MAX-PLANCK-INSTITUT FÜR INFORMATIK

Theorem Proving in Cancellative Abelian Monoids

Harald Ganzinger
Uwe Waldmann

MPI–I–96–2–001 January 1996

Im Stadtwald
D 66123 Saarbrücken
Germany

Authors' Addresses

Max-Planck-Institut für Informatik
Im Stadtwald
66123 Saarbrücken
Germany
E-Mail: {hg,uwe}@mpi-sb.mpg.de
Phone: +49 681 302 {5361,5431}

Publication Notes

An extended abstract of this report has been submitted.

Acknowledgements

The research described in this paper was supported in part by the ESPRIT Basic Research Working Group 6028 (Construction of Computational Logics). We wish to thank Leo Bachmair, Deepak Kapur, David Plaisted, Georg Struth, and Jürgen Stuber for helpful discussions on the subject of this paper.

Abstract

We describe a refined superposition calculus for cancellative abelian monoids. They encompass not only abelian groups, but also such ubiquitous structures as the natural numbers or multisets. Both the AC axioms and the cancellation law are difficult for a general purpose superposition theorem prover, as they create many variants of clauses which contain sums. Our calculus requires neither explicit inferences with the theory clauses for cancellative abelian monoids nor extended equations or clauses. Improved ordering constraints allow us to restrict to inferences that involve the maximal term of the maximal sum in the maximal literal. Furthermore, the search space is reduced drastically by certain variable elimination techniques.

Keywords

Automated Theorem Proving, First-Order Logic, Superposition, Cancellative Abelian Monoids, Associativity, Commutativity, Variable Elimination, Term Rewriting.

1 Introduction

To be useful in applications such as program verification and synthesis, a theorem prover must combine mathematical with meta-mathematical reasoning. Theories from which program properties are to be derived are divided into parts that specify standard mathematical structures, including numbers, lists, multisets, graphs, and into other parts that provide the axioms for additional free function and predicate symbols. The latter describe, in a more or less ad hoc manner, objects, and their properties, of the particular domain of application for which a program is to be written.

There are several distinct lines of investigation along which previous attempts have been made for attacking this problem. In hierarchic situations where free function symbols either do not exist or where they are defined in a "sufficiently complete" manner on top of the primitive structures, mathematical knowledge can be incorporated by using decision procedures and constraint solvers as black boxes. This is the case, for instance, in constraint logic programming, in resolution theorem proving with theory unification [10, 11], or in theorem proving with constraints [19]. The case of sufficient-completely defined free functions has been investigated e.g. by Bachmair, Ganzinger, and Waldmann [6]; Avenhaus and Becker [1] have described a related order-sorted approach.

In an algebraic context, however, sufficient completeness practically excludes uninterpreted function symbols. The situation is similar when one considers, for instance, theorem provers based on extensions of Presburger arithmetic. As they lack the possibility to handle existential quantification, they are suited for verification (where instantiations are known), but not for proof discovery and program synthesis (Hines [16]). Furthermore, experiments by Boyer and Moore [9] show that mathematical routines rarely have a chance to contribute to a proof, unless they are very tightly interwoven with the rest of the prover. Heuristic approaches that involve such a more sophisticated combination of modules, e.g., in the Boyer-Moore prover or in the PVS-system [22], have proven to be very useful in practice.

Our own line of research is based on the integrating approach, where the inference rules are adapted to the theory in a specific way. Equality is the prime example for this technique: In paramodulation and superposition, there are no resolution inferences with the transitivity or symmetry axioms. Rather, the equality axioms are coded as inference rules, which can be subject to specific ordering restrictions. Other transitive relations, in particular orderings, can be handled by related rewriting techniques yielding chaining calculi as they have been investigated by Bledsoe, Kunen, and Shostak [8], Hines [15], and

Bachmair and Ganzinger [3], among others. The laws of associativity and commutativity of binary functions have previously been integrated into paramodulation calculi: Extended clauses and unification modulo AC make explicit inferences with the AC axioms unnecessary (see Peterson and Stickel [23] for the case of unit equations, and Wertz [29] and Bachmair and Ganzinger [2], for the clausal case). Unfortunately, AC-unification is doubly exponential, and it shows in practice. By using constraints, AC-unification can be replaced by the "only" simply exponential problem of AC-unifiability, as described by Nieuwenhuis and Rubio [21], and Vigneron [28]. On the other hand, (some) equality constraints need to be solved for simplification. One of the results of the present paper is that there are other ways to avoid especially prolific cases of AC-unification in the presence of more algebraic structure.

The present paper describes a refined superposition calculus for cancellative abelian monoids. They encompass not only abelian groups, but also such ubiquitous structures as the natural numbers or multisets. Like the AC axioms, the cancellation law is difficult for a general purpose superposition theorem prover, as it creates many variants of clauses which contain sums. The main highlights of our calculus are the following. (i) There is no need for explicit inferences with the theory clauses for cancellative abelian monoids. Hence, there is no generation and recombination of different variants of one and the same clause. (ii) There is no need for extended equations or clauses. By virtue of this fact, many especially prolific instances of AC-unification are avoided. In fact, AC-unification can be replaced by ACU-unification. (iii) The ordering constraints for superposition can be further refined. As in previous calculi one may ignore non-maximal literals, as well as the smaller side of an equation for chaining. Here, in addition, only the maximal term of a (maximal) sum can participate in an inference. (iv) The general notion of redundancy as it was introduced by Bachmair and Ganzinger in [4] can be appropriately refined. It forms the basis for developing specific forms of the usual simplification techniques. Simplification is an indispensable component of any saturation-based theorem prover. In this paper, redundancy will in particular allow to show admissibility of certain variable elimination techniques.

A crucial indication of the practicality of any approach for integrating mathematical theories by specifically refined forms of resolution and paramodulation or chaining is the extent to which they avoid a certain especially prolific form of inference. It is the type of inference in which the main term in one of the premises is a variable, and, hence, unification is no longer an effective filter. For the equational case it is known that paramodulation into or below variables is not needed. For dense, total orderings without endpoints similar results have been obtained in [3, 8, 15, 24]. In the calculus to be described below, inferences are computed with maximal terms in sums. The sum itself, obviously, is not a variable, but the term may be one. The fact that certain inferences involving such variables cannot be avoided is an indication for why non-refined superposition strategies can perform badly for monoids with cancellation. In our refined calculus, only unshielded variables, i.e., variables that do not occur somewhere else as arguments of free function symbols, pose a problem in this regard. That leaves us with the problem of how to deal with unshielded variables. Fortunately unshielded variables can be eliminated in many, though not all cases without affecting refutational completeness.

With these characteristics, the present approach is a considerable improvement over Hsiang, Rusinowitch, and Sakai's extension of ordered paramodulation [17, 26] to handle cancellation laws. Our approach is related to normalized rewriting modulo the group axioms (Marché [20]) and superposition for integer modules (Stuber [27]). Both handle only the stronger case of groups. As will become apparent, working in "non-groups" makes some aspects of equational theorem proving significantly more difficult while others are simplified. For instance, there will sometimes be a need for abstraction[1]; in other situations, the number of overlaps is reduced, and in non-groups it may be easier to eliminate positive occurrences of unshielded variables.

2 Algebraic Foundations

An abelian semigroup is an algebraic structure consisting of a non-empty set G and a binary relation $+$ that satisfies the associativity axiom and the commutativity axiom

$$\text{(A)} \qquad \forall x, y, z: \qquad x + (y + z) = (x + y) + z$$
$$\text{(C)} \qquad \forall x, y: \qquad x + y = y + x\,.$$

An abelian monoid is an abelian semigroup $(G, +)$ with a constant $0 \in G$ such that the identity axiom holds:

$$\text{(U)} \qquad \forall x: \qquad x + 0 = x\,.$$

An abelian semigroup $(G, +)$ or monoid $(G, +, 0)$ is called cancellative, if it satisfies additionally the cancellation axiom

$$\text{(K)} \qquad \forall x, y, z: \qquad x + z = y + z \;\Rightarrow\; x = y\,,$$

or in other words, if the difference between two elements is uniquely defined whenever it exists. A cancellative abelian semigroup or monoid is an abelian

[1] This would not even be possible in the purely equational framework of [20].

group if and only if difference is a total function. Typical examples of cancellative abelian monoids where difference is partial are the natural numbers under addition, the non-zero integers under multiplication, and multisets under union.

We denote the set of the axioms (A), (C), (U), and (K) by ACUK, and the respective subsets by AC, ACU, and ACK. For $m \in \mathbf{N}^+$, $\sum_{i\in\{1,\ldots,m\}} x_i$ is an abbreviation for the sum $x_1 + \cdots + x_m$ and mx is an abbreviation for the m-fold sum $x + \cdots + x$. In a monoid, $\sum_{i\in\emptyset} x_i$ and $0x$ are defined as 0.

Lemma 2.1 *Let $(G, +)$ be a cancellative abelian semigroup and let $b, c \in G$. If $b + c = b$, then G is a monoid and c is its unique identity element.*

Proof. For every $x \in G$, $x + b + c = x + b$, hence by cancellation $x + c = x$. If furthermore $b' + c' = b'$ for some $b', c' \in G$, then $b' + c' = b' = b' + c$, hence by cancellation $c' = c$. □

It is a well-known result that every finite cancellative abelian semigroup is a group. The following lemma generalizes this fact.

Lemma 2.2 *Let $(G, +)$ be a cancellative abelian semigroup and I be a finite non-empty set of indices. For all $i \in I$, let $a_i \in G$ and let f_i be a function from G to G. If*

$$\bigvee_{i\in I} x + f_i(x) = a_i$$

holds for every $x \in G$, then G is a group.

Proof. Let $b = \sum_{j\in I} 2a_j$, then

$$\bigvee_{i\in I} \sum_{j\in I} 2a_j + f_i(b) = a_i\,.$$

Pick any $k \in I$ such that

$$\sum_{j\in I} 2a_j + f_k(b) = a_k$$

and let $K = I \setminus \{k\}$. By Lemma 2.1, $\sum_{j\in K} 2a_j + a_k + f_k(b)$ is an identity element of G. We denote it by 0.

$$\sum_{j\in K} 2a_j + a_k + f_k(b) = 0$$

As every a_i occurs at least once in this sum, every a_i has an inverse, namely the remainder of the sum. Let us denote the inverse of a_i by a'_i, then for every $x \in G$

$$\bigvee_{i \in I} x + f_i(x) + a'_i = 0\,.$$

Thus for every $x \in G$ there is an $i \in I$ such that $f_i(x) + a'_i$ is the inverse of x. As every x has an inverse, G is a group. □

3 The Cancellative Superposition Calculus

3.1 Preliminaries

We will develop our calculus in the framework of equational clauses. An equation e is an unordered pair $\{t, t'\}$, usually written as $t \approx t'$. A literal is either an equation e (also called a positive literal) or a negated equation $\neg\, e$ (also called a negative literal). The symbol $[\neg]\, e$ denotes either of these. Instead of $\neg\, t \approx t'$, we sometimes write $t \not\approx t'$. A clause is a finite multiset of literals. Terms are formed over a many-sorted signature (without subsorts or overloading), so every variable x comes with a unique declaration $x : S$ and every function symbol with a unique declaration $f : S_1 \ldots S_n \to S$ (for $n \in \mathbf{N}$). In particular, 0 and + have the declarations $0 : \to S_{\mathrm{CAM}}$ and $+ : S_{\mathrm{CAM}} S_{\mathrm{CAM}} \to S_{\mathrm{CAM}}$.[2] We consider only well-formed terms and equations and assume that every sort is inhabited, i.e., that for every sort there exists at least one ground term. Predicates p different from $\approx$ are coded using function symbols p', so the literal $[\neg]\, p(t)$ is represented by the equation $[\neg]\, p'(t) \approx \mathit{true}$. We assume that the reader is familiar with standard concepts and notations in the area of rewriting (to be found for instance in the survey of Dershowitz and Jouannaud [12]), and in the area of superposition-based theorem proving (see Bachmair and Ganzinger [4]).

The symbol $=_{\mathrm{ACU}}$ denotes the congruence generated by ACU. The ACU-congruence class of a term t is $[t]_{\mathrm{ACU}} = \{\, t' \mid t =_{\mathrm{ACU}} t' \,\}$.

Definition 3.1 *A function symbol that is different from* 0 *and* + *is called a free function symbol. A term is called atomic, if it is not a variable and its top symbol is different from* +. *A term* t *is called a proper sum, if* $t = t_1 + t_2$ *and* $t_1 \neq_{\mathrm{ACU}} 0$, $t_2 \neq_{\mathrm{ACU}} 0$.

The set of all terms is the disjoint union of the three sets $\{\, t \mid \exists x\colon\ x$ is a variable, $t =_{\mathrm{ACU}} x \,\}$, $\{\, t \mid \exists s\colon\ s$ is atomic, $t =_{\mathrm{ACU}} s \,\}$, and $\{\, t \mid \exists s\colon\ s$ is a

[2] There is no scalar multiplication in our signature, so mt is merely an abbreviation for $t + \cdots + t$.

proper sum, $t =_{\mathrm{ACU}} s$ }. We can therefore extend the terminology above to ACU-congruence classes, and say that $[t]_{\mathrm{ACU}}$ is a variable (an atomic term, a proper sum), if there is some $s \in [t]_{\mathrm{ACU}}$ with this property.

We say that a term t occurs in s at the top, if there is a position $o \in \mathrm{Pos}(s)$ such that $s|_o = t$ and for every proper prefix o' of o, $s(o')$ equals $+$. We say that t occurs in s below a free function symbol, if there is a position $o \in \mathrm{Pos}(s)$ such that $s|_o = t$ and $s(o')$ is a free function symbol for some proper prefix o' of o. We extend this terminology to ACU-congruence classes, and say that $[t]_{\mathrm{ACU}}$ occurs in $[s]_{\mathrm{ACU}}$ at the top (below a free function symbol), if there are some $t' \in [t]_{\mathrm{ACU}}$ and $s' \in [s]_{\mathrm{ACU}}$ with this property. For instance, $[2a]_{\mathrm{ACU}}$ and $[a + f(a + b)]_{\mathrm{ACU}}$ occur at the top of $[3a + 2f(a + b)]_{\mathrm{ACU}}$, $[a]_{\mathrm{ACU}}$ occurs both at the top and below a free function symbol.

Definition 3.2 *A reduction ordering $\succ$ is called* ACU*-compatible, if $s' =_{\mathrm{ACU}} s \succ t =_{\mathrm{ACU}} t'$ implies $s' \succ t'$.*

Every ACU-compatible reduction ordering extends naturally to a reduction ordering on ACU-congruence classes.

For ground terms, we can obtain an ACU-compatible reduction ordering $\succ$ from an arbitrary AC-compatible ordering $\succ_1$ by defining $s \succ t$ if $s{\downarrow} \succ_1 t{\downarrow}$, where $s{\downarrow}$ denotes the normal form of s under rewriting with the rule $x + 0 \to x$. We can lift this ordering to non-ground terms by defining $s \succ t$ if $s\theta \succ t\theta$ for all ground instances $s\theta$ and $t\theta$. However, as shown by Jouannaud and Marché [18], it happens quite frequently that $\succ$ orders a pair of terms in an operationally undesirable way, or that $s[x]$ and $t[x]$ are uncomparable because $s[0] \succ t[0]$ but $s[u] \prec t[u]$ for all non-zero ground terms u.[3] This is a serious problem, if one is interested in classical rewriting. It is not a hindrance, though, for calculi like superposition or unfailing completion, which are preferably implemented using constraints. (In fact, Jouannaud and Marché's method can be considered as a variant of unfailing completion with constraints.)

Definition 3.3 *We say that an* ACU*-compatible ordering has the multiset property, if whenever a ground atomic term u is greater than v_i for every i in a finite index set $I \neq \emptyset$, then $u \succ \sum_{i \in I} v_i$.*

From now on, $\succ$ will always denote an ACU-compatible ordering that has the multiset property and is total on ACU-congruence classes.[4] Examples of

[3] Jouannaud and Marché's statement that "AC1-rewrite orderings cannot really exist" [18] should be taken with a grain of salt, however.

[4] In practice, it is sufficient if the ordering can be extended to a total ordering.

orderings with these properties are obtained from the associative path ordering (Bachmair and Plaisted [7]) or the ordering of Rubio and Nieuwenhuis [25] with precedence $f_n \succ \ldots \succ f_1 \succ + \succ 0$ by comparing $s{\downarrow}$ and $t{\downarrow}$ as described above.

A ground literal e is called true in a set E of ground equations, if $e \in E$. A ground literal $\neg\, e$ is called true in E, if $e \notin E$. A ground clause is called true in E, if at least one of its literals is true in E. If a ground literal or clause is not true in E, it is called false in E.

A set E of ground equations is called a model of a clause C, if every ground instance $C\theta$ of C is true in E; it is called a model of a set N of clauses, if it is a model of every $C \in N$. If N and N' are sets of clauses, we write $N \models N'$ if every model of N, ACUK, and the equality axioms is a model of N'. In other words, $\models$ denotes entailment modulo ACUK. If C is a clause, $N \models C$ is a shorthand for $N \models \{C\}$.

Convention 3.4 *For the remainder of this paper, we will work* only *with* ACU*-congruence classes, rather than with terms. To simplify notation, we will omit the* $[\_]_{\mathrm{ACU}}$ *and drop the subscript of* $=_{\mathrm{ACU}}$*. So* all *terms, equations, substitutions, inference rules, etc., are to be taken modulo* ACU*, that is, as representatives of their congruence classes.*

The ordering restrictions that we use are more refined than usual, in that they are based primarily on maximal terms in sums.

Definition 3.5 *Let e be a ground equation, then the maximal atomic term of e (with or without multiplicity) is defined in the following way:*

- *If e is an equation of the form $nu + \sum_{i \in I} s_i \approx mu + \sum_{j \in J} t_j$, where u, s_i, and t_j are atomic terms, $n \geq m \geq 0$, $n \geq 1$, and $u \succ s_i$ and $u \succ t_j$ for all $i \in I$, $j \in J$, then* $\mathrm{mt}(e) = u$ *and* $\mathrm{mt}_{\#}(e) = nu$.

- *If e is an equation of the form $u \approx v$, where u doesn't have sort S_{CAM} and $u \succeq v$, then* $\mathrm{mt}(e) = \mathrm{mt}_{\#}(e) = u$.

Definition 3.6 *The symbol* $\mathrm{ms}(t)$ *denotes the multiset of all non-zero atomic terms occurring at the top of a ground term t, i.e.,*

- $\mathrm{ms}(t) = \{\, v_j \mid j \in J \,\}$*, if $t = \sum_{j \in J} v_j$ and all v_j are non-zero atomic terms. (In particular* $\mathrm{ms}(0) = \emptyset$*, as J may be empty.)*

- $\mathrm{ms}(t) = \{t\}$*, if t doesn't have sort S_{CAM}.*

If e is a ground equation $t \approx t'$, then $\mathrm{ms}(e)$ *is the multiset union of* $\mathrm{ms}(t)$ *and* $\mathrm{ms}(t')$*.*

Definition 3.7 *The ordering* $\succ_X$ *on ground terms is defined as follows:* $s \succ_X t$ *if there is an* $s' \in \mathrm{ms}(s)$ *such that* $s' \succ t'$ *for all* $t' \in \mathrm{ms}(t)$*. For arbitrary terms,* $s \succ_X t$ *if* $s\theta \succ_X t\theta$ *for all ground instances* $s\theta$ *and* $t\theta$*.*

Definition 3.8 *The ordering* $\succ$ *on terms is extended to an ordering* $\succ_L$ *on literals as follows: Every ground literal* $[\neg]\ s \approx t$ *is mapped to the quadruple*

$$(\mathrm{mt}_{\#}(s \approx t), pol, \mathrm{ms}(s \approx t), \{s, t\}),$$

where pol is 1 *for negative literals and* 0 *for positive ones. Two ground literals are compared by comparing their associated quadruples using the lexicographic combination of the ordering* $\succ$ *on terms, the ordering* $>$ *on* $\mathbf{N}$*, the multiset extension of* $\succ$ *and the multiset extension of* $\succ$*. The ordering is lifted to possibly non-ground literals in the usual way, so* $[\neg]\ e_1 \succ_L [\neg]\ e_2$ *if and only if* $[\neg]\ e_1\theta \succ_L [\neg]\ e_2\theta$ *for all ground instances* $[\neg]\ e_1\theta$ *and* $[\neg]\ e_2\theta$*. In order to use the ordering* $\succ_L$ *to compare equations, the latter are identified with positive literals.*

The ordering $\succ_C$ *on clauses is the multiset extension of the literal ordering* $\succ_L$*.*

As $\succ_L$ and $\succ_C$ are obtained from noetherian orderings by multiset extension and lexicographic combination, they are noetherian, too.

Definition 3.9 *A substitution* σ *is called t-preserving, if* $t\sigma = t$*.*

Any ACU-unification algorithm (e.g., Herold and Siekmann [14]) can be used to compute complete sets of t-preserving ACU-unifiers by simply taking all variables in t as constants.

We need the concept of t-preservation when we deal with variables that are known to correspond to atomic terms on the ground level. Consider the two terms $s_1 = 2x$ and $s_2 = 2y + z$. There are unifiers of s_1 and s_2, e.g., $\{x \mapsto y' + z',\ y \mapsto y',\ z \mapsto 2z'\}$. There is no unifier, though, that maps z to a non-zero atomic ground term. Neither does a z-preserving unifier of s_1 and s_2 exist.

Lemma 3.10 *Let* $s_1 = \sum_{i \in I} m_i x_i + m'u$ *and* $s_2 = \sum_{j \in J} n_j y_j + n'u$*, where the* x_i *and* y_j *are pairwise distinct variables, u is either atomic or a variable, and none of the* x_i *and* y_j *occurs in u. Let U be a complete set of u-preserving* ACU*-unifiers of* s_1 *and* s_2*. Suppose that* θ *is an* ACU*-unifier of* s_1 *and* s_2 *and that* $u\theta$ *is a non-zero atomic ground term. Then* $\theta = \sigma\rho$ *for some* $\sigma \in U$ *and some substitution* ρ*.*

Proof. Let U' be a complete set of ACU-unifiers of s_1 and s_2, and let $\sigma' \in U'$ such that $\theta = \sigma'\rho'$.

If u is atomic, we may assume without loss of generality that σ' maps all variables in u to fresh variables. So there is a variable renaming τ such that $\sigma'\tau$ is a u-preserving ACU-unifier of s_1 and s_2. Hence $\sigma'\tau = \sigma\tau'$ for some $\sigma \in U$, and $\theta = \sigma\tau'\tau^{-1}\rho'$.

If u is a variable, we may assume without loss of generality that $u\sigma' = \sum_{k\in K} \mu_k z_k$ for $\mu_k \in \mathbf{N}^+$ and fresh variables z_k. As $u\theta = u\sigma'\rho'$ is a non-zero atomic ground term, $K = \{k_0\} \cup K'$, where $z_k\rho' = 0$ for $k \in K'$ and $\mu_{k_0} = 1$. Let ρ'' be the substitution that maps z_k to 0 for every $k \in K'$ and let τ be the variable renaming $\{z_{k_0} \mapsto u\}$. Then $\theta = \sigma'\rho' = \sigma'\rho''\tau\tau^{-1}\rho'$. As $\sigma'\rho''\tau$ is a u-preserving ACU-unifier of s_1 and s_2, we have $\sigma'\rho''\tau = \sigma\tau'$ for some $\sigma \in U$. Thus $\theta = \sigma\tau'\tau^{-1}\rho'$. □

Definition 3.11 *Let x be a variable occurring in some literal or clause. We say that x is shielded in the literal or clause, if it occurs at least once below a free function symbol. Otherwise, x is called unshielded.*

For example, the variables x and z are shielded in $x + y + f(x) \approx g(z)$, whereas y is unshielded.

We assume to be given a selection function that assigns to every clause a (possibly empty) subset of its negative literals.

Definition 3.12 *A variable x that occurs in a literal $[\neg]\, e$ of a clause C is called eligible, if either C has no selected literals and x is unshielded in C, or $\neg\, e$ is a selected literal of C and x is unshielded in $\neg\, e$.*

The importance of unshielded variables stems from the fact that they may correspond to maximal atomic subterms in a ground instance. If a variable x is shielded in a clause (or selected literal), then the clause or literal contains an atomic subterm $t[x]$. As $x\theta \prec (t[x])\theta$, an atomic subterm of $x\theta$ cannot be maximal.

3.2 The Ideas

We will describe a refutationally complete theorem proving method for first-order theories that include the axioms of cancellative abelian monoids. As the precise rules, to be given in section 3.3, turn out to be rather complex, we will start with a somewhat informal step-by-step presentation of the essential ideas.

The Superposition Calculus The superposition calculus of Bachmair and Ganzinger [4] is a refutationally complete theorem proving method for arbitrary first-order clauses with equality. Starting from an initial set of formulae, a superposition-based theorem prover computes inferences, e.g.,

$$\frac{D \vee s \approx s' \qquad C \vee [\neg]\, t[u] \approx t'}{(D \vee C \vee [\neg]\, t[s'] \approx t')\sigma} \qquad \sigma = \mathrm{mgu}(s, u)$$

and adds their conclusions to the set of formulae. If the initial set is inconsistent, then every fair derivation will eventually produce a contradiction (i.e., the empty clause). The inference rules are restricted forms of paramodulation, resolution, and factoring. For instance, it is sufficient to perform only those superposition inferences that involve maximal literals and maximal sides of equalities. Besides, a global redundancy criterion allows to discard certain inferences and formulae. More precisely, a formula is redundant in a set of formulae N and may be deleted, if it is entailed by smaller clauses in N. An inference is redundant in N (and may be omitted in a fair derivation), if its conclusion follows from clauses in N that are smaller than the largest premise.

Our goal is to develop a refined and otherwise similar calculus for cancellative abelian monoids that makes superpositions with the ACUK axioms superfluous.

Cancellative Superposition Let us first restrict to the case that $+$ is the only non-constant function symbol. In a cancellative abelian monoid, the congruence law and the cancellation law are in a certain sense complementary. The congruence law states that adding equal terms on both sides of an equation preserves truth, and conversely, that dropping equal terms on both sides of an equation preserves falsity. The cancellation law states that dropping equal terms on both sides of an equation preserves truth, and that adding equal terms on both sides of an equation preserves falsity. Hence, if we have an equation $u + s \approx s'$ where the atomic term u is larger than s and s', then we can infer $s' + u + t \approx u + s + t'$ from $u + t \approx t'$ by congruence, and $s' + t \approx s + t'$ by cancellation. Similarly, we can infer $s' + u + t \not\approx u + s + t'$ from $u + t \not\approx t'$ by cancellation, and $s' + t \not\approx s + t'$ by congruence. Intuitively, this means that rather than replacing the left hand side of a rewrite rule by the right hand side, we replace the maximal atomic part by the remainder: We rewrite u to s' while adding s to the other side of the (possibly negated) equation. This is the essential reason why extended rules are unnecessary in our calculus.

The method can be generalized to equational clauses. Taking into account that u might occur more than once in a sum we get the (ground) inference rule

$$\underline{\text{Cancellative Superposition}} \qquad \frac{D \vee mu + s \approx s' \qquad C \vee [\neg]\, nu + t \approx t'}{D \vee C \vee [\neg]\, (n{-}m)u + t + s' \approx t' + s}$$

where $n \geq m \geq 1$.[5]

Together with the cancellation, equality resolution, and cancellative equality factoring rules, this rule is refutationally complete for sets of ground clauses, provided that $+$ is the only non-constant function symbol.

$$\underline{\text{Cancellation}} \qquad \frac{D \vee [\neg]\, nu + t \approx mu + s}{D \vee [\neg]\, (n{-}m)u + t \approx s}$$

$$\underline{\text{Equality Resolution}}^{6} \qquad \frac{C \vee \neg\, 0 \approx 0}{C}$$

$$\underline{\text{Cancellative Eq. Factoring}} \qquad \frac{D \vee nu + s \approx s' \vee nu + t \approx t'}{D \vee \neg\, s + t' \approx s' + t \vee nu + t \approx t'}$$

The inference system remains refutationally complete if we add ordering restrictions, such that inferences are computed only if the literals involved are maximal (or selected) in their clauses and u is atomic and strictly larger than s, s', t, and t'.

Example 3.13 *Suppose that the ordering on constant symbols is given by $b \succ b' \succ c \succ d \succ d'$. We will show that the following four clauses are contradictory with respect to* ACUK. *(The maximal part(s) of every clause are underlined.)*

$$\underline{2b} + c \approx d \tag{1}$$

$$\underline{b'} + c \approx d' \tag{2}$$

$$\underline{d} \approx d' \tag{3}$$

$$\underline{2b} \not\approx b' \tag{4}$$

Cancellative superposition of (1) and (4) yields

$$d \not\approx \underline{b'} + c \tag{5}$$

Cancellative superposition of (2) and (5) yields

$$d + \underline{c} \not\approx d' + \underline{c} \tag{6}$$

[5] Recall that we are working with terms modulo ACU. In particular, this implies that s and t may be missing (i.e., zero).

[6] As the cancellation rule transforms $C \vee \neg\, s \approx s$ into $C \vee \neg\, 0 \approx 0$, it suffices to handle only the latter by equality resolution.

By cancellation of (6) we obtain

$$\underline{d} \not\approx d' \tag{7}$$

Cancellative superposition of (3) and (7) produces

$$\underline{d'} \not\approx \underline{d'} \tag{8}$$

which by cancellation and equality resolution yields the empty clause.

Speaking in terms of AG-normalized completion (Marché [20]), we can work directly with the symmetrisation; Marché's Ψ_{AG} and Θ_{AG} have no counterpart in our framework. Consequently, the number of overlaps that have to be considered is reduced. On the other hand, we lack an inverse, which will lead to certain problems once free function symbols are introduced.

The Non-Ground Case Lifting the calculus to non-ground clauses turns out to be a non-trivial task. In the standard superposition calculus, for lifting one simply needs to replace equality in the ground inference by unifiability (or by equality constraints). The situation is similar here, as long as all variables in our clauses are shielded. If a variable is unshielded, however, we have to take into account that it might be instantiated with a sum and that only the maximal atomic part of the sum takes part in the unification. Consider the clauses $D = D'(x) \lor 3x + d \approx e$ and $C = C'(y) \lor 5y + 2b \approx c$, where b is the maximal constant. A substitution may map x to $\mu b + s$ and y to $\nu b + t$, for some $\nu, \mu \in \mathbf{N}$ and arbitrary ground terms s and t. If $3\mu \leq 5\nu + 2$, then the resulting ground clauses allow the cancellative superposition inference

$$\frac{D'(\mu b + s) \lor 3\mu b + 3s + d \approx e \qquad C'(\nu b + t) \lor (5\nu{+}2)b + 5t \approx c}{D'(\mu b + s) \lor C'(\nu b + t) \lor (5\nu{+}2{-}3\mu)b + 5t + e \approx c + 3s + d}$$

How can we represent this infinite number of ground clauses finitely on the non-ground level without introducing second-order variables?

For the left premise, it's easy. We map the variable x to a sum of two fresh variables, $\hat{x} + \tilde{x}$. The variable $\hat{x}$ is meant to subsume the part of $\mu b + s$ that is consumed during the ground inference (namely μb); the second variable is meant to subsume the part that is left over (namely s).

For the right premise, the situation is a bit more complicated, since not all b's need to be consumed. Each of the ν b's produces 5 copies. If $\mu = 2$, then there are 6 b's to be consumed, thus $\nu \geq 1$. We may assume that these 6 b's consist of 5 copies of one of the ν b's and one of the 2 b's that were present in the non-ground clause. As for the left premise, it suffices to map y to a sum of two fresh variables, $\hat{y} + \tilde{y}$.

If $\mu = 3$, then there are 9 b's to be consumed, thus $\nu \geq 2$. Again we may assume that 5 of the 9 b's are copies of one of the ν b's. In this case, however, the remaining 4 b's have to be taken from the 5 copies of the second of the ν b's. Therefore, we have to map y to the sum $\hat{y} + b + \tilde{y}$. The variable $\hat{y}$ is meant to subsume those b's, whose 5 copies are completely consumed. The variable $\tilde{y}$ is meant to subsume the part of $\nu b + t$ that is left over. Finally, there is one b, for which λ of its 5 copies are consumed, where $3 \leq \lambda \leq 4$.

We obtain two kinds of non-ground inferences: First, for the case that each of the ν b's is either completely consumed or left over:

$$\frac{D'(x) \lor 3x + d \approx e \qquad C'(y) \lor 5y + 2b \approx c}{(D'(\hat{x} + \tilde{x}) \lor C'(\hat{y} + \tilde{y}) \lor (2{-}\lambda)b + 5\tilde{y} + e \approx c + d + 3\tilde{x})\sigma}$$

where $0 \leq \lambda \leq 2$ and σ is a most general ACU-unifier of $3\hat{x}$ and $5\hat{y} + \lambda b$. Second, for the case that one of the ν b's is only partially consumed:

$$\frac{D'(x) \lor 3x + d \approx e \qquad C'(y) \lor 5y + 2b \approx c}{(D'(\hat{x} + \tilde{x}) \lor C'(\hat{y} + b + \tilde{y}) \lor (7{-}\lambda)b + 5\tilde{y} + e \approx c + d + 3\tilde{x})\sigma}$$

where $3 \leq \lambda \leq 4$ and σ is a most general ACU-unifier of $3\hat{x}$ and $5\hat{y} + \lambda b$.

In the general case, the maximal equations of the two clauses may have the form

$$\sum_{j \in J} n_j y_j + \sum_{l \in L} n'_l v_l + t \approx t'$$

with variables y_j and unifiable non-variable terms v_l. Using the idea above we can still obtain an inference rule that produces only finitely many conclusions. Let v be a most general common instance of all v_l (modulo ACU). In the left premise, we map every variable x_i to $\hat{x}_i + \tilde{x}_i$. In the right premise, there are two possibilities: If every v that is substituted into the y_j's is either completely consumed or left over, then it suffices to map every y_j to $\hat{y}_j + \tilde{y}_j$. Otherwise, we have to map some y_{j_0} to the sum $\hat{y}_{j_0} + v + \tilde{y}_{j_0}$, where λ of the n_{j_0} copies of v are consumed. (It is in fact sufficient to pick *one* $j_0 \in J$ and to restrict to $\sum_{l \in L} n'_l < \lambda < n_{j_0}$. Furthermore, we will show that we may set $\hat{y}_j = 0$ if either $n_j \leq \sum_{l \in L} n'_l + n_{j_0} - \lambda$ or $n_j > n_{j_0}$.) A similar technique is also used for cancellation inferences and cancellative equality factoring inferences.

It is obvious that inferences involving unshielded variables may be very prolific. In practice, they should be avoided whenever possible. In Section 5 we will discuss suitable techniques to do this.

Free Function Symbols So far, we have considered signatures where $+$ is the only non-constant function symbol. If we add free function symbols, and possibly other sorts, then we also have to use the inference rules of the traditional superposition calculus, that is equality resolution, standard superposition, and standard equality factoring. But this is not sufficient, as shown by the following example.

Example 3.14 *Suppose that the ordering on constant symbols is given by $b \succ b' \succ c \succ d \succ d'$. In every model of the three clauses*

$$\underline{2b} + c \approx d \tag{1}$$

$$\underline{b'} + c \approx d' \tag{2}$$

$$\underline{d} \approx d' \tag{3}$$

the terms $2b$ and b' are equal. As we have shown in Example 3.13 we can thus refute the set of clauses (1)–(4).

$$\underline{2b} \not\approx b' \tag{4}$$

However, the cancellative superposition rule is limited to superpositions at the top of a term. There is no way to perform a cancellative superposition inference below a free function symbol, hence there is no way to derive the empty clause from the clauses (1), (2), (3), and (9).

$$\underline{f(2b)} \not\approx f(b') \tag{9}$$

If we were working in groups, we could simply derive $f(d - c) \not\approx f(b')$. But this is impossible in our context.

Hsiang, Rusinowitch, and Sakai [17, 26] have solved this problem by introducing the following inference rule:

$$\frac{D \vee u + s \approx s' \qquad C \vee v + s \approx s'}{D \vee C \vee u \approx v}$$

In the example above, this rule allows to derive $\underline{2b} \approx b'$ from the first three clauses, which can then be applied to (9) by standard superposition. However, there is a drawback of this approach. Before we can apply the rule of Hsiang, Rusinowitch, and Sakai, we have to use clause (3) to replace d by d' in (1). Since the term d is not maximal in (1), the rule can be only used in conjunction with ordered paramodulation (where inferences may involve *smaller* parts of maximal literals), but does not work together with strict superposition (where such inferences are excluded).

The concept of abstraction yields another solution for the problem, which fits more smoothly into the superposition calculus. Abstracting out an occurrence of a term v in a clause $C[v]$ means replacing v by a new variable x and adding $x \not\approx v$ as a new condition to the clause. In our case, we have to abstract out a term v of sort S_{CAM} occurring immediately below a free function symbol, if there is some other clause $D \vee mu + s \approx s'$ such that (i) mu occurs at the top of v, but (ii) a standard superposition of $mu + s$ into v is impossible. We emphasize that the new variable x is shielded in the resulting clause.

Abstraction
$$\frac{D \vee mu + s \approx s' \qquad C \vee [\neg]\, w[nu + t] \approx w'}{C \vee \neg\, x \approx nu + t \vee [\neg]\, w[x] \approx w'}$$

where $n \geq m$.

Using this inference rule, the set of clauses (1), (2), (3), and (9) of Example 3.14 can be refuted as follows:

Example 3.15 *Abstraction of (1) and (9) yields*

$$x \not\approx 2b \vee \underline{f(x)} \not\approx \underline{f(b')} \tag{10}$$

By (non-ground) cancellation of (10) with the unifier $\{x \mapsto b'\}$ *we obtain*

$$b' \not\approx \underline{2b} \vee 0 \not\approx 0 \tag{11}$$

which can be refuted in the same way as (4) in Example 3.13.

The abstraction rule is extended to non-ground premises in the same way as the cancellative superposition rule.

3.3 Inference System

General Remarks

In an expression like $\sum_{i \in I} m_i x_i + \sum_{k \in K} m'_k u_k + s$, every x_i is a variable of sort S_{CAM}, every u_k is an atomic term of sort S_{CAM}, and s is an arbitrary term of sort S_{CAM}. The coefficients m_i and m'_k are elements of $\mathbf{N}^+$. Both I and K are linearly ordered sets of indices; I and K may be empty, s may be 0, unless explicitly said otherwise.

To simplify the presentation, we give the inference rules for unconstrained clauses. However, it should be mentioned that in fact all ordering conditions may be turned into constraints which are inherited by the conclusions of the inferences. Similarly, it may be advisable in an implementation to work with unifiability constraints as much as possible rather than computing unifiers eagerly.

We use the phrase "most general (t-preserving) ACU-unifier of u and v" to denote some member of a fixed complete set of (t-preserving) ACU-unifiers of u and v.

Cancellation

$$\frac{D \vee [\neg]\, e_1}{(D \vee [\neg]\, e_0)\sigma}$$

if the following conditions are satisfied:

- $e_1 \;=\; \sum_{j\in J} n_j x_j + \sum_{l\in L} n'_l v_l + t \approx \sum_{i\in I} m_i x_i + \sum_{k\in K} m'_k u_k + s.$
- $e_0 \;=\; \mu z + \sum_{j\in \tilde{J}} \tilde{n}_j \tilde{x}_j + t \approx \sum_{i\in \tilde{I}} \tilde{m}_i \tilde{x}_i + s.$
- $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
- $\{\, x_i \mid i \in I \,\}$ is the set of all eligible variables of $[\neg]\, e_1$ that occur in the right hand side of e_1; $\{\, x_j \mid j \in J \,\}$ is the set of all eligible variables of $[\neg]\, e_1$ that occur in the left hand side of e_1.
- $\tilde{I} = I \setminus \{\, i \in I \cap J \mid n_i \geq m_i \,\}$, $\tilde{J} = J \setminus \{\, i \in I \cap J \mid m_i \geq n_i \,\}$, $\tilde{m}_i = m_i$ if $i \in I \setminus J$, $\tilde{m}_i = m_i - n_i$ if $i \in I \cap J$ and $m_i > n_i$, $\tilde{n}_i = n_i$ if $i \in J \setminus I$, $\tilde{n}_i = n_i - m_i$ if $i \in I \cap J$ and $n_i > m_i$.
- $m' = \sum_{k\in K} m'_k$ and $n' = \sum_{l\in L} n'_l$.
- Either $J_- = \tilde{J}$, $J_0 = J_= = J_+ = \emptyset$, $0 \leq \lambda \leq n'$, and $\mu = n' - \lambda$; or $J_0 = \{j_0\} \subseteq \tilde{J}$, $n' < \lambda < \tilde{n}_{j_0}$, $J_= = \{\, j \in \tilde{J} \mid \tilde{n}_j \leq n' + \tilde{n}_{j_0} - \lambda \,\}$, $J_- = \{\, j \in \tilde{J} \mid n' + \tilde{n}_{j_0} - \lambda < \tilde{n}_j < \tilde{n}_{j_0} \vee (\tilde{n}_j = \tilde{n}_{j_0} \wedge j < j_0) \,\}$, $J_+ = \{\, j \in \tilde{J} \mid \tilde{n}_j > \tilde{n}_{j_0} \vee (\tilde{n}_j = \tilde{n}_{j_0} \wedge j > j_0) \,\}$, and $\mu = n' + \tilde{n}_{j_0} - \lambda$.
- σ_1 maps x_i to $\hat{x}_i + \tilde{x}_i$ for $i \in \tilde{I} \cup J_-$, to $\hat{x}_i + z + \tilde{x}_i$ for $i \in J_0$, and to $\tilde{x}_i$ for $i \in J_= \cup J_+$; σ_2 is a most general ACU-unifier of all u_k, v_l, and z ($k \in K$, $l \in L$); σ_3 is a most general $z\sigma_2$-preserving ACU-unifier of $\sum_{i\in\tilde{I}} \tilde{m}_i \hat{x}_i + m' z\sigma_2$ and $\sum_{j\in J_-\cup J_0} \tilde{n}_j \hat{x}_j + \lambda z \sigma_2$; and $\sigma = \sigma_1\sigma_2\sigma_3$.
- Either $\neg\, e_1$ is a selected literal, or the premise has no selected literals and $[\neg]\, e_1\sigma$ is a maximal literal in $(D \vee [\neg]\, e_1)\sigma$.
- $z\sigma \not\preceq s\sigma$, $z\sigma \not\preceq t\sigma$, $z\sigma \not\preceq \tilde{x}_i\sigma$ for $i \in \tilde{I} \cup J_+$, $z\sigma \not\prec_{\mathrm{x}} \tilde{x}_i\sigma$ for $i \in J_= \cup J_- \cup J_0$.

Equality Resolution

$$\frac{C \vee \neg\, u \approx v}{C\sigma}$$

if the following conditions are satisfied:

- Either $u\sigma = v\sigma = 0$ or u and v don't have sort S_{CAM} and σ is a most general ACU-unifier of u and v.

- Either $\neg\, u \approx v$ is a selected literal, or $\neg\,(u \approx v)\sigma$ is a maximal literal in $(C \vee \neg\, u \approx v)\sigma$.

Standard Superposition

$$\frac{D \vee u \approx v \qquad C \vee [\neg]\, t[w] \approx t'}{(D \vee C \vee [\neg]\, t[v] \approx t')\sigma}$$

if the following conditions are satisfied:

- w is not a variable.

- If t has sort S_{CAM}, then w occurs below a free function symbol in t. If t is a proper sum, then w occurs in a maximal atomic subterm of t.

- σ is a most general ACU-unifier of u and w.

- $u\sigma \not\preceq v\sigma$ and $t\sigma \not\preceq t'\sigma$.

- The first premise has no selected literals.

- $u\sigma \approx v\sigma$ is a strictly maximal literal in $(D \vee u \approx v)\sigma$.

- Either $\neg\, t[w] \approx t'$ is a selected literal, or the second premise has no selected literals and $([\neg]\, t[w] \approx t')\sigma$ is a maximal literal in $(C \vee [\neg]\, t[w] \approx t')\sigma$ (strictly maximal, if it is a positive literal).

- $(D \vee u \approx v)\sigma \not\succeq_{\text{C}} (C \vee [\neg]\, t[w] \approx t')\sigma$.

Cancellative Superposition

$$\frac{D \vee e_1 \qquad C \vee [\neg]\, e_2}{(D \vee C \vee [\neg]\, e_0)\sigma}$$

if the following conditions are satisfied:

- $e_1 \;=\; \sum_{i \in I} m_i x_i + \sum_{k \in K} m'_k u_k + s \approx s'$.
- $e_2 \;=\; \sum_{j \in J} n_j y_j + \sum_{l \in L} n'_l v_l + t \approx t'$.
- $e_0 \;=\; \mu z + \sum_{j \in J} n_j \tilde{y}_j + t + s' \approx \sum_{i \in I} m_i \tilde{x}_i + s + t'$.
- $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
- $\{\, x_i \mid i \in I \,\}$ is the set of all eligible variables of e_1 that occur in the left hand side but not in the right hand side of e_1; $\{\, y_j \mid j \in J \,\}$ is the set of all eligible variables of $[\neg]\, e_2$ that occur in the left hand side but not in the right hand side of e_2.
- The left hand side of e_2 is not a variable (i.e., either $\sum_{j \in J} n_j > 1$ or $\sum_{l \in L} n'_l v_l + t \neq 0$).
- $m' = \sum_{k \in K} m'_k$ and $n' = \sum_{l \in L} n'_l$.
- Either $J_- = J$, $J_0 = J_= = J_+ = \emptyset$, $0 \leq \lambda \leq n'$, and $\mu = n' - \lambda$; or $J_0 = \{j_0\} \subseteq J$, $n' < \lambda < n_{j_0}$, $J_= = \{\, j \in J \mid n_j \leq n' + n_{j_0} - \lambda \,\}$, $J_- = \{\, j \in J \mid n' + n_{j_0} - \lambda < n_j < n_{j_0} \vee (n_j = n_{j_0} \wedge j < j_0) \,\}$, $J_+ = \{\, j \in J \mid n_j > n_{j_0} \vee (n_j = n_{j_0} \wedge j > j_0) \,\}$, and $\mu = n' + n_{j_0} - \lambda$.
- σ_1 maps x_i to $\hat{x}_i + \tilde{x}_i$ for $i \in I$ and y_j to $\hat{y}_j + \tilde{y}_j$ for $j \in J_-$, to $\hat{y}_j + z + \tilde{y}_j$ for $j \in J_0$, and to $\tilde{y}_j$ for $i \in J_= \cup J_+$; σ_2 is a most general ACU-unifier of all u_k, v_l, and z ($k \in K$, $l \in L$); σ_3 is a most general $z\sigma_2$-preserving ACU-unifier of $\sum_{i \in I} m_i \hat{x}_i + m' z\sigma_2$ and $\sum_{j \in J_- \cup J_0} n_j \hat{y}_j + \lambda z \sigma_2$; and $\sigma = \sigma_1\sigma_2\sigma_3$.
- The first premise has no selected literals.
- $e_1\sigma$ is a strictly maximal literal in $(D \vee e_1)\sigma$.
- Either $\neg\, e_2$ is a selected literal, or the second premise has no selected literals and $[\neg]\, e_2\sigma$ is a maximal literal in $(C \vee [\neg]\, e_2)\sigma$ (strictly maximal, if e_2 occurs positively).

- $(D \vee e_1)\sigma \not\succeq_{\mathrm{C}} (C \vee [\neg]\, e_2)\sigma$.
- $z\sigma \not\succeq s\sigma$, $z\sigma \not\succeq s'\sigma$, $z\sigma \not\succeq t\sigma$, $z\sigma \not\succeq t'\sigma$, $z\sigma \not\succeq \tilde{x}_i\sigma$ for $i \in I$, $z\sigma \not\succeq \tilde{y}_j\sigma$ for $j \in J_+$, $z\sigma \not\succ_{\mathrm{X}} \tilde{y}_j\sigma$ for $j \in J_= \cup J_- \cup J_0$.

Abstraction

$$\frac{D \vee e_1 \qquad C \vee [\neg]\, w[\hat{w}] \approx w'}{C \vee \neg\, x \approx \hat{w} \vee [\neg]\, w[x] \approx w'}$$

if the following conditions are satisfied:

- $e_1 \;=\; \sum_{i \in I} m_i x_i + \sum_{k \in K} m'_k u_k + s \approx s'$.
- $\hat{w} \;=\; \sum_{j \in J} n_j y_j + \sum_{l \in L} n'_l v_l + t$.
- $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
- $\{\, x_i \mid i \in I \,\}$ is the set of all eligible variables of e_1 that occur in the left hand side but not in the right hand side of e_1.
- None of the variables y_j occurs in the non-variable terms v_l. (The y_j may occur in t, however.)
- $\hat{w}$ occurs in w immediately below some free function symbol. If w is a proper sum, then $\hat{w}$ occurs in a maximal atomic subterm of w.
- $m' = \sum_{k \in K} m'_k$ and $n' = \sum_{l \in L} n'_l$.
- Either $J_- = J$, $J_0 = J_= = J_+ = \emptyset$, and $0 \leq \lambda \leq n'$; or $J_0 = \{j_0\} \subseteq J$, $n' < \lambda < n_{j_0}$, $J_= = \{\, j \in J \mid n_j \leq n' + n_{j_0} - \lambda \,\}$, $J_- = \{\, j \in J \mid n' + n_{j_0} - \lambda < n_j < n_{j_0} \vee (n_j = n_{j_0} \wedge j < j_0) \,\}$, and $J_+ = \{\, j \in J \mid n_j > n_{j_0} \vee (n_j = n_{j_0} \wedge j > j_0) \,\}$.
- σ_1 maps x_i to $\hat{x}_i + \tilde{x}_i$ for $i \in I$ and y_j to $\hat{y}_j + \tilde{y}_j$ for $j \in J_-$, to $\hat{y}_j + z + \tilde{y}_j$ for $j \in J_0$, and to $\tilde{y}_j$ for $i \in J_= \cup J_+$; σ_2 is a most general ACU-unifier of all u_k, v_l, and z ($k \in K$, $l \in L$); σ_3 is a most general $z\sigma_2$-preserving ACU-unifier of $\sum_{i \in I} m_i \hat{x}_i + m' z\sigma_2$ and $\sum_{j \in J_- \cup J_0} n_j \hat{y}_j + \lambda z\sigma_2$; and $\sigma = \sigma_1\sigma_2\sigma_3$.
- The first premise has no selected literals.
- $e_1\sigma$ is a strictly maximal literal in $(D \vee e_1)\sigma$.

- Either $\neg\, w[\hat{w}] \approx w'$ is a selected literal, or the second premise has no selected literals and $[\neg]\,(w[\hat{w}] \approx w')\sigma$ is a maximal literal in $(C \vee [\neg]\, w[\hat{w}] \approx w')\sigma$ (strictly maximal, if it is a positive literal).
- If $J = \{j_1\}$, $n_{j_1} = 1$, and $L = \emptyset$, then $t = t_1 + t_2$, where t_1 is non-zero and either t_1 is a variable or $y_{j_1}\sigma \not\prec t_1\sigma$.
- If $I = \emptyset$, then there exists no t' such that $t\sigma = s\sigma + t'$.
- $(D \vee e_1)\sigma \not\succeq_{\mathrm{C}} (C \vee [\neg]\, w[\hat{w}] \approx w')\sigma$.
- $z\sigma \not\preceq s\sigma$, $z\sigma \not\preceq s'\sigma$, $(w[\hat{w}])\sigma \not\preceq w'\sigma$.

Standard Equality Factoring

$$\frac{D \vee u \approx v \vee u' \approx v'}{(D \vee \neg\, v \approx v' \vee u' \approx v')\sigma}$$

if the following conditions are satisfied:

- u, u', v, and v' don't have sort S_{CAM}.
- The premise has no selected literals.
- σ is a most general ACU-unifier of u and u'.
- $u\sigma \not\preceq v\sigma$ and $u\sigma \not\preceq v'\sigma$.
- $u\sigma \approx v\sigma$ is a maximal literal in $(D \vee u \approx v \vee u' \approx v')\sigma$.
- $u\sigma$ is maximal in $(D \vee u \approx v \vee u' \approx v')\sigma$.

Cancellative Equality Factoring

$$\frac{D \vee e_1 \vee e_2}{(D \vee \neg\, e_0 \vee e_2)\sigma}$$

if the following conditions are satisfied:

- $e_1 \;=\; \sum_{i\in I} m_i x_i + \sum_{k\in K} m'_k u_k + s \approx s'$.
- $e_2 \;=\; \sum_{j\in J} n_j x_j + \sum_{l\in L} n'_l v_l + t \approx t'$.

- $e_0 = \sum_{i \in I} m_i \tilde{x}_i + s + t' \approx \sum_{j \in J} n_j \tilde{x}_j + s' + t.$

- $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.

- $\{ x_i \mid i \in I \}$ is the set of all eligible variables of e_1 that occur in the left hand side but not in the right hand side of e_1; $\{ x_j \mid j \in J \}$ is the set of all eligible variables of e_2 that occur in the left hand side but not in the right hand side of e_2.

- $m' = \sum_{k \in K} m'_k$ and $n' = \sum_{l \in L} n'_l$.

- σ_1 maps x_i to $\hat{x}_i + \tilde{x}_i$ for $i \in I \cup J$; σ_2 is a most general ACU-unifier of all u_k, v_l, and z ($k \in K$, $l \in L$); σ_3 is a most general $z\sigma_2$-preserving ACU-unifier of $\sum_{i \in I} m_i \hat{x}_i + m' z \sigma_2$ and $\sum_{j \in J} n_j \hat{x}_j + n' z \sigma_2$.

- The premise has no selected literals.

- $e_1 \sigma$ is a maximal literal in $(D \vee e_1 \vee e_2)\sigma$.

- $z\sigma \not\preceq s\sigma$, $z\sigma \not\preceq s'\sigma$, $z\sigma \not\preceq t\sigma$, $z\sigma \not\preceq t'\sigma$, $z\sigma \not\preceq \tilde{x}_i\sigma$ for $i \in I \cup J$.

Theorem 3.16 *For every inference rule*

$$\frac{C_1 \ldots C_n}{C_0}$$

of the cancellative superposition calculus, we have $\{C_1, \ldots, C_n\} \models C_0$.

Proof. By routine computation. □

3.4 Redundancy and Saturation

To make a saturation-based theorem proving technique practically useful, the inference system has to be complemented with a redundancy criterion. Both clauses and inferences may be redundant. A redundant clause can be deleted from the current set of clauses at any point of the saturation process. A redundant inference may be ignored during the saturation process without endangering the fairness of the derivation. (See [4, 6] for a more detailed discussion of these aspects of superposition. Note that "redundancy" is called "compositeness" in [4].)

Definition 3.17 *Let N be a set of clauses. A ground clause D is called* ACUK*-redundant with respect to N, if ground instances $D_1, \dots, D_n$ of clauses in N exist such that $\{D_1, \dots, D_n\} \models D$ and $D_i \prec_{\mathrm{C}} D$ for every $i \in \{1, \dots, n\}$. A non-ground clause is called* ACUK*-redundant with respect to N, if all its ground instances are.*

Traditionally, an inference is considered to be redundant, if its conclusion follows from clauses that are smaller than the maximal premise. Since the conclusion of an abstraction inference is non-ground even if the premises are ground, we have to modify this scheme slightly.

Definition 3.18 *Let N be a set of clauses. A ground inference with conclusion C_0 and maximal premise C is called* ACUK*-redundant with respect to N in the following two cases:*

- *If it is not an abstraction inference and if there are ground instances $D_1, \dots, D_n$ of clauses in N such that $\{D_1, \dots, D_n\} \models C_0$ and $D_i \prec_{\mathrm{C}} C$ for every $i \in \{1, \dots, n\}$.*
- *If it is an abstraction inference*

$$\frac{D' \vee mu + s \approx s' \qquad C' \vee [\neg]\, w[nu + t] \approx w'}{C' \vee \neg\, x \approx nu + t \vee [\neg]\, w[x] \approx w'}$$

  and if for every substitution ρ that maps x to a ground term $r \prec nu + t$ there exist ground instances $D_1, \dots, D_n$ of clauses in N such that $\{D_1, \dots, D_n\} \models C_0\rho$ and $D_i \prec_{\mathrm{C}} C$ for every $i \in \{1, \dots, n\}$.

To extend this definition to non-ground inferences, we need the concept of an instance of an inference.

Definition 3.19 *Let $C_1, \dots, C_n$ be clauses and let θ be a substitution such that all $C_i\theta$ are ground. If there are inferences (modulo* ACU*)*

$$\frac{C_1\theta \dots C_n\theta}{D\theta}$$

and

$$\frac{C_1 \dots C_n}{D}$$

then the former inference is called a ground instance of the latter.

It should be noted that not every inference from $C_1\theta \ldots C_n\theta$ is a ground instance of an inference from $C_1 \ldots C_n$. As in the standard superposition calculus [4], a ground superposition may take place below a variable position of C_n. For instance, if $C_1 = a \approx b$ and $C_2 = f(x,x) \approx x$, then there is a standard superposition inference from the ground instances $C_1\theta = a \approx b$ and $C_2\theta = f(a,a) \approx a$, but not from C_1 and C_2 themselves.

Cancellation inferences pose an additional difficulty. If a variable x occurs on both sides of an equation in a clause C, then in a non-ground cancellation inference it is removed completely, whereas a ground inference from $C\theta$ may remove $x\theta$ only partially. Consider the following example. Let b, c, d be constants such that $b \succ c \succ d$, and let C be a clause $x + c \approx x + d$. Then there is a cancellation inference from C, namely

$$\frac{x + c \approx x + d}{c \approx d}$$

In a ground instance of C, x might be instantiated with a sum $b + d$. But

$$\frac{b + c + d \approx b + 2d}{c \approx d}$$

does not qualify as a cancellation inference. Conversely,

$$\frac{b + c + d \approx b + 2d}{c + d \approx 2d}$$

qualifies as a cancellation inference, but it is not an instance of the non-ground inference above. The notion of a weak instance will encompass such cases.

Definition 3.20 *Let $C = C' \vee [\neg]\, e$ be a clause and let θ be a substitution such that $C\theta$ is ground. If there are cancellation inferences (modulo* ACU*)*

$$\frac{C'\theta \vee [\neg]\, e\theta}{C'\theta \vee [\neg]\, s + t \approx s + t'}$$

and

$$\frac{C' \vee [\neg]\, e}{D}$$

and $D\theta = C'\theta \vee [\neg]\, t \approx t'$, then the former inference is called a weak ground instance of the latter.

Definition 3.21 *A non-ground inference is called* ACUK*-redundant with respect to N in the following two cases:*

- *If it is not a cancellation inference and if all its ground instances are* ACUK-*redundant.*

- *If it is a cancellation inference and if all its weak ground instances are* ACUK-*redundant.*

One way to make an inference from clauses in N redundant is to add its conclusion to N.

Definition 3.22 *We say that a set N of clauses is saturated up to* ACUK-*redundancy, if every inference from non-*ACUK-*redundant clauses in N is* ACUK-*redundant.*

Under which conditions is an inference from ground clauses $C_i\theta$ a (weak) ground instance of an inference from C_i? This question will be answered by the so-called "lifting lemmas". To prove them, we need the following technical lemma.

Lemma 3.23 *Let J be a finite and linearly ordered set of indices. Suppose that $m, n \in \mathbf{N}$, and $n_j \in \mathbf{N}^+$, $\nu_j \in \mathbf{N}$ for every $j \in J$. Furthermore, assume that $m \leq \sum_{j \in J} n_j \nu_j + n$. Then one of the following two properties holds:*

(i) There exist $\lambda \in \mathbf{N}$ and $\kappa_j \in \mathbf{N}$, such that $\lambda \leq n$, $\kappa_j \leq \nu_j$ for $j \in J$, and $m = \sum_{j \in J} n_j \kappa_j + \lambda$.

(ii) There exist $\lambda \in \mathbf{N}$, $\kappa_j \in \mathbf{N}$, and $j_0 \in J$ such that $n < \lambda < n_{j_0}$, $\kappa_j \leq \nu_j$ for $j \in J$, $\kappa_{j_0} < \nu_{j_0}$, $\nu_j = 0$ whenever $n_j > n_{j_0} \vee (n_j = n_{j_0} \wedge j > j_0)$, $\kappa_j = 0$ whenever $n_j \leq n + n_{j_0} - \lambda$, and $m = \sum_{j \in J} n_j \kappa_j + \lambda$.

Proof. Let $m_0 = \sum_{j \in J} n_j \nu_j$. Assume that property (i) does not hold. Then obviously $n < m < m_0$, and hence $\nu_j > 0$ for some $j \in J$. Let $n_{\max}$ be the maximum of $\{\, n_j \mid j \in J,\, \nu_j > 0 \,\}$, and let j_0 be the maximum of $\{\, j \in J \mid n_j = n_{\max},\, \nu_j > 0 \,\}$.

Let M denote the set of all $\mu \in \mathbf{N}$ such that there are $\kappa_j \in \mathbf{N}$ with $\kappa_j \leq \nu_j$ for $j \in J$, $\kappa_{j_0} < \nu_{j_0}$, and $\mu = \sum_{j \in J} n_j \kappa_j$. It is easy to see that $\{0, m_0 - n_{j_0}\} \subseteq M \subseteq \{0, 1, \ldots, m_0 - n_{j_0}\}$. Furthermore, if μ_1 and μ_2 are elements of M such that $\mu_1 < \mu_2$ and $\{\, \mu \in M \mid \mu_1 < \mu < \mu_2 \,\}$ is empty, then $\mu_2 - \mu_1 \leq n_{j_0}$. Therefore, there exist $\mu \in M$ such that $\mu \leq m < \mu + n_{j_0}$. Let μ_0 be the smallest $\mu \in M$ with this property, and let $\lambda = m - \mu_0$, then $m = \sum_{j \in J} n_j \kappa_j + \lambda$ as required. Obviously, $\lambda = m - \mu_0 < n_{j_0}$. Besides, $\lambda \leq n$ would imply that property (i) holds, so $n < \lambda < n_{j_0}$.

It remains to show that $\kappa_j = 0$ whenever $n_j \leq n + n_{j_0} - \lambda$. Assume there were a $j_1 \in J$ such that $n_{j_1} \leq n + n_{j_0} - \lambda$ and $\kappa_{j_1} > 0$. Then define $\kappa'_j = \kappa_j$ for $j \neq j_1$ and $\kappa'_{j_1} = \kappa_{j_1} - 1$. Consequently, $m = \sum_{j\in J} n_j\kappa_j + \lambda = \sum_{j\in J} n_j\kappa'_j + n_{j_1} + \lambda$. If $n_{j_1} < n_{j_0} - \lambda$, then this contradicts the minimality of μ_0. Otherwise $n_{j_0} - \lambda \leq n_{j_1} \leq n + n_{j_0} - \lambda$. Let $\lambda' = n_{j_1} + \lambda - n_{j_0}$, $\kappa''_j = \kappa'_j$ for $j \neq j_0$, and $\kappa''_{j_0} = \kappa'_{j_0} + 1$, then $m = \sum_{j\in J} n_j\kappa'_j + n_{j_0} + \lambda' = \sum_{j\in J} n_j\kappa''_j + \lambda'$. As $\lambda' \leq n$, this contradicts our assumption that property (i) does not hold. $\square$

Lemma 3.24 *Let C be a clause and let θ be a substitution such that $C\theta$ is ground. Then every cancellation inference from $C\theta$ is a weak ground instance of a cancellation inference from C.*

Proof. Suppose that $C\theta = C'\theta \vee [\neg]\, e_1\theta$ and $e_1\theta = n\bar{u} + \bar{t} \approx m\bar{u} + \bar{s}$, such that $\bar{u}$ is an atomic ground term, $\bar{u} \succ \bar{s}$, $\bar{u} \succ \bar{t}$, and $n \geq m \geq 1$. If either $\neg\, e_1\theta$ is a selected literal, or $C\theta$ has no selected literals and $[\neg]\, e_1\theta$ is maximal in $C\theta$, then this clause allows a cancellation inference

$$\frac{C'\theta \vee [\neg]\, n\bar{u} + \bar{t} \approx m\bar{u} + \bar{s}}{C'\theta \vee [\neg]\, (n{-}m)\bar{u} + \bar{t} \approx \bar{s}}$$

By maximality, $\bar{u}$ may only result from instantiating eligible variables or atomic terms. Let $\{\, x_j \mid j \in J \,\}$ and $\{\, x_i \mid i \in I \,\}$ be the sets of all eligible variables of $[\neg]\, e_1$ that occur in the left or right hand side of e_1, respectively. We may assume that

$$e_1 \;=\; \sum_{j\in J} n_j x_j + \sum_{l\in L} n'_l v_l + t \approx \sum_{i\in I} m_i x_i + \sum_{k\in K} m'_k u_k + s$$

such that

$$\begin{aligned}
x_i\theta &= \nu_i\bar{u} + \bar{s}_i \;\text{ for } i \in I \cup J, \text{ where } \nu_i \in \mathbf{N} \text{ and } \bar{u} \succ \bar{s}_i,\\
v_l\theta &= \bar{u} \;\text{ for } l \in L,\\
u_k\theta &= \bar{u} \;\text{ for } k \in K,\\
n' &= \textstyle\sum_{l\in L} n'_l,\\
m' &= \textstyle\sum_{k\in K} m'_k,\\
n &= \textstyle\sum_{j\in J} n_j\nu_j + n',\\
m &= \textstyle\sum_{i\in I} m_i\nu_i + m',\\
\bar{t} &= \textstyle\sum_{j\in J} n_j\bar{s}_j + t\theta,\\
\bar{s} &= \textstyle\sum_{i\in I} m_i\bar{s}_i + s\theta.
\end{aligned}$$

Let $\tilde{I} = I \setminus \{\, i \in I \cap J \mid n_i \geq m_i \,\}$, $\tilde{J} = J \setminus \{\, i \in I \cap J \mid m_i \geq n_i \,\}$, $\tilde{m}_i = m_i$ if $i \in I \setminus J$, $\tilde{m}_i = m_i - n_i$ if $i \in I \cap J$ and $m_i > n_i$, $\tilde{n}_i = n_i$ if $i \in J \setminus I$,

$\tilde{n}_i = n_i - m_i$ if $i \in I \cap J$ and $n_i > m_i$, $p_i = \min\{m_i, n_i\}$ for $i \in I \cap J$. Note that $\tilde{I}$ and $\tilde{J}$ are disjoint. Let $\tilde{n} = \sum_{j\in\tilde{J}} \tilde{n}_j\nu_j + n'$, $\tilde{m} = \sum_{i\in\tilde{I}} \tilde{m}_i\nu_i + m'$, and $\tilde{p} = \sum_{i\in I\cap J} p_i\nu_i$, then $\tilde{n} + \tilde{p} = n$ and $\tilde{m} + \tilde{p} = m$. As $m \leq n$, we obtain

$$\tilde{m} \leq \tilde{n} = \sum_{j\in\tilde{J}} \tilde{n}_j\nu_j + n' .$$

We have to distinguish two cases. If there exist $\lambda, \kappa_j \in \mathbf{N}$ such that $\lambda \leq n'$, $\kappa_j \leq \nu_j$ for $j \in \tilde{J}$, and $\tilde{m} = \sum_{j\in\tilde{J}} \tilde{n}_j\kappa_j + \lambda$, then let $\mu = n' - \lambda$, $J_- = \tilde{J}$ and $J_0 = J_= = J_+ = \emptyset$.

Otherwise, by Lemma 3.23, there exist $\lambda, \kappa_j \in \mathbf{N}$ and $j_0 \in \tilde{J}$ such that $n' < \lambda < \tilde{n}_{j_0}$, $\kappa_{j_0} < \nu_{j_0}$, $\kappa_j \leq \nu_j$ for $j \in J$, $\nu_j = 0$ for $j \in J_+$, $\kappa_j = 0$ for $j \in J_=$, and $\tilde{m} = \sum_{j\in\tilde{J}} \tilde{n}_j\kappa_j + \lambda$, where $J_= = \{\, j \in \tilde{J} \mid \tilde{n}_j \leq n' + \tilde{n}_{j_0} - \lambda \,\}$ and $J_+ = \{\, j \in \tilde{J} \mid \tilde{n}_j > \tilde{n}_{j_0} \vee (\tilde{n}_j = \tilde{n}_{j_0} \wedge j > j_0) \,\}$. Let $\mu = n' + \tilde{n}_{j_0} - \lambda$, $J_0 = \{j_0\}$, and $J_- = \{\, j \in \tilde{J} \mid n' + \tilde{n}_{j_0} - \lambda < \tilde{n}_j < \tilde{n}_{j_0} \vee (\tilde{n}_j = \tilde{n}_{j_0} \wedge j < j_0) \,\}$.

We define two substitutions σ_1 and ρ_1 as follows: Let σ_1 map x_i to $\hat{x}_i + \tilde{x}_i$ for $i \in \tilde{I} \cup J_-$, to $\hat{x}_i + z + \tilde{x}_i$ for $i \in J_0$, and to $\tilde{x}_i$ for $i \in J_= \cup J_+$. Let ρ_1 map $\hat{x}_i$ to $\nu_i\bar{u}$ for $i \in \tilde{I}$ and to $\kappa_i\bar{u}$ for $i \in J_- \cup J_0$, $\tilde{x}_i$ to $\bar{s}_i$ for $i \in \tilde{I} \cup J_+$, to $(\nu_i - \kappa_i)\bar{u} + \bar{s}_i$ for $i \in J_= \cup J_-$, and to $(\nu_i - \kappa_i - 1)\bar{u} + \bar{s}_i$ for $i \in J_0$, x_i to $\nu_i\bar{u} + \bar{s}_i$ for $i \in (I \cup J) \setminus (\tilde{I} \cup \tilde{J})$, z to $\bar{u}$, and every other variable $y \in \mathrm{Var}(C)$ to $y\theta$. It is easy to verify that θ equals $\sigma_1\rho_1$ over $\mathrm{Var}(C)$.

As ρ_1 equals θ over all variables occurring in u_k and v_l, ρ_1 is a unifier of u_k, v_l, and z ($k \in K$, $l \in L$). Hence there is a most general ACU-unifier σ_2 of all u_k, v_l, and z and a substitution ρ_2 such that $\rho_1 = \sigma_2\rho_2$ over $\mathrm{Var}(C) \cup \mathrm{Ran}(\sigma_1)$. We may assume that $\mathrm{Dom}(\sigma_2) \subseteq \mathrm{Var}(\{\, u_k, v_l, z \mid k \in K,\ l \in L \,\})$.

Now consider the terms $r = \sum_{i\in\tilde{I}} \tilde{m}_i\hat{x}_i + m'z\sigma_2$ and $r' = \sum_{j\in J_-\cup J_0} \tilde{n}_j\hat{x}_j + \lambda z\sigma_2$. As $\hat{x}_i\rho_2 = \hat{x}_i\rho_1$ and $z\sigma_2\rho_2 = z\rho_1 = \bar{u}$, we obtain $r\rho_2 = r'\rho_2 = \tilde{m}\bar{u}$. Besides, ρ_2 maps $z\sigma_2$ to the non-zero atomic ground term $\bar{u}$. By Lemma 3.10, there is a most general $z\sigma_2$-preserving ACU-unifier σ_3 of r and r', such that $\rho_2 = \sigma_3\rho_3$ over $\mathrm{Var}(C) \cup \mathrm{Ran}(\sigma_1) \cup \mathrm{Ran}(\sigma_2)$. Let $\sigma = \sigma_1\sigma_2\sigma_3$.

We define

$$e_0 \;=\; \mu z + \sum_{j\in\tilde{J}} \tilde{n}_j\tilde{x}_j + t \approx \sum_{i\in\tilde{I}} \tilde{m}_i\tilde{x}_i + s$$

then

$$\frac{C' \vee [\neg]\, e_1}{(C' \vee [\neg]\, e_0)\sigma}$$

is a cancellation inference from C that we denote by ι. It is easy to see that the ordering conditions of the inference rule are satisfied.

It remains to show that the cancellation inference ι' given by

$$\frac{C'\theta \vee [\neg]\, n\bar{u} + \bar{t} \approx m\bar{u} + \bar{s}}{C'\theta \vee [\neg]\, (n{-}m)\bar{u} + \bar{t} \approx \bar{s}}$$

is a weak ground instance of ι. We apply the substitution $\sigma\rho_3$ to the premise and the conclusion of ι. First, $\sigma\rho_3$ equals θ over $\mathrm{Var}(C)$ and σ is idempotent, so $C'\sigma\sigma\rho_3 = C'\sigma\rho_3 = C'\theta$ and $e_1\sigma\rho_3 = n\bar{u} + \bar{t} \approx m\bar{u} + \bar{s}$.

Second, $\sigma\sigma\rho_3$ equals ρ_1 over $\mathrm{Var}(e_0)$. Thus, after some computation, $e_0\sigma\sigma\rho_3 = e_0\rho_1$ turns out to be $(\tilde{n}{-}\tilde{m})\bar{u} + \sum_{j\in\tilde{J}} \tilde{n}_j\bar{s}_j + t\theta \approx \sum_{i\in\tilde{I}} \tilde{m}_i\bar{s}_i + s\theta$.

Now $n - m = (\tilde{n}{+}\tilde{p}) - (\tilde{m}{+}\tilde{p}) = \tilde{n} - \tilde{m}$. Furthermore, if $w = \sum_{i\in I\cup J} p_i\bar{s}_i$, then $\bar{t} = \sum_{j\in\tilde{J}} \tilde{n}_j\bar{s}_j + t\theta + w$ and $\bar{s} = \sum_{i\in\tilde{I}} \tilde{m}_i\bar{s}_i + s\theta + w$, hence ι' is a weak ground instance of ι. ☐

Lemma 3.25 *Let C be a clause and let θ be a substitution such that $C\theta$ is ground. Then every equality resolution, standard equality factoring, or cancellative equality factoring inference from $C\theta$ is a ground instance of a cancellation inference from C.*

Proof. For equality resolution and standard equality factoring inferences, this is proved as in the classical case. The proof for cancellative equality factoring inferences is similar to the proof of Lemma 3.24. The main differences are that we can work directly with I and J (so there is no need to compute $\tilde{I}$ and $\tilde{J}$ first), and that m and n in the proof of Lemma 3.24 are now identical (so *every* variable x_i is mapped to $\hat{x}_i + \tilde{x}_i$). ☐

Lemma 3.26 *Let $C_1 = C_1' \vee s \approx s'$ and $C_2 = C_2' \vee [\neg]\, t \approx t'$ be two clauses (without common variables) and let θ be a substitution such that $C_1\theta$ and $C_2\theta$ are ground.*

If there is a cancellative superposition inference

$$\frac{C_1'\theta \vee s\theta \approx s'\theta \qquad C_2'\theta \vee [\neg]\, t\theta \approx t'\theta}{D}$$

(where the maximal atomic subterms of $s\theta$ and $t\theta$ are overlapped), and t is not a variable, then the inference is a ground instance of a cancellative superposition inference from C_1 and C_2.

If there is a standard superposition inference

$$\frac{C_1'\theta \vee s\theta \approx s'\theta \qquad C_2'\theta \vee [\neg]\, t\theta[s\theta] \approx t'\theta}{D}$$

and $s\theta$ *does not occur in* $t\theta$ *at or below a variable position of* t *(i.e.,* $x\theta = w[s\theta]$ *for some* $x \in \mathrm{Var}(t)$*), then the inference is a ground instance of a standard superposition inference from* C_1 *and* C_2*.*

If there is an abstraction inference

$$\frac{C_1'\theta \vee s\theta \approx s'\theta \qquad C_2'\theta \vee [\neg]\, t\theta[\hat{w}\theta] \approx t'\theta}{C_2'\theta \vee \neg\, x \approx \hat{w}\theta \vee [\neg]\, t\theta[x] \approx t'\theta}$$

where $\hat{w}\theta = t_0 + t_1$ *doesn't occur below a variable position of* t*, the maximal atomic subterms of* $s\theta$ *overlap with* t_0*, and if* $\hat{w} = z + \hat{t}$ *and* t_0 *occurs below* z*, then* $\hat{t} = \hat{t}' + \hat{t}''$ *and* $\hat{t}'$ *is a variable or* $0 \prec \hat{t}'\theta \preceq t_0$*, then the inference is a ground instance of an abstraction inference from* C_1 *and* C_2*.*

Proof. For standard superposition inferences, this is proved as in the classical case. For cancellative superposition and abstraction inferences, the proof is again similar to the proof of Lemma 3.24. The main difference is that we can work directly with I and J (so there is no need to compute $\tilde{I}$ and $\tilde{J}$ first). □

4 Refutational Completeness

4.1 Rewriting on Equations

To show that the inference system described so far is refutationally complete we have to demonstrate that every saturated clause set that doesn't contain the empty clause has a model. The traditional approach to construct such a model is rewrite-based: First an ordering is imposed on the set of all ground instances of clauses in the set. Starting with an empty interpretation all such instances are inspected in ascending order. If a reductive clause is false and irreducible in the partial interpretation constructed so far, its maximal equation is turned into a rewrite rule and added to the interpretation. If the original clause set is saturated and doesn't contain the empty clause, then the final interpretation is a model of all ground instances, and thus of the original clause set (Bachmair and Ganzinger [4]).

In our case, we have to modify this scheme. For an adequate treatment of cancellative superposition it is not sufficient to be able to replace equals by equals inside a term. Rather, we need a rewrite relation on equations.

Definition 4.1 *A ground equation* e *is called a rewrite rule if* $\mathrm{mt}(e)$ *doesn't occur on both sides of* e*.*

Every rewrite rule has either the form $nv + s \approx s'$, where v is an atomic term, $n \in \mathbf{N}^+$, $v \succ s$, and $v \succ s'$, or the form $v \approx s'$, where v (and thus s') does not have sort S_{CAM}. This is an easy consequence of the multiset property of $\succ$.

At the top of a term, we will use rewrite rules in a specific way: Application of a rule $nv + s \approx s'$ to an equation $nv + t \approx t'$ means to replace nv by s' and simultaneously to add s to the other side, obtaining $s' + t \approx t' + s$.[7] However, this is only possible at the top of an equation, not below a free symbol. Consequently, there may be equations $t \approx t'$ that can be reduced to $0 \approx 0$, whereas $f(t) \approx f(t')$ is irreducible. To compensate for this fact our rewrite relation takes two sets of rewrite rules as parameters: one set of rules generated directly from the clause set, and a second set of "derived" rules, which are applied only below a free symbol and will fix the problem above.

Definition 4.2 *Given a pair (R, R') of sets of rewrite rules, the three relations $\rightarrow_{\gamma,(R,R')}$, $\rightarrow_{\delta,(R,R')}$, and $\rightarrow_{\kappa}$ are defined (modulo* ACU*) as follows:*

- $nv + t \approx t' \ \rightarrow_{\gamma,(R,R')} \ s' + t \approx t' + s$,

  if $nv + s \approx s'$ is a rule in R.

- $t[s] \approx t' \ \rightarrow_{\delta,(R,R')} \ t[s'] \approx t'$,

  if (i) $s \approx s'$ is a rule in $R \cup R'$ and (ii) s doesn't have sort S_{CAM} or s occurs in t below some free function symbol.

- $u + t \approx u + t' \ \rightarrow_{\kappa} \ t \approx t'$,
  $u \approx u \ \rightarrow_{\kappa} \ 0 \approx 0$,

  if u is different from 0.

The union of $\rightarrow_{\gamma,(R,R')}$, $\rightarrow_{\delta,(R,R')}$, and $\rightarrow_{\kappa}$ is denoted by $\rightarrow_{(R,R')}$.

We say that an equation e is γ-reducible, if $e \rightarrow_{\gamma} e'$ (analogously for δ and κ). It is called reducible, if it is γ-, δ-, or κ-reducible.

Unlike κ-reducibility, γ- and δ-reducibility can be extended to terms: A term t is called γ-reducible, if $t \approx t' \rightarrow_{\gamma} e'$, where the rewrite step takes place at the left hand side (analogously for δ). It is called reducible, if it is γ- or δ-reducible.

[7] While we have the restriction $v \succ s$, $v \succ s'$ for the rewrite rules, there is no such restriction for the equations to which rules are applied.

Lemma 4.3 *The relation $\rightarrow_{(R,R')}$ is contained in $\succ_{\mathrm{L}}$ and thus noetherian.*

Definition 4.4 *Given a pair (R, R') of sets of rewrite rules, the truth set $\mathrm{tr}(R, R')$ of (R, R') is the set of all equations $t \approx t'$ for which there exists a derivation $t \approx t' \rightarrow^{*}_{(R,R')} 0 \approx 0$.*

4.2 Model Construction

Definition 4.5 *A ground clause $C \vee e$ is called reductive for e, if e is a rewrite rule and strictly maximal in $C \vee e$.*

Definition 4.6 *Let N be a set of (possibly non-ground) clauses and let $\overline{N}$ be the set of all ground instances of clauses in N. Using induction on the clause ordering we define sets of rules R_C, R_C^{K}, E_C, and E_C^{K}, for all clauses $C \in \overline{N}$. Let C be such a clause and assume that R_D, R_D^{K}, E_D, and E_D^{K} have already been defined for all $D \in \overline{N}$ such that $C \succ_{\mathrm{C}} D$. Then*

$$R_C = \bigcup_{D \prec_{\mathrm{C}} C} E_D \qquad \text{and} \qquad R_C^{\mathrm{K}} = \bigcup_{D \prec_{\mathrm{C}} C} E_D^{\mathrm{K}} .$$

E_C is the singleton set $\{e\}$, if C is a clause $C' \vee e$ such that (i) C is reductive for e, (ii) C is false in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$, (iii) C' is false in $\mathrm{tr}(R_C \cup \{e\}, R_C^{\mathrm{K}})$, and (iv) $\mathrm{mt}_{\#}(e)$ is irreducible with respect to (R_C, R_C^{K}). Otherwise, E_C is empty.

The set E_C^{K} is non-empty only if E_C contains a rule $nv + s \approx s'$. In this case, E_C^{K} is the set of all $knv + r \approx r' \in \mathrm{tr}(R_C \cup E_C, R_C^{\mathrm{K}})$ such that $v \succ r$, $v \succ r'$, and $knv + r \approx r'$ is δ_{K}-irreducible with respect to (R_C, R_C^{K}).

Finally, the sets R_∞ and R_∞^{K} are defined by

$$R_\infty = \bigcup_{D \in \overline{N}} E_D \qquad \text{and} \qquad R_\infty^{\mathrm{K}} = \bigcup_{D \in \overline{N}} E_D^{\mathrm{K}} .$$

Our goal is to show that $\mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$ is a model of the axioms of cancellative abelian monoids, and, for certain clause sets N, also a model of N. To this end, we will first put together some basic properties of R_C and R_C^{K}. In Section 4.3 we prove that the rewrite relation associated with (R_C, R_C^{K}) satisfies a restricted confluence property. The equality axioms and the laws of cancellative abelian monoids follow as easy corollaries. Then we show in Section 4.4 that $\mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$ is in fact a model of N, provided that N is saturated and does not contain the empty clause.

Lemma 4.7 *Let C and D be two clauses from $\overline{N}$ such that $C \succ_{\mathrm{C}} D$. If $E_C = \{e_1\}$ and $E_D = \{e_2\}$, then $\mathrm{mt}(e_1) \succ \mathrm{mt}(e_2)$.*

Proof. Suppose that $\mathrm{mt}(e_1) \preceq \mathrm{mt}(e_2)$. Then either $\mathrm{mt}_\#(e_1) \prec \mathrm{mt}_\#(e_2)$, so by the definition of the clause ordering, we would have $C \prec_{\mathrm{C}} D$. Or $\mathrm{mt}(e_1) = \mathrm{mt}(e_2)$ and $\mathrm{mt}_\#(e_1) \succeq \mathrm{mt}_\#(e_2)$, then $\mathrm{mt}_\#(e_1)$ could be γ- or δ-reduced using e_2. This is impossible, however. □

Lemma 4.8 *For every $C \in \overline{N}$, $R_C \cup R_C^{\mathrm{K}} \subseteq \mathrm{tr}(R_C, R_C^{\mathrm{K}})$ and $R_\infty \cup R_\infty^{\mathrm{K}} \subseteq \mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$.*

Lemma 4.9 *If the rewrite rule e is contained in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$, then $\mathrm{mt}_\#(e)$ is reducible by (R_C, R_C^{K}).*

Proof. Suppose that $e = nv + s \approx s'$, where $v = \mathrm{mt}(e)$, $n \in \mathbf{N}^+$, $v \succ s$, and $v \succ s'$. Then there is a derivation

$$nv + s \approx s' \ \rightarrow^*_{(R_C, R_C^{\mathrm{K}})} \ 0 \approx 0\,.$$

During this derivation, all occurrences of v are deleted eventually. As s and s' are smaller than v, it is impossible to derive an occurrence of v on the right-hand side. Therefore, the occurrences of v cannot be deleted by κ-steps, but only by γ- or δ-steps, so nv is reducible.

The case that $e = v \approx s'$ and v doesn't have sort S_{CAM} is proved in the same way. □

Lemma 4.10 *If $nv + s = s'$ is a rewrite rule from E_C, then for every $knv + r \approx r' \in E_C \cup E_C^{\mathrm{K}}$ there is an $(R_C \cup E_C, R_C^{\mathrm{K}})$-derivation*

$$knv + r \approx r' \ \rightarrow^+_\gamma \ r + ks' \approx r' + ks \ \rightarrow^* \ 0 \approx 0$$

starting with k-fold application of $nv + s = s'$.

Proof. If $knv + r \approx r' \in E_C$, then this is obvious, as E_C is a singleton. If $knv + r \approx r' \in E_C^{\mathrm{K}}$, then by definition, there is a derivation

$$knv + r \approx r' \ \rightarrow^*_{(R_C \cup E_C, R_C^{\mathrm{K}})} \ 0 \approx 0\,.$$

During this derivation, all occurrences of v are deleted eventually. As $knv + r \approx r'$ is $\delta\kappa$-irreducible with respect to (R_C, R_C^{K}), this can only happen by k-fold γ-application of $nv + s = s'$. These γ-steps are independent of any preceding rewrite steps, hence we can shift them to the front of the derivation. □

Lemma 4.11 *Let v be an atomic term, $n \in \mathbf{N}^+$. If nv is γ-reducible by $(R_\infty, R_\infty^{\mathrm{K}})$ or (R_C, R_C^{K}) for some $C \in \overline{N}$ then nv is δ-irreducible by $(R_\infty, R_\infty^{\mathrm{K}})$ and (R_D, R_D^{K}) for every $D \in \overline{N}$.*

Proof. If nv is γ-reducible, then there exists a rule $kv + s \approx s' \in E_{C'}$, where $k \leq n$. Suppose that nv were δ-reducible using a rule $t \approx t' \in E_{D'} \cup E_{D'}^{\mathrm{K}}$. We distinguish between three cases:

If $D' \prec_{\mathrm{C}} C'$, then t would have to be a subterm of v. Consequently, kv would be reducible with respect to $(R_{C'}, R_{C'}^{\mathrm{K}})$, which is impossible by the definition of $E_{C'}$.

If $D' \succ_{\mathrm{C}} C'$, then t is strictly larger than $k'v$ for any $k' \in \mathbf{N}^+$, hence nv cannot be δ-reduced by $t \approx t'$.

If $D' = C'$ then t has the form $k'kv + r$, and a δ-reduction using $t \approx t'$ may take place only below a free function symbol. Again, it is impossible to δ-reduce nv by $t \approx t'$. ☐

Lemma 4.12 *If $C \in \overline{N}$ and if $e \in \mathrm{tr}(R_C \cup E_C, R_C^{\mathrm{K}})$ is $\delta\kappa$-irreducible with respect to $(R_C \cup E_C, R_C^{\mathrm{K}})$, then $e \in R_C^{\mathrm{K}} \cup E_C^{\mathrm{K}} \cup \{0 \approx 0\}$.*

If $C \in \overline{N}$ and if $e \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$ is $\delta\kappa$-irreducible with respect to (R_C, R_C^{K}), then $e \in R_C^{\mathrm{K}} \cup \{0 \approx 0\}$. Similarly, if $e \in \mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$ is $\delta\kappa$-irreducible with respect to $(R_\infty, R_\infty^{\mathrm{K}})$, then $e \in R_\infty^{\mathrm{K}} \cup \{0 \approx 0\}$.

Proof. Suppose that e is different from $0 \approx 0$ and let $v = \mathrm{mt}(e)$. Without loss of generality we may assume that $e = nv + s \approx s'$, $v \succ s$, and $v \succ s'$. During the derivation

$$nv + s \approx s' \ \rightarrow^*_{(R_C \cup E_C, R_C^{\mathrm{K}})} \ 0 \approx 0\,,$$

all occurrences of v are deleted eventually. This can be done only by (possibly several) γ-rewriting steps. Therefore, $R_C \cup E_C$ contains (exactly!) one rule $kv + t = t' \in E_D$ for some $D \preceq_{\mathrm{C}} C$. Since nv is deleted completely, n must be equal to $k'k$ where k' is the number of γ-rewriting steps. The remaining subterms in the equation are smaller than v, so all further reduction steps can use only rules from (R_D, R_D^{K}). Thus $e \in E_D^{\mathrm{K}} \subseteq R_C^{\mathrm{K}} \cup E_C^{\mathrm{K}}$.

The second part of the lemma is proved analogously. ☐

4.3 Confluence

It is easy to see that the relation $\rightarrow_{(R_C, R_C^{\mathrm{K}})}$ is in general not confluent:

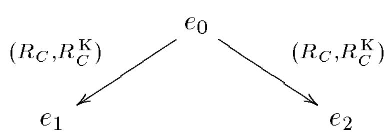
Example 4.13 *Let $N = \overline{N} = \{C\}$ where C is the clause $2c \approx d$. Then $R_\infty = \{2c \approx d\}$ and $R_\infty^{\mathrm{K}} = \emptyset$. Now the equation $2c \approx c$ can be rewritten to*

$d \approx c$, using an γ-step, and also to $c \approx 0$, using a κ-step. Both equations are irreducible.

We can merely show that $\rightarrow_{(R_C, R_C^{\mathrm{K}})}$ is confluent on $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$, that is, that any two derivations starting from an equation e can be joined, provided that there is a derivation $e \rightarrow^{+} 0 \approx 0$. In fact, this will be sufficient for our purposes.

Theorem 4.14 *The relation $\rightarrow_{(R_C, R_C^{\mathrm{K}})}$ is confluent on $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$ for every $C \in \overline{N}$. The relation $\rightarrow_{(R_\infty, R_\infty^{\mathrm{K}})}$ is confluent on $\mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$.*

Proof. Let us consider the relation $\rightarrow_{(R_C, R_C^{\mathrm{K}})}$. (The case of $\rightarrow_{(R_\infty, R_\infty^{\mathrm{K}})}$ is similar.) We have to prove that for any two (R_C, R_C^{K})-derivations $e \rightarrow^{+} 0 \approx 0$ and $e \rightarrow^{+} e'$ there exists a derivation $e' \rightarrow^{*} 0 \approx 0$. As usual, this is done by analyzing critical pairs and by noetherian induction over the size of e. However, we need the induction hypothesis not only to show that local confluence implies global confluence, but even to prove local confluence (in particular in Case 2 below). Let us consider a peak



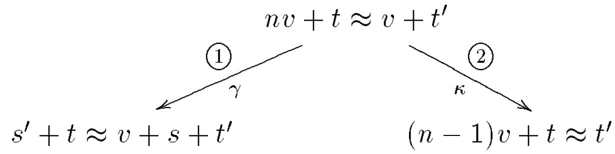
where either e_1 or e_2 can be reduced to $0 \approx 0$.

Case 1: Trivial peaks.

As in the traditional term rewriting framework, every peak converges if the two rewrite steps take place at disjoint redexes. Furthermore, local confluence is obvious, if both steps are κ-steps, if both steps are γ-steps (since there is at most one rule that can be applied in an γ-step at some nv), or if there is one δ- and one κ-step. Finally, Lemma 4.11 shows that γ- and δ-steps can only take place at disjoint redexes. It remains thus to consider γ/κ-peaks and δ/δ-peaks.

Case 2: γ/κ-peaks.

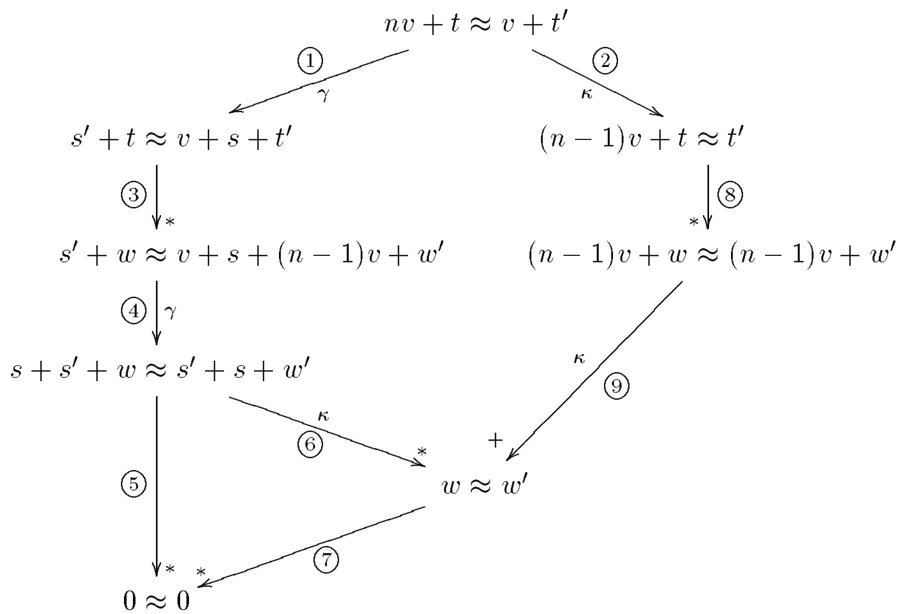
Closing a peak between a κ-step and an γ-step is trivial if the latter takes place at some free function symbol. It suffices therefore to consider the situation where a rewrite rule $nv + s \approx s' \in E_D \subseteq R_C$ with $n \geq 2$ is applied at the top of an equation $nv + t \approx v + t'$. This yields a peak



where either $s' + t \approx v + s + t'$ or $(n-1)v + t \approx t'$ can be rewritten to $0 \approx 0$ by (R_C, R_C^{K}).

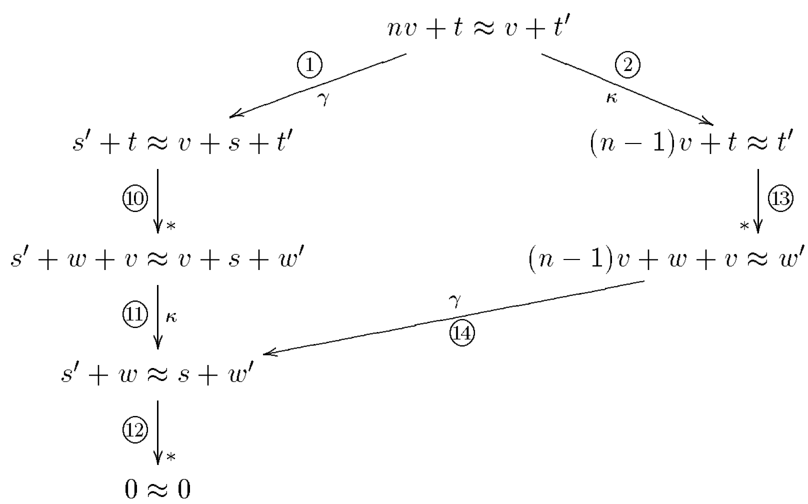
Case 2.1: $s' + t \approx v + s + t' \to^* 0 \approx 0$.

At some step of the (R_C, R_C^{K})-derivation $s' + t \approx v + s + t' \to^* 0 \approx 0$ the term v must be eventually deleted. Suppose that the deletion happens by another application of $nv + s \approx s'$. Such a step requires the presence of $n - 1$ further occurrences of v. As these occurrences cannot be derived from s or s', we may assume without loss of generality that the derivation has the form ③-④-⑤:



On the other hand, we can use κ-steps ⑥ to rewrite $s + s' + w \approx s' + s + w'$ to $w \approx w'$. By the induction hypothesis, the peak between ⑤ and ⑥ can be joined, hence there is a derivation ⑦. As the steps ③ take place only at t and t', we can simulate them by ⑧. Finally, the diagram is closed by using $n - 1$ κ-steps ⑨.

If v is not deleted by another application of $nv + s \approx s'$, then it must be deleted by a κ-step. Again, this requires the existence of another occurrence of v, which cannot be derived from s or s'. We may thus assume that the derivation has the form ⑩-⑪-⑫:



As the steps ⑩ take place only at t and t', we can simulate them by ⑬. Finally, we can close the diagram using γ-rewriting ⑭ by $nv + s \approx s'$.

Case 2.2: $(n-1)v + t \approx t' \to^* 0 \approx 0$.

This is proved in nearly the same way as Case 2.1.

Case 3: δ/δ-peaks.

It remains to show that every δ/δ-peak converges. Suppose that the first rewrite step uses a rule from some E_D or E_D^{K}, and that the second rewrite step uses a rule from some $E_{D'}$ or $E_{D'}^{\mathrm{K}}$, where $D \succeq_{\mathrm{C}} D'$. We have to distinguish two cases: Either the two steps rewrite overlapping parts of a sum, or they take place one below the other.

Case 3.1: Overlaps below a free symbol.

Suppose that there are two rules $s[t] \approx s' \in E_D \cup E_D^{\mathrm{K}}$ and $t \approx t' \in E_{D'} \cup E_{D'}^{\mathrm{K}}$, such that t occurs in s below a free function symbol. Obviously, D' must be smaller than D. As all rules from E_D^{K} and the maximal terms of rules from E_D are δ-irreducible with respect to (R_D, R_D^{K}), $s[t] \approx s'$ must be a rule $nv + u[t] \approx s' \in E_D$ and the peak has the form

$$
\begin{array}{ccc}
 & e[nv + u[t]] & \\
\overset{①}{\underset{\delta}{\swarrow}} & & \overset{②}{\underset{\delta}{\searrow}} \\
e[nv + u[t']] & & e[s']
\end{array}
$$

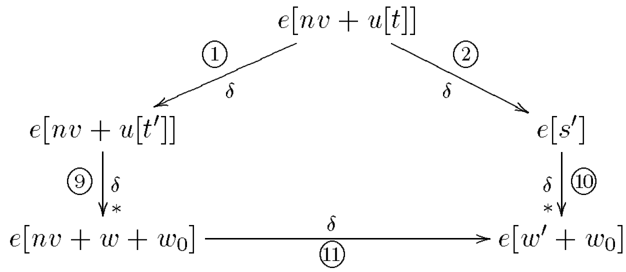
where $nv + u[t]$ occurs in e below a free function symbol.

Starting from the equation $nv + u[t'] \approx s'$ we construct a derivation as follows:

$$
\begin{array}{c}
nv + u[t'] \approx s' \\
③ \Big\downarrow {\scriptstyle \delta\, *} \\
nv + w + w_0 \approx w' + w_0 \\
④ \Big\downarrow {\scriptstyle \kappa\, *} \\
nv + w \approx w' \\
⑤ \Big\downarrow {\scriptstyle \gamma} \\
s' + w \approx w' + u[t] \\
⑥ \Big\downarrow {\scriptstyle \delta} \\
s' + w \approx w' + u[t'] \\
⑦ \Big\downarrow {\scriptstyle \delta\, *} \\
w' + w_0 + w \approx w' + w + w_0 \\
⑧ \Big\downarrow {\scriptstyle \kappa\, *} \\
0 \approx 0
\end{array}
$$

First we δ-normalize the equation with respect to (R_D, R_D^{K}) by rewriting $u[t']$ to $w + w_0$ and s' to $w' + w_0$ ③. Then we use κ-steps ④ to cancel the common part w_0 (if any) on both sides of the equation. Using an γ-step by $nv + u[t] \approx s'$ ⑤ and a δ-step by $t \approx t'$ ⑥ we obtain $s' + w \approx w' + u[t']$, As the steps ③ take place only at $u[t']$ and s', we can simulate them by ⑦. Finally, we can use κ-rewriting ⑧ to reduce the equation to $0 \approx 0$.

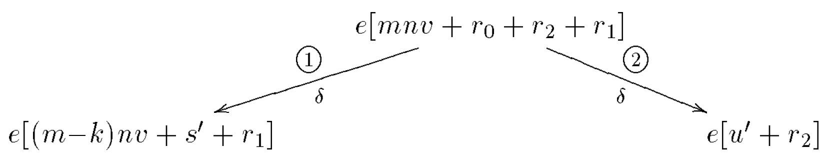
By the construction of w and w' and by Lemma 4.11, the equation $nv + w \approx w'$ is $\delta\kappa$-irreducible with respect to $(R_D \cup E_D, R_D^{\mathrm{K}})$. By Lemma 4.12, it is contained in E_D^{K}. Hence we can join the peak between ① and ② as follows:



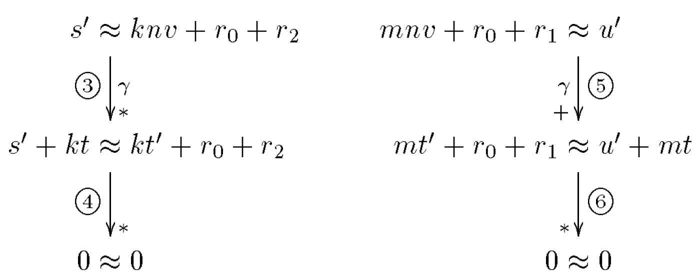
where steps ⑨ and ⑩ simulate ③, and step ⑪ uses $nv + w \approx w'$.

Case 3.2: Overlaps at the top.

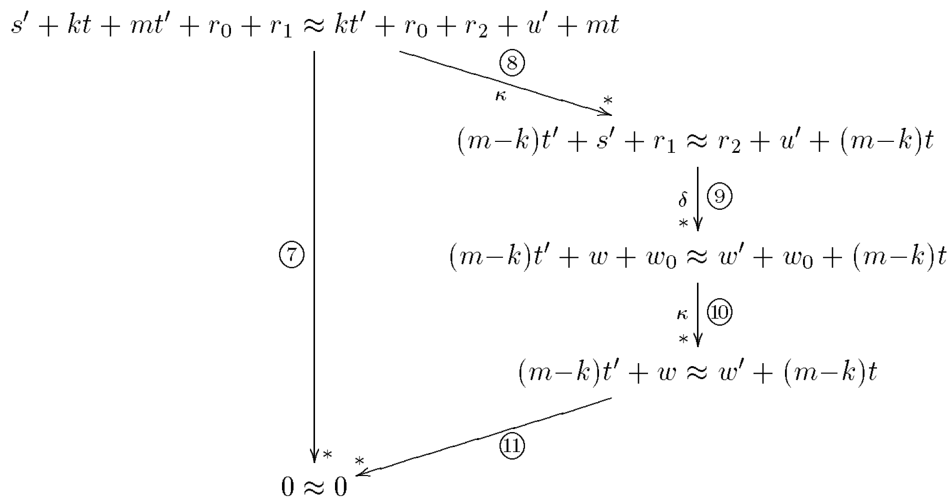
Suppose that two rules $u \approx u' \in E_D \cup E_D^{\mathrm{K}}$ and $s \approx s' \in E_{D'} \cup E_{D'}^{\mathrm{K}}$ are used in δ-steps to rewrite the same redex or overlapping parts of a sum in an equation e. Without loss of generality let $D \succeq_{\mathrm{C}} D'$. If u or s doesn't have sort S_{CAM}, then D and D' and thus $s \approx s'$ and $u \approx u'$ must be identical, so the peak converges trivially. We can therefore assume without loss of generality that $u = mnv + r_0 + r_1$ and $s = knv + r_0 + r_2$, where E_D contains the rule $nv + t \approx t'$. Deviating from our standard notational convention we allow $k = 0$ (if and only if $D \succ_{\mathrm{C}} D'$) so that we can handle the cases $D \succ_{\mathrm{C}} D'$ and $D = D'$ simultaneously. If $D = D'$, we assume by symmetry that $m \geq k$. The peak has the form



As $s \approx s'$ and $u \approx u'$ are contained in $E_D \cup R_D \cup E_D^{\mathrm{K}} \cup R_D^{\mathrm{K}}$, there exist $(R_D \cup E_D, R_D^{\mathrm{K}})$-derivations ③-④ and ⑤-⑥ starting with k- or m-fold γ-application of $nv + t \approx t'$.



Consider the two equations $s' + kt \approx kt' + r_0 + r_2$ and $mt' + r_0 + r_1 \approx u' + mt$. If we add the left-hand sides and right-hand sides, respectively, we obtain a new equation that can be rewritten to $0 \approx 0$ using a combination ⑦ of ④ and ⑥:



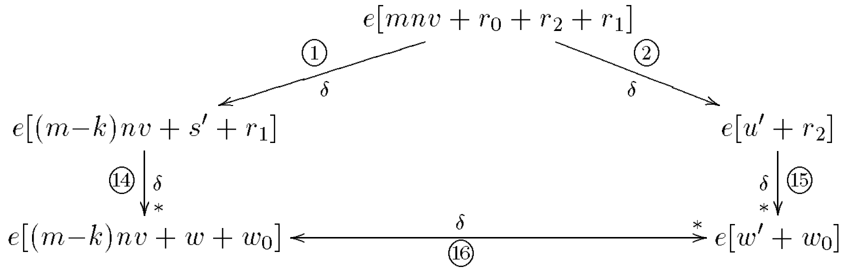
We will now show how to construct an alternative derivation starting from $s' + kt + mt' + r_0 + r_1 \approx kt' + r_0 + r_2 + u' + mt$. First we use κ-steps ⑧ to cancel kt, kt' and r_0 on both sides of the equation. Then the resulting equation is partially $\delta\kappa$-normalized, first by δ-rewriting $s' + r_1$ to $w + w_0$ and $r_2 + u'$ to $w' + w_0$ ⑨, then by cancelling ⑩ the common part w_0 (if different from 0). Using the induction hypothesis, we see that the peak between ⑦ and ⑧-⑨-⑩ can be joined, thus there is a derivation ⑪ which closes the diagram.

Consider now the equation $(m-k)nv + w \approx w'$. Using $(m-k)$-fold γ-rewriting by $nv + t \approx t'$ and continuing as in ⑪, we obtain a derivation

$$\begin{array}{c} (m-k)nv + w \approx w' \\ \text{⑫} \Big\downarrow{}^{\gamma}_{*} \\ (m-k)t' + w \approx w' + (m-k)t \\ \text{⑬} \Big\downarrow{}_{*} \\ 0 \approx 0 \end{array}$$

Furthermore, by the construction of w and w' and by Lemma 4.11, the equation $(m-k)nv + w \approx w'$ is $\delta\kappa$-irreducible with respect to $(R_D \cup E_D, R_D^{\mathrm{K}})$.

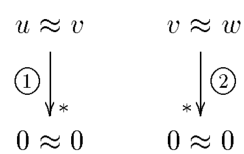
By Lemma 4.12, $(m{-}k)nv + w \approx w'$ is contained in $R_D^{\mathrm{K}} \cup E_D^{\mathrm{K}} \cup \{0 \approx 0\}$. Therefore, we can join the peak between ① and ② as follows:



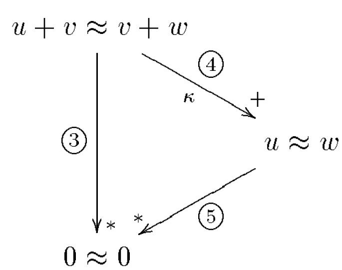
where steps ⑭ and ⑮ simulate ⑨, and step ⑯ uses $(m{-}k)nv + w \approx w'$. $\square$

Corollary 4.15 *For every* $C \in \overline{N}$, $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$ *and* $\mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$ *are models of the equality axioms.*

Proof. We consider only $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$; the proof for $\mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$ is similar. It is obvious that $u \approx u \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$ for every term u, and that $u \approx v \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$ implies $v \approx u \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$. For the transitivity axiom, consider two equations $u \approx v$ and $v \approx w \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$.

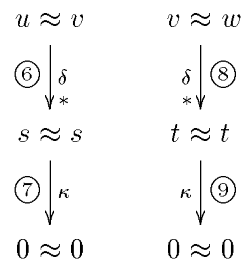


If u, v and w have sort S_{CAM}, we can combine the derivations ① and ② and obtain a derivation ③:

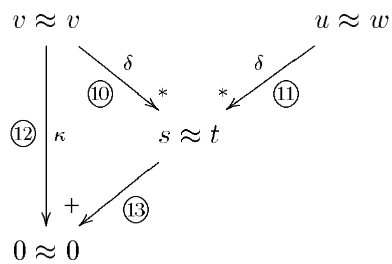


On the other hand, we can use κ-steps ④ to cancel v on both sides of the equation. By Theorem 4.14, there is a derivation ⑤, hence $u \approx w \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$.

If u, v and w don't have sort S_{CAM}, the derivations ① and ② must have the form ⑥-⑦ and ⑧-⑨:

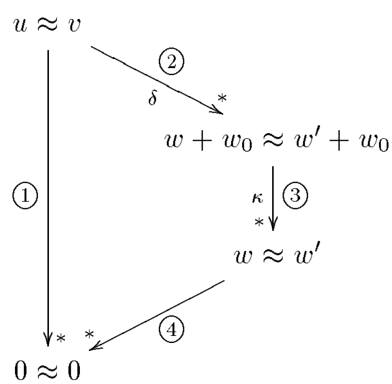


As the δ-steps in ⑥ and ⑧ rewrite each side of the equations separately, we can use the same rules to rewrite $v \approx v$ ⑩ and $u \approx w$ ⑪.



On the other hand, we can rewrite $v \approx v$ immediately to $0 \approx 0$ ⑫. By confluence, there is a derivation ⑬ and $u \approx w \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$.

It remains to show that $u \approx v \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$ entails $t[u] \approx t[v] \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$. If there is no free symbol in t above u, this is trivial, so let us assume that u occurs in t below a free symbol. Consider the derivation ①:



We can $\delta\kappa$-normalize $u \approx v$, first by δ-rewriting u to $w + w_0$ and v to $w' + w_0$ ②, then by cancelling ③ the common part w_0. (If u and v don't have sort S_{CAM}, ② yields $w \approx w'$ directly.) According to Theorem 4.14, there exists a derivation ④. The equation $w \approx w'$ is $\delta\kappa$-irreducible with respect to (R_C, R_C^{K}), hence it is contained in $R_C^{\mathrm{K}} \cup \{0 \approx 0\}$ by Lemma 4.12. Without loss of generality we assume $w \succeq w'$. This allows us to construct the following derivation:

$$
\begin{array}{c}
t[u] \approx t[v] \\
⑤ \Big\downarrow {}^{\delta}_{*} \\
t[w + w_0] \approx t[w' + w_0] \\
⑥ \Big\downarrow {}^{\delta}_{*} \\
t[w' + w_0] \approx t[w' + w_0] \\
⑦ \Big\downarrow {}^{\kappa}_{+} \\
0 \approx 0
\end{array}
$$

where step ⑤ simulates ② and step ⑥ uses $w \approx w'$ (if different from $0 \approx 0$). Summarizing we get $t[u] \approx t[v] \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$. □

Corollary 4.16 *For every $C \in \overline{N}$, $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$ and $\mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$ are models of* ACUK*.*

Proof. The cancellation axiom is proved in the same way as the transitivity axiom (Corollary 4.15). The associative, commutative, and identity axioms are obvious. □

4.4 Completeness

Lemma 4.17 *Let C be a clause from $\overline{N}$. If C is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$, then it is also true in $\mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$ and $\mathrm{tr}(R_D, R_D^{\mathrm{K}})$ for any $D \succ_{\mathrm{C}} C$.*

Lemma 4.18 *Let $C = C' \vee e$ be a clause from $\overline{N}$ such that $E_C = \{e\}$. Then C is true and C' is false in $\mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$ and $\mathrm{tr}(R_D, R_D^{\mathrm{K}})$ for any $D \succeq_{\mathrm{C}} C$.*

Lemma 4.19 *Let N be a set of clauses that is saturated up to* ACUK*-redundancy and doesn't contain the empty clause. Then for every ground clause $C \in \overline{N}$ we have:*

(i) If C is an instance of a clause in N with selected literals, then C is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$.

(ii) If C is redundant, then it is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$.

(iii) $E_C = \emptyset$ if and only if C is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$.

(iv) C is true in $\mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$ and in $\mathrm{tr}(R_D, R_D^{\mathrm{K}})$ for every $D \succ_{\mathrm{C}} C$.

Proof. We use induction on the clause ordering $\succ_{\rm C}$ and assume that (i)–(iv) are already satisfied for all $C' \in \overline{N}$ with $C \succ_{\rm C} C'$. Note that the "if" part of (iii) is obvious from the model construction and that condition (iv) follows immediately from (iii), Lemma 4.18, and Lemma 4.17. Let $C = \hat{C}\theta$ for some $\hat{C} \in N$.

Case 1: C is redundant.

If C is redundant, then there are clauses $C_1, \dots, C_n \in \overline{N}$ such that $C_i \prec_{\rm C} C$ for every $i \in \{1, \dots, n\}$ and $\{C_1, \dots, C_n\} \models C$. By part (iv) of the induction hypothesis, all C_i are true in $\mathrm{tr}(R_C, R_C^{\rm K})$. As $\mathrm{tr}(R_C, R_C^{\rm K})$ satisfies ACUK, C is true in $\mathrm{tr}(R_C, R_C^{\rm K})$.

Case 2: $x\theta$ equals some smaller term.

Suppose there is a variable x in $\hat{C}$ and a ground term w such that $x\theta \succ w$ and $x\theta \approx w \in \mathrm{tr}(R_C, R_C^{\rm K})$. Let the substitution ρ be defined by $x\rho = w$ and $y\rho = y\theta$ for every variable $x \neq y$. The clause $\hat{C}\rho$ is smaller than C. By part (iv) of the induction hypothesis, it is true in $\mathrm{tr}(R_C, R_C^{\rm K})$. As $\mathrm{tr}(R_C, R_C^{\rm K})$ satisfies the equality axioms, every literal of $\hat{C}\theta$ is true in $\mathrm{tr}(R_C, R_C^{\rm K})$ if and only if the corresponding literal of $\hat{C}\rho$ is true; hence C is true in $\mathrm{tr}(R_C, R_C^{\rm K})$.

Case 3: C contains a selected or maximal negative literal.

Suppose that $C = C' \vee \neg\, e$ doesn't fall into the preceding two categories and that the literal $\neg\, e$ is either an instance of a selected literal in $\hat{C}$ or or that $\hat{C}$ contains no selected literal and $\neg\, e$ is maximal in C. If $e \notin \mathrm{tr}(R_C, R_C^{\rm K})$, there is nothing to show, so assume that there is a $(R_C, R_C^{\rm K})$-derivation from e to $0 \approx 0$. Let $v = \mathrm{mt}(e)$.

Case 3.1: v occurs on both sides of e.

If e equals $v \approx v$ where v either doesn't have sort $S_{\rm CAM}$ or equals 0, then there is an equality resolution inference from $C = C' \vee \neg\, e$ yielding C'. This inference is a ground instance of an equality resolution inference from $\hat{C}$. By saturation up to ACUK-redundancy, it is redundant, hence there are clauses $C'_1, \dots, C'_n \in \overline{N}$ such that $C'_i \prec_{\rm C} C$ and $\{C'_1, \dots, C'_n\} \models C'$. By the induction hypothesis, all C'_i and thus C' and C are true in $\mathrm{tr}(R_C, R_C^{\rm K})$.

If e equals $nv + t \approx mv + t'$ with $n \geq m \geq 1$ then there is a cancellation inference from $C = C' \vee \neg\, e$ yielding $C' \vee \neg\, (n{-}m)v + t \approx t'$. This inference is a weak ground instance of a cancellation inference from $\hat{C}$. As above, we deduce that $C' \vee \neg\, (n{-}m)v + t \approx t'$ and C are true in $\mathrm{tr}(R_C, R_C^{\rm K})$.

Case 3.2: v occurs on only one side of e.

If v occurs only on one side of e, then e has either the form $nv + s \approx s'$ or $v \approx s'$ and v doesn't have sort S_{CAM}. We write $e[v]$ if the distinction between these two cases is irrelevant.[8] By Lemma 4.9 we may assume that the reduction from e to $0 \approx 0$ starts with an γ- or δ-step at nv or v.

Case 3.2.1: Reduction by R_C.

Suppose that the reduction from e to $0 \approx 0$ starts with the application of a rule $e' \in E_D \subseteq R_C$, where $D = \hat{D}\theta = D' \vee e'$ for some $\hat{D} \in N$. By part (i)-(iii) of the induction hypothesis and Lemma 4.18, D is not redundant, $\hat{D}$ has no selected literals, and D' is false in $\text{tr}(R_C, R_C^{\text{K}})$.

Case 3.2.1.1: Reduction at the top and v has sort S_{CAM}.

If v has sort S_{CAM} and the reduction takes place at the top, then e' is a rewrite rule $mv + t \approx t'$. Consider the cancellative superposition inference

$$\frac{D' \vee mv + t \approx t' \qquad C' \vee \neg\, nv + s \approx s'}{D' \vee C' \vee \neg\, (n-m)v + s + t' \approx s' + t}$$

If $nv + s$ were equal to $x\theta$ for some variable x occurring in $\hat{C}$, then $x\theta \approx s' \in \text{tr}(R_C, R_C^{\text{K}})$, so C would be subject to Case 2 above. By Lemma 3.26 the inference is therefore a ground instance of a cancellative superposition inference from $\hat{D}$ and $\hat{C}$. As N is saturated, it is redundant, thus its conclusion is true in $\text{tr}(R_C, R_C^{\text{K}})$. Both D' and $\neg\, (m-n)v + s + t' \approx s' + t$ are false in $\text{tr}(R_C, R_C^{\text{K}})$, so C' and C must be true in $\text{tr}(R_C, R_C^{\text{K}})$.

Case 3.2.1.2: Reduction below a free function symbol or v doesn't have sort S_{CAM}.

If v does not have sort S_{CAM} or if the reduction takes place below a free function symbol, then e' is a rewrite rule $t \approx t'$ and $e = e[v[t]]$. If t occurred in e at or below a variable position of $\hat{C}$, say, $x\theta = w[t]$, then $x\theta \approx w[t'] \in \text{tr}(R_C, R_C^{\text{K}})$, so C would be subject to Case 2 above. Consequently, the standard superposition inference

$$\frac{D' \vee t \approx t' \qquad C' \vee \neg\, e[v[t]]}{D' \vee C' \vee \neg\, e[v[t']]}$$

is a ground instance of a standard superposition inference from $\hat{D}$ and $\hat{C}$. Again, by saturation, its conclusion is true in $\text{tr}(R_C, R_C^{\text{K}})$; and since D' and $\neg\, e[v[t']]$ are false in $\text{tr}(R_C, R_C^{\text{K}})$, both C' and C must be true.

[8] Recall that nv is merely an abbreviation for the n-fold sum $v + \cdots + v$. If $e = nv + s \approx s'$, then the hole in $e[\_]$ is the position of *one* of the n v's.

Case 3.2.2: Reduction by R_C^{K}.

Suppose that e can be reduced using a rule $e'' = kmu + r \approx r'$ from R_C^{K}, but not using a rule from R_C. Let $D = \hat{D}\theta = D' \vee e'$ such that $e' = mu + s \approx s'$, $E_D = \{e'\}$, and $e'' \in E_D^{\mathrm{K}}$. We may assume that some sum $kmu + r + r''$ occurs in v immediately below a free function symbol. As e can't be reduced using e', there is no s'' such that $r + r'' = s + s''$, hence there is an abstraction inference

$$\frac{D' \vee mu + s \approx s' \qquad C' \vee \neg\, e[v[kmu + r + r'']]}{C' \vee \neg\, y \approx kmu + r + r'' \vee \neg\, e[v[y]]}$$

If $kmu + r$ occurred in e at or below a variable position of $\hat{C}$, then C would be subject to Case 2 above. Hence let $\hat{C} = \hat{C}' \vee \neg\, \hat{e}[\hat{v}[\hat{t}]]$, where $\hat{e}[\_]\theta = e[\_]$, $\hat{v}[\_]\theta = v[\_]$, and $\hat{t}\theta = kmu + r + r''$. Assume that $\hat{t}$ had the form $z + \sum_{j\in J} \hat{t}_j$, where $z\theta \succeq mu$ and all $\hat{t}_j$ are atomic terms and are either zero or $\hat{t}_j\theta \succ z\theta$. Then $z\theta$ could be written as $kmu + r + r'''$, since $u \succ r$. This is impossible, though, as $kmu + r$ must not occur at or below a variable position. By Lemma 3.26, the inference is therefore a ground instance of an abstraction inference from $\hat{D}$ and $\hat{C}$. By saturation,

$$C' \vee \neg\, r' + r'' \approx kmu + r + r'' \vee \neg\, e[v[r' + r'']]$$

is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$; and since $r' + r'' \approx kmu + r + r'' \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$, C must be true likewise.

Case 4: C doesn't contain a selected or maximal negative literal.

Suppose that C doesn't fall into one of the cases 1–3. Then C can be written as $C' \vee e$, where e is a maximal literal of C. If $E_C = \{e\}$ or C' is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$, then there is nothing to show, so assume that $E_C = \emptyset$ and that C' is false in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$. Let $v = \mathrm{mt}(e)$.

Case 4.1: v occurs on both sides of e.

Obviously e cannot have the form $v \approx v$, since then C would be a tautology and thus redundant. If e equals $mv + t \approx nv + t'$ with $n \geq m \geq 1$, then there is a cancellation inference from C. As in case 3.1, we can show that C is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$.

Case 4.2: v occurs on only one side of e.

If v occurs only on one side of e, then either $e = nv + s \approx s'$, or $e = v \approx s'$ and v doesn't have sort S_{CAM}.

Case 4.2.1: C' is true in $\mathrm{tr}(R_C \cup \{e\}, R_C^{\mathrm{K}})$.

Suppose that C' is false in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$, but true in $\mathrm{tr}(R_C \cup \{e\}, R_C^{\mathrm{K}})$. (This includes in particular the case that e is maximal, but not strictly maximal in C.) Then $C' = C'' \vee e_1$, where the literal e_1 in C' is not larger than e and has an $(R_C \cup \{e\}, R_C^{\mathrm{K}})$-derivation to $0 \approx 0$ in which e is used. Since the latter condition implies $\mathrm{mt}_\#(e_1) \succeq \mathrm{mt}_\#(e)$, e_1 must be a rewrite rule and $\mathrm{mt}_\#(e_1) = \mathrm{mt}_\#(e)$.

Case 4.2.1.1: v has sort S_{CAM}.

If v has sort S_{CAM}, let $e = nv + s \approx s'$ and $e_1 = nv + t \approx t'$. Without loss of generality, the $(R_C \cup \{e\}, R_C^{\mathrm{K}})$-derivation of e_1 starts with an γ-application of e and has the form

$$nv + t \approx t' \;\rightarrow\; s' + t \approx s + t' \;\rightarrow^*_{(R_C, R_C^{\mathrm{K}})}\; 0 \approx 0$$

The cancellative equality factoring inference

$$\frac{C'' \vee nv + s \approx s' \vee nv + t \approx t'}{C'' \vee \neg\, s + t' \approx s' + t \vee nv + t \approx t'}$$

is a ground instance of a cancellative equality factoring inference from $\hat{C}$. By saturation, its conclusion is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$. As $s + t' \approx s' + t \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$, $C'' \vee s' + t \vee nv + t \approx t'$ and thus C must be true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$.

Case 4.2.1.2: v doesn't have sort S_{CAM}.

If v doesn't have sort S_{CAM}, then let $e = v \approx s'$ and $e_1 = v \approx t'$. Analogously to case 4.2.1.1 we can show that the conclusion of the standard equality factoring inference

$$\frac{C'' \vee v \approx s' \vee v \approx t'}{C'' \vee \neg\, s' \approx t' \vee v \approx t'}$$

is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$ and that $s' \approx t' \in \mathrm{tr}(R_C, R_C^{\mathrm{K}})$. Again, C must be true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$.

Case 4.2.2: C' is false in $\mathrm{tr}(R_C \cup \{e\}, R_C^{\mathrm{K}})$.

So far we have considered the cases that C is not reductive for e, or that C is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$, or that C' is true in $\mathrm{tr}(R_C \cup \{e\}, R_C^{\mathrm{K}})$. If none of these conditions applies but E_C is empty, then $\mathrm{mt}_\#(e)$ must be reducible with respect to (R_C, R_C^{K}). We can then show that C is true in $\mathrm{tr}(R_C, R_C^{\mathrm{K}})$, using essentially the same techniques as in case 3.2. This concludes the proof of this lemma. □

Theorem 4.20 *Let N be a set of clauses that is saturated up to* ACUK-*redundancy. Then $N \cup \mathrm{ACUK}$ is equality unsatisfiable if and only if N contains the empty clause.*

Proof. If N contains the empty clause, then it is unsatisfiable. Otherwise, $\mathrm{tr}(R_\infty, R_\infty^{\mathrm{K}})$ is a model of the equality axioms (by Corollary 4.15), of ACUK (by Corollary 4.16), and of N (by part (iv) of Lemma 4.19). □

5 Simplification Techniques

Let N be a set of clauses. We say that $M \subseteq N$ is simplified to another set of clauses M', if $N \models M'$ (so that we may add M', once we have got N), and if M is ACUK-redundant with respect to $N \cup M'$ (so that we may delete M, once we have added M').

For example, every clause

$$C_1 \;=\; C \vee [\neg]\, s + t \approx s + t'$$

can be simplified to

$$C_0 \;=\; C \vee [\neg]\, t \approx t'$$

(independently of N). We can also extend the classical simplification by demodulation to cancellative superposition, so that a clause

$$C_1 \;=\; C \vee C' \vee [\neg]\, s + t \approx t'$$

can be simplified to

$$C_0 \;=\; C \vee C' \vee [\neg]\, w' + t \approx t' + w$$

provided that N contains a clause D such that

$$D\sigma \;=\; C' \vee s + w \approx w'$$

where $s \succ_{\mathrm{X}} w$, $s \succ_{\mathrm{X}} w'$, and $C_1 \succ_{\mathrm{C}} D\sigma$. For instance, every inference in Example 3.13 is a simplification of the maximal premise. In particular, this technique can in nearly all cases be used to eliminate any remaining occurrences of the redex in the conclusion of a cancellative superposition inference.

Bachmair and Ganzinger [4] list a number of general simplification techniques such as case analysis or contextual rewriting. These methods can easily be extended to our framework. In this section, we will concentrate on techniques that help to reduce the number of clauses with eligible variables.

Whereas the ordering conditions of our inference rules make cancellative superposition inferences into shielded variables superfluous, cancellative superposition inferences into unshielded variables cannot generally be avoided. As an example, consider the clauses $b + c \approx d$ and $x + c \not\approx d$ with the ordering $b \succ c \succ d$. Since unification is not an effective filter, clauses with eligible variables are extremely prolific. Simplification techniques offer the possibility to remove certain clauses with eligible variables from the clause set, or at least to render inferences with them redundant.

Lemma 5.1 *Let* $C \in N$ *be a clause*

$$\bigvee_{i \in I} n_i x + s_i \approx t_i$$

where x *occurs in neither of the* t_i*. Then every model of* $\mathrm{ACK} \cup N$ *can be extended to a model of* $x + 0 \approx x$ *and* $x + (-x) \approx 0$*, where* 0 *and* $-$ *are new function symbols.*

Proof. Let θ be a substitution that maps all variables in C except x to constants. Let M be an arbitrary model of $\mathrm{ACK} \cup N$, where M_{CAM} is the carrier set of the sort S_{CAM}. Then t_i corresponds to a constant in M_{CAM} and $(n_i - 1)x + s_i$ to a (possibly constant) function from M_{CAM} to M_{CAM}. By Lemma 2.2, M_{CAM} is a group. Hence we can interpret 0 and $-$ appropriately in M so that M becomes a model of $x + 0 \approx x$ and $x + (-x) \approx 0$. □

Lemma 5.2 *Let* C *be a clause*

$$C' \vee \bigvee_{i \in I} n_i x + s_i \approx t_i$$

where x *occurs neither in* C' *nor in the* t_i*. Let* N *be a set of clauses and* 0 *and* $-$ *be new function symbols. Then every model of* $\{C\} \cup \mathrm{ACK}$ *is either a model of* C' *or it can be extended to a model of* $x + 0 \approx x$ *and* $x + (-x) \approx 0$*. In particular,* $N \cup \{C\} \cup \mathrm{ACK}$ *is satisfiable if and only if* $N \cup \{C'\} \cup \mathrm{ACK}$ *or* $N \cup \{C\} \cup \{x + (-x) \approx 0\} \cup \mathrm{ACUK}$ *is.*

Proof. Suppose that there is a model M of $\{C\} \cup \mathrm{ACK}$ that is not a model of C'. Let θ be a substitution with domain $\mathrm{Var}(C')$ such that the ground clause $C'\theta$ is false in M. Consequently,

$$\bigvee_{i \in I} n_i x + s_i\theta \approx t_i\theta$$

must be true in M. By Lemma 5.1, M can be extended to a model of $x + 0 \approx x$ and $x + (-x) \approx 0$. □

In its most general form, we can use this lemma to split one theorem proving derivation into two branches in a tableaux-like manner (cf. [5]). It is particularly useful if one of the two branches can immediately be seen to fail. This happens in two situations: First, if C' is empty, the first branch can be closed immediately. In this case C implies the identity and inverse axioms, and, although it not required by fairness, it may be wise to add them to find an easier proof.[9] Second, if N contains some subset N' that implies that S_{CAM} is not a group, the second branch can be closed immediately. (For instance, N' might consist of the single clause $y + a \not\approx b$.) In this case, C can be simplified to C'. In "non-groups" it is thus always possible to get rid of unshielded variables that occur only positively.

Unshielded variables occurring negatively are somewhat harder to handle. There is a variant of "rewriting with equations of conditions" which can sometimes be applied if an unshielded variable occurs in two different literals. A clause

$$C_1 \;=\; C \vee \neg\, mx + s \approx s' \vee [\neg]\, nx + t \approx t'$$

with $n \geq m \geq 1$ is equivalent to

$$C_0 \;=\; C \vee \neg\, mx + s \approx s' \vee [\neg]\, (n{-}m)x + t + s' \approx t' + s\,.$$

Repeated use of this inference leads to a clause in which x occurs only in one negative literal $kx + w \approx w'$ and possibly in some positive literals (with coefficients smaller than k). Unfortunately, this is not a simplification for all instances of $C_1\theta$, but just for those that satisfy $x\theta \succ_{\mathrm{x}} s\theta$ and $x\theta \succ_{\mathrm{x}} s'\theta$.[10] Hence adding C_0 makes it unnecessary to consider inferences with C_1 that involve only x but no subterm of s or s'. Inferences that involve both x and a subterm of s or s' are still necessary, though.

If an unshielded variable occurs only in one negative literal (and no positive one), we can eliminate it, provided that the inverse axiom has been derived and that the coefficient of the variable is 1. In a group, every clause of the form

$$C \vee \neg\, x + s \approx s'\,,$$

where x doesn't occur in C, can be simplified to C.

What can be done to eliminate unshielded variables that occur negatively with a coefficient k larger than 1, and possibly also positively with coefficients

[9] The reader might try to refute $\mathrm{ACUK} \cup \{\, 2x \approx a \vee x \approx b,\ 2b \not\approx a \,\}$ without first deducing the inverse axiom.

[10] It is not possible to weaken this condition to $x\theta \not\prec_{\mathrm{x}} s\theta$ and $x\theta \not\prec_{\mathrm{x}} s'\theta$. For instance, if C_1 is the clause $2x + 4b \not\approx c \vee 3x \not\approx d$ and $b \succ c \succ d$, then C_0 is $2x + 4b \not\approx c \vee x + c \not\approx d + 4b$, and $C_0\theta$ is larger than $C_1\theta$ if θ maps x to b.

smaller than k? A clause like $2x \not\approx a$ is true in some groups and some non-groups (take the integers or the naturals and $a = 1$), and false in others. To remove such a literal, additional properties of the model class are required, namely, that for every $k \in \mathbf{N}^+$ and u there is a v such that $kv \approx u$.

If a clause has the form

$$C \lor \neg\, kx + s \approx s' \lor nx + t \approx t'$$

with $n < k$, then it is possible to derive

$$C \lor \neg\, kx + s \approx s' \lor kt + ns' \approx kt' + ns\,.$$

In general, this is not a simplification, since the second clause does not imply the first one. To show the equivalence of the two clauses, one needs that $ky \approx kz$ implies $y \approx z$. This property does not hold in $\mathbf{Z}/4$, for instance. On the other hand, it does hold in totally ordered cancellative abelian monoids, if the ordering is compatible with addition. It is still to be investigated to which degree we can exploit this.

6 Conclusions

We have presented a calculus for first-order equational theorem proving in the presence of the axioms of cancellative abelian monoids. The calculus is refutationally complete without requiring extended clauses or explicit inferences with the theory clauses. Compared with the conventional superposition calculus, on which it is based, the ordering constraints are strengthened in such a way that we may not only restrict to inferences that involve the maximal side of the maximal literal, but even to inferences that involve the maximal atomic terms. The calculus may further be furnished with selection functions.

In traditional AC-superposition, extended rules show a rather prolific behaviour. In our approach, AC-unification is replaced by ACU-unification. Furthermore, cancellative superposition makes extended rules superfluous, and the ordering constraints mentioned above allow to exclude inferences involving shielded variables altogether. Many occurrences of unshielded variables can be eliminated by appropriate simplification techniques. Unfortunately cancellative superpositions into variables in sums cannot be completely avoided. More ways to eliminate unshielded variables are possible in the presence of further algebraic structure. This is still a matter of further investigation.

At the time of writing this paper, we cannot yet report about practical experiences with our calculus. An implementation in the Saturate system [13] is under way.

References

[1] Jürgen Avenhaus and Klaus Becker. Conditional rewriting modulo a built-in algebra. SEKI Report SR-92-11, Fachbereich Informatik, Universität Kaiserslautern, 1992.

[2] Leo Bachmair and Harald Ganzinger. Associative-commutative superposition. In Nachum Dershowitz and Naomi Lindenstrauss, editors, *Conditional and Typed Rewrite Systems, 4th International Workshop, CTRS-94*, LNCS 968, pages 1–14, Jerusalem, Israel, July 13–15, 1994. Springer-Verlag.

[3] Leo Bachmair and Harald Ganzinger. Ordered chaining for total orderings. In Alan Bundy, editor, *Twelfth International Conference on Automated Deduction*, LNAI 814, pages 435–450, Nancy, France, June 26–July 1, 1994. Springer-Verlag.

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

[5] Leo Bachmair, Harald Ganzinger, and Uwe Waldmann. Superposition with simplification as a decision procedure for the monadic class with equality. In Georg Gottlob, Alexander Leitsch, and Daniele Mundici, editors, *Computational Logic and Proof Theory, Third Kurt Gödel Colloquium*, LNCS 713, pages 83–96, Brno, Czech Republic, August 24–27, 1993. Springer-Verlag.

[6] Leo Bachmair, Harald Ganzinger, and Uwe Waldmann. Refutational theorem proving for hierarchic first-order theories. *Applicable Algebra in Engineering, Communication and Computing*, 5(3/4):193–212, April 1994.

[7] Leo Bachmair and David A. Plaisted. Termination orderings for associative-commutative rewriting systems. *Journal of Symbolic Computation*, 1:329–349, 1985.

[8] W. W. Bledsoe, K. Kunen, and R. Shostak. Completeness results for inequality provers. *Artificial Intelligence*, 27:255–288, 1985.

[9] R[obert] S. Boyer and J S[trother] Moore. Integrating decision procedures into heuristic theorem provers: A case study of linear arithmetic. In Jean E. Hayes, Donald Michie, and Judith Richards, editors, *Machine*

Intelligence 11: Logic and the acquisition of knowledge, chapter 5, pages 83–124. Oxford University Press, 1988.

[10] Hans-Jürgen Bürckert. A resolution principle for clauses with constraints. In Mark E. Stickel, editor, *10th International Conference on Automated Deduction*, LNAI 449, pages 178–192, Kaiserslautern, FRG, July 24–27, 1990. Springer-Verlag.

[11] Hans-Jürgen Bürckert. *A Resolution Principle for a Logic with Restricted Quantifiers*. LNAI 568. Springer-Verlag, Berlin, Heidelberg, New York, 1991.

[12] Nachum Dershowitz and Jean-Pierre Jouannaud. Rewrite systems. In Jan van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B: Formal Models and Semantics, chapter 6, pages 243–320. Elsevier Science Publishers B.V., Amsterdam, New York, Oxford, Tokyo, 1990.

[13] Harald Ganzinger. The Saturate system. Available on the world-wide web under URL http://www.mpi-sb.mpg.de/SATURATE/Saturate.html, 1994.

[14] Alexander Herold and Jörg H. Siekmann. Unification in Abelian semigroups. *Journal of Automated Reasoning*, 3(3):247–283, 1987.

[15] Larry M. Hines. Completeness of a prover for dense linear logics. *Journal of Automated Reasoning*, 8:45–75, 1992.

[16] Larry M. Hines. Str$\dotplus$ve and integers. In Alan Bundy, editor, *Twelfth International Conference on Automated Deduction*, LNAI 814, pages 416–430, Nancy, France, June 26–July 1, 1994. Springer-Verlag.

[17] Jieh Hsiang, Michaël Rusinowitch, and Ko Sakai. Complete inference rules for the cancellation laws (extended abstract). In John McDermott, editor, *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, volume 2, pages 990–992, Milan, Italy, August 23–28 1987. Morgan Kaufmann Publishers, Inc.

[18] Jean-Pierre Jouannaud and Claude Marché. Termination and completion modulo associativity, commutativity and identity. *Theoretical Computer Science*, 104(1):29–51, October 1992.

[19] Claude Kirchner, Hélène Kirchner, and Michaël Rusinowitch. Deduction with symbolic constraints. *Revue Française d'Intelligence Artificielle*, 4(3):9–52, 1990. Special issue on automatic deduction.

[20] Claude Marché. Normalised narrowing and normalised completion. In *Ninth Annual IEEE Symposium on Logic in Computer Science*, pages 394–403, Paris, France, July 4–7, 1994. IEEE Computer Society Press, Los Alamitos, CA, USA.

[21] Robert Nieuwenhuis and Albert Rubio. AC-superposition with constraints: no AC-unifiers needed. In Alan Bundy, editor, *Twelfth International Conference on Automated Deduction*, LNAI 814, pages 545–559, Nancy, France, June 26–July 1, 1994. Springer-Verlag.

[22] S. Owre, J. M. Rushby, and N. Shankar. PVS: A prototype verification system. In Deepak Kapur, editor, *11th International Conference on Automated Deduction*, LNAI 607, pages 748–752, Saratoga Springs, New York, USA, June 15–18, 1992. Springer-Verlag.

[23] Gerald E. Peterson and Mark E. Stickel. Complete sets of reductions for some equational theories. *Journal of the ACM*, 28(2):233–264, April 1981.

[24] M. M. Richter. Some reordering properties for inequality proof trees. In E. Börger, G. Hasenjäger, and D. Rödding, editors, *Logic and Machines: Decision Problems and Complexity*, LNCS 171, pages 183–197. Springer-Verlag, 1984.

[25] Albert Rubio and Robert Nieuwenhuis. A precedence-based total AC-compatible ordering. In Claude Kirchner, editor, *Rewriting Techniques and Applications, 5th International Conference, RTA-93*, LNCS 690, pages 374–388, Montreal, Canada, June 16–18, 1993. Springer-Verlag.

[26] Michaël Rusinowitch. *Démonstration Automatique: Techniques de Réécriture*, chapitre 7: Ensembles complets de règles d'inférence pour les axiomes de régularité, pages 111–127. InterEditions, Paris, 1989.

[27] Jürgen Stuber. Superposition theorem proving for integer modules. To be submitted, 1996.

[28] Laurent Vigneron. Associative-commutative deduction with constraints. In Alan Bundy, editor, *Twelfth International Conference on Automated Deduction*, LNAI 814, pages 530–544, Nancy, France, June 26–July 1, 1994. Springer-Verlag.

[29] Ulrich Wertz. First-order theorem proving modulo equations. Technical Report MPI-I-92-216, Max-Planck-Institut für Informatik, Saarbrücken, Germany, April 1992.

INFORMATIK

Below you find a list of the most recent technical reports of the research group *Logic of Programming* at the Max-Planck-Institut für Informatik. They are available by anonymous ftp from our ftp server `ftp.mpi-sb.mpg.de` under the directory `pub/papers/reports`. Most of the reports are also accessible via WWW using the URL `http://www.mpi-sb.mpg.de`. If you have any questions concerning ftp or WWW access, please contact `reports@mpi-sb.mpg.de`. Paper copies (which are not necessarily free of charge) can be ordered either by regular mail or by e-mail at the address below.

Max-Planck-Institut für Informatik
Library
attn. Birgit Hofmann
Im Stadtwald
D-66123 Saarbrücken
GERMANY
e-mail: `library@mpi-sb.mpg.de`

MPI-I-96-2-001	H. Ganzinger, U. Waldmann	Theorem Proving in Cancellative Abelian Monoids
MPI-I-95-2-011	P. Barth, A. Bockmayr	Modelling Mixed-Integer Optimisation Problems in Constraint Logic Programming
MPI-I-95-2-010	D. A. Plaisted	Special Cases and Substitutes for Rigid E-Unification
MPI-I-95-2-009	L. Bachmair, H. Ganzinger	Ordered Chaining Calculi for First-Order Theories of Binary Relations
MPI-I-95-2-008	H. J. Ohlbach, R. A.Schmidt, U. Hustadt	Translating Graded Modalities into Predicate Logic
MPI-I-95-2-007	A. Nonnengart, A. Szalas	A Fixpoint Approach to Second-Order Quantifier Elimination with Applications to Correspondence Theory
MPI-I-95-2-006	D. Basin, H. Ganzinger	Automated Complexity Analysis Based on Ordered Resolution
MPI-I-95-2-005	F. Baader, H. J. Ohlbach	A Multi-Dimensional Terminological Knowledge Representation Language
MPI-I-95-2-003	P. Barth	A Davis-Putnam Based Enumeration Algorithm for Linear Pseudo-Boolean Optimization
MPI-I-95-2-002	H. J. Ohlbach, R. A. Schmidt	Functional Translation and Second-Order Frame Properties of Modal Logics
MPI-I-95-2-001	S. Vorobyov	Proof normalization and subject reduction in extensions of Fsub
MPI-I-94-261	P. Barth, A. Bockmayr	Finite Domain and Cutting Plane Techniques in CLP($\mathcal{PB}$)
MPI-I-94-257	S. Vorobyov	Structural Decidable Extensions of Bounded Quantification
MPI-I-94-254	P. Madden	Report and abstract not published
MPI-I-94-252	P. Madden	A Survey of Program Transformation With Special Reference to *Unfold/Fold* Style Program Development
MPI-I-94-251	P. Graf	Substitution Tree Indexing
MPI-I-94-246	M. Hanus	On Extra Variables in (Equational) Logic Programming