Draft of "Cancellative Abelian Monoids and Related Structures in Refutational Theorem Proving, Part I and II", to appear in *Journal of Symbolic Computation*

Cancellative Abelian Monoids and Related Structures in Refutational Theorem Proving

Uwe Waldmann

MPI für Informatik
Stuhlsatzenhausweg 85
66123 Saarbrücken
Germany
e-mail: uwe@mpi-sb.mpg.de

September 5, 2001

Abstract: We present superposition calculi in which the axioms of cancellative abelian monoids and, optionally, further axioms (e. g., the inverse or the torsion-freeness axiom) are integrated. Cancellative abelian monoids comprise abelian groups, but also such ubiquitous structures as the natural numbers or multisets. Our calculi require neither extended clauses nor explicit inferences with the theory axioms. The number of variable overlaps is significantly reduced by strong ordering restrictions and powerful variable elimination techniques; in divisible torsion-free abelian groups, variable overlaps, AC-unification and AC-orderings can even be avoided completely.

1. Introduction

If we want to tackle real life problems with an automated theorem prover, the prover must operate in a heterogeneous world. In fields of application like program verification, it has to deal with uninterpreted function and predicate symbols that are specific for a particular domain, as well as with standard algebraic structures or theories, such as the natural numbers, or abelian groups, or orderings. Unfortunately, axioms like associativity or commutativity are difficult for a general-purpose theorem prover, as they allow a huge number of inferences and tend to generate numerous equivalent formulae. A sophisticated treatment of the standard theories is therefore crucial to the performance of the prover. For this purpose, mathematical and meta-mathematical techniques have to be combined.

There have been some attempts to integrate (fragments of) first-order logic into mathematical systems, for instance by Shostak (1979), who has demonstrated that decision procedures for universally-quantified Presburger arithmetic can be extended to universally-quantified function symbols. A proposal to build a theorem prover within a computer algebra system can be found in (Buchberger 1996). More often, however, the problem has been tackled from the other side, by integrating mathematical knowledge into a general-purpose theorem prover.

It is difficult to couple a decision procedure for a decidable theory to a prover as a black box: The requirement of sufficient completeness (Bachmair, Ganzinger, and Waldmann 1994) practically excludes uninterpreted function symbols; and even in situations where sufficient completeness is not a too restrictive requirement, insufficient communication between the general prover and the external decision procedure makes the latter almost useless (Boyer and Moore 1988). Consequently, the integration has to be achieved on the level of the inference system and simplification techniques.

Paramodulation and superposition calculi illustrate the advantages of such an integration (Peterson 1983, Zhang and Kapur 1988, Rusinowitch 1991, Hsiang and Rusinowitch 1991, Bachmair and Ganzinger 1994c). These calculi can be seen as the result of building-in the equality axioms into resolution. Hence resolution inferences with the equality axioms become unnecessary, besides, paramodulation inferences at or below variables can be shown to be superfluous. Furthermore, compared with ordered resolution using explicit equality axioms, the ordering restrictions and the redundancy criterion can be strengthened significantly.

Similar techniques can be used for other theories. The integration of the associativity and commutativity axioms into the paramodulation calculus has been investigated already by Plotkin (1972) and Slagle (1974). Paul (1992), Rusinowitch and Vigneron (1995), Wertz (1992), and Bachmair and Ganzinger (1994a) have built associativity and commutativity into superposition using AC-unification and extended clauses, developed for the equational case by Peterson and Stickel (1981). In this way, inferences with the AC axioms become superfluous. New sources of inefficiency emerge, however, as a minimal complete set of AC-unifiers may have doubly exponential size. Using constraints, the enumeration of unifiers can be avoided (Nieuwenhuis and Rubio 1994, Vigneron 1994); one still has to solve the unifiability problem, though, which is NP-complete (Kapur and Narendran 1992a).

The problem can be mitigated by integrating more axioms. If our theory contains also the identity law, then AC-unification can be replaced by ACU-unification (Boudet, Contejean, and Marché 1996, Jouannaud and Marché 1992), which is only simply exponential (Kapur and Narendran 1992b), and even unitary for the special case that sums of variables are to be unified. We observe a much more radical improvement when we switch over from the abelian semigroup axioms AC or the abelian monoid axioms ACU to the axioms of abelian groups. To see the operational difference between these theories in a rewrite or superposition-based calculus, consider the two unit clauses $u_1 + \cdots + u_k \approx s$ and $v_1 + \cdots + v_l \approx t$. In AC-superposition, there is an inference between these two clauses (via extended clauses), whenever *some* u_i is unifiable with *some* v_j. In the presence of the inverse axiom, extended rules become superfluous, and the number of AC-unifications (and unifiers) is dramatically reduced, as an inference is required only if the *maximal* u_i is unifiable with the *maximal* v_j. This technique can be found for instance in normalized rewriting (Marché 1996) and in the superposition calculi for abelian groups presented by Stuber (1996) and Godoy and Nieuwenhuis (2000); the latter calculus furthermore uses unification modulo abelian groups (which is even more efficient than ACU-unification).

In groups the difference of any two elements exists and is unique. Operationally, the second property is much more important than the first one. We can thus employ similar techniques for abelian monoids satisfying the cancellation axiom $x + y \not\approx x + z \vee y \approx z$. Cancellative abelian monoids are in some sense the most general algebraic structure where such an "abelian-group-like" reasoning is possible. They comprise abelian groups, but also such ubiquitous structures as the natural numbers or multisets. In this paper, we present a refined superposition calculus for sets of clauses including the axioms of cancellative abelian monoids. As in the abelian group calculi above, ordering restrictions can be strengthened and explicit inferences with the theory axioms and extended clauses are superfluous. In particular, the restriction to overlaps of maximal summands in maximal sides of maximal literals excludes superpositions with shielded variables. While inferences with unshielded variables cannot generally be avoided, the number of unshielded variables can be reduced using suitable simplification techniques. Furthermore our calculus offers the possibility to integrate the torsion-freeness axioms $\psi x \not\approx \psi y \vee x \approx y$ for all $\psi \in \mathbf{N}^{>0}$ (or even generalized forms of torsion-freeness, where ψ ranges over some subset $\Psi \subseteq \mathbf{N}^{>0}$).

In divisible torsion-free abelian groups (e. g., the rational numbers), the abelian group axioms are extended by the torsion-freeness axioms, the divisibility axioms $\forall x\ \exists y\colon ky \approx x$ for all $k \in \mathbf{N}^{>0}$, and the non-triviality axiom $\exists y\colon y \not\approx 0$. In such structures every clause can be transformed into an equivalent clause without unshielded variables. When we try to integrate variable elimination into the superposition calculus, however, a typical discrepancy between computer algebra methods and first-order theorem proving methods becomes visible. In the variable elimination algorithm, the clauses are ordered with respect to the number of their variables: the less variables a clause has, the smaller it is. In superposition and related calculi, variables are objects that may be instantiated with small or with large terms, and the clause ordering has to be stable under instantiation. Consequently, the variable elimination algorithm is not necessarily a simplification in the superposition calculus: some ground instances of the transformed clause may be too large. It turns out, however, that all the critical instances can be handled by case analysis. The resulting calculus requires neither variable overlaps nor explicit inferences with the theory axioms. If clauses are fully abstracted eagerly, even AC-unifications can be avoided.

The outline of this work is as follows: In Section 2, we provide the logical prerequisites of this work, recapitulate some properties of saturation-based theorem proving, and fix the necessary notation. In Section 3, we present and explain the inference rules of the cancellative superposition calculus. The refutational completeness proof for this calculus, which follows in Section 4, is based on the model construction technique of Bachmair and Ganzinger (1994c), using a novel kind of rewriting on equations. Section 5 starts with a presentation of simplification techniques. We discuss several refinements of the cancellative superposition calculus and its use as a decision procedure. Divisible torsion-free abelian groups are treated in Section 6; the conclusions follow in Section 7. In Appendix A, we provide some algebraic background information concerning cancellative (abelian) semigroups and monoids and their relationship to abelian groups.

This paper is a revised version of (Waldmann 1997). A precursor of the cancellative superposition calculus that did not include torsion-freeness has been sketched in (Ganzinger and Waldmann 1996b). Note that the model construction described there cannot be extended to the torsion-free case. The results of Section 6 have been presented in (Waldmann 1998b). For lack of space, these two papers did not include formal proofs.

2. Saturation-Based Theorem Proving

2.1. Logical Foundations

We start this section by briefly summarizing the logical foundations of refutational first-order theorem proving. A more detailed introduction can be found in (Fitting 1990). Some differences between Fitting's presentation and ours are due to the fact that we develop our calculus not in a single-sorted but in a many-sorted framework (without subsorts or overloading) and that we restrict ourselves to clauses over the single predicate symbol $\approx$, rather than dealing with arbitrary first-order formulae over arbitrary sets of predicates.

We assume a signature $(\mathcal{S}, \Sigma)$ consisting of a set of sorts $\mathcal{S}$ and a set of function symbols Σ, and a set of variables $\mathcal{V}$. The sets Σ and $\mathcal{V}$ are disjoint. Every function symbol $f \in \Sigma$ comes with a unique arity $n \in \mathbf{N}$ and a unique declaration $f : S_1 \dots S_n \to S_0$,[1] every variable $x \in \mathcal{V}$ comes with a unique declaration $x : S_0$, where $S_0, \dots, S_n \in \mathcal{S}$. The set of terms of sort S is the least set containing x whenever $x : S \in \mathcal{V}$, and containing $f(t_1, \dots, t_n)$ whenever each t_i is a term of sort S_i and $f : S_1 \dots S_n \to S \in \Sigma$. Throughout this paper we assume that function symbols and variables are declared appropriately such that all syntactic objects (terms, equations, etc.) are well-formed.

The set of variables occurring in a syntactic object Q is denoted by $\mathrm{Var}(Q)$. If $\mathrm{Var}(Q)$ is empty, then Q is called ground. We require that for every sort there exist infinitely many variables and at least one ground term (that is, that every sort is inhabited).

An equation e is an ordered pair of terms, usually written as $t \approx t'$, where t and t' have the same sort. The left-hand and right-hand side of e are denoted by $\mathrm{lhs}(e)$ and $\mathrm{rhs}(e)$.

To simplify the presentation we confine ourselves to equality as the only predicate of our logical language. This does not restrict its expressivity: A predicate P different from $\approx$ can be coded using a function symbol p, so that $P(t_1, \dots, t_n)$ is to be taken as an abbreviation for $p(t_1, \dots, t_n) \approx \mathrm{true}_p$, where $p(t_1, \dots, t_n)$ and true_p have a new sort S_p.

A literal is either an equation e (also called a positive literal) or a negated equation $\neg\, e$ (also called a negative literal). A clause is a finite multiset of literals, usually written as a disjunction. The symbol $[\neg]\, e$ denotes either e or $\neg\, e$. Instead of $\neg\, t \approx t'$, we sometimes write $t \not\approx t'$. The submultiset of all negative literals of a clause C is abbreviated by $\mathrm{neg}(C)$. The symbol $\bot$ denotes the empty clause, that is, the empty multiset of literals.

A substitution σ is a sort-preserving mapping from $\mathcal{V}$ into the set of terms over Σ and $\mathcal{V}$. Substitutions are homomorphically extended to terms, and likewise to equations, literals, or clauses. We use postfix notation for substitutions and write $t\sigma$ instead of $\sigma(t)$; $\sigma\sigma'$ is the substitution that maps every x to $(x\sigma)\sigma'$. A syntactic object Q' is called an instance of an object Q, if $Q\sigma = Q'$ for some substitution σ. For a set N of clauses, the set of all ground instances of clauses in N is denoted by $\overline{N}$.

The set $\mathrm{Dom}(\sigma) = \{\, x \in \mathcal{V} \mid x\sigma \neq x \,\}$ is called the domain of the substitution σ, $\mathrm{Ran}(\sigma) = \bigcup_{x \in \mathrm{Dom}(\sigma)} \mathrm{Var}(x\sigma)$ is called its range. A substitution with domain $\{x_1, \dots, x_n\}$ that maps the variables $x_1, \dots, x_n$ to the terms $t_1, \dots, t_n$, respectively, is denoted by $\{x_1 \mapsto t_1, \dots, x_n \mapsto t_n\}$. A substitution σ is said to be idempotent, if $\sigma\sigma = \sigma$, that is, if $\mathrm{Dom}(\sigma) \cap \mathrm{Ran}(\sigma) = \emptyset$. If σ and σ' are substitutions and $\mathcal{V}' \subseteq \mathcal{V}$, we say that $\sigma = \sigma'$ over $\mathcal{V}'$ if $x\sigma = x\sigma'$ for every $x \in \mathcal{V}'$.

