ may be generated too, due to other equations. The variable x with the substitution σ is reducible by such a rule $b \to 0$. So with the standard notion of irreducibility, rules generated later on could reduce the substitution of clauses generating smaller ones. Therefore this classical notion is not adequate in our context. Roughly speaking, we need to allow such big summands that are cancelled out to be reducible.

Indeed, with the notion used here, the one of Definition 27, we will see in Lemma 30 that $x\sigma$ will be irreducible w.r.t. all generated rules with maximal summand smaller than or equal to $-c$. And indeed this irreducibility is preserved in the conclusions of inferences. Assume we want to refute $2c + y \not\simeq 0$, where $y\sigma$ is 0 with the rule $c \to 0$. Observe that $(2c+y, \sigma)$ is irreducible w.r.t. the generated R. At the non-ground level, the reduction with $c \to 0$ corresponds to an inference with the orientation $c \simeq 2a+2b-2x$, and the resulting clause is $c + 2a + 2b - 2x + y \not\simeq 0$. Observe that $(c + 2a + 2b - 2x + y, \sigma)$ is irreducible w.r.t. R, since the maximal summand of AG-$nf(c+2a+2b-2x\sigma+y\sigma)$ is c. Here, some constraints can be added, like for example $c > 2a + 2b - 2x$. Such constraints can be handled with the methods presented in Godoy and Nieuwenhuis (2001). In this case, the only possible solution σ is $x\sigma = a + b$. $\square$

The following lemma shows that our notion of orientation for left premises of direct AG-superposition fulfills the requirements.

Lemma 30. *Let $C \mid T$ be a clause whose instance $C\sigma$ with reductive form C_r generates the rule $nu \to r$.*

Then there exists an orientation $l_1 \simeq r_1$ of the positive equation $e \simeq 0$ of C, and some extension of σ in AG-normal form satisfying the splitting constraint of the orientation, and AG-$nf(l_1\sigma) = nu$ and AG-$nf(r_1\sigma) = r$. Furthermore, all variables x in r_1 satisfy that $x\sigma$ is irreducible w.r.t. $R_S^{\preceq -u}$.

Proof. W.l.o.g., let e be of the form $k_1x_1 + \cdots + k_px_p + ku + v$ where the k_i and k are (possibly zero) integers, the x_i are variables, and v is the (possibly 0) sum of constants different from u. Now consider the orientation of $e \simeq 0$ into $l_1 \simeq r_1$ where

$$l_1 = k_1y_1 + \cdots + k_py_p \; + \; ku$$
$$r_1 = -k_1z_1 - \cdots - k_pz_p \; - \; v$$

i.e. where each x_i has been split into $y_i + z_i$. Furthermore, consider the extension of σ where $y_i\sigma$ consists of all (positive or negative) u in $x_i\sigma$, and z_i is the sum of the remaining constants, that is, if $x_i\sigma =_{\mathrm{AC0}} m_iu + v_i$ where u does not occur in v_i, then $y_i\sigma = m_iu$ and $z_i\sigma = v_i$. Note that in v_i constants larger or smaller than u may appear, but not u itself.

Then AG-$nf(l_1\sigma) = nu$ and AG-$nf(r_1\sigma) = r$. It remains to be shown that every variable z_i in r_1 satisfy that $x\sigma$ is irreducible w.r.t. $R_S^{\preceq -u}$. We know that (e, σ) is irreducible w.r.t R_{C_r}, i.e. $x_i\sigma$ is irreducible w.r.t. $R_{C_r}^{\preceq -u}$. Then, since $z_i\sigma$ is a sum of constants that already appear in $x_i\sigma$, we have that $z_i\sigma$ is irreducible w.r.t. $R_{C_r}^{\preceq -u}$. $\square$

Note that in this case where all free symbols are constants, for a given clause with positive equation $e \simeq 0$ there are at most two orientations $l_1 \simeq r_1$: one where the maximal constant symbol of e (if there is any) is in l_1, and another one where there is no constant symbol at all in l_1 (if there is any variable in e).

Lemma 31. *Let e be a term such that (e, σ) is irreducible w.r.t. R_S, and let $e\sigma \simeq 0$ in reductive form be $mu \simeq v$. Furthermore, let $nu \to r$ be a rule in R_S with $1 \leq n \leq m$.*

Then there exists a splitting $e_1 + e_2$ of e and an extension of σ in AG-normal form satisfying the corresponding splitting constraint, and $(e_1+e_2)\sigma =_{\mathrm{AG}} e\sigma$ and $e_1\sigma =_{\mathrm{AG}} nu$. Moreover, all variables x in e_2 satisfy that $x\sigma$ is irreducible w.r.t. $R_S^{\preceq -u}$.

Proof. For every variable x_i in e, w.l.o.g. we have $x_i\sigma =_{\mathrm{AC0}} m_i u + v_i$ where u does not occur in v_i, and where $m_i \geq 0$ because (e, σ) is irreducible w.r.t. R_S (which contains $nu \to r$ and hence if $n > 1$ also $-u \to (n-1)u - r$).

Therefore, since $e\sigma \simeq 0$ in reductive form is $mu \simeq v$, and $m \geq n$, we can assume that $e \simeq 0$ (in one of its one-sided forms) is of the form

$$k_1x_1 + \cdots + k_px_p + ku + e' \simeq 0$$

where $k \geq 0$, e' is the (possibly 0) sum of the remaining constants and variables, and $\{x_1, \ldots, x_p\}$ is a minimal set of variables with positive coefficients k_i such that $k_1m_1 + \cdots + k_pm_p + k \geq n$ or, if k is negative, $k_1m_1 + \cdots + k_pm_p \geq n$.

Now we distinguish three possible situations:

1. $k \geq n$, and hence p is 0. Then some splitting of the form

$$e_1 = nu$$
$$e_2 = (k-n)u + e'$$

   fulfills the requirements. Note that (e_2, σ) is irreducible w.r.t. R_S since e_2 has the same variables as e and the maximal summand of AG-$nf(e_2\sigma)$ is smaller than or equal to u, the maximal summand of AG-$nf(e\sigma)$.
2. $n > k > 0$. Then $k_1m_1 + \cdots + k_pm_p + k \geq n > k_2m_2 + \cdots + k_pm_p + k$ (the latter relation by minimality of the set $\{x_1, \ldots, x_p\}$). Now let l be $n - (k_2m_2 + \cdots + k_pm_p + k)$, i.e. intuitively, l is the number of u's we need from the k_1m_1 u's in $x_1\sigma$. We assume that l mod k_1 is not 0 (the case of l mod $k_1 = 0$ is analogous and the differences are commented on below). Now let m' be l div k_1, let k' be l mod k_1, and consider the splitting

$$e_1 = ku + k_1y + k'y' + k_2y_2 + \cdots + k_py_p$$
$$e_2 = (k_1 - k')y' + k_1z + k_2z_2 + \cdots + k_pz_p + e'$$

   where every x_i is split into $y_i + z_i$, except for x_1 that is split into $y + y' + z$ (if l mod k_1 is 0 then the variable y' is not needed in the splitting and x is split into $y + z$) and let $y\sigma$ be $m'u$, let $y'\sigma$ be u, let $z\sigma$ be $(m_1 - m' - 1)u + v_1$, and for i in $2 \ldots p$, let $y_i\sigma$ be m_iu, and let $z_i\sigma$ be v_i. This fulfills the requirements, and, for similar reasons as in Lemma 30 we have that every variable x in e_2 satisfies that $x\sigma$ is irreducible w.r.t. $R_S^{\preceq -u}$.
3. $k \leq 0$. Then $k_1m_1 + \cdots + k_pm_p \geq n > k_2m_2 + \cdots + k_pm_p$. As in the previous case, assume that l mod k_1 is not 0, and let l be $n - k_2m_2 + \cdots + k_pm_p$, let m' be l div k_1, let k' be l mod k_1, and consider the splitting

$$e_1 = k_1y + k'y' + k_2y_2 + \cdots + k_py_p$$
$$e_2 = (k_1 - k')y' + k_1z + k_2z_2 + \cdots + k_pz_p + ku + e'$$

and let $y\sigma$ be $m'u$, let $y'\sigma$ be u, let $z\sigma$ be $(m_1 - m' - 1)u + v_1$, and for i in $2 \ldots p$, let $y_i\sigma$ be $m_i u$, and let $z_i\sigma$ be v_i. This fulfills the requirements, and, for similar reasons as in Lemma 30, every variable x in e_2 satisfies that $x\sigma$ is irreducible w.r.t. $R_S^{\preceq -u}$. $\square$

The proof of the previous lemma reveals that the definition of splitting of right premises (Definition 20) could be made more restrictive. Indeed this is possible, thus reducing the number of inferences that need to be considered. In fact, the following more restrictive definition is also adequate for the general case handled in the next section, where we consider arbitrary free symbols. We decided to give Definition 20 as it is because it is simpler, but here we give the more restrictive alternative (it can be skipped by all readers except the ones interested in implementing these techniques in the most optimized way).

Let t be a non-variable subterm of e in a literal $e \simeq 0$ or $e \not\simeq 0$ where t is not immediately below an AG-symbol and the head symbol of t is free or $+$. W.l.o.g., let t be of the form

$$n_1 s_1 + \cdots + n_p s_p + m_1 x_1 + \cdots + m_q x_q + t'$$

where all s_i are summands, all x_i variables, all n_i and m_i are positive coefficients, and t' contains only negative summands and negative variables.

We choose a subset of the s_i, say $\{s_1 \ldots s_{p'}\}$ with $p' \leq p$, and a subset of the x_i, say $\{x_1, \ldots, x_{q'}\}$ with $q' \leq q$. The case where the subset of summands is empty, the subset of variables contains only x_1 and m_1 is 1 is not accepted (no inferences in variables are permitted). If (i) the subset of variables is empty, we choose a summand in $\{s_1 \ldots s_{p'}\}$, say s_1, and a number $n'_1 \leq n_1$. Otherwise, if (ii) the subset of variables is non-empty we choose one of those variables, say x_1 and a number $m'_1 < m_1$.

In case (i), $t_1 + t_2$ is a *splitting* for t if t_1 and t_2 are of the form

$$n'_1 s_1 + \cdots + n_{p'} s_{p'}$$
$$(n_1 - n'_1)s_1 + n_{p'+1} s_{p'+1} \ldots + n_p s_p + m_1 x_1 + \cdots + m_q x_q + t'$$

respectively. In case (ii), split every variable x_i of $\{x_2, \ldots, x_{q'}\}$ into $y_i + z_i$. If (ii.1) m'_1 is 0, then split x_1 into $y_1 + z_1$, and otherwise, if (ii.2) m'_1 is not 0, then split x_1 into $y_1 + y'_1 + z_1$. In case (ii.1), $t_1 + t_2$ is a *splitting* for t if t_1 and t_2 are of the form

$$n_1 s_1 + \cdots + n_{p'} s_{p'} m_1 y_1 + \cdots + m_{q'} y_{q'} + t'$$
$$n_{p'+1} s_{p'+1} \ldots + n_p s_p + m_1 z_1 + \cdots + m_{q'} z_{q'} + m_{q'+1} x_{q'+1} + \cdots + m_q x_q + t'$$

respectively. In case (ii.2), $t_1 + t_2$ is a *splitting* for t if t_1 and t_2 are of the form

$$n_1 s_1 + \cdots + n_{p'} s_{p'} m_1 y_1 + m'_1 y'_1 + m_2 y_2 + \cdots + m_{q'} y_{q'} + t'$$
$$n_{p'+1} s_{p'+1} \ldots + n_p s_p + (m_1 - m'_1) y'_1 + m_1 z_1 + \cdots + m_{q'} z_{q'}$$
$$+ m_{q'+1} x_{q'+1} + \cdots + m_q x_q + t'$$

respectively.

Theorem 32. *$\mathcal{H}$ is refutation complete for constrained Horn clauses where all free symbols are constants and the initial set of clauses has only empty constraints.*

Proof. In fact, we will show that in this case where all free symbols are constants, no inferences by inverse superposition are needed. Let S be the closure under $\mathcal{H}$ of a set of

Horn clauses S_0 without constraints, and assume $\square \notin S$. Again we prove that then the equality Herbrand interpretation I defined as the congruence on ground terms generated by $R_S \cup AG$ is an AG-model for S. But now this is done in two steps. Let $Ir_{R_S}(S)$ denote the set of ground instances $C\sigma$ of $C \mid T$ in S such that (C, σ) is irreducible w.r.t. R_S.

1. First, it is proved that $I \models Ir_{R_S}(S)$, in a very similar way as for the ground case, by deriving a contradiction from the existence of such a $C\sigma$ whose reductive form is minimal w.r.t. $\succ_c$. This is done in detail below.
2. Second, from $I \models Ir_{R_S}(S)$ it follows that $I \models S$ for the following reasons. For each ground instance $C\sigma$ of a clause $C \mid T$ in S_0, consider another instance $C\sigma'$ of C, where $x\sigma'$ is the normal form w.r.t. R_S of $x\sigma$ for every variable x of C. Since T is empty (as S_0 has no constraints), $C\sigma'$ is also an instance of S_0. It is also in $Ir_{R_S}(S_0)$, since (C, σ') is obviously irreducible. Since $S_0 \subseteq S$ and $I \models Ir_{R_S}(S)$ we have $I \models Ir_{R_S}(S_0)$ and hence $I \models C\sigma'$, which implies $I \models C\sigma$, and hence we also have $I \models S_0$. But since $S_0 \models S$, this gives us $I \models S$.

We now prove the first part. Let C_r be the minimal, w.r.t $\succ_c$, reductive form of some $C\sigma$ in $Ir_{R_S}(S)$ that is an instance of a clause $C \mid T_C$ such that $I \not\models C_r$.

If C_r is a disjunction of literals of the form $0 \not\simeq 0$, then an inference by AG-zero-instance applies to any one of these literals, eliminating it, and its conclusion has a smaller false counter example.

Otherwise, as in the ground case (the proof of Theorem 11), let s be the maximal summand in C_r. Then C_r is either of the form $C'_r \vee ms \simeq t$ with $s \succ C'_r$, or else it is $C'_r \vee ms \not\simeq t$ with $s \succeq C'_r$. Then C is of the form $C' \vee e \simeq 0$ or $C' \vee e \not\simeq 0$, where the reductive forms of $C'\sigma$ and $e\sigma \simeq 0$ are C'_r and $ms \simeq t$ respectively.

As in Theorem 11, in both cases ms is reducible by R_S. Since all free symbols are constants, the rule reducing ms must be of the form $ns \to r$, with $m \geq n \geq 1$. This rule has been generated by the reductive form D_r of an instance $D\sigma$ of a clause $D \mid T_D$. Let D be of the form $D' \vee e' \simeq 0$.

Then by Lemma 30 there exists an orientation $l_1 \simeq r_1$ of $e' \simeq 0$ and an extension of σ preserving AG-equality such that AG-$nf(l_1\sigma)$ is ns and AG-$nf(r_1\sigma)$ is r, and such that every variable x in r_1 satisfies that $x\sigma$ is irreducible w.r.t. $R_S^{\preceq -u}$.

Furthermore, by Lemma 31, there exists a splitting $e_1 + e_2$ of e and a new extension of σ (here we assume as usual that both clauses C and D contain different variables and that the splittings in them are done also with different variables) that is AG-preserving such that $(e_1 + e_2)\sigma =_{\mathrm{AG}} e\sigma$, and $e_1\sigma =_{\mathrm{AG}} ns$, and where every variable x in e_2 satisfies that $x\sigma$ is irreducible w.r.t. $R_S^{\preceq -u}$.

Now, since every variable x of $r_1 + e_1$ satisfies that $x\sigma$ is irreducible w.r.t. $R_S^{\preceq -u}$, and since the maximal summand of AG-$nf((r_1 + e_1)\sigma)$ is smaller than or equal to u, it holds that $(r_1 + e_1, \sigma)$ is irreducible w.r.t. R_S.

Now, the following inference exists:

$$\frac{D' \vee l_1 \simeq r_1 \mid T_D \qquad C' \vee e_1 + e_2 \simeq 0 \mid T_C}{C' \vee r_1 + e_2 \simeq 0 \mid T_D \wedge T_C \wedge l = e_1 \wedge \tau}.$$

Its conclusion belongs to S, since S is closed under $\mathcal{H}$, and it has an instance with σ that contradicts the minimality of C_r. $\square$

6. Completeness for arbitrary Horn clauses

In this section we drop the restriction that all free symbols are constants. All definitions and proofs that are needed for this purpose follow the same intuition as in its analogue for the constants-only case, but several aspects become technically a bit more involved.

Example 33. This example shows that in the presence of arbitrary free symbols a more refined notion of irreducibility than the one of Definition 27 is needed. We continue with Example 29, and consider new problems due to the non-constant symbols. Suppose we have a unary symbol f bigger than a, b and c, and an equation $f(c) \simeq 0$. It is reducible with the rule $c \to 0$, that, at the non-ground level, is $c \to 2a + 2b - 2x$, with the substitution $\{x \mapsto a + b\}$. By the corresponding direct AG-superposition inference we obtain $f(2a + 2b - 2x) \simeq 0$. At the ground level it is of the form $f(0) \simeq 0$. Observe that $f(0) \succ x\sigma$, and hence, $x\sigma$ would be reducible by a rule with left-hand side b, that is smaller than the maximal summand of the equation. For this reason, we need a more complex notion of irreducibility, where the irreducibility of a variable x in an AG-context is only necessary for summands in $x\sigma$ that are smaller than or equal to the maximal reducible summand of such an AG-context, and not to the maximal summand in the equation. $\square$

The following definitions are parameterized by the given rewrite system R, and we always denote (possibly with subscripts) terms by s, t, u, v, positions by p, q and variables by x, y, z.

We first define irreducibility for pairs (s, σ) where s is a term and σ a substitution, both in normal form w.r.t. R_{AG}. Then still $s\sigma$ needs not be in normal form w.r.t. R_{AG}, because the following two kinds of steps may be applicable: (i) if x is a variable occurring immediately below a $-$ in s and $x\sigma$ is headed by $+$, then this $-$ is “moved inwards”; (ii) after this, some “complementary” pairs u and $-u'$ below the same $+$ are cancelled if u and u' are summands with $u =_{\mathrm{AG}} u'$.

Definition 34. Let s be a ground term, and let R be a ground TRS. We define $\mathrm{maxred}_R(s)$ to be the maximal summand u such that either:

- AG-$nf(s)$ is of the form $nu + v$ and $nu \to r \in R$; In this case we say that u is determined by a *top-level positive reduction*.
- AG-$nf(s)$ is of the form $-u + v$, and $-u \to r \in R$; Then u is determined by a *top-level negative reduction*.
- AG-$nf(s)$ is of the form $u + v$ or $-u + v$ and u is reducible at non-top-level by R; Then u is determined by a *non-top-level reduction*.
- AG-$nf(s)$ is irreducible and u is 0.

Definition 35. Let s be a term and let σ be a substitution, both in normal form w.r.t. R_{AG}, and let R be a ground TRS.

The pair (s, σ) is called *recursively irreducible* w.r.t. R if the following conditions hold. Let u be $\mathrm{maxred}_R(s\sigma)$.

1. For all x such that s is of the form $x + s'$, and all summands v with $u \succeq v$ and such that $x\sigma$ is of the form $mv + v'$,

- if u is determined by a top-level negative reduction, then either $u \succ v$ and mv is irreducible w.r.t. R, or v is u and m is positive;
- otherwise (top-level positive or non-top-level reduction) mv is irreducible w.r.t. R.

2. For all x such that s is of the form $-x + s'$, and all summands v with $u \succeq v$ and such that $x\sigma$ is of the form $mv + v'$,

   - if u is determined by a top-level negative reduction, then either $u \succ v$ and mv is irreducible w.r.t. R, or v is u;
   - otherwise (top-level positive or non-top-level reduction) mv is irreducible w.r.t. R.

3. For all t of the form $f(t_1, \ldots, t_m)$ such that AG-$nf(s\sigma)$ is of the form $t+v$ or $-t+v$ and $u \succeq$ AG-$nf(t\sigma)$, each (t_i, σ) is recursively irreducible w.r.t. R.

Definition 36. Let s be a term, let u be a summand, and let σ be a substitution, both in normal form w.r.t. R_{AG}, let C be a clause, and let R be a ground TRS.

The pair (s, σ) is called $(u\ \succeq)$-*irreducible* (resp. $(u\ \succ)$-*irreducible*) w.r.t. R if the following conditions hold.

1. For all x such that s is of the form $x + s'$ or $-x + s$, and all summands v with $u \succeq v$ (resp. $u \succ v$) and such that $x\sigma$ is of the form $mv + v'$, the term mv is irreducible w.r.t. R.
2. For all t of the form $f(t_1, \ldots, t_m)$ such that AG-$nf(s\sigma)$ is of the form $t+v$ or $-t+v$ and $u \succeq$ AG-$nf(t\sigma)$, each (t_i, σ) is recursively irreducible.

If u is the maximal summand of AG-$nf(s\sigma)$ w.r.t. $\succ$, then, we simply say that the pair (s, σ) is *irreducible* w.r.t. R.

The pair (C, σ) is irreducible w.r.t. R if (e, σ) is irreducible for all its equations $e \simeq 0$ (note that this notion does not depend on which one of the two possibilities of writing the equation like $e \simeq 0$ is chosen).

6.1. Model generation

As in the case where all free symbols are constants, which was explained in Section 5, now the AG-model induced by R is built. Again the rules are generated, exactly as in Definition 28 of Section 5, by the reductive forms of instances $C\sigma$ of clauses $C \mid T$ of S, where (C, σ) is irreducible. But now the notion of irreducibility is according to Definition 36. The main theorem of this section says that $\mathcal{H}$ is refutation complete for constrained Horn clauses if the initial set of clauses has only empty constraints. Its proof follows the same arguments as its analogue in the constants-only case, Theorem 32. Lemma 44 finds, for a given term that is reducible by R, a context inside it where the maximal summand is reducible at the top. This gives us an inference at the ground level. Lemmas 45–48 justify that there exist orientations and splittings at the non-ground level corresponding to the inference at the ground level. This new inference at the non-ground level has to satisfy some conditions of irreducibility that are justified by Lemmas 51, 53 and 54.

Lemma 37. *Let u be the maximal summand in AG-$nf(s\sigma)$, let R_1 be a rewrite system and let R_2 be a rewrite system with left hand sides of the form nw or $-w$, where $n > 0$ and w is a summand such that $w \succ u$. Let (s, σ) be recursively irreducible w.r.t. R_1.*

Then, (s, σ) is recursively irreducible w.r.t. $R_1 \cup R_2$.

Proof. We prove it by induction on the size of s. Let v be $\mathrm{maxred}_{R_1}(s\sigma)$. Observe that $u \succeq v$. Since u is the maximal summand of AG-$nf(s\sigma)$, and for all the w, we have that $w \succ u$, then $\mathrm{maxred}_{R_1 \cup R_2}(s\sigma)$ is v. Moreover, the sets $R_1^{\prec mv}$ and $(R_1 \cup R_2)^{\prec mv}$ coincide for any m. Therefore, the conditions of recursive irreducibility for variables x such that s is of the form $x + s'$ or $-x + s'$ are satisfied. Let s be of the form $t + s'$ or $-t + s'$, for a summand t of the form $f(t_1, \ldots, t_n)$, and such that $v \succeq$ AG-$nf(t\sigma)$. Then, we have that $v \succ$ AG-$nf(t_i\sigma)$. Therefore, for all the w, we have that w is greater than the maximal summand in AG-$nf(t_i\sigma)$. By induction hypothesis, (t_i, σ) is recursively irreducible w.r.t. $R_1 \cup R_2$. $\square$

Lemma 38. *Let u be the maximal summand in AG-$nf(s\sigma)$, let R_1 be a rewrite system and let R_2 be a rewrite system with left hand sides of the form nw or $-w$, where $n > 0$ and w is a summand. Let v be a ground summand in AG-normal form such that $v \succeq u$ and (s, σ) is $(v \succeq)$-irreducible w.r.t. R_1.*

If all such w satisfy that $w \succ v$, then, (s, σ) is $(v \succeq)$-irreducible w.r.t. $R_1 \cup R_2$.
If all such w satisfy that $w \succeq v$, then, (s, σ) is $(v \succ)$-irreducible w.r.t. $R_1 \cup R_2$.

Proof. We only prove the first statement (the second one is analogous). Observe that the sets $R_1^{\preceq mv}$ and $(R_1 \cup R_2)^{\preceq mv}$ coincide for any m. Therefore, the conditions of $(v \succeq)$-irreducibility for variables x such that s is of the form $x + s'$ or $-x + s'$ are satisfied. Let s be of the form $t + s'$ or $-t + s'$, for a summand t of the form $f(t_1, \ldots, t_n)$, and such that $v \succeq$ AG-$nf(t\sigma)$. Then, we have that $v \succ$ AG-$nf(t_i\sigma)$. Therefore, for all the w, we have that w is greater than the maximal summand in AG-$nf(t_i\sigma)$. By Lemma 37, (t_i, σ) is recursively irreducible w.r.t. $R_1 \cup R_2$. $\square$

Lemma 39. *If, as in the definition of R, the reductive form $C\mathrm{red}$ of $C\sigma$ generates the rules $nu \to r$ and $-u \to (n-1)u - r$, then (C, σ) is irreducible not only w.r.t. $R_{C\mathrm{red}}$, but w.r.t. $R \backslash \{nu \to r, -u \to (n-1)u - r\}$. Moreover, if $e \simeq 0$ is a negative equation of C, then (e, σ) is irreducible w.r.t. R.*

Proof. Let $e \simeq 0$ be an equation of C. Let $R^{C\mathrm{red}}$ be the set of rules generated by reductive forms bigger than Cred w.r.t. $\succ_C$. Then, $R^{C\mathrm{red}}$ is of the form $\bigcup_{i \in I}\{n_i u_i \to r_i, -u_i \to (n_i - 1)u_i - r_i\}$. All these u_i's are larger than the maximal summand of AG-$nf(e\sigma)$. Moreover, if $e \simeq 0$ is a negative equation, u is larger than the maximal summand AG-$nf(e\sigma)$. By applying Lemma 38 with $R^{C\mathrm{red}}$ and $R^{C\mathrm{red}} \cup \{nu \to r, -u \to (n-1)u - r\}$ for negative equations, the lemma follows. $\square$

Lemma 40. *Let s be a term of the form $s_1 + s_2$. Let (s, σ) be $(u \succeq)$-irreducible (resp. $(u \succ)$-irreducible) w.r.t. R', for a given summand u. Then, (s_1, σ) and (s_2, σ) are $(u \succeq)$-irreducible (resp. $(u \succ)$-irreducible) w.r.t. R'.*

Lemma 41. *Let s be a term of the form $s_1 + s_2$. Let (s, σ) be recursively irreducible w.r.t. R'. Let* $\text{maxred}_R(s\sigma) \succ \text{maxred}_R(s_1\sigma)$.

Then, (s_1, σ) is recursively irreducible w.r.t. R'.

Lemma 42. *Let (s, σ) be $(u \succeq)$-irreducible (resp. $(u \succ)$-irreducible or recursively irreducible) w.r.t. R. Let s be of the form* (i) $nx + s'$ *or* (ii) $-nx + s$. *Let x_1 and x_2 be variables not in s such that $x_1\sigma$ and $x_2\sigma$ are in AG-normal form, and $(x_1+x_2)\sigma =_{\text{AC0}} x\sigma$.*

Then, we have that (i) (nx_1+nx_2+s', σ) *or* (ii) $(-nx_1-nx_2+s', \sigma)$ *is $(u \succeq)$-irreducible (resp. $(u \succ)$-irreducible or recursively irreducible) w.r.t. R.*

Lemma 43. *Let (s, σ) be $(u \succ)$-irreducible w.r.t. R. Let s be of the form $nx + s'$. Let $x\sigma$ be:* (i) $u_1 + \cdots + u_m - v_1 - \cdots - v_k$, *or* (ii) $u_1 + \cdots + u_m$ *or* (iii) $-v_1 - \cdots - v_k$, *where the u_i and v_i are summands. Let x_1 and x_2 be variables not in s. Let $x_1\sigma$ be $u_1 + \cdots + u_m$ in cases* (i) *and* (ii), *and* 0 *otherwise. Let $x_2\sigma$ be $v_1 + \cdots + v_k$ in cases* (i) *and* (iii), *and* 0 *otherwise.*

Then, we have that $(nx_1 - nx_2 + s', \sigma)$ is $(u \succ)$-irreducible w.r.t. R.

Proof. Since s' is a subsum of s and $x_1\sigma$ is a subsum of $x\sigma$, the only doubt for reducibility is what happens with $x_2\sigma$. If $x_2\sigma$ is of the form $mv_i + v'$ for some v_i such that $u \succ v_i$, then $x\sigma$ is of the form $-mv_i + v''$. Since (s, σ) is $(u \succ)$-irreducible, v_i is irreducible w.r.t. R, and no rule with left-hand side $-v_i$ nor v_i appears in R, and hence, a term of the form $n'v_i$ is not a left-hand side of a rule of R. Hence, such variables x_2 satisfy the conditions for irreducibility, and $(nx_1 - nx_2 + s', \sigma)$ is $(u \succ)$-irreducible w.r.t. R. □

Lemma 44. *Let t be a term in AG-normal form and reducible by R. Then, there exists an AG-context t' of t, and a summand u such that u is* $\text{maxred}_R(t')$ *by top-level reduction.*

Proof. This can be proved by induction on the size of t. The term t by itself is an AG-context of t. Let v be $\text{maxred}_R(t)$. If it is by top-level reduction, then, u is v, and we are done. Otherwise, it is by non-top-level reduction, and then, v is of the form $f(v_1, \ldots, v_n)$, and one of the v_i is reducible. Then, by induction hypothesis, this v_i contains the t' and u satisfying the required condition. □

Lemma 45. *Let the reductive form* Cred *of $C\sigma$ generate the rule $nu \to r'$.*

Then there exists an orientation $l \simeq r$ of the positive equation $e \simeq 0$ of C, and an extension of σ satisfying the splitting constraint of the orientation, such that AG-$nf(l\sigma)$ is nu, AG-$nf(r\sigma)$ is r', and (r, σ) is $(u \succeq)$-irreducible w.r.t. R.

Proof. By Lemma 39, (e, σ) is irreducible w.r.t. $R \backslash \{nu \to r, -u \to (n-1)u - r\}$. In fact, it is $(u \succeq)$-irreducible w.r.t. $R \backslash \{nu \to r, -u \to (n-1)u - r\}$, since u is the maximal summand of AG-$nf(e\sigma)$. Observe that AC-changes in the substitution do not affect irreducibility. Hence we can suppose that $x_i\sigma$ is of the form (i) l_iu, or (ii) $l_iu + v_i$, or (iii) v_i, for all variables x_i in e, where v_i has no occurrences of u at top-level position. Let e' be the result of replacing each occurrence of x_i at top-level position by $y_i + z_i$, where y_i and z_i are new variables. Let σ be extended such that $y_i\sigma$ is l_iu (in cases i and ii) or 0 (in case iii), and $z_i\sigma$ is 0 (case i) and v_i otherwise. By Lemma 42, (e', σ) is $(u \succeq)$-irreducible w.r.t. $R \backslash \{nu \to r, -u \to (n-1)u - r\}$.

Now, we may write e' as $l + l'$, for terms l and l' such that l contains all the y_i, and all the summands t at top-level position in e' such that AG-$nf(t\sigma)$ is u; and l' contains all the z_i, and the rest of the summands. By Lemma 40, (l', σ) is $(u \succeq)$-irreducible w.r.t. $R\backslash\{nu \to r, -u \to (n-1)u - r\}$.

We have that AG-$nf(l\sigma)$ is nu, and AG-$nf(l'\sigma)$ is AG-$nf(-r')$. Moreover, if l' is of the form $x + l''$ or $-x + l''$, and $x\sigma$ is of the form $mv + v'$ for some summand v with $u \succeq v$, then, necessarily $u \succ v$, due to the way we have extended σ to the variables in l'. Therefore, mv is irreducible w.r.t. R, because it is irreducible w.r.t. $\{nu \to r, -u \to (n-1)u - r\}$, and w.r.t. $R\backslash\{nu \to r, -u \to (n-1)u - r\}$, since (l', σ) is $(u \succeq)$-irreducible w.r.t. $R\backslash\{nu \to r, -u \to (n-1)u - r\}$.

Furthermore, if l' is of the form $t + l''$ or $-t + l''$, for some summand t of the form $f(t_1, \ldots, t_n)$ such that $u \succeq$ AG-$nf(t\sigma)$, we have that $u \succ$ AG-$nf(t_i\sigma)$, and, by Lemma 37, (t_i, σ) is recursively irreducible w.r.t. R.

Therefore, (l', σ) is $(u \succeq)$-irreducible w.r.t. R. And, if we take r as AG-$nf(-l')$, (r, σ) is $(u \succeq)$-irreducible w.r.t. R. $\square$

Lemma 46. *Let the reductive form* Cred *of* $C\sigma$ *generate the rule* $-u \to (n-1)u - r'$.

Then there exists an orientation $l \simeq r$ *of the positive equation* $e \simeq 0$ *of* C, *and an extension of* σ *satisfying the splitting constraint of the orientation, such that AG-*$nf(l\sigma)$ *is* $-u$, *AG-*$nf(r\sigma)$ *is* $(n-1)u - r'$, *and* (r, σ) *is* $(u \succ)$*-irreducible w.r.t.* R. *Moreover, for all* x *such that* r *is of the form* $x + s$, *we have that* $x\sigma$ *is not of the form* $-u + s'$.

Proof. By Lemma 39, (e, σ) is irreducible (in fact $(u \succeq)$-irreducible) w.r.t. $R\backslash\{nu \to r, -u \to (n-1)u - r\}$. This implies that if e is of the form $x + e_2$ or $-x + e_2$ and $x\sigma$ is of the form $mv + v'$ for some summand v with $u \succ v$, then mv is irreducible w.r.t. $R\backslash\{nu \to r, -u \to (n-1)u - r\}$, and, in fact, w.r.t. R. Additionally, if e is of the form $t + e_2$ or $-t + e_2$ for some summand t of the form $f(t_1, \ldots, t_n)$ such that $u \succeq$ AG-$nf(t\sigma)$, we have that (t_i, σ) is recursively irreducible w.r.t. $R\backslash\{nu \to r, -u \to (n-1)u - r\}$. But observe that, since $u \succ$ AG-$nf(t_i\sigma)$, by Lemma 37, (t_i, σ) is recursively irreducible w.r.t. R. Altogether this implies that (e, σ) is $(u \succ)$-irreducible w.r.t. R.

Let us consider now a certain variable x that appears in e at top-level positive variable position. AC-changes in the substitution do not affect irreducibility. Hence we can assume that $x\sigma$ is of the form (i) v, or (ii) $v + w$, or (iii) w, where v (resp. w) contains only positive (resp. negative) summands at their top-level positions. Let e' be the result of replacing each occurrence of x at top-level positive variable position by $y - z$, where y and z are new variables. Let σ be extended such that $y\sigma$ is v (in cases i and ii) or 0 (in case iii), and $z\sigma$ is 0 (case i) and AG-$nf(-w)$ otherwise. By Lemma 43, (e', σ) is $(u \succ)$-irreducible w.r.t. R. We can repeat this process with all the variables in e at top-level position. Let the resulting term be e'. Again, (e', σ) is $(u \succ)$-irreducible w.r.t. R.

Since AG-$nf(e'\sigma)$ is $nu - r$, either (i) e' is of the form $x + e''$ for some variable x and $x\sigma$ is of the form $u + e'''$ or u, or (ii) e' is of the form $-x + e''$, and $x\sigma$ is of the form $-u + e'''$ or $-u$, or (iii) e' is of the form $t + e''$ for some summand t such that $t\sigma =_{\mathrm{AG}} u$.

In case (i), we replace this occurrence of x by $x_1 + x_2$, where x_1 and x_2 are new variables, and we extend σ such that $x_1\sigma$ is u, and $x_2\sigma$ is e''' or 0, depending on the case. By Lemma 42, we have that $(x_1 + x_2 + e'', \sigma)$ is $(u \succ)$-irreducible w.r.t. R. By Lemma 40

$(x_2 + e'', \sigma)$ is $(u \succ)$-irreducible w.r.t. R. Therefore $-x_1 \simeq x_2 + e''$ is an orientation that satisfies the required conditions.

Case (ii) is identical to case (i), but now, the obtained orientation is $x_1 \simeq -x_2 + e''$.

In case (iii), by Lemma 40, we have that (e'', σ) is $(u \succ)$-irreducible w.r.t. R. Therefore $-t \simeq e''$ is an orientation that satisfies the required conditions. $\Box$

Lemma 47. *Let $nu \to r'$ be a rule of R. Let (s, σ) be* (i) *irreducible or* (ii) *recursively irreducible w.r.t. R. Let AG-$nf(s\sigma)$ be of the form $nu + s'$. Let u be in case* (i) *the maximal summand of AG-$nf(s\sigma)$, or in case* (ii) $\mathrm{maxred}_R(s\sigma)$.

Then, there exists a splitting $s_1 + s_2$ of s, and an extension of σ satisfying the corresponding splitting constraint, such that $(s_1 + s_2)\sigma =_{\mathrm{AG}} s\sigma$, and $s_1\sigma$ is nu, and (s_2, σ) is $(u \succeq)$-irreducible w.r.t. R.

Moreover, in case (i), *the maximal summand of AG-$nf(s_2\sigma)$, and, in case* (ii), *the summand* $\mathrm{maxred}_R(s_2\sigma)$, *is smaller than or equal to u w.r.t. $\succ$.*

Proof. From our hypothesis, it follows that (s, σ) is $(u \succeq)$-irreducible w.r.t. R (observe that for the case (ii) u is not determined by top-level negative reduction). Moreover, if s is of the form $-x + t$, then $x\sigma$ is not of the form $-u + t'$. Since AG-$nf(s\sigma)$ is $nu + s'$, these nu's can not be provided by negative variables at top-level position. Thus s has to be of the form $m_1 x_1 + \cdots + m_q x_q + n_1 t_1 + \cdots + n_p t_p + s''$, where the m_i and n_i are positive, AG-$nf(t_i\sigma)$ is u for i in $1 \ldots p$, and $x_i\sigma$ is of the form $k_i u + v_i$ for i in $1 \ldots q$ with positive k_i, and $m_1 * k_1 + \cdots + m_q * k_q + n_1 + \cdots n_p \geq n$. Moreover, such x_i and t_i can be chosen to satisfy the following conditions: the x_i and t_i do not appear in s'', and p is maximal (i.e. if q is not 0, that is there is at least one chosen variable x_1, then no summand t such that s'' is of the form $t + s'''$ satisfies $t\sigma =_{\mathrm{AG}} u$), and q is minimal (i.e. by eliminating one variable, say x_1, we have that $m_2 * k_2 + \cdots + m_q * k_q + n_1 + \cdots + n_p < n$). The case where p is 0 and q is 1 and m_1 is 1 is not possible, since $x_1\sigma$ cannot contain more than $n - 1$ u's, because it would be reducible w.r.t. R, contradicting the $(u \succeq)$-irreducibility of (s, σ).

For facility of explanations, we assume that $m_1 * k_1 + \cdots + m_q * k_q + n_1 + \cdots n_p$ is exactly n. Other situations are treated analogously, by doing the corresponding additional splittings as explained in Lemma 31.

Now, we split every x_i into $y_i + z_i$, where y_i and z_i are new variables, and σ is extended such that $y_i\sigma$ is $k_i u$, and $z_i\sigma$ is v_i. Thanks to Lemma 42, the obtained term is $(u \succeq)$-irreducible w.r.t. R. It may be written $s_1 + s_2$, where s_1 contains all the t_i and all the y_i, and s_2 contains the rest of summands and variables. The AG-normal form of $s_1\sigma$ is nu, and of $s_2\sigma$ is s', and $s_1 + s_2$ is a splitting for s. By Lemma 40, (s_2, σ) is $(u \succeq)$-irreducible w.r.t. R.

Moreover, in case (i), the maximal summand of AG-$nf(s_2\sigma)$, and, in case (ii), the summand $\mathrm{maxred}_R(s_2\sigma)$, is smaller than or equal to u w.r.t. $\succ$. Observe that, omitting the summand u, the AG-normal forms of $s\sigma$ and $s_2\sigma$ coincide. $\Box$

Lemma 48. *Let $n > 1$ and $-u \to (n-1)u - r'$ be a rule of R. Let (s, σ) be recursively irreducible w.r.t. R. Let AG-$nf(s\sigma)$ be of the form $-u + s'$. Let u be* $\mathrm{maxred}_R(s\sigma)$.

Then, there exists a splitting $s_1 + s_2$ of s, and an extension of σ satisfying the corresponding splitting constraint, such that $(s_1 + s_2)\sigma =_{\mathrm{AG}} s\sigma$, and $s_1\sigma$ is $-u$, and (s_2, σ) is $(u \succ)$-irreducible and recursively irreducible w.r.t. R. Moreover, we have that $u \succeq$ $\mathrm{maxred}_R(s_2\sigma)$.

Proof. From our hypothesis, it follows that (s, σ) is $(u \succ)$-irreducible w.r.t. R. Moreover, if s is of the form $x + s''$, then $x\sigma$ is not of the form $-u + s'''$. Since the AG-normal form of $s\sigma$ is $-u + s'$, then, either (i) s is of the form $-v + t$ for some summand v such that AG-$nf(v\sigma)$ is u, or (ii) s is of the form $-x + t$ for some variable x such that $x\sigma$ is of the form $u + s'''$ or u.

In case (i), we may take $-v$ as s_1, and t as s_2. Then $s_1 + s_2$ is a splitting for s, and $u \succeq \mathrm{maxred}_R(s_2\sigma)$.

In case (ii), we may split x into $x_1 + x_2$, for new variables x_1 and x_2, and extend σ such that $x_1\sigma$ is u and $x_2\sigma$ is s''' or 0, depending on the case. By Lemma 42, $-x_1 - x_2 + t$ is $(u \succ)$-irreducible and recursively irreducible w.r.t. R. We may take $-x_1$ as s_1, and $-x_2 + t$ as s_2. Then $s_1 + s_2$ is a splitting for s, and $u \succeq \mathrm{maxred}_R(s_2\sigma)$.

In both cases, by Lemmas 40 and 41, we have that (s_2, σ) is $(u \succ)$-irreducible and recursively irreducible w.r.t. R. $\square$

Lemma 49. *Let t be a summand. Let $(t + s, \sigma)$ be recursively irreducible w.r.t. R. Let AG-$nf(t\sigma)$ be smaller than or equal to* $\mathrm{maxred}_R((t + s)\sigma)$ *w.r.t. $\succ$. Let t' be a summand such that (t', σ) is recursively irreducible w.r.t. R, and AG-$nf(t\sigma) \succ$ AG-$nf(t'\sigma)$.*

Then (AG-$nf(t' + s), \sigma)$ is recursively irreducible w.r.t. R.

Proof. Let u be $\mathrm{maxred}_R(t\sigma + s\sigma)$. After replacing t by t', this maximal reducible summand does not increase. Moreover, if u is $\mathrm{maxred}(t'\sigma + s\sigma)$, then it is due to the same reason as before (top-level positive reduction, or top-level negative reduction or non-top-level reduction). Except for t', the variables and summands that appear in AG-$nf(t' + s)$ at top-level position are the same ones that appear in $t + s$ at top-level position, and with the same sign. Therefore, the conditions for irreducibility are satisfied for the variables at top-level position. But also for t', since it is recursively irreducible w.r.t. R. $\square$

Lemma 50. *Let (s, σ) be recursively irreducible w.r.t. R. Let t be an AG-normal form of $s\sigma$. Let p be an AG-context position in t such that $t|_p$ is reducible w.r.t. R.*

Then there exists an AG-context position q in s such that $s\sigma|_q =_{\mathrm{AG}} t|_p$, and $(s|_q, \sigma)$ is recursively irreducible, and for all terms r, $s[r]_q\sigma =_{\mathrm{AG}} t[r\sigma]_p$.

Moreover, let (r, σ) be recursively irreducible w.r.t. R, and let $t|_p \succ$ AG-$nf(r\sigma)$.

Then (AG-$nf(s[r]_q), \sigma)$ is recursively irreducible w.r.t. R.

Proof. This is proved by induction on the size of s. In the case where p is λ, q is λ, and all the results are obvious. Therefore, suppose that p is not λ. Then, p is of the form $p'.p''$, where $t|_{p'}$ is a summand of the form $f(t_1, \ldots, t_n)$ at the AG-context λ. Let u be $\mathrm{maxred}(t)$. Since $t|_{p'}$ is reducible, we have that $u \succeq t|_{p'}$. An AG-context of $t|_{p'}$ is reducible, and therefore there is an i such that t_i is reducible, and p is of the form $p'.i.p'''$.

Since t is an AG-normal form of $s\sigma$, we have that, either (i) s is of the form $x + s'$ or $-x + s'$, and $x\sigma$ is of the form $t|_{p'} + s''$ or $-t|_{p'} + s''$; or (ii) s is of the form $v + s'$ for some summand v such that $v\sigma =_{\mathrm{AG}} t|_{p'}$ and t is of the form $t|_{p'} + t'$; or (iii) s is of the form $-v + s'$ for some summand v such that $v\sigma =_{\mathrm{AG}} t|_{p'}$ and t is of the form $-t|_{p'} + t'$.

In case (i), $x\sigma$ is of the form $mt|_{p'} + s''$, and $t|_{p'}$ is reducible at non-top position by R, and $\mathrm{maxred}_R(t) \succeq t|_{p'}$, but $t|_{p'}$ cannot be $\mathrm{maxred}_R(t)$ by top-level reduction (observe that for the rules $nu \to r$ of R such u's are irreducible at non-top by R). Altogether this contradicts the hypothesis of recursive irreducibility, and therefore, only cases (ii) and (iii) are possible. In fact, we consider only case (ii), since case (iii) is analogous. The summand v has to be of the form $f(v_1, \ldots, v_n)$, and $v_i\sigma =_{\mathrm{AG}} t_i$. By induction hypothesis, there exists an AG-context position q' in v_i such that $v_i\sigma|_{q'} =_{\mathrm{AG}} t_i|_{p'''}$, and for all terms r, $v_i[r]_{q'}\sigma =_{\mathrm{AG}} t[r\sigma]_{p'''}$. Moreover, if (r, σ) is recursively irreducible w.r.t. R, and $t|_p \succ$ AG-$nf(r\sigma)$, we have that (AG-$nf(v_i[r]_{q'}), \sigma)$ is recursively irreducible w.r.t. R. Moreover, $(f(v_1, \ldots,$ AG-$nf(v_i[r]_{q'}), v_n), \sigma)$ is recursively irreducible w.r.t. R. Finally, by Lemma 49, $(f(v_1, \ldots,$ AG-$nf(v_i[r]_{q'}), v_n) + s', \sigma)$ is recursively irreducible w.r.t. R. □

Lemma 51. *Let (s, σ) be irreducible w.r.t. R. Let t be an AG-normal form of $s\sigma$. Let p be an AG-context position in t different from λ such that $t|_p$ is reducible w.r.t. R.*

Then there exists an AG-context position q in s different from λ, such that $s|_q\sigma =_{\mathrm{AG}} t|_p$, and $(s|_q, \sigma)$ is recursively irreducible, and for all terms r we have $s[r]_q\sigma =_{\mathrm{AG}} t[r\sigma]_p$.

Moreover, let (r, σ) be recursively irreducible w.r.t. R, and let $t|_p \succ$ AG-$nf(r\sigma)$.

Then (AG-$nf(s[r]_q), \sigma)$ is irreducible w.r.t. R.

Proof. The proof is analogous to the previous one, except for the fact that, instead of doing induction, it refers to the previous lemma, and that we need a modification of Lemma 49 for dealing with irreducible pairs instead of recursively irreducible pairs w.r.t. R. □

Lemma 52. *Let (s, σ) be $(u \succeq)$-irreducible w.r.t. R.*

If $\mathrm{maxred}_R(s\sigma)$ *is smaller than or equal to u, then (s, σ) is recursively irreducible w.r.t. R.*

If the maximal summand of AG-$nf(s\sigma)$ is smaller than or equal to u, then (s, σ) is irreducible w.r.t. R.

Proof. Direct by applying the definition. □

Lemma 53. *Let (r, σ) and (t, σ) be $(u \succeq)$-irreducible w.r.t. R.*

Then, (AG-$nf(r + t), \sigma)$ is $(u \succeq)$-irreducible w.r.t. R.

Additionally, suppose that $\mathrm{maxred}_R((r + t)\sigma)$ *is smaller than or equal to u w.r.t. $\succ$. Then (AG-$nf(r + t), \sigma)$ is recursively irreducible w.r.t. R.*

Moreover, if the maximal summand of AG-$nf((r + t)\sigma)$ is smaller than or equal to u, then (AG-$nf(r + t), \sigma)$ is irreducible w.r.t. R.

Proof. Observe that r and t are in AG-normal form. Therefore, the AG-normal form of $r+t$ is obtained by eliminating some summands at the AG-context λ, by the inverse rule. If AG-$nf(r+t)$ is of the form $x+s$ or $-x+s$, then, either r or t is of the form $x+s'$ or $-x+s'$, and, therefore, $x\sigma$ satisfies the corresponding requirements. If AG-$nf(r + t)$ is of the form $v + s$ or $-v + s$ for a given summand $v = f(v_1, \ldots, v_n)$ such that $u \succeq$ AG-$nf(v\sigma)$, then, either r or t is of the form $v + s'$ or $-v + s'$, and hence such a v satisfies the corresponding requirements. Therefore (AG-$nf(r + t), \sigma)$ is $(u \succeq)$-irreducible w.r.t. R.

The rest of the proof is a direct consequence of Lemma 52. □

Lemma 54. *Let $n > 1$, and $-u \to (n - 1)u - r'$ be a rule of R. Let AG-$nf(r\sigma)$ be $(n - 1)u - r'$.*

Let (r, σ) be $(u \succ)$-irreducible w.r.t. R, and if r is of the form $x + s$, then, $x\sigma$ is not of the form $-u + s'$.

Let (t, σ) *be* $(u \succ)$*-irreducible, and recursively irreducible w.r.t.* R*, and AG-*$nf(t\sigma)$ *is not of the form* $u + s'$.

Let $\mathrm{maxred}_R(t\sigma)$ *be smaller than or equal to* u *w.r.t.* $\succ$.

Then, $(\text{AG-}nf(r+t), \sigma)$ *is recursively irreducible w.r.t.* R.

Proof. Since r and t are in AG-normal form, the AG-normal form of $r + t$ is obtained by eliminating some summands at the AG-context λ, by the inverse rule.

Observe that, since AG-$nf(r\sigma)$ is $(n-1)u - r'$, and AG-$nf(t\sigma)$ is not of the form $u + s'$, it holds that AG-$nf((r+t)\sigma)$ is of the form $mu + s''$ or s'', where s'' does not contain u's at the AG-context λ, and m is negative, or positive but smaller than n.

If m is positive, or AG-$nf((r+t)\sigma$ is of the form s'', then, $\mathrm{maxred}_R((r+t)\sigma)$ is a certain v smaller than u w.r.t. $\succ$, and $(\text{AG-}nf(r+t), \sigma)$ is $(v \succeq)$-irreducible w.r.t. R, since both (r, σ) and (t, σ) are $(u \succ)$-irreducible. By Lemma 52, $(\text{AG-}nf(r+t), \sigma)$ is recursively irreducible w.r.t. R.

From now on, we assume that AG-$nf((r+t)\sigma)$ is of the form $mu + s''$, for a given negative m. In this case, AG-$nf(t\sigma)$ contains more than $n-1$ negative u's, and hence $\mathrm{maxred}_R(t\sigma)$ and $\mathrm{maxred}_R(r + t\sigma)$ has to be u by top-level negative reduction.

If AG-$nf(r+t)$ is of the form $x + s$ and $x\sigma$ is of the form $kv + v'$ for a given summand v with $u \succeq v$, then, either r or t is of the form $x + s_1$. In both cases, if $u \succ v$, then, since both (r, σ) and (t, σ) are $(u \succ)$-irreducible, we have that kv is irreducible w.r.t. R. Therefore assume that $x\sigma$ is of the form $ku + v'$ (i.e. v is u), and then, for satisfying the recursive-irreducibility conditions it is enough to show that k is positive. If r is of the form $x + s_1$, by our hypothesis k is positive. If t is of the form $x + s_1$ then, k is positive due to the fact that (t, σ) is recursively irreducible, and u is $\mathrm{maxred}_R(t\sigma)$ determined by top-level negative reduction.

If AG-$nf(r+t)$ is of the form $-x+s$, and $x\sigma$ is of the form $kv+v'$ for a given summand v with $u \succeq v$, then, either r or t is of the form $x + s_1$. In both cases, if $u \succ v$, then, since both (r, σ) and (t, σ) are $(u \succ)$-irreducible, we have that kv is irreducible w.r.t. conditions for such $-x + s$ and kv are satisfied trivially.

If AG-$nf(r+t)$ is of the form $v + s$ or $-v + s$ for a given summand $v = f(v_1, \ldots, v_n)$ such that $u \succeq$ AG-$nf(v\sigma)$, then, either r or t is of the form $v + s'$ or $-v + s'$. Since both (r, σ) and (t, σ) are $(u \succ)$-irreducible, it holds that all the (v_i, σ) are recursively irreducible w.r.t. R. $\square$

Theorem 55. *$\mathcal{H}$ is refutation complete for constrained Horn clauses if the initial set of clauses has only empty constraints.*

Proof. This proof is analogous to the one for Theorem 32. The differences are in how it is proved that $I \vDash Ir_{R_S}(S)$.

Let $C\mathrm{red}$ be the minimal, w.r.t $\succ_c$, reductive form of some $C\sigma$ in $Ir_{R_S}(S)$ that is an instance of a clause $C \mid T_C$ such that $I \nvDash C\mathrm{red}$.

If $C\mathrm{red}$ is a disjunction of literals of the form $0 \not\simeq 0$, then an inference by AG-zero-instance applies to any one of these literals, eliminating it, and its conclusion has a smaller false counter example.

Otherwise, as in the ground case (the proof of Theorem 11), let s be the maximal summand in $C\mathrm{red}$. Then $C\mathrm{red}$ is either of the form $C\mathrm{red}' \vee ms \simeq t$ with $s \succ C\mathrm{red}'$ (a), or else it is $C\mathrm{red}' \vee ms \not\simeq t$ with $s \succeq C\mathrm{red}'$ (b). As in Theorem 11, in both cases ms is reducible by R. Then, by Lemma 44 there exists an AG-context s' that is a subterm of ms, and a summand u such that u is $\mathrm{maxred}_R(s')$ by top-level reduction. Therefore, a rule in R of the form $nu \to r'$ or $-u \to (n-1)u - r'$ reduces s', and it has to be $nu \to r'$ if s' is ms; and moreover, no rule with bigger left-hand side reduces s'.

Therefore, C is of the form $C' \vee e \simeq 0$ or $C' \vee e \not\simeq 0$, where $ms - t$ is an AG-normal form of $e\sigma$.

The rule reducing $ms - t$ (at the AG-context $ms - t$ or in an AG-context inside s), has been generated by the reductive form Dred of an instance $D\sigma$ of a clause $D \mid T_D$. Let D be of the form $D' \vee d \simeq 0$. Now, we distinguish two cases:

(a) If the rule reducing ms is $nu \to r'$, then, by Lemma 45, there exists an orientation $l \simeq r$ of $d \simeq 0$ such that AG-$nf(l\sigma)$ is nu and AG-$nf(r\sigma)$ is r'. Moreover, (r, σ) is $(u \succeq)$-irreducible w.r.t. R. Now, we analyze two possibilities:

(a.1) If s' is ms, then s is u, and AG-$nf(e\sigma)$ is $mu - t$, for $m \geq n$. Moreover, u is the maximal summand of $ms - t$ and (e, σ) is irreducible w.r.t. R. By Lemma 47, there exists a splitting $e_1 + e_2$ of e such that $(e_1 + e_2)\sigma =_{\mathrm{AG}} e\sigma$, and $e_1\sigma$ is nu, and (e_2, σ) is $(u \succeq)$-irreducible w.r.t. R, and the maximal summand of AG-$nf(e_2\sigma)$ is smaller than or equal to u. By Lemma 53, (AG-$nf(r + e_2), \sigma)$ is irreducible w.r.t. R. Now, the following inference exists:

$$\frac{D' \vee l \simeq r \mid T_D \qquad C' \vee e_1 + e_2 \simeq 0 \mid T_C}{C' \vee r + e_2 \simeq 0 \mid T_D \wedge T_C \wedge l = e_1. \wedge \tau}$$

Its conclusion belongs to S, since S is closed under $\mathcal{H}$, and it has an instance with σ contradicting the minimality of $C\mathrm{red}$.

(a.2) If s' is inside s, i.e. $(ms - t)|_p$ is s' for some position p below some s, then, by Lemma 51, there exists an AG-context position q in e such that $e|_q\sigma =_{\mathrm{AG}} s'$, and $(e|_q, \sigma)$ is recursively irreducible w.r.t. R, and for all terms r'', $e[r'']_q\sigma =_{\mathrm{AG}} (ms - t)[r''\sigma]_p$. Moreover, if (r'', σ) is recursively irreducible and $s' \succ$ AG-$nf(r''\sigma)$, then, $(e[r'']_q, \sigma)$ is irreducible w.r.t. R.

Now, we will obtain the concrete r'' that is interesting for us. Denote $e|_q$ by e'. Observe that e' is recursively irreducible w.r.t. R, and s' is of the form $nu + s''$, and $\mathrm{maxred}_R(s')$ is u. By Lemma 47, there exists a splitting $e'_1 + e'_2$ of e' such that $(e'_1 + e'_2)\sigma =_{\mathrm{AG}} e'\sigma$, and $e'_1\sigma$ is nu, and (e'_2, σ) is $(u \succeq)$-irreducible w.r.t. R, and $\mathrm{maxred}_R(e'_2\sigma)$ is smaller than or equal to u. By Lemma 53, (AG-$nf(r + e'_2), \sigma)$ is recursively irreducible w.r.t. R. This AG-$nf(r + e'_w)$ is the r'' we wanted. Now, the following inference exists:

$$\frac{D' \vee l \simeq r \mid T_D \qquad C' \vee e[e'_1 + e'_2]_q \simeq 0 \mid T_C}{C' \vee e[r + e'_2]_q \simeq 0 \mid T_D \wedge T_C \wedge l = e'_1 \wedge \tau}.$$

Its conclusion belongs to S, since S is closed under $\mathcal{H}$, and it has an instance with σ contradicting the minimality of $C\mathrm{red}$.

(b) If the rule reducing ms is $-u \to (n-1)u - r'$, then, the contradiction of the minimality of $C\mathrm{red}$ follows, now, from Lemmas 46, 48, 51 and 54; in a similar way to case (a.2). □

7. General clauses

The inference system is extended to non-Horn clauses in the standard way, with (equality) factoring, which in the ground case is

$$\text{AG-}\textit{factoring}: \quad \frac{C \vee nu \simeq r \vee nu \simeq r'}{C \vee r \not\simeq r' \vee nu \simeq r'}$$

with the ordering restrictions that u is the maximal summand in the clause, which does not appear in a negative equation, and where $nu \simeq r$ is maximal w.r.t. $\succ_e$.

For the non-ground case, the two equations involved have to be oriented as the left premises of AG-superposition (note that if both orientations require to split a certain variable x, then it needs to be split only once). Let us denote by $\mathcal{I}$ the rules of $\mathcal{H}$ (with the same ordering restrictions as the factoring rule) plus this additional rule. By a relatively standard adaptation of the rule generation with respect to the Horn case (i.e. as for standard superposition, see Bachmair and Ganzinger, 1994), we obtain the following:

Theorem 56. *The inference system $\mathcal{I}$ is refutation complete for general clauses.*

8. Conclusions

A new technique has been presented for superposition with first-order clauses with built-in AG. Compared with previous approaches, it is simpler, and AG-unification is used instead of the computationally more expensive unification modulo AC. Furthermore, no inferences with the AG axioms or abstraction rules are needed; in this sense this is the first approach where AG is completely built in. It may be possible to extend our techniques to other built-in theories, like rings or fields, provided suitable convergent term rewrite systems (possibly modulo AC) exist.

On the theoretical side, we believe that our techniques and results may lead to logic-based decidability and complexity results, along the lines of, e.g. Basin and Ganzinger (1996), Nieuwenhuis (1998), Ganzinger and de Nivelle (1999) and Waldmann (1999).

On the practical side, due to the simplicity and restrictiveness of our inference system, its compatibility with redundancy notions and constraints, and the fact that standard term orderings like RPO can be used, we believe that our techniques will become the method of choice for practice. However, it is clear that much work remains to be done in order to make the techniques described in this article ready for practice, in spite of the fact that, in the meantime, some of the problems for dealing with AG-ordering constraints have been solved (Godoy and Nieuwenhuis, 2001). The authors plan to develop a first experimental implementation in the coming years in order to obtain more insight in aspects like how and when to compute redundancies, or orientations and splittings.

We now very briefly comment on a few aspects that have not been treated yet in this article.

Our completeness proofs are compatible with the notions for redundancy and saturation as in the *basic* framework of Nieuwenhuis and Rubio (1995) and Bachmair et al. (1995). Note that, by dealing with constrained clauses, no AG-unifiers are computed. Instead, the unification problems are stored in the constraints and a constrained clause $C \mid T$

is redundant if T is unsatisfiable. Apart from the well-known basicness restriction, an additional advantage is that only one conclusion is generated, instead of one conclusion for each AG-unifier Vigneron (1994) and Nieuwenhuis and Rubio (1997).

Checking the ordering restrictions in our framework is different from the usual situation. Instead of checking whether, say, for given terms s and t, there exists some ground σ such that $s\sigma \succ_{\text{rpo}} t\sigma$, we need to check whether this holds after normalizing both sides by R_{AG}, that is, whether AG-$nf(s\sigma) \succ_{\text{rpo}}$ AG-$nf(t\sigma)$. Deciding the satisfiability of such constraints is NP-complete Godoy and Nieuwenhuis (2001). One can also add information to the constraint language of Godoy and Nieuwenhuis (2001) for stating that if $n_1 s_1 + \cdots + m_1 y_1 + \cdots$ is the left hand side of an orientation (Definition 16) then all s_i are equal and all summands in the y_i are equal to these s_i.

It is also possible to find sufficient conditions for ruling out redundant inferences without fully deciding satisfiability. In practice, for efficiency reasons, such approximations are used as well for standard superposition. Neither soundness nor completeness require to actually decide ordering constraints.

Example 57. Suppose s is $f(f(0) - x)$ and t is x. It is easy to see that $s\sigma \succ_{\text{rpo}} t\sigma$ for all σ. But if σ is $\{x \to f(0)\}$, both terms normalize w.r.t. R_{AG} into $f(0)$. □

The fact that ordering restrictions are checked after normalization w.r.t. R_{AG} complicates optimizations related to the analysis of the so-called *shielded* variables of a clause C, that is, variables that occur below a free symbol in C.

Example 58. In the context of Ganzinger and Waldmann (1996) and Stuber (1998), shieldedness of variables like x in the clause $f(x - f(a)) \not\simeq 0 \vee 2x \simeq b$ allow one to conclude that $2x$ cannot contain the maximal summand of $C\sigma$ for any σ and hence $2x$ need not be used as left premise in any inference. In our case, the instance where $x\sigma$ is $f(a)$ may generate the rule $2f(a) \to b$, and hence we can rule out the inferences only for other instances. Similar optimizations apply to right premises. □

Also other shieldedness-related optimizations can be used. For example, let $e \simeq 0$ be an equation of a clause C where e is of the form $s + n_1 x_1 + \cdots + n_k x_k \simeq 0$ and the distinct variables x_i do not occur elsewhere in s or in C. If $n_i = 1$ (or $n_i = -1$) for some i, then such an equation $e \simeq 0$ collapses the theory: $s + x \simeq 0$ implies $s + (-s + t) \simeq 0$ and hence $t \simeq 0$ for every t. Hence one can assume that any such a clause $C \vee s + x \simeq 0$ is eagerly replaced by C. This can be combined with the fact that $e \simeq 0$ is logically equivalent modulo AG to $s + nz \simeq 0$, where $n = \gcd(n_1, \ldots, n_k)$ and z is a new variable.

Acknowledgements

Both authors have been partially supported by the Spanish CICYT project MAVERISH ref. TIC2001-2476-C03-01.

References

Baader, F., Siekmann, J., 1993. Unification theory. In: Gabbay, D.M., Hogger, C.J., Robinson, J.A. (Eds.), Handbook of Logic in Artificial Intelligence and Logic Programming. Oxford University Press.

Baader, F., Snyder, W., 2001. Unification theory. In: Robinson, J.A., Voronkov, A. (Eds.), Handbook of Automated Reasoning, vol. 1, Elsevier Science Publishers and MIT Press, pp. 445–532.

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

Bachmair, L., Ganzinger, H., Lynch, Chr., Snyder, W., 1995. Basic paramodulation. Information and Computation 121 (2), 172–192.

Basin, D., Ganzinger, H., 1996. Complexity analysis based on ordered resolution. In: LICS, Eleventh Annual IEEE Symposium on Logic in Computer Science, IEEE Computer Society Press, New Brunswick, New Jersey, USA, pp. 456–465.

Le Chenadec, P., 1986. Canonical forms in finitely presented algebras, Research Notes in Theoretical Computer Science. Pitman, London; Wiley, New York, Toronto.

Dershowitz, N., 1982. Orderings for term-rewriting systems. Theoretical Computer Science 17 (3), 279–301.

Dershowitz, N., Jouannaud, J-P., 1990. Rewrite systems. In: van Leeuwen, J. (Ed.), Handbook of Theoretical Computer Science, vol. B, Formal Models and Semantics. Elsevier Science Publishers B.V., Amsterdam, New York, Oxford, Tokyo, pp. 244–320 (Chapter 6).

Ganzinger, H., de Nivelle, H., 1999. A superposition decision procedure for the guarded fragment with equality. In: LICS, 14th IEEE Symposium on Logic in Computer Science. Trento, Italy, pp. 295–305.

Ganzinger, H., Meyer, C., Veanes, M., 1999. The two-variable guarded fragment with transitive relations. In: LICS, 14th IEEE Symposium on Logic in Computer Science, Trento, Italy, pp. 24–34.

Ganzinger, H., Waldmann, U., 1996. Theorem proving in cancellative Abelian monoids. In: McRobbie, M.A., Slaney, J.K. (Eds.), CADE, Thirteenth International Conference on Automated Deduction, LNAI, vol. 1104. Springer, New Brunswick, pp. 388–402.

Godoy, G., Nieuwenhuis, R., 2000. Paramodulation with built-in Abelian groups. In: LICS, 15th IEEE Symposium on Logic in Computer Science, IEEE Computer Society Press, Santa Barbara, USA, pp. 413–424.

Godoy, G., Nieuwenhuis, R., 2001. Ordering constraints for deduction with built-in Abelian semigroups, monoids and groups. In: LICS, 16th IEEE Symposium on Logic in Computer Science, IEEE Computer Society Press, Boston, USA, pp. 38–47.

Lankford, D.S., Ballantyne, A.M., 1977. Decision procedures for simple equational theories with commutative-associative axioms: complete sets of commutative-associative reductions, Technical Report Memo ATP-39. Department of Mathematics and Computer Science, University of Texas, Austin, TX.

Marché, 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.

Marché, C., 1996. Normalized rewriting: an alternative to rewriting modulo a set of equations. Journal of Symbolic Computation 21 (3), 253–288.

Nieuwenhuis, R., 1996. Basic paramodulation and decidable theories. In: Eleventh Annual IEEE Symposium on Logic in Computer Science. IEEE Computer Society Press. New Brunswick, New Jersey, USA, pp. 473–482.

Nieuwenhuis, R., 1998. Decidability and complexity analysis by basic paramodulation. Information and Computation 147, 1–21.

Nieuwenhuis, R., Rubio, A., 1995. Theorem proving with ordering and equality constrained clauses. Journal of Symbolic Computation 19 (4), 321–351.

Nieuwenhuis, R., Rubio, A., 1997. Paramodulation with built-in AC-theories and symbolic constraints. Journal of Symbolic Computation 23 (1), 1–21.

Nieuwenhuis, R., Rubio, A., 2001. Paramodulation-based theorem proving. In: Robinson, J.A., Voronkov, A. (Eds.), Handbook of Automated Reasoning. Elsevier Science Publishers and MIT Press, pp. 371–444.

Peterson, G.E., Stickel, M.E., 1981. Complete sets of reductions for some equational theories. Journal of the Association for Computing Machinery 28 (2), 233–264.

Plotkin, G., 1972. Building in equational theories. Machine Intelligence 7, 73–90.

Slagle, J.R., 1974. Automated theorem-proving for theories with simplifiers, commutativity, and associativity. Journal of the ACM 21 (4), 622–642.

Stuber, J., 1998. Superposition theorem proving for Abelian groups represented as integer modules. Theoretical Computer Science 208 (1–2), 149–177.

Stuber, J., 1999. Superposition Theorem Proving for Commutative Algebraic Theories, Dissertation. Technische Fakultät, Universität des Saarlandes, Saarbrücken, Germany.

Vigneron, L., 1994. Associative Commutative Deduction with constraints. In: Bundy, A. (Ed.), CADE, 12th International Conference on Automated Deduction, LNAI, vol. 814. Springer-Verlag, Nancy, France, pp. 530–544.

Waldmann, U., 1997. Cancellative Abelian Monoids in Refutational Theorem Proving, Dissertation. Universität des Saarlandes, Saarbrücken, Germany.

Waldmann, U., 1998. Superposition for divisible torsion-free Abelian monoids. In: Kirchner, C., Kirchner, H. (Eds.), CADE-98, Proceedings of the Fifteenth International Conference on Automated Deduction, LNAI, vol. 1421. Springer, Lindau, Germany, pp. 144–159.

Waldmann, U., 1999. Cancellative superposition decides the theory of divisible torsion-free Abelian groups. In: Logic Programming and Automated Reasoning, Int. Conf., LNAI, vol. 1705. Springer-Verlag, Tbilisi, Georgia, pp. 131–147.

Zhang, H., 1993. A case study of completion modulo distributivity and Abelian groups. In: Kirchner, C. (Ed.), RTA-93, Proceedings of the 5th International Conference on Rewriting Techniques and Applications, LNCS, vol. 690. Springer-Verlag, pp. 32–46.