An interpretation $\mathfrak{M}$ for the signature $(\mathcal{S}, \Sigma)$ is a mapping that assigns to every sort $S \in \mathcal{S}$ a non-empty set $S^{\mathfrak{M}}$, to every function symbol $f : S_1 \dots S_n \to S_0 \in \Sigma$ a function $f^{\mathfrak{M}} : S_1^{\mathfrak{M}} \times \dots \times S_n^{\mathfrak{M}} \to S_0^{\mathfrak{M}}$, and to the equality predicate $\approx$ a binary relation $\approx^{\mathfrak{M}} \subseteq \bigcup_{S \in \mathcal{S}} S^{\mathfrak{M}} \times S^{\mathfrak{M}}$. We assume that the sets $S_1^{\mathfrak{M}}$ and $S_2^{\mathfrak{M}}$ are disjoint for any $S_1, S_2 \in \mathcal{S}$, $S_1 \neq S_2$. The union $\bigcup_{S \in \mathcal{S}} S^{\mathfrak{M}}$ is called the domain of the interpretation. An $\mathfrak{M}$-assignment α is a sort-preserving mapping from the set of variables $\mathcal{V}$ into the domain of $\mathfrak{M}$. Assignments can be homomorphically extended to terms over Σ and $\mathcal{V}$.

[1] This includes constant declarations $b : \to S_0$. The set of natural numbers (starting with 0) is denoted by $\mathbf{N}$, the set of positive integers (starting with 1) by $\mathbf{N}^{>0}$.

A positive literal $t \approx t'$ is called true with respect to $\mathfrak{M}$ and α if $\alpha(t) \approx^{\mathfrak{M}} \alpha(t')$, A negative literal $\neg\, t \approx t'$ is called true with respect to $\mathfrak{M}$ and α if $\alpha(t) \not\approx^{\mathfrak{M}} \alpha(t')$.

A clause C is called true with respect to $\mathfrak{M}$ and α if at least one of its literals is true. An interpretation $\mathfrak{M}$ is a model of C, if C is true with respect to $\mathfrak{M}$ and α for every $\mathfrak{M}$-assignment α. It is a model of a set N of clauses, if it is a model of every $C \in N$. If $\mathfrak{M}$ is a model of N, we also say that it satisfies N. A set of clauses is called satisfiable if it has a model. Obviously every set of clauses containing $\bot$ is unsatisfiable.

In refutational theorem proving, one is primarily interested in the question whether or not a given set of clauses is satisfiable. For this purpose we may confine ourselves to term-generated interpretations, that is, to interpretations $\mathfrak{M}$ where every element of some $S^{\mathfrak{M}}$ is the image of some ground term of sort S.[2] We may even confine ourselves to Herbrand interpretations, that is, to term-generated interpretations whose domain is the set of ground terms, and where every ground term is interpreted by itself: A set of clauses has a model, if and only if it has a term-generated model, if and only if it has a Herbrand model.

As long as we restrict ourselves to term-generated models we may think of a non-ground clause as a finite representation of the set of all its ground instances: A term-generated interpretation is a model of a clause C if and only it is a model of all ground instances of C.

Every Herbrand interpretation is completely characterized by the interpretation $\approx^{\mathfrak{M}}$ of the equality predicate $\approx$. For any set $E_{\mathfrak{M}}$ of ground equations there is exactly one Herbrand interpretation $\mathfrak{M}$ in which the equations in $E_{\mathfrak{M}}$ are true and all other ground equations are false. We will usually identify $\mathfrak{M}$ and $E_{\mathfrak{M}}$. A positive ground literal e is thus true in $E_{\mathfrak{M}}$, if $e \in E_{\mathfrak{M}}$; a negative ground literal $\neg\, e$ is true in $E_{\mathfrak{M}}$, if $e \notin E_{\mathfrak{M}}$.

When one uses the equality symbol $\approx$ in a logical language, one is commonly interested in interpretations $\mathfrak{M}$ in which $\approx^{\mathfrak{M}}$ is not an arbitrary binary relation but actually the equality relation on the domain of $\mathfrak{M}$. We refer to such interpretations as normal. A normal interpretation that is a model of a set N of clauses is called a normal model of N.

If we want to recover the intuitive semantics of the equality symbol while working with Herbrand interpretations, we have to encode the intended properties of the equality symbol explicitly. The clauses

$$x \approx x \qquad \text{(Reflexivity)}$$

$$x \not\approx y \,\vee\, y \approx x \qquad \text{(Symmetry)}$$

$$x \not\approx y \,\vee\, y \not\approx z \,\vee\, x \approx z \qquad \text{(Transitivity)}$$

$$x_1 \not\approx y_1 \,\vee\, \ldots \,\vee\, x_n \not\approx y_n \,\vee\, f(x_1, \ldots, x_n) \approx f(y_1, \ldots, y_n) \qquad \text{(Congruence)}$$

(for every n-ary function symbol $f \in \Sigma$) are called equality axioms. If N is a set of clauses, then an interpretation that is a model of N and of the equality axioms is called an equality model of N. A set of clauses has a normal model, if and only if it has a term-generated normal model, if and only if it has an equality Herbrand model.

Let N and N' be sets of clauses. If every equality Herbrand model of N is a model of N', we say that N entails N' modulo equality and denote this by $N \models_{\approx} N'$.

In the rest of the paper, we will almost exclusively work with (equality) Herbrand interpretations and models, or more precisely, with the set $E_{\mathfrak{M}}$ of equations correspond-

[2] Recall that we require every sort to be inhabited, so the sets $S^{\mathfrak{M}}$ of a term-generated interpretation are in fact non-empty.

ing to a Herbrand interpretation $\mathfrak{M}$. For simplicity, we will usually drop the attribute "Herbrand". The dualism between term-generated normal models and equality Herbrand models will only be exploited in Subsection 5.2.

2.2. Rewrite Systems

To prove the completeness of our calculus, we have to construct Herbrand interpretations and to check whether a given equation is contained in such an interpretation. Rewriting techniques are our main tool for this task. The rest of this subsection serves mainly to fix the necessary notation; for more detailed information about rewrite systems we refer the reader to Dershowitz and Jouannaud (1990).

As usual, positions (also known as occurrences) of a term are denoted by strings of natural numbers. The set of all positions of a term t is $\mathrm{pos}(t)$. If o is a position of t, then $t|_o$ is the subterm of t at o, $t(o)$ is the function symbol of t at o, and $t[t']_o$ is the result of the replacement of the subterm at o in t by t'. We write $t[t']$ if o is clear from the context.

A binary relation $\rightarrow$ is called a rewrite relation, if it is stable under substitutions and contexts, that is, if $t_1 \rightarrow t_2$ implies $t_1\sigma \rightarrow t_2\sigma$ and $s[t_1]_o \rightarrow s[t_2]_o$ for all terms t_1, t_2, and s, such that $s[t_1]_o$ and $s[t_2]_o$ are well-formed, and for all substitutions σ.

For a binary relation $\rightarrow$, we commonly use the symbol $\leftarrow$ for its inverse relation, $\leftrightarrow$ for its symmetric closure, $\rightarrow^+$ for its transitive closure, and $\rightarrow^*$ for its reflexive-transitive closure (and thus $\leftrightarrow^*$ for its reflexive-symmetric-transitive closure).

A binary relation $\rightarrow$ is called noetherian (or terminating), if there is no infinite chain $t_1 \rightarrow t_2 \rightarrow t_3 \rightarrow \cdots$. We say that t is a normal form (or irreducible) with respect to $\rightarrow$ if there is no t' such that $t \rightarrow t'$; t is called a normal form of s if $s \rightarrow^* t$ and t is a normal form. We say that $\rightarrow \subseteq \Pi \times \Pi$ is confluent on $\Pi' \subseteq \Pi$, if for every $t_0 \in \Pi'$ and $t_1, t_2 \in \Pi$ such that $t_1 \leftarrow^* t_0 \rightarrow^* t_2$ there exists a $t_3 \in \Pi$ such that $t_1 \rightarrow^* t_3 \leftarrow^* t_2$; the relation $\rightarrow$ is called confluent, if it is confluent on its carrier set Π.

A transitive and irreflexive binary relation $\succ$ is called an ordering. An ordering on terms is called a reduction ordering, if it is a noetherian rewrite relation. We use the symbol $\succeq$ to denote the reflexive closure of an ordering $\succ$. If $(\Pi_0, \succ)$ is an ordered set, $\Pi \subseteq \Pi_0$, and $s \in \Pi_0$, then $\Pi^{\prec s}$ is an abbreviation for $\{\, t \in \Pi \mid t \prec s \,\}$.

In the sequel, we will need the following variation on the familiar "diamond lemma".

Lemma 2.1. *Let $\succ$ be a noetherian ordering over Π, let $\rightarrow\ \subseteq\ \succ$. Let s and r be two elements of Π, such that r is irreducible with respect to $\rightarrow$ and define $\Pi_r^s = \{\, t \in \Pi \mid s \succeq t,$ $t \rightarrow^* r \,\}$. If for every $t_0, t_1, t_2 \in \Pi$ such that $s \succeq t_0$ and $t_1 \leftarrow t_0 \rightarrow t_2 \rightarrow^* r$ there exists a $t_3 \in \Pi$ such that $t_1 \rightarrow^* t_3 \leftarrow^* t_2$, then $\rightarrow$ is confluent on Π_r^s and Π_r^s is closed under $\rightarrow$.*

Proof. It is obviously sufficient to prove that for every $t_0 \in \Pi_r^s$ and $t_1' \in \Pi$, $t_0 \rightarrow^* t_1'$ implies $t_1' \in \Pi_r^s$. We use noetherian induction over the size of t_0: Let $t_0 \in \Pi_r^s$ and $t_1' \in \Pi$ such that $t_0 \rightarrow^* t_1'$. If this derivation is empty, there is nothing to show, so suppose that $t_0 \rightarrow t_1 \rightarrow^* t_1'$. As $t_0 \in \Pi_r^s$ is reducible, it is different from r, hence there is a non-empty derivation $t_0 \rightarrow t_2 \rightarrow^* r$. By assumption, there exists a $t_3 \in \Pi$ such that $t_1 \rightarrow^* t_3 \leftarrow^* t_2$. Now $t_0 \succ t_2$ and $t_2 \in \Pi_r^s$, hence by the induction hypothesis, $t_3 \in \Pi_r^s$ and thus $t_1 \in \Pi_r^s$. Since $t_0 \succ t_1$, we can use the induction hypothesis once more and obtain $t_1' \in \Pi_r^s$. $\square$

A rewrite rule e is a pair (t, t') of terms, usually written as $t \rightarrow t'$, where t and t' have the same sort. A rewrite system is a set of rewrite rules. If R is a rewrite system, then the rewrite relation $\rightarrow_R$ associated with R is the smallest rewrite relation containing $t \rightarrow_R t'$ for every rule $t \rightarrow t' \in R$.

2.3. Saturation and Redundancy

Most automated theorem provers for first-order logic are refutational provers. To show that a formula C' follows from a formula C, they try to refute $C \wedge \neg C'$. Often the formula $C \wedge \neg C'$ is further normalized, for instance by skolemization and transformation into clause form. The problem to prove arbitrary theorems is thus reduced to the problem to refute sets of clauses. The prover is called refutationally complete, if it finds a refutation whenever the input is contradictory.

Theorem proving methods such as resolution or superposition aim at deducing a contradiction from a set of formulae by recursively inferring new formulae from given ones. The deductive inference system that computes these new formulae is the central part of a saturation-based theorem prover. We may think of an inference system as a function *Inf* that maps a set N of formulae to a set of inferences

$$\iota \;=\; \frac{C_k \;\ldots\; C_1}{C_0}\;,$$

where $\{C_1, \ldots, C_k\} \subseteq N$. The formulae $C_k, \ldots, C_1$ are called premises of ι. The formula C_0 is called the conclusion and is denoted by $\mathrm{concl}(\iota)$. Typically, an inference system is sound with respect to a given semantical consequence relation $\models$, that is, $\{C_1, \ldots, C_k\} \models \{C_0\}$ for all inferences ι. The consequence relation $\models$ may for example be the relation $\models_{\approx}$ of Subsection 2.1, or another relation with the properties (i) $N_1 \cup N_2 \models N_1$, (ii) if $N_1 \models N_2$ and $N_1 \models N_3$, then $N_1 \models N_2 \cup N_3$, and (iii) if $N_1 \models N_2$ and $N_2 \models N_3$, then $N_1 \models N_3$.

A theorem prover computes one of the possible inferences of the current set of formulae and adds its conclusion to the current set, until a "saturated" set N^* is reached, where the conclusion of every inference in $\mathit{Inf}(N^*)$ is already contained in N^*. The concept of saturation allows us to define refutational completeness as a static property, rather than a dynamic one: We say that *Inf* is refutationally complete, if a saturated set is unsatisfiable if and only if it contains a contradictory formula, say the empty clause $\bot$.

To keep the search space as small as possible, inference rules are equipped with strong local restrictions. Nevertheless, most of the generated formulae are not actually needed for deriving a contradiction, and saturated sets tend to be very large, often infinite. For that reason, we introduce a global concept of redundancy that allows us to weaken the notion of saturation and to discard useless formulae. Let $\mathit{Red}^{\mathrm{C}}$ be a mapping from sets of formulae to sets of formulae and $\mathit{Red}^{\mathrm{I}}$ be a mapping from sets of formulae to sets of inferences. The sets $\mathit{Red}^{\mathrm{C}}(N)$ and $\mathit{Red}^{\mathrm{I}}(N)$ specify formulae and inferences considered unnecessary in the context of a given set N. For instance, $\mathit{Red}^{\mathrm{C}}(N)$ may consist of all tautologies and formulae subsumed by N. Formulae in $\mathit{Red}^{\mathrm{C}}(N)$ may be removed from N, while inferences in $\mathit{Red}^{\mathrm{I}}(N)$ may be ignored. We emphasize that $\mathit{Red}^{\mathrm{C}}(N)$ need not be a subset of N and that $\mathit{Red}^{\mathrm{I}}(N)$ will usually also contain inferences whose premises are not in N.

Definition 2.2. *A pair $\mathit{Red} = (\mathit{Red}^{\mathrm{I}}, \mathit{Red}^{\mathrm{C}})$ is called a redundancy criterion (with respect to an inference system Inf and a consequence relation $\models$), if the following conditions are satisfied for all sets of formulae N and N':*

(i) $N \setminus Red^{\mathrm{C}}(N) \models Red^{\mathrm{C}}(N)$.
(ii) *If* $N \subseteq N'$, *then* $Red^{\mathrm{C}}(N) \subseteq Red^{\mathrm{C}}(N')$ *and* $Red^{\mathrm{I}}(N) \subseteq Red^{\mathrm{I}}(N')$.
(iii) *If* $N' \subseteq Red^{\mathrm{C}}(N)$, *then* $Red^{\mathrm{C}}(N) \subseteq Red^{\mathrm{C}}(N \setminus N')$ *and* $Red^{\mathrm{I}}(N) \subseteq Red^{\mathrm{I}}(N \setminus N')$.
(iv) *If* $\iota \in Inf(N')$ *and* $\mathrm{concl}(\iota) \in N$, *then* $\iota \in Red^{\mathrm{I}}(N)$.

Inferences in $Red^{\mathrm{I}}(N)$ *and formulae in* $Red^{\mathrm{C}}(N)$ *are called redundant with respect to* N.

Condition (i) requires that redundant formulae logically follow from the non-redundant ones. Conditions (ii) and (iii) indicate that redundant formulae and inferences must remain redundant if formulae are added or if redundant formulae are deleted.[3] Finally, condition (iv) states that an inference is redundant with respect to N if its conclusion is already present in N (regardless of whether or not the premises are in N).

DEFINITION 2.3. *A binary relation* $\vdash$ *on sets of formulae is called a derivation relation (with respect to an inference system Inf, a redundancy criterion Red, and a consequence relation* $\models$*), if the following conditions are satisfied for all sets of formulae* N *and* N'*:*

(i) *If* $N \vdash N'$, *then* $N \models N'$.
(ii) *If* $N \vdash N'$, *then* $N \setminus N' \subseteq Red^{\mathrm{C}}(N')$.
(iii) *If* $\iota \in Inf(N)$, *then* $N \vdash N \cup \{\mathrm{concl}(\iota)\}$.

The triple $(Inf, Red, \vdash)$ *is called a theorem proving calculus.*

Note that $N \vdash N'$ implies $N' \models N$ by condition (i) of Def. 2.2.

In a theorem proving calculus, we are not only allowed to add the conclusions of inferences to the current set of formulae; we may also add lemmas (provided they follow logically from the old formulae) or delete redundant formulae. Furthermore, if formulae do not become redundant accidentally, we still have the chance to *make* them redundant. This process is called simplification.

DEFINITION 2.4. *Let* N *be a set of clauses. We say that* $M \subseteq N$ *is simplified to a set* M' *of clauses, if* $N \models M'$ *and if* M *is redundant with respect to* $N \cup M'$.

LEMMA 2.5. *If* $M \subseteq N$ *is simplified to* M', *then* $N \vdash (N \cup M') \setminus M$ *is an admissible derivation step according to Def. 2.3.*

PROOF. As $N \models M'$, we have $N \models (N \cup M') \setminus M$, so condition (i) of Def. 2.3 is satisfied. To prove condition (ii), we note that $N \setminus ((N \cup M') \setminus M) = M$, and $M \subseteq Red^{\mathrm{C}}(N \cup M')$ by assumption, and $Red^{\mathrm{C}}(N \cup M') \subseteq Red^{\mathrm{C}}((N \cup M') \setminus M)$ by part (iii) of Def. 2.2. □

DEFINITION 2.6. *A set* N *of formulae is called saturated with respect to a theorem proving calculus* $(Inf, Red, \vdash)$, *if* $Inf(N) \subseteq Red^{\mathrm{I}}(N)$.

A finite or infinite sequence $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$ is called an $(Inf, Red, \vdash)$-derivation, or simply derivation, if the theorem proving calculus is clear from the context. The set

[3] These conditions are slightly stronger than those of (Bachmair, Ganzinger, and Waldmann 1994).

$N_\infty = \bigcup_i \bigcap_{j \geq i} N_j$ of all persisting formulae is called the limit of the derivation. In particular, the limit of a finite sequence $N_0 \vdash^* N_k$ equals N_k.

A derivation $N_0 \vdash N_1 \vdash \ldots$ is called fair, if $\mathit{Inf}(N_\infty) \subseteq \bigcup_j \mathit{Red}^{\mathrm{I}}(N_j)$. If a derivation is fair, then its limit is saturated (Waldmann 1997).

Definition 2.7. *A theorem proving calculus* $(\mathit{Inf}, \mathit{Red}, \vdash)$ *is called refutationally complete, if for every saturated set* N *of formulae we have* $N \models \{\bot\}$ *if and only if* $\bot \in N$.

Like the notion of saturation, most results on theorem proving calculi do not depend on the particular choice of a derivation relation. In such cases, we will omit the derivation relation and write $(\mathit{Inf}, \mathit{Red})$ instead of $(\mathit{Inf}, \mathit{Red}, \vdash)$.

3. Cancellative Superposition

3.1. Preliminaries

Throughout the paper we assume that the set of sorts $\mathcal{S}$ contains a sort S_{CAM} and that Σ contains function symbols 0 and + with the declarations $0 : \to S_{\mathrm{CAM}}$ and $+ : S_{\mathrm{CAM}} S_{\mathrm{CAM}} \to S_{\mathrm{CAM}}$. There is no scalar multiplication in our signature; if t is a term of sort S_{CAM} and $m \in \mathbf{N}$, then mt is merely an abbreviation for the m-fold sum $t + \cdots + t$. (As usual we define $0t = 0$ and $1t = t$.)

Definition 3.1. *Let* $\Psi \subseteq \mathbf{N}^{>0}$. *The clauses*

$$(x + y) + z \approx x + (y + z) \qquad \text{(Associativity)}$$

$$x + y \approx y + x \qquad \text{(Commutativity)}$$

$$x + 0 \approx x \qquad \text{(Identity)}$$

$$x + y \not\approx x + z \ \vee\ y \approx z \qquad \text{(Cancellation)}$$

$$\psi x \not\approx \psi y \ \vee\ x \approx y \qquad \text{(Ψ-Torsion-Freeness)}$$

(for every $\psi \in \Psi$*) are the axioms of* Ψ*-torsion-free cancellative abelian monoids. The first four clauses are denoted by* A, C, U, *and* K, *the set of* Ψ*-torsion-freeness axioms by* T_Ψ. *We write* ACUKT_Ψ *for the whole set and* AC, ACU, ACK *for the respective subsets.*

By Lemma A.8, we assume always without loss of generality that Ψ contains 1 and is closed under multiplication and factors. In practice, Ψ will usually be either $\{1\}$ (so Ψ-torsion-freeness is void) or $\mathbf{N}^{>0}$ (so Ψ-torsion-freeness is ordinary torsion-freeness).

Definition 3.2. *For sets of clauses* N *and* N', *we write* $N \models_\Psi N'$ *if* $N \cup \mathrm{ACUKT}_\Psi \models_\approx N'$. *In other words,* $\models_\Psi$ *denotes entailment modulo* ACUKT_Ψ *and equality. If* C *is a clause,* $N \models_\Psi C$ *is a shorthand for* $N \models_\Psi \{C\}$.

Definition 3.3. *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 3.4. *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 $\{\, t \mid \exists x\colon x \text{ is a variable}, t =_{\mathrm{ACU}} x \,\}$, $\{\, t \mid \exists s\colon s \text{ is atomic}, t =_{\mathrm{ACU}} s \,\}$, and $\{\, t \mid \exists s\colon s \text{ is a proper sum}, t =_{\mathrm{ACU}} s \,\}$. 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 some $s \in [t]_{\mathrm{ACU}}$ has this property.

We say that the term t occurs in the term s at the top, if there is an $o \in \mathrm{pos}(s)$ such that $s|_o = t$ and $s(o')$ equals $+$ for every proper prefix o' of o. We say that t occurs in s below a free function symbol, if there is an $o \in \mathrm{pos}(s)$ such that $s|_o = t$ and $s(o')$ is a free function symbol for some proper prefix o' of o; if additionally $|o'| + 1 = |o|$, we say that t occurs in s immediately below a free function symbol. We extend this terminology to ACU-congruence classes, and say that $[t]_{\mathrm{ACU}}$ occurs in $[s]_{\mathrm{ACU}}$ at the top or (immediately) 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, $[2b]_{\mathrm{ACU}}$ and $[f(2c) + b]_{\mathrm{ACU}}$ occur at the top of $[c + 2b + 3f(2c)]_{\mathrm{ACU}}$; $[c]_{\mathrm{ACU}}$ occurs both at the top and below a free function symbol, but not immediately below a free function symbol; $[2c]_{\mathrm{ACU}}$ occurs immediately below a free function symbol.

A substitution σ is called an ACU-unifier of the terms $t_1, \dots, t_n$, if $t_i\sigma =_{\mathrm{ACU}} t_j\sigma$ for all $i, j \in \{1, \dots, n\}$. A set U of ACU-unifiers of $t_1, \dots, t_n$ is called complete, if for every ACU-unifier θ of $t_1, \dots, t_n$ there exists a $\sigma \in U$ and a substitution ρ such that $x\theta =_{\mathrm{ACU}} x\sigma\rho$ for all $x \in \mathrm{Var}(\{t_1, \dots, t_n\})$. ACU-unification is finitary: for every set of terms $\{t_1, \dots, t_n\}$ there exists a (possibly empty) finite minimal complete set of idempotent ACU-unifiers. We use the phrase "most general ACU-unifier of s and t" to denote some member of a fixed complete set of ACU-unifiers of s and t. Without loss of generality we assume that all unifiers in this complete set are idempotent.

DEFINITION 3.5. *A reduction ordering $\succ$ is* 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.

As observed by Jouannaud and Marché (1992), we can obtain an ACU-compatible reduction ordering for ground terms from an AC-compatible ordering:

LEMMA 3.6. *Let $\succ_1$ be an* AC*-compatible reduction ordering, such that* 0 *is minimal with respect to $\succ_1$. Let the ordering $\succ$ on ground terms be defined by $s \succ t$ if $s{\downarrow} \succ_1 t{\downarrow}$, where $s{\downarrow}$ denotes the normal form of s under rewriting with the rules $x + 0 \to x$ and $0 + x \to x$. Then $\succ$ is an* ACU*-compatible reduction ordering on ground terms.*

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 (1992), it happens quite frequently that $\succ$ orders a pair of terms in an operationally undesirable way, or that $s[x]_o$ and $t[x]_o$ are uncomparable because $s[0]_o \succ t[0]_o$ but $s[u]_o \prec t[u]_o$ for all non-zero ground terms u.[4] 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.[5] In fact, Jouannaud and Marché's method can be considered as a variant of unfailing completion with constraints.

[4] Jouannaud and Marché's statement that "AC1-rewrite orderings cannot really exist" (Jouannaud and Marché, 1992) should be taken with a grain of salt, however.

[5] Constraint versions of the calculi presented in this paper can be found in (Waldmann 1997).

DEFINITION 3.7. *An* ACU*-compatible ordering has the multiset property, if whenever a ground atomic term* $u \neq 0$ *is greater than* t_i *for all* i *in a finite set* I*, then* $u \succ \sum_{i \in I} t_i$.

From now on, $\succ$ will always denote an ACU-compatible reduction ordering that has the multiset property, is total on ACU-congruence classes,[6] and satisfies $t \not\succ s[t]_o$ for every term $s[t]_o$.[7] An example of an ordering with these properties is obtained from the recursive path ordering with precedence $f_n \succ \ldots \succ f_1 \succ + \succ 0$ and multiset status for $+$ by comparing $s{\downarrow}$ and $t{\downarrow}$ as described in Lemma 3.6.[8]

CONVENTION 3.8. *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* $[\_]_{\text{ACU}}$ *and drop the subscript of* $=_{\text{ACU}}$. *So* all *terms, equations, substitutions, inference rules, etc., are to be taken modulo* ACU*, that is, as representatives of their congruence classes. Furthermore, we will use the equality predicate as a symmetric operator, thus ignoring the difference between* $t \approx t'$ *and* $t' \approx t$.

DEFINITION 3.9. *Let* t *be a ground term, then the maximal atomic subterm of* t *(with or without multiplicity) is defined in the following way:*

(i) If t *is a term of the form* $nu + \sum_{i \in I} v_i$*, where* u *and* v_i *are atomic terms,* $n \geq 1$*, and* $u \succ v_i$ *for all* $i \in I$ *then* $\mathrm{mt}(t) = u$ *and* $\mathrm{mt}_{\#}(t) = nu$.
(ii) If t *does not have sort* S_{CAM}*, then* $\mathrm{mt}(t) = \mathrm{mt}_{\#}(t) = t$.

If e *is a ground equation* $t \approx t'$*, then* $\mathrm{mt}(e) = \max\{\mathrm{mt}(t), \mathrm{mt}(t')\}$ *and* $\mathrm{mt}_{\#}(e) = \max\{\mathrm{mt}_{\#}(t), \mathrm{mt}_{\#}(t')\}$.

The symbol $\mathrm{ms}(t)$ denotes the multiset of all non-zero atomic terms occurring at the top of a ground term t. More precisely, $\mathrm{ms}(t) = \{t\}$, if t does not have sort S_{CAM}, and $\mathrm{ms}(t) = \{\, v_j \mid j \in J \,\}$, 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.) 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 3.10. *The ordering* $\succ$ *on terms is extended to an ordering* $\succ_{\text{L}}$ *on literals as follows: Every ground literal* $[\neg]\ s \approx t$ *is mapped to the quintuple*

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

where pol *is* 1 *for negative literals and* 0 *for positive ones, and where* bal *is* 1*, if* $\mathrm{mt}(s \approx t)$ *occurs on both sides of the literal, and otherwise* 0. *Two ground literals are compared by comparing their associated quintuples using the lexicographic combination of the ordering* $\succ$ *on terms, the ordering* $>$ *on* $\mathbf{N}$*, 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*

[6] In practice, it is sufficient if the ordering can be extended to a total ordering.
[7] In a many-sorted framework this property does not follow automatically from $\succ$ being total and noetherian. As an example let $\mathcal{S} = \{S, S'\}$ and $\Sigma = \{b : \to S,\ f : S \to S'\}$ with the ordering $b \succ f(b)$.
[8] Note that polynomial orderings (Ben Cherifa and Lescanne 1987) are unsuited, as they do not have the multiset property.

usual way, so $[\neg]\, e_1 \succ_{\mathrm{L}} [\neg]\, e_2$ *if and only if* $[\neg]\, e_1\theta \succ_{\mathrm{L}} [\neg]\, e_2\theta$ *for all ground instances* $[\neg]\, e_1\theta$ *and* $[\neg]\, e_2\theta$. *To compare equations using the ordering* $\succ_{\mathrm{L}}$, *they are identified with positive literals.*

The ordering $\succ_{\mathrm{C}}$ *on clauses is the multiset extension of the literal ordering* $\succ_{\mathrm{L}}$.

As $\succ_{\mathrm{L}}$ and $\succ_{\mathrm{C}}$ are obtained from noetherian orderings by multiset extension and lexicographic combination, they are noetherian, too. Furthermore, they are total on ground literals/clauses, thanks to the last component of the associated quintuples.

DEFINITION 3.11. *Let* C *be a clause. A variable* $x \in \mathrm{Var}(C)$ *is called shielded in* C, *if it occurs at least once below a free function symbol. Otherwise,* x *is called unshielded.*

For example, x and z are shielded in $x + y + f(x) \approx g(z)$, whereas y is unshielded. The importance of unshielded variables stems from the fact that they may correspond to maximal atomic subterms in a ground instance. If x is shielded in C, then C contains an atomic subterm $t[x]$, and as $x\theta \prec (t[x])\theta$, an atomic subterm of $x\theta$ cannot be maximal.

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

DEFINITION 3.12. *A variable* x *occurring in a clause* C *is called eligible, if* x *has sort* S_{CAM} *and either* C *has no selected literals and* x *is unshielded in* C, *or* x *occurs in some selected literal and* x *is unshielded in the subclause consisting of all selected literals of* C. *The set of all eligible variables of a clause* C *is denoted by* $\mathrm{elig}(C)$.

3.2. IDEAS AND CONCEPTS

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

Our goal is to develop a superposition-like calculus for Ψ-torsion-free cancellative abelian monoids that makes superpositions with the ACUKT_Ψ axioms superfluous.

Cancellative superposition.

Let us first restrict to the case that S_{CAM} is the only sort, $+$ is the only non-constant function symbol, and that $\Psi = \{1\}$ (in other words, T_Ψ is void). In a cancellative abelian monoid, the congruence law and the cancellation law are in some sense complementary: If we have an equation $u + t \approx t'$, then we can infer $t' + u + s \approx u + t + s'$ from $u + s \approx s'$ by congruence, and $t' + s \approx t + s'$ by cancellation. Similarly, we can infer $t' + u + s \not\approx u + t + s'$ from $u + s \not\approx s'$ by cancellation, and $t' + s \not\approx t + s'$ 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 summand by the remainder: We rewrite u to t' while adding t to the other side of the equation and obtain an equivalent equation.

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

Pos. Canc. Superposition
$$\frac{D' \vee nu + t \approx t' \qquad C' \vee mu + s \approx s'}{D' \vee C' \vee (m{-}n)u + s + t' \approx s' + t}$$
where $m \geq n \geq 1$.[9]

If the equation $mu + s \approx s'$ occurs negatively, the rule is similar. In fact, in this case we have to perform an inference only if, by repeated replacement of nu, mu is eliminated completely. In other words, an inference is only necessary if $m = \chi n$ for some $\chi \in \mathbf{N}^{>0}$.

Neg. Canc. Superposition
$$\frac{D' \vee nu + t \approx t' \qquad C' \vee \neg\, mu + s \approx s'}{D' \vee C' \vee \neg\, s + \chi t' \approx s' + \chi t}$$
where $m = \chi n$, $n \geq 1$, $\chi \in \mathbf{N}^{>0}$.

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

Cancellation
$$\frac{C' \vee [\neg]\, mu + s \approx m'u + s'}{C' \vee [\neg]\, (m{-}m')u + s \approx s'}$$
where $m \geq m' \geq 1$.

Equality Resolution[10]
$$\frac{C' \vee \neg\, 0 \approx 0}{C'}$$

Canc. Eq. Factoring
$$\frac{C' \vee mu + t \approx t' \vee mu + s \approx s'}{C' \vee \neg\, t + s' \approx t' + s \vee mu + t \approx t'}$$

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[11] 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 $\mathrm{ACUKT}_{\{1\}}$. (The maximal parts of every clause are underlined.)

$$\underline{4b} + c \approx 4d \tag{1}$$
$$\underline{2b'} + c \approx 2d' \tag{2}$$
$$\underline{2d} \approx d' \tag{3}$$
$$\underline{4b} \not\approx 2b' \tag{4}$$

Cancellative superposition of (1) and (4) yields

$$4d \not\approx \underline{2b'} + c \tag{5}$$

Cancellative superposition of (2) and (5) yields

$$4d + \underline{c} \not\approx 2d' + \underline{c} \tag{6}$$

[9] Recall that we are working with terms modulo ACU, so s and t may be missing (that is, zero).
[10] 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.
[11] Except for the literal $mu + t \approx t'$ in the *cancellative equality factoring* rule.

By *cancellation* of (6) we obtain

$$\underline{4d} \not\approx 2d' \tag{7}$$

Cancellative superposition of (3) and (7) produces

$$\underline{2d'} \not\approx \underline{2d'} \tag{8}$$

which by *cancellation* and *equality resolution* yields the empty clause.

Speaking in terms of AG-normalized completion (Marché 1996), we can work directly with the symmetrization (if it exists); Marché's Ψ_{AG} and Θ_{AG} have no counterpart in our framework. On the other hand, we lack an inverse, which will lead to certain problems once free function symbols are introduced.

Torsion-freeness.

Until now, we have only considered the case $\Psi = \{1\}$. What has to be changed if Ψ is an arbitrary subset of $\mathbf{N}^{>0}$ closed under multiplication and factors? Nothing, as far as *positive cancellative superposition*, *cancellation*, and *equality resolution* inferences are concerned. The main modification is necessary for the *negative cancellative superposition* rule. So far, we had to perform an inference between $D' \vee nu + t \approx t'$ and $C' \vee \neg\, mu + s \approx s'$ only if $m = \chi n$. However, by Ψ-torsion-freeness and congruence, the literals $\neg\, mu + s \approx s'$ and $\neg\, \psi mu + \psi s \approx \psi s'$ are equivalent for each $\psi \in \Psi$. Therefore, an inference between $D' \vee nu + t \approx t'$ and $C' \vee \neg\, mu + s \approx s'$ is now necessary whenever $\psi m = \chi n$ for some $\psi \in \Psi$ and $\chi \in \mathbf{N}^{>0}$. The general version of the inference rule is thus:

Neg. Canc. Superposition
$$\frac{D' \vee nu + t \approx t' \qquad C' \vee \neg\, mu + s \approx s'}{D' \vee C' \vee \neg\, \psi s + \chi t' \approx \psi s' + \chi t}$$
where $\psi m = \chi n$, $n \geq 1$, $\psi \in \Psi$, $\chi \in \mathbf{N}^{>0}$.

With the additional condition that $\gcd(\psi, \chi) = 1$, there exists at most one pair (ψ, χ) for any given combination of m and n: If $n/\gcd(m,n) \in \Psi$, then $\psi = n/\gcd(m,n)$ and $\chi = m/\gcd(m,n)$; otherwise, no ψ and χ with the desired properties exist.

A similar modification applies to the *cancellative equality factoring* rule.

Canc. Eq. Factoring
$$\frac{C' \vee nu + t \approx t' \vee mu + s \approx s'}{C' \vee \neg\, \psi t + \chi s' \approx \psi t' + \chi s \vee nu + t \approx t'}$$
where $\psi n = \chi m$, $n \geq 1$, $\psi \in \Psi$, $\chi \in \mathbf{N}^{>0}$.

Again, requiring that $\gcd(\psi, \chi) = 1$ ensures that there is at most one pair (ψ, χ).

The non-ground case (I).

So far, we have confined ourselves to ground clauses containing $+$ as the only non-constant function symbol. Giving up these restrictions, we have to find a way to lift the inference rules developed above to clauses with variables. In the standard superposition calculus, lifting means replacing equality in the ground inference by unifiability. As long as all variables in our clauses are shielded, the situation is similar here: In a clause $C = C' \vee [\neg]\, e_1$, a maximal equation e_1 need no longer have the form $mu + s \approx s'$ with a unique maximal atomic term u. Rather, it may contain several (distinct but ACU-unifiable) maximal atomic terms u_k with multiplicities m_k^*, where k ranges over some finite non-empty index set K. We obtain thus $e_1 = \sum_{k\in K} m_k^* u_k + s \approx s'$. In the inference rule, the substitution σ that unifies all u_k (and the corresponding terms v_l from the other premise) is applied to the conclusion. For instance, the *positive cancellative superposition* rule now has the following form:

Pos. Canc. Superposition $$\frac{D' \vee e_2 \qquad C' \vee e_1}{(D' \vee C' \vee e_0)\sigma}$$

where

(i) $e_1 = \sum_{k\in K} m_k^* u_k + s \approx s'$.
(ii) $e_2 = \sum_{l\in L} n_l^* v_l + t \approx t'$.
(iii) $m = \sum_{k\in K} m_k^* \geq n = \sum_{l\in L} n_l^*$.
(iv) u is one of the u_k or v_l $(k \in K, l \in L)$.
(v) $e_0 = (m-n)u + s + t' \approx t + s'$.
(vi) σ is a most general ACU-unifier of all u_k and v_l $(k \in K, l \in L)$.
(vii) $u\sigma \not\preceq s\sigma$, $u\sigma \not\preceq s'\sigma$, $u\sigma \not\preceq t\sigma$, $u\sigma \not\preceq t'\sigma$.

The other inference rules can be lifted in a similar way, again under the condition that all variables in the clauses are shielded. If unshielded variables occur, the situation becomes significantly more complicated. This case will be treated below.

Free function symbols.

As soon as the clauses contain non-constant free function symbols, and possibly other sorts, we also have to use the inference rules of the traditional superposition calculus, that is, *standard superposition*, *equality resolution*, and *standard equality factoring.*

Standard Superposition $$\frac{D' \vee t \approx t' \qquad C' \vee [\neg]\, s[t] \approx s'}{D' \vee C' \vee [\neg]\, s[t'] \approx s'}$$
where $t \succ t'$, $s \succ s'$.

Equality Resolution $$\frac{C' \vee \neg\, u \approx u}{C'}$$
where u is either 0 or does not have sort S_{CAM}.

Std. Eq. Factoring $$\frac{C' \vee u \approx v' \vee u \approx u'}{C' \vee \neg\, u' \approx v' \vee u \approx v'}$$
where $u \succ u' \succeq v'$.

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 ACUKT$_\Psi$-model of the three clauses

$$\underline{4b} + c \approx 4d \tag{1}$$

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

$$\underline{2d} \approx d' \tag{3}$$

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

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

If $2 \in \Psi$, we can even refute the set of clauses (1), (2), (3), and (9).

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

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 $\perp$ from the clauses (1), (2), (3), and (10).

$$\underline{f(4b)} \not\approx f(2b') \tag{10}$$

Nor is it possible to derive $\perp$ from the clauses (1), (2), (3), and (11), if $2 \in \Psi$.

$$\underline{f(2b)} \not\approx f(b') \tag{11}$$

If we were working in groups, we could simply derive $f(4d-c) \not\approx f(2b')$ from clause (10). But first, our framework is more general than groups, and second, even this method would not be usable to refute clause (11).

Hsiang, Rusinowitch, and Sakai (1987) and Rusinowitch (1989) have solved this problem by introducing the following inference rule:

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

This rule allows us to derive $\underline{4b} \approx 2b'$ from clauses (1)–(3) of Ex. 3.14, which can then be applied to (10) by *standard superposition*. However, before we can apply the rule of Hsiang, Rusinowitch, and Sakai, we have to use clause (3) to replace $4d$ by $2d'$ in (1). Since $4d$ is not maximal in (1), the rule can only be used in conjunction with ordered paramodulation (where inferences involving *smaller* parts of maximal literals are required), but does not work together with strict superposition (where such inferences are unnecessary).[12] Furthermore, this method would again be limited to the $\Psi = \{1\}$ case.

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 w in a clause $C[w]$ means replacing w by a new variable y and adding $y \not\approx w$ as a new condition to the clause. In our case, we have to abstract out a term w of sort S_{CAM} occurring immediately below a free function symbol, if there is some other clause $D' \vee nu + t \approx t'$ such that u occurs at the top of w.

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

[12] This has been pointed out to me by Leo Bachmair.

The *abstraction* rule has some peculiar properties that distinguish it from the other rules of our calculus. It is the only inference rule whose conclusion is non-ground, even if the premises are ground. We emphasize that the new variable y is shielded in the resulting clause. Besides, it should be noted that the first premise is completely irrelevant for the correctness of the inference: whenever the second premise is true in some interpretation, the conclusion is true. The first premise serves only to determine whether an *abstraction* inference is necessary, it does not influence the result of the inference.

Using the *abstraction* rule, the set of clauses (1), (2), (3), and (11) of Ex. 3.14 (assuming $2 \in \Psi$) can be refuted as follows:

EXAMPLE 3.15. *Abstraction* of (1) and (11) yields

$$y \not\approx 2b \vee \underline{f(y)} \not\approx \underline{f(b')} \tag{12}$$

By *cancellation*[13] of (12) with the unifier $\{y \mapsto b'\}$ we obtain

$$b' \not\approx \underline{2b} \vee 0 \not\approx 0 \tag{13}$$

Cancellative superposition of (1) and (13) yields

$$c + \underline{2b'} \not\approx 4d \vee 0 \not\approx 0 \tag{14}$$

which can be refuted in the same way as clause (5) in Ex. 3.13.

The relationship between the coefficients m and n in the *abstraction* rule above is not completely obvious. Intuitively, an *abstraction* inference between $D = D' \vee nu + t \approx t'$ and $C = C' \vee [\neg]\, s[mu + q] \approx s'$ is needed, if there is some clause D_0 such that (i) D_0 is entailed by D and some other clauses, (ii) D_0 is not derivable using the inference rules, (iii) a *standard superposition* of D into C is impossible, (iv) if D_0 were contained in the clause set, a *standard superposition* of D_0 into C would be necessary. In Ex. 3.14, this clause D_0 is either $4b \approx 2b'$ (for arbitrary Ψ) or $2b \approx b'$ (if $2 \in \Psi$); it follows from clauses (1)–(3), but is not derivable. A detailed analysis shows that it suffices to consider the case that $D_0 = D'_0 \vee m'u + r \approx r'$ has the same maximal term u as D, that $\psi m' = \chi n$ for some $\psi \in \Psi$, $\chi \in \mathbf{N}^{>0}$, and that $m' \leq m$ (otherwise, D_0 could not be superposed on $s[mu + q]$).

The *abstraction* rule is extended to non-ground premises in essentially the same way as the *cancellative superposition* rule. The new question that arises here is: Do we have a similar situation as for the *standard superposition* rule, where superpositions at or below variable positions are superfluous? Can we avoid an *abstraction* if the maximal term of D overlaps with a variable in C, rather than with an atomic term? The answer is negative, in general. This is due to the fact that, even if D overlaps only at a variable, the "hypothetical superposition" with the entailed clause D_0 may take place at a non-variable position. As an example, consider $D = \underline{b} + c \approx d$ and $C = \underline{f(x + c')} \approx g(c')$. The maximal term b of D overlaps only with the variable x in C. However, D, together with some other clause, may entail $D_0 = \underline{b} + c' \approx d'$, allowing a superposition on C at a non-variable position. Only if the variable x occurs immediately below the free function symbol or if it occurs in a sum $x + t_1[x] + \cdots + t_n[x]$, where every other summand contains x as a proper subterm, we can be sure that the hypothetical superposition would take place at a variable position. This is therefore the only situation where *abstraction* is superfluous.

[13] Using the lifted version of the *cancellation* rule described above.

The non-ground case (II).

When we discussed the lifting of the inference rules to non-ground clauses, we left out the handling of unshielded variables. Recall that a variable z is shielded in a clause C, if C contains some atomic subterm $t[z]$. Shielded variables are easy to handle because they cannot correspond to maximal terms in a ground instance $C\theta$. An unshielded variable x, on the other hand, can be instantiated with an atomic term $x\theta = \bar{u}$ that is maximal in $C\theta$. Even worse, it can be instantiated with a sum $x\theta = \mu\bar{u} + \bar{s}$ that contains an unknown number of occurrences of the maximal term $\bar{u}$ and a likewise unknown sum $\bar{s}$ of non-maximal terms. Now $\mu\bar{u}$ may be involved in a ground *positive cancellative superposition* inference from $C\theta$. How can we represent this ground inference on the non-ground level without introducing second-order variables?

The solution for this problem is to map x to a sum of two fresh variables, $\hat{x} + \check{x}$. The variable $\hat{x}$ is meant to subsume the maximal part of $x\theta$, that is $\mu\bar{u}$, the second variable $\check{x}$ is meant to subsume the rest, that is $\bar{s}$. As μ is unknown, we can no longer *count* the numbers of occurrences of the maximal terms in order to compute their difference in the *positive cancellative superposition* rule. We can, however, use ACU-unification to "subtract" the terms: Suppose that the maximal literal of the left premise contains the unshielded variables y_1 and y_2 and that the maximal literal of the right premise contains the unshielded variable x_1 and the maximal terms u_1 and $2u_2$, where $u_1\theta = u_2\theta = \bar{u}$. The variables $\hat{y}_1$, $\hat{y}_2$, and $\hat{x}_1$ represent the occurrences of $\bar{u}$ in $y_1\theta$, $y_2\theta$, and $x_1\theta$.[14] To compute the difference $k\bar{u}$ of $(\hat{x}_1 + u_1 + 2u_2)\theta$ and $(\hat{y}_1 + \hat{y}_2)\theta$, we introduce a new variable z and compute a complete set U of ACU-unifiers of $\hat{x}_1 + u_1 + 2u_2$ and $z + \hat{y}_1 + \hat{y}_2$. For one $\sigma \in U$, we have $\theta = \sigma\rho$ over $\mathrm{Var}(u_1) \cup \mathrm{Var}(u_2) \cup \{\hat{x}_1, \hat{y}_1, \hat{y}_2, z\}$, so $k\bar{u}$ is an instance of $z\sigma$.

In general, we may assume that a literal e has the form

$$\sum_{i\in I} m_i x_i + \sum_{k\in K} m_k^* u_k + s \approx s'\,,$$

where every x_i is an unshielded variable for $i \in I$ and every u_k is a maximal atomic term for $k \in K$. Then the sum $\sum_{i\in I} m_i\hat{x}_i + \sum_{k\in K} m_k^* u_k$ takes the role of mu in the ground inference rule; the sum $\sum_{i\in I} m_i\check{x}_i$ is joined with s. We may leave out unshielded variables that occur also in the right-hand side of e or in some negative literal – if the maximal atomic term (on the ground level) occurred also on the right-hand side, then the *positive cancellative superposition* rule would not be applicable, if it occurred in some negative literal $\neg\, e'$, then $\neg\, e'$ would be even larger than e.

The lifting of the *negative cancellative superposition* rule happens in a similar way. Again, every unshielded variable x_i is mapped to $\hat{x}_i + \check{x}_i$, such that $\hat{x}_i$ represents the occurrences of the maximal atomic term of the ground clause and $\check{x}_i$ represents the rest of the term. The additional problem that arises here is that we can no longer compute a unique pair (ψ, χ). There is no universal solution for this problem. The general form of the *negative cancellative superposition* rule, that we will give in the sequel, may therefore produce infinitely many inferences for a given pair of premises. In Section 5, we will show how the general system can be refined to specialized finitely branching systems for the two most important cases of Ψ, that is $\Psi = \{1\}$ and $\Psi = \mathbf{N}^{>0}$.

[14] Note that it is not required that the maximal term occurs in *all* unshielded variables. It is thus possible that $\hat{y}_1\theta$, $\hat{y}_2\theta$, or $\hat{x}_1\theta$ is zero.

Redundancy.

The inference rules described so far are only one of the components of the cancellative superposition calculus. The other one is the associated redundancy criterion. Since understanding the latter requires to some extent understanding the idea of the completeness proof, we will postpone its definition until Section 4.

3.3. The Inference System

Let us start the presentation of the inference rules with a few general conventions.

Every term occurring in a sum is assumed to have sort S_{CAM}. The letters u and v, possibly with indices, denote atomic terms unless explicitly said otherwise; x, y, and z denote variables. In an expression like $\sum_{i\in I} m_i x_i + \sum_{k\in K} m'_k u_k + s$, both I and K are finite sets of indices; I and K may be empty, s may be 0, unless explicitly said otherwise. The coefficients m_i and m'_k are elements of $\mathbf{N}^{>0}$.

If a literal of a clause is selected, then an inference must not involve non-selected literals of this clause. If an inference involves a non-selected literal, then it must be maximal in the respective clause (except for the literal $v \approx v'$ in *standard equality factoring* and the literal e_2 in *cancellative equality factoring*). If an inference involves a selected literal, then it must be maximal among the selected literals of this clause. A positive literal that is involved in a *superposition* or *abstraction* inference must be strictly maximal in the respective clause. In all *superposition* and *abstraction* inferences, the left premise is smaller than the right premise. In *standard superposition* and *abstraction* inferences, if $s[w]$ is a proper sum, then w occurs in a maximal atomic subterm of s.

Inference System 3.16. The inference system $CInf_\Psi$ of the cancellative superposition calculus consists of the inference rules *cancellation*, *equality resolution*, *standard superposition*, *negative cancellative superposition*, *positive cancellative superposition*, *abstraction*, *standard equality factoring*, and *cancellative equality factoring*, as described below.

Cancellation

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

if the following conditions are satisfied:

(i) $e_1 \;=\; \sum_{i\in I} m_i x_i + \sum_{k\in K} m^*_k u_k + s \approx \sum_{i\in I'} m'_i x_i + \sum_{k\in K'} m^{*\prime}_k u_k + s'$.

(ii) $e_0 \;=\; z + \sum_{i\in I} m_i \check{x}_i + s \approx \sum_{i\in I'} m'_i \check{x}_i + s'$.

(iii) $I \cup K \neq \emptyset$ and $I' \cup K' \neq \emptyset$.

(iv) If $[\neg]\, e_1$ is a positive literal:
$\{\, x_i \mid i \in I \,\} = \mathrm{elig}(C' \lor e_1) \cap \mathrm{Var}(\mathrm{lhs}(e_1)) \setminus \mathrm{Var}(\mathrm{neg}(C'))$,
$\{\, x_i \mid i \in I' \,\} = \mathrm{elig}(C' \lor e_1) \cap \mathrm{Var}(\mathrm{rhs}(e_1)) \setminus \mathrm{Var}(\mathrm{neg}(C'))$.
Otherwise:
$\{\, x_i \mid i \in I \,\} = \mathrm{elig}(C' \lor \neg e_1) \cap \mathrm{Var}(\mathrm{lhs}(e_1))$,
$\{\, x_i \mid i \in I' \,\} = \mathrm{elig}(C' \lor \neg e_1) \cap \mathrm{Var}(\mathrm{rhs}(e_1))$.

(v) If $K \cup K' \neq \emptyset$, u is one of the u_k ($k \in K \cup K'$), otherwise, u is a new variable.

(vi) σ_1 maps x_i to $\hat{x}_i + \check{x}_i$ for $i \in I \cup I'$; σ_2 is a most general ACU-unifier of all u_k ($k \in K \cup K'$); σ_3 is a most general ACU-unifier of $\sum_{i\in I} m_i \hat{x}_i + \big(\sum_{k\in K} m^*_k\big) u\sigma_2$ and $z + \sum_{i\in I'} m'_i \hat{x}_i + \big(\sum_{k\in K'} m^{*\prime}_k\big) u\sigma_2$ such that $\big(\sum_{i\in I'} m'_i \hat{x}_i + \big(\sum_{k\in K'} m^{*\prime}_k\big) u\sigma_2\big)\sigma_3$ is not identical to 0; and $\sigma = \sigma_1\sigma_2\sigma_3$.

(vii) $u\sigma \not\preceq s\sigma$, $u\sigma \not\preceq s'\sigma$, $u\sigma \not\preceq \check{x}_i\sigma$ for $i \in I \cup I'$.

Equality Resolution

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

if the following condition is satisfied:

(i) Either σ is a most general ACU-unifier of u, u', and 0; or u and u' do not have sort S_{CAM} and σ is a most general ACU-unifier of u and u'.

Standard Superposition

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

if the following conditions are satisfied:

(i) w is not a variable.
(ii) If s has sort S_{CAM}, then w occurs below a free function symbol in s.
(iii) σ is a most general ACU-unifier of t and w.
(iv) $s[w]\sigma \not\preceq s'\sigma$, $t\sigma \not\preceq t'\sigma$.

Negative Cancellative Superposition

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

if the following conditions are satisfied:

(i) $e_1 \;=\; \sum_{i\in I} m_i x_i + \sum_{k\in K} m_k^* u_k + s \approx s'$.
(ii) $e_2 \;=\; \sum_{j\in J} n_j y_j + \sum_{l\in L} n_l^* v_l + t \approx t'$.
(iii) $e_0 \;=\; \sum_{i\in I} \psi m_i \check{x}_i + \psi s + \chi t' \approx \sum_{j\in J} \chi n_j \check{y}_j + \chi t + \psi s'$.
(iv) $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
(v) $\{\, x_i \mid i \in I \,\} = \text{elig}(C' \vee \neg e_1) \cap \text{Var}(\text{lhs}(e_1)) \setminus \text{Var}(\text{rhs}(e_1))$,
$\{\, y_j \mid j \in J \,\} = \text{elig}(D' \vee e_2) \cap \text{Var}(\text{lhs}(e_2)) \setminus \text{Var}(\text{rhs}(e_2)) \setminus \text{Var}(\text{neg}(D'))$.
(vi) $\text{lhs}(e_1)$ is not a variable (i. e., either $\sum_{i\in I} m_i > 1$ or $\sum_{k\in K} m_k^* u_k + s \neq 0$).
(vii) If $I = \{i_1\}$, $m_{i_1} = 1$, and $K = \emptyset$, then $\text{lhs}(e_2)$ is not an atomic term. If additionally $\Psi = \{1\}$, then $J \neq \emptyset$ or $t \neq 0$.
(viii) $\psi \in \Psi$ and $\chi \in \mathbf{N}^{>0}$, such that $\gcd(\psi, \chi) = 1$.
(ix) If $K \cup L \neq \emptyset$, u is one of the u_k or v_l ($k \in K$, $l \in L$), otherwise, u is a new variable.
(x) σ_1 maps x_i to $\hat{x}_i + \check{x}_i$ and y_j to $\hat{y}_j + \check{y}_j$ for $i \in I$, $j \in J$; σ_2 is a most general ACU-unifier of all u_k and v_l ($k \in K$, $l \in L$); σ_3 is a most general ACU-unifier of $\sum_{i\in I} \psi m_i \hat{x}_i + \big(\sum_{k\in K} \psi m_k^*\big) u\sigma_2$ and $\sum_{j\in J} \chi n_j \hat{y}_j + \big(\sum_{l\in L} \chi n_l^*\big) u\sigma_2$; and $\sigma = \sigma_1\sigma_2\sigma_3$.
(xi) $u\sigma \not\preceq s\sigma$, $u\sigma \not\preceq s'\sigma$, $u\sigma \not\preceq t\sigma$, $u\sigma \not\preceq t'\sigma$, $u\sigma \not\preceq \check{x}_i\sigma$ for $i \in I$, $u\sigma \not\preceq \check{y}_j\sigma$ for $j \in J$.

Positive Cancellative Superposition

$$\frac{D' \vee e_2 \qquad C' \vee e_1}{(D' \vee C' \vee e_0)\sigma}$$

if the following conditions are satisfied:

(i) $e_1 = \sum_{i\in I} m_i x_i + \sum_{k\in K} m_k^* u_k + s \approx s'$.
(ii) $e_2 = \sum_{j\in J} n_j y_j + \sum_{l\in L} n_l^* v_l + t \approx t'$.
(iii) $e_0 = z + \sum_{i\in I} m_i \check{x}_i + s + t' \approx \sum_{j\in J} n_j \check{y}_j + t + s'$.
(iv) $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
(v) $\{\, x_i \mid i \in I \,\} = \mathrm{elig}(C' \vee e_1) \cap \mathrm{Var}(\mathrm{lhs}(e_1)) \setminus \mathrm{Var}(\mathrm{rhs}(e_1)) \setminus \mathrm{Var}(\mathrm{neg}(C'))$, $\{\, y_j \mid j \in J \,\} = \mathrm{elig}(D' \vee e_2) \cap \mathrm{Var}(\mathrm{lhs}(e_2)) \setminus \mathrm{Var}(\mathrm{rhs}(e_2)) \setminus \mathrm{Var}(\mathrm{neg}(D'))$.
(vi) If $K \cup L \neq \emptyset$, u is one of the u_k or v_l ($k \in K$, $l \in L$), otherwise, u is a new variable.
(vii) σ_1 maps x_i to $\hat{x}_i + \check{x}_i$ and y_j to $\hat{y}_j + \check{y}_j$ for $i \in I$, $j \in J$; σ_2 is a most general ACU-unifier of all u_k and v_l ($k \in K$, $l \in L$); σ_3 is a most general ACU-unifier of $\sum_{i\in I} m_i \hat{x}_i + \big(\sum_{k\in K} m_k^*\big) u\sigma_2$ and $z + \sum_{j\in J} n_j \hat{y}_j + \big(\sum_{l\in L} n_l^*\big) u\sigma_2$ such that $\big(\sum_{j\in J} n_j \hat{y}_j + \big(\sum_{l\in L} n_l^*\big) u\sigma_2\big)\sigma_3$ is not identical to 0; and $\sigma = \sigma_1\sigma_2\sigma_3$.
(viii) $u\sigma \not\preceq s\sigma$, $u\sigma \not\preceq s'\sigma$, $u\sigma \not\preceq t\sigma$, $u\sigma \not\preceq t'\sigma$, $u\sigma \not\preceq \check{x}_i\sigma$ for $i \in I$, $u\sigma \not\preceq \check{y}_j\sigma$ for $j \in J$.

Abstraction

$$\frac{D' \vee e_2 \qquad C' \vee [\neg]\, s[w] \approx s'}{C' \vee \neg\, y \approx w \vee [\neg]\, s[y] \approx s'}$$

if the following conditions are satisfied:

(i) $w = \sum_{i\in I} m_i x_i + \sum_{k\in K} m_k^* u_k + q$.
(ii) $e_2 = \sum_{j\in J} n_j y_j + \sum_{l\in L} n_l^* v_l + t \approx t'$.
(iii) $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
(iv) None of the variables x_i occurs in the non-variable terms u_k or at the top of q.
(v) $\{\, y_j \mid j \in J \,\} = \mathrm{elig}(D' \vee e_2) \cap \mathrm{Var}(\mathrm{lhs}(e_2)) \setminus \mathrm{Var}(\mathrm{rhs}(e_2)) \setminus \mathrm{Var}(\mathrm{neg}(D'))$.
(vi) w occurs in s immediately below some free function symbol.
(vii) $\psi \in \Psi$ and $\chi \in \mathbf{N}^{>0}$, such that $\gcd(\psi, \chi) = 1$.
(viii) If $K \cup L \neq \emptyset$, u is one of the u_k or v_l ($k \in K$, $l \in L$), otherwise, u is a new variable.
(ix) σ_1 maps x_i to $\hat{x}_i + \check{x}_i$ and y_j to $\hat{y}_j + \check{y}_j$ for $i \in I$, $j \in J$; σ_2 is a most general ACU-unifier of all u_k and v_l ($k \in K$, $l \in L$); σ_3 is a most general ACU-unifier of $\sum_{i\in I} \psi m_i \hat{x}_i + \big(\sum_{k\in K} \psi m_k^*\big) u\sigma_2$ and $\psi z + \sum_{j\in J} \chi n_j \hat{y}_j + \big(\sum_{l\in L} \chi n_l^*\big) u\sigma_2$ such that $\big(\sum_{j\in J} \chi n_j \hat{y}_j + \big(\sum_{l\in L} \chi n_l^*\big) u\sigma_2\big)\sigma_3$ is not identical to 0; and $\sigma = \sigma_1\sigma_2\sigma_3$.
(x) If $I = \{i_1\}$, $m_{i_1} = 1$, and $K = \emptyset$, then $q = q_1 + q_2$, where q_1 is a non-zero atomic term not containing x_{i_1} as a subterm.
(xi) $\mathrm{lhs}(e_2)\sigma$ is not a subterm of $w\sigma$.
(xii) $w\sigma \not\preceq y\sigma$, $s[w]\sigma \not\preceq s'\sigma$, $u\sigma \not\preceq t\sigma$, $u\sigma \not\preceq t'\sigma$.

Standard Equality Factoring

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

if the following conditions are satisfied:

(i) u, u', v, and v' do not have sort S_{CAM}.
(ii) σ is a most general ACU-unifier of u and v.
(iii) $u\sigma \not\preceq u'\sigma$, $u\sigma \not\preceq v'\sigma$.

Cancellative Equality Factoring

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

if the following conditions are satisfied:

(i) $e_1 \;=\; \sum_{i\in I} m_i x_i + \sum_{k\in K} m_k^* u_k + s \approx s'$.
(ii) $e_2 \;=\; \sum_{j\in J} n_j x_j + \sum_{l\in L} n_l^* v_l + t \approx t'$.
(iii) $e_0 \;=\; \sum_{j\in J} \psi n_j \check{x}_j + \psi t + \chi s' \approx \sum_{i\in I} \chi m_i \check{x}_i + \chi s + \psi t'$.
(iv) $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
(v) $\{\, x_i \mid i \in I \,\} = \text{elig}(C' \vee e_2 \vee e_1) \cap \text{Var}(\text{lhs}(e_1)) \setminus \text{Var}(\text{rhs}(e_1)) \setminus \text{Var}(\text{rhs}(e_2)) \setminus \text{Var}(\text{neg}(C'))$,
$\{\, x_j \mid j \in J \,\} = \text{elig}(C' \vee e_2 \vee e_1) \cap \text{Var}(\text{lhs}(e_2)) \setminus \text{Var}(\text{rhs}(e_1)) \setminus \text{Var}(\text{rhs}(e_2)) \setminus \text{Var}(\text{neg}(C'))$.
(vi) $\psi \in \Psi$ and $\chi \in \mathbf{N}^{>0}$, such that $\gcd(\psi, \chi) = 1$.
(vii) If $K \cup L \neq \emptyset$, u is one of the u_k or v_l $(k \in K,\, l \in L)$, otherwise, u is a new variable.
(viii) σ_1 maps x_i to $\hat{x}_i + \check{x}_i$ for $i \in I \cup J$; σ_2 is a most general ACU-unifier of all u_k and v_l $(k \in K,\, l \in L)$; σ_3 is a most general ACU-unifier of $\sum_{i\in I} \chi m_i \hat{x}_i + \bigl(\sum_{k\in K} \chi m_k^*\bigr) u\sigma_2$ and $\sum_{j\in J} \psi n_j \hat{x}_j + \bigl(\sum_{l\in L} \psi n_l^*\bigr) u\sigma_2$; and $\sigma = \sigma_1\sigma_2\sigma_3$.
(ix) $u\sigma \not\preceq s\sigma$, $u\sigma \not\preceq s'\sigma$, $u\sigma \not\preceq t\sigma$, $u\sigma \not\preceq t'\sigma$, $u\sigma \not\preceq \check{x}_i\sigma$ for $i \in I \cup J$.

THEOREM 3.17. *The inference rules of the cancellative superposition calculus are sound with respect to* $\models_\Psi$.

PROOF. By routine computation. □

The inference rules remain sound if we ignore the ordering conditions. The *abstraction* rule remains sound, even if all its conditions are ignored.

4. Refutational Completeness

4.1. Ideas and Concepts

In the previous section we have presented the inference system of the cancellative superposition calculus. We will now define the associated redundancy criterion and demonstrate that the resulting calculus is refutationally complete. Again, we start with an informal explanation of the ideas of the proof, before we present the formal details.

Constructing an interpretation.

A theorem proving calculus is refutationally complete, if every saturated set of formulae either contains a contradictory formula or has a model. If the formulae in question are clauses, then the only contradictory formula is the empty clause $\bot$.

It is obvious that a set N of clauses does not have a model if it contains the empty clause. Our task is to show the reverse: Whenever N is saturated and $\bot \notin N$, then we will construct a model for the set $\overline{N}$ of all ground instances of clauses in N (and thus for N). The essential idea goes back to Zhang and Kapur (1988) and Pais and Peterson (1991) and was extended by Bachmair and Ganzinger (1994c): We inspect all clauses in $\overline{N}$ in ascending order and construct a sequence of interpretations, starting with the empty interpretation. If a clause $C \in \overline{N}$ is false in the current interpretation and has a positive and strictly maximal literal e, and if some additional conditions are satisfied, then a new interpretation is created extending the current one in such a way that e becomes true. We say that C is productive. Otherwise, the current interpretation is left unchanged. In this way we generate a sequence of interpretations with the following monotonicity properties:

(i) If a positive literal is true in some interpretation, then it remains true in all future interpretations.
(ii) If a clause is true at the time where it is inspected, then it remains true in all future interpretations.
(iii) If a clause $C = C' \vee e$ is productive, then C remains true and C' remains false in all future interpretations.

It is clear from (ii) and (iii) that every clause in $\overline{N}$ is true in the limit interpretation, if it is either true at the time where it is inspected or if it is productive. It remains to show that, by saturation, every ground instance in $\overline{N}$ falls into one of these two classes.

Standard superposition.

The scheme described so far characterizes most model construction proofs for superposition-like calculi. Before we explain our own refinements, let us recapitulate the standard superposition calculus of Bachmair and Ganzinger (1994c). Here the interpretations are the sets of all equations $t \approx t'$, such that $t \to^* t'' \leftarrow^* t'$ using the previously collected maximal literals as rewrite rules. A clause $C' \vee e$ may be productive only if the left-hand (that is, larger) side of e is irreducible with respect to the current set of rules, and if C' remains false after e has been included in the rewrite system.

If the clause $C = C' \vee e$ is productive, then $\mathrm{lhs}(e)$ is larger than every term occurring in negative literals of clauses smaller than C or in positive literals of productive clauses smaller than C. Consequently, the rule e cannot be used to rewrite such literals. This guarantees that the above-mentioned properties (ii) and (iii) hold. Furthermore, as every newly added rule is irreducible with respect to the old rules, and as its left-hand side is larger than the left-hand sides of the old rules, the resulting rewrite systems are confluent and terminating, hence the interpretations are equality interpretations.

According to the redundancy criterion of Bachmair and Ganzinger (1994c),[15] a ground clause C_0 is redundant with respect to a set of clauses N, if there is a subset $\{D_1, \ldots, D_n\}$ of $\overline{N}$ such that C_0 is entailed by $D_1, \ldots, D_n$ and each D_i is smaller than C_0. Similarly a ground inference is redundant, if its conclusion C_0 is entailed by $D_1, \ldots, D_n$ and each D_i is smaller than the maximal premise C_1. These definitions guarantee that C_0 is true in an interpretation whenever the clauses $D_1, \ldots, D_n$ are true.

[15] Note that "redundancy" is called "compositeness" in (Bachmair and Ganzinger 1994c). In later papers the standard terminology has changed.

If the non-maximal premises (if any) of a redundant inference are productive and $D_1, \ldots, D_n$ are true at the time where C_1 is inspected, we can show even more: As above, the conclusion C_0 is true in the current interpretation. Furthermore, by property (iii), the subclause C' of a productive clause $C' \vee e$ is false, hence all literals of C_0 that come from a non-maximal premise are false in the current interpretation. Together, these two facts allow us to prove that the maximal premise C_1 must be true in the current interpretation.

The theory axioms.

Cancellation is an operation that inherently involves both sides of an equation. To adapt the model construction depicted above to the cancellative superposition calculus, we generalize the notion of rewriting in such a way that simultaneous changes on both sides of an equation become possible. We do not rewrite each side of the equation by its own any longer, using a rewrite relation on *terms*: $t \to^* t'' \leftarrow^* t'$. Rather we use a rewrite relation on *equations*: $t \approx t' \to^* t'' \approx t''$. In this way, we can not only replace equals by equals in one side of the equation, we can also transform an equation $u + t \approx u + t'$ into the equivalent $t \approx t'$. Even more important, we can use a "rewrite rule" $mu + s \approx s'$ to transform $mu + t \approx t'$ into $s' + t \approx t' + s$. The interpretation induced by a set R of "rewrite rules" is now the set of all equations that can be rewritten to $0 \approx 0$ using R.

Example 4.1. Suppose that the rules $\underline{2b} + c \approx d$ and $\underline{2b'} + c \approx d$ are elements of our set of rewrite rules. Then we can, for instance, rewrite $2b \approx 2b'$ to $0 \approx 0$: We apply the first rule ①, then the second rule ②, and finally we cancel d and c ③.

$$\begin{array}{c}
2b \approx 2b' \\
①\big\downarrow \\
d \approx 2b' + c \\
②\big\downarrow \\
d + c \approx d + c \\
③\big\downarrow^{+} \\
0 \approx 0
\end{array}$$

The technique sketched so far would be sufficient to prove the completeness of our calculus, if $+$ were the only non-constant function symbol and $\Psi = \{1\}$. In the presence of free function symbols and the torsion-freeness axioms, however, further problems arise.

Example 4.2. In the previous example, we have shown that we can rewrite $2b \approx 2b'$ to $0 \approx 0$ using the two rules $\underline{2b} + c \approx d$ and $\underline{2b'} + c \approx d$.

However, there is no way to rewrite $f(2b) \approx f(2b')$ to $0 \approx 0$, although this equation is a consequence from the two rewrite rules and the theory axioms, just as $2b \approx 2b'$ is.

Similarly, if $2 \in \Psi$, then $3b \approx 3b'$ follows from the current set of rules and ACUKT_Ψ, hence it should be true in the current interpretation. Nevertheless, even our generalized form of rewriting is not powerful enough to rewrite $3b \approx 3b'$ to $0 \approx 0$. Consequently, the set of equations that can be rewritten to $0 \approx 0$ is neither a model of the equality axioms nor of the theory axioms ACUKT_Ψ.

A two-step approach solves the problem. Rather than constructing one set of rewrite rules to determine the truth or falsehood of an equation, we construct two such sets. Let us call the elements of these sets primary and secondary rules, respectively. In the beginning of the model construction both sets are empty. We use the current set of secondary rules to check whether some clause is true. If it is false, and if the other conditions for productivity are satisfied, then two things happen: First we turn the maximal positive literal e of the clause into a primary rule. Afterwards, we determine a certain set of rules $s \approx s'$ such that $\psi s \approx \psi s'$ can be rewritten to $0 \approx 0$ using the primary rule e and all the current secondary rules. This set of rules is added to the current set of secondary rules.

EXAMPLE 4.3. Let $\Psi = \mathbf{N}^{>0}$ and consider the set of ground clauses (1)–(4) ordered by (1) $\prec_{\mathrm{C}}$ (2) $\prec_{\mathrm{C}}$ (3) $\prec_{\mathrm{C}}$ (4):

$$\underline{2b'} + c \approx d \tag{1}$$
$$\underline{2b} + c \approx d \tag{2}$$
$$\underline{3b} \approx 3b' \tag{3}$$
$$\underline{f(2b)} \approx f(2b') \tag{4}$$

We start the model construction with empty sets of primary and secondary rules. Clause (1) is false in the empty interpretation, so $\underline{2b'} + c \approx d$ becomes a primary rule. As this rule can be rewritten to $0 \approx 0$ by itself, it is also added to the set of secondary rules.

Clause (2) is false in the interpretation generated by the current set of secondary rules. So $\underline{2b} + c \approx d$ becomes a primary rule and, again, a secondary rule. Furthermore, it is now possible to rewrite $\underline{2b} \approx 2b'$ to $0 \approx 0$ using the primary rule $\underline{2b} + c \approx d$ and the secondary rule $\underline{2b'} + c \approx d$, therefore $b \approx b'$ is turned into a secondary rule.[16]

As $\underline{b} \approx b'$ is now a secondary rule, $\underline{3b} \approx 3b'$ and $\underline{f(2b)} \approx f(2b')$ have a derivation to $0 \approx 0$ using the secondary rules. Clauses (3) and (4) are thus true in the current interpretation.

The integration of the theory axioms allows a generalization of the redundancy criterion. In the definition of redundancy, we can replace the usual entailment relation $\models_\approx$ by the theory entailment relation $\models_\Psi$. A clause is therefore redundant with respect to N, if it follows from smaller ground instances of clauses in N and from ground instances of the equality and ACUKT_Ψ axioms (analogously for inferences).[17]

It remains to show that the interpretations generated this way are in fact equational models of the given clauses and the theory axioms. To this end, we prove first that the generalized rewrite relation is confluent on the set of all equations that allow a derivation to $0 \approx 0$ (by analysis of various kinds of critical pairs). As an easy corollary we obtain that the interpretations satisfy ACUKT_Ψ and the equality axioms. Finally, we prove that the limit interpretation is a model of $\overline{N}$ whenever N is saturated and does not contain $\perp$.

[16] These two rules are not the only new secondary rules generated but the only ones that are relevant for this example. In fact the sets of secondary rules that are added in each step are usually infinite.

[17] In some theory superposition calculi, for example in AC-superposition (Wertz 1992, Bachmair and Ganzinger 1994a), there are ordering conditions not only for the instances of clauses of N, but also for the instances of the theory axioms. In our case there are no such requirements; the instances of the equality and ACUKT_Ψ axioms may be arbitrarily large.

This proof proceeds in essentially the same way as for standard superposition: Whenever we encounter a clause that is neither true in the current interpretation nor productive, then we can show that there is some non-redundant ground inference with this clause, which violates saturation.

Lifting.

The refutational completeness proof that we have sketched so far is based not on the clauses in N but on their ground instances: The interpretation is constructed from the set of ground instances, the proof that it is in fact a model proceeds by inspecting the ground instances. When we encounter a ground instance that it neither true in the current interpretation nor productive, we can show that some non-redundant ground inference with this clause is possible. In the calculus, however, we want to work with non-ground clauses, each of whom may represent an infinite number of ground instances. To connect these two worlds, we have to extend the definition of redundancy (and hence, of saturation) to non-ground clauses, and we have to relate inferences between clauses in N with inferences between instances of these clauses.

We have already defined ground instances of clauses, and we can do the same for inferences. If there is an inference from non-ground clauses and an inference from ground instances of these clauses, then the latter is called a ground instance of the former. The redundancy of non-ground clauses and inferences can now be defined via lifting: A non-ground clause or inference is redundant, if all its ground instances have this property.

As in the standard superposition calculus, not every inference from ground instances $C_k\theta, \dots, C_1\theta$ is a ground instance of an inference from $C_k, \dots, C_1$ (for instance, because the ground instances overlap at a variable position of C_1). The so-called lifting lemmas will show that all ground inferences that are actually needed in the refutational completeness proof are in fact instances of inferences from clauses in N.

4.2. Rewriting on Equations

When we want to show that the inference system described in Section 3 is refutationally complete we have to demonstrate that every saturated clause set not containing the empty clause has a model. To construct this model we need a rewrite relation on equations.

DEFINITION 4.4. *A ground equation e is called a cancellative rewrite rule with respect to $\succ$, if* $\mathrm{mt}(e)$ *does not occur on both sides of e.*

We will usually drop the attributes "cancellative" and "with respect to $\succ$", speaking simply of "rewrite rules".

Every rewrite rule has either the form $mu + s \approx s'$, where u is an atomic term, $m \in \mathbf{N}^{>0}$, $u \succ s$, and $u \succ s'$, or the form $u \approx s'$, where u (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 $mu + s \approx s'$ to an equation $mu + t \approx t'$ means to replace mu by s' and simultaneously to add s to the other side, obtaining $s' + t \approx t' + s$.[18]

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

DEFINITION 4.5. *Given a set R of rewrite rules, the three binary relations $\to_{\gamma,R}$, $\to_{\delta,R}$, and $\to_{\kappa}$ on ground equations are defined (modulo* ACU*) as follows:*

(i) $mu + t \approx t' \;\to_{\gamma,R}\; s' + t \approx t' + s$,
if $mu + s \approx s'$ *is a rule in* R.

(ii) $t[s] \approx t' \;\to_{\delta,R}\; t[s'] \approx t'$,
if (i) $s \approx s'$ *is a rule in* R *and (ii)* s *does not have sort* S_{CAM} *or* s *occurs in* t *below some free function symbol.*

(iii) $u + t \approx u + t' \;\to_{\kappa}\; t \approx t'$,
$u \approx u \;\to_{\kappa}\; 0 \approx 0$,
if u *is atomic and different from* 0.

The union of $\to_{\gamma,R}$, $\to_{\delta,R}$, and $\to_{\kappa}$ is denoted by $\to_R$.[19]

We say that an equation e is γ-reducible, if $e \to_{\gamma} e'$ (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 $t \approx t' \to_{\gamma} e'$, 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.

LEMMA 4.6. *Let R be a set of rewrite rules, s a term of sort S_{CAM}, $m \in \mathbf{N}^{>0}$. Then s is δ-reducible with respect to R if and only if ms is; $s \approx s'$ is δ-reducible (κ-reducible, $\delta\kappa$-reducible), if and only if $ms \approx ms'$ is.*

PROOF. The "only if" part is trivial, the "if" part follows from the fact that δ-steps can only take place at or below a free function symbol. $\square$

LEMMA 4.7. *The relation $\to_R$ is contained in $\succ_{\mathrm{L}}$ and thus noetherian.*

DEFINITION 4.8. *Given a set R of rewrite rules, the truth set $\mathrm{tr}(R)$ of R is the set of all equations $s \approx s'$ for which there exists a derivation $s \approx s' \;\to_R^*\; 0 \approx 0$. The Ψ-truth set $\mathrm{tr}_{\Psi}(R)$ of R is the set of all equations $s \approx s'$, such that either $s \approx s' \in \mathrm{tr}(R)$ and s does not have sort S_{CAM} or $\psi s \approx \psi s' \in \mathrm{tr}(R)$ for some $\psi \in \Psi$.*

LEMMA 4.9. *Let e be a rewrite rule and R be a set of rewrite rules. If e is contained in $\mathrm{tr}(R)$, then $\mathrm{mt}_{\#}(e)$ is reducible with respect to R.*

PROOF. Suppose that $e = mu + s \approx s'$, where $u = \mathrm{mt}(e)$, $m \in \mathbf{N}^{>0}$, $u \succ s$, and $u \succ s'$. Then there is a derivation

$$mu + s \approx s' \;\to_R^*\; 0 \approx 0\,.$$

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

The case that $e = u \approx s'$ and u does not have sort S_{CAM} is proved in the same way. $\square$

[19] As we deal only with ground terms and as there are no non-trivial contexts around equations, this operation does indeed satisfy the definition of a rewrite relation, albeit in an unorthodox way.

DEFINITION 4.10. *A ground clause $C' \vee e$ is called reductive for e, if e is a cancellative rewrite rule and strictly maximal in $C' \vee e$.*

DEFINITION 4.11. *Let N be a set of (possibly non-ground) clauses that does not contain the empty clause. Using induction on the clause ordering we define sets of rules E_C, E_C^Ψ, and R_C^Ψ, for all clauses $C \in \overline{N}$. Let C be such a clause and assume that E_D, E_D^Ψ, and R_D^Ψ have already been defined for all $D \in \overline{N}$ such that $C \succ_{\mathrm{C}} D$. Then the set R_C^Ψ is given by*

$$R_C^\Psi = \bigcup_{D \prec_{\mathrm{C}} C} E_D^\Psi .$$

The set E_C of primary rules for the clause 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 $\mathrm{tr}(R_C^\Psi)$, (iii) C' is false in $\mathrm{tr}_\Psi(R_C^\Psi \cup \{e\})$, and (iv) $\mathrm{mt}_\#(e)$ is irreducible with respect to R_C^Ψ. Otherwise, E_C is empty.

If $E_C = \{e\}$, then the set E_C^Ψ of secondary rules for C is the set of all rewrite rules $e' \in \mathrm{tr}_\Psi(R_C^\Psi \cup E_C)$ such that $\mathrm{mt}(e') = \mathrm{mt}(e)$ and e' is $\delta\kappa$-irreducible with respect to R_C^Ψ. Otherwise, E_C^Ψ is empty.

Finally, the set R_∞^Ψ is defined by

$$R_\infty^\Psi = \bigcup_{D \in \overline{N}} E_D^\Psi .$$

Primary rules are obtained directly from productive clauses in a similar way as in the model construction for standard superposition (Bachmair and Ganzinger 1994c). However, as we have seen in Subsection 4.1, the set of equations that can be rewritten to $0 \approx 0$ using primary rules is neither a model of the equality axioms nor of the theory axioms ACUKT_Ψ. Therefore primary rules are only an intermediate step in our model construction: At any stage of the model construction, we use the newly generated primary rule and the previously generated secondary rules to obtain new secondary rules. An equation may become a new secondary rule, if it is a rewrite rule, and if it is true with respect to the old secondary rules, the new primary rule and Ψ-torsion-freeness. Two more conditions imposed on secondary rules are not strictly speaking necessary but they simplify certain parts of the proof: We require that a new secondary rule has the same maximal term as the new primary rule and that it is $\delta\kappa$-irreducible with respect to the old secondary rules.

EXAMPLE 4.12. Suppose that $b \succ b' \succ c \succ d$ and let $\Psi = \mathbf{N}^{>0}$ and $N = \overline{N} = \{D, C\}$ where D is the clause $2b' + c \approx d$ and C is the clause $2b + c \approx d$.

The clause D is minimal in $\overline{N}$, hence $R_D^\Psi = \emptyset$.

As D is productive, $E_D = \{\, 2b' + c \approx d \,\}$. Now every equation of the form $2mb' + mc \approx md$ with $m \geq 1$ has an $(R_D^\Psi \cup E_D)$-derivation to $0 \approx 0$ (using m-fold γ-application of $2b' + c \approx d$ and $2m$ κ-steps), hence $R_C^\Psi = E_D^\Psi = \{\, 2mb' + mc \approx md \mid m \in \mathbf{N}^{>0} \,\}$.

The clause C is again productive, hence $E_C = \{\, 2b + c \approx d \,\}$. Then E_C^Ψ contains all equations of the form $mb + m'b' + \frac{1}{2}(m{+}m')c \approx \frac{1}{2}(m{+}m')d$ for $m \geq 1$, $m' \geq 0$, and $m + m'$ even, all equations of the form $mb + \frac{1}{2}(m{-}m')c \approx m'b' + \frac{1}{2}(m{-}m')d$ for $m \geq 1$, $m \geq m' \geq$

0, and $m + m'$ even, and all equations of the form $mb + \frac{1}{2}(m'{-}m)d \approx m'b' + \frac{1}{2}(m'{-}m)c$ for $m' \geq m \geq 1$ and $m + m'$ even.

Let us show this for the last case (the other cases are proved analogously). Consider the equation $e' = mb + \frac{1}{2}(m'{-}m)d \approx m'b' + \frac{1}{2}(m'{-}m)c$. Obviously e' has the maximal atomic term b, furthermore, e' is $\delta\kappa$-irreducible with respect to R_C^{Ψ}. It remains to show that $e' \in \mathrm{tr}_{\Psi}(R_C^{\Psi} \cup E_C)$, that is, that there exists a $\psi \in \Psi$ such that $m\psi b + \frac{1}{2}(m'{-}m)\psi d \approx m'\psi b' + \frac{1}{2}(m'{-}m)\psi c$ is contained in $\mathrm{tr}(R_C^{\Psi} \cup E_C)$. Let $\psi = 2$. Then there exists an $(R_C^{\Psi} \cup E_C)$-derivation from $2mb + (m'{-}m)d \approx 2m'b' + (m'{-}m)c$ to $0 \approx 0$:

$$
\begin{array}{c}
2mb + (m'{-}m)d \approx 2m'b' + (m'{-}m)c \\
① \Big\downarrow {}^{\gamma}_{+} \\
md + (m'{-}m)d \approx 2m'b' + (m'{-}m)c + mc \\
② \Big\downarrow {}^{\gamma}_{+} \\
m'c + md + (m'{-}m)d \approx m'd + (m'{-}m)c + mc \\
③ \Big\downarrow {}^{\kappa}_{+} \\
0 \approx 0
\end{array}
$$

This derivation consists of m-fold γ-application of $2b + c \approx d \in E_C$ ①, m'-fold γ-application of $2b' + c \approx d \in R_C^{\Psi}$ ②, and $2m'$ κ-steps to cancel $m'c + m'd$ on both sides of the equation ③.

Our goal is to show that $\mathrm{tr}(R_{\infty}^{\Psi})$ is an equality model of the axioms of Ψ-torsion-free 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^{Ψ} and R_{∞}^{Ψ}. In Subsection 4.4 we prove that the rewrite relations associated with R_C^{Ψ} and R_{∞}^{Ψ} satisfy a restricted confluence property. The equality and ACUKT_{Ψ} axioms follow as easy corollaries. Then we show in Subsection 4.7 that $\mathrm{tr}(R_{\infty}^{\Psi})$ is a model of N, provided that N is saturated and does not contain $\bot$.

Lemma 4.13. *If $E_C = \{mu + s \approx s'\}$, then there exist terms r and r' such that $mu + r \approx r'$ is contained in E_C^{Ψ}. If $E_C = \{u \approx s'\}$ and u does not have sort S_{CAM}, then there exists a term r' such that $u \approx r'$ is contained in E_C^{Ψ}.*

Proof. We prove the first part of the lemma, the second one being similar. Let $mu + r \approx r'$ be the result of $\delta\kappa$-normalizing $mu + s \approx s'$ with respect to R_C^{Ψ}.

$$
\begin{array}{c}
mu + s \approx s' \\
① \Big\downarrow {}^{\delta\cup\kappa}_{*} \\
mu + r \approx r'
\end{array}
$$

Then $u \succ s \succ r$ and $u \succ s' \succ r'$. Starting from $mu + r \approx r'$ we can construct a derivation

$$
\begin{array}{c}
mu + r \approx r' \\
\text{②}\big\downarrow \gamma \\
s' + r \approx r' + s \\
\text{③}\big\downarrow {}^{\delta\cup\kappa}_{*} \\
r' + r \approx r' + r \\
\text{④}\big\downarrow {}^{\kappa}_{*} \\
0 \approx 0
\end{array}
$$

where ② uses $mu + s \approx s'$ and ③ simulates ①. Hence $mu + r \approx r'$ is contained in $\mathrm{tr}_\Psi(R_C^\Psi \cup E_C)$ and thus in E_C^Ψ. □

The following lemma is proved in a similar way.

LEMMA 4.14. *For every $C \in \overline{N}$ we have $E_C \cup E_C^\Psi \subseteq \mathrm{tr}(R_C^\Psi \cup E_C^\Psi) \subseteq \mathrm{tr}(R_\infty^\Psi)$.*

LEMMA 4.15. *Let C and D be two clauses from $\overline{N}$ such that $C \succ_{\mathrm{C}} D$. If $e_1 \in E_C \cup E_C^\Psi$ and $e_2 \in E_D \cup E_D^\Psi$, then $\mathrm{mt}(e_1) \succ \mathrm{mt}(e_2)$.*

PROOF. As $\mathrm{mt}(e') = \mathrm{mt}(e'')$ for any two rules $e', e'' \in E_C \cup E_C^\Psi$, it suffices to consider the case that $e_1 \in E_C$ and $e_2 \in E_D$. 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 $E_D^\Psi \subseteq R_C^\Psi$, owing to Lemma 4.13. This is impossible, however. □

LEMMA 4.16. *Let u be atomic. If mu is γ-reducible with respect to E_C^Ψ for some $m \in \mathbf{N}^{>0}$ and $C \in \overline{N}$, then nu is δ-irreducible with respect to E_D^Ψ for every $n \in \mathbf{N}^{>0}$ and $D \in \overline{N}$.*

PROOF. If mu is γ-reducible, then there is a rule $ku + r \approx r' \in E_C^\Psi$ with $k \leq m$. Suppose that nu were δ-reducible by a rule $t \approx t' \in E_D^\Psi$. We distinguish between three cases:

If $D \prec_{\mathrm{C}} C$, then t would have to be a subterm of u. Consequently, u would be reducible with respect to R_C^Ψ, which is impossible by the definition of E_C.

If $D \succ_{\mathrm{C}} C$, then $t \succ k'u$ for every $k' \in \mathbf{N}^{>0}$, hence nu cannot be δ-reduced by $t \approx t'$.

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

4.4. CONFLUENCE

It is easy to see that the relations $\rightarrow_{R_C^\Psi}$ and $\rightarrow_{R_\infty^\Psi}$ are in general not confluent: For example let $N = \overline{N} = \{D, C\}$ where D is the clause $2c \approx d$ and C is the clause $b \not\approx 0$. Given the ordering $b \succ c \succ d$, we obtain $E_D = \{2c \approx d\}$, $E_C = \emptyset$, and $E_D^\Psi = R_C^\Psi = R_\infty^\Psi = \{\, 2mc \approx md \mid m \in \mathbf{N}^{>0} \,\}$. Now the equation $2c \approx c$ can be rewritten to $d \approx c$, using a γ-step, and also to $c \approx 0$, using a κ-step. Both equations are irreducible.

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

By definition, for every rewrite rule $ku + r \approx r'$ in some E_C^Ψ there is a $\psi \in \Psi$ and an $(R_C^\Psi \cup E_C)$-derivation from $\psi ku + \psi r \approx \psi r'$ to $0 \approx 0$. The following lemma shows that we may enforce a particular structure upon this derivation. Besides, it shows that for every finite set of rules from E_C^Ψ we may choose a single $\psi \in \Psi$ for all rules.

Lemma 4.17. *Let $E_C = \{mu + s = s'\}$, let I be a finite set of indices, and for every $i \in I$, let $k_i u + r_i \approx r_i'$ be a rule from E_C^Ψ. Then there is a $\psi \in \Psi$ such that for every $i \in I$ there is an $(R_C^\Psi \cup E_C)$-derivation*

$$\begin{array}{c}
\psi k_i u + \psi r_i \approx \psi r_i' \\
\textcircled{1}\Big\downarrow{}^{\gamma}_{+} \\
\chi_i s' + \psi r_i \approx \psi r_i' + \chi_i s \\
\textcircled{2}\Big\downarrow{}_{*} \\
0 \approx 0
\end{array}$$

where $\psi k_i = \chi_i m$. This derivation starts with χ_i-fold application of $mu + s \approx s'$ ①; the remaining steps use only rules from R_C^Ψ ②.

Proof. For every $i \in I$ there exists a $\psi_i \in \Psi$ and an $(R_C^\Psi \cup E_C)$-derivation ③.

$$\begin{array}{c}
\psi_i k_i u + \psi_i r_i \approx \psi_i r_i' \\
\textcircled{3}\Big\downarrow{}_{+} \\
0 \approx 0
\end{array}$$

by definition of E_C^Ψ. As Ψ is closed under multiplication, $\psi = \prod_{i \in I} \psi_i$ is contained in Ψ. From ③ it is easy to construct an $(R_C^\Psi \cup E_C)$-derivation ④.

$$\begin{tikzcd}
\psi k_i u + \psi r_i \approx \psi r_i' \arrow[dd, "\textcircled{4}"', "+"] \arrow[dr, "\textcircled{1}", "\gamma\ +"'] & \\
& \chi_i s' + \psi r_i \approx \psi r_i' + \chi_i s \arrow[dl, "\textcircled{2}", "*"'] \\
0 \approx 0 &
\end{tikzcd}$$

During ④ all occurrences of u are deleted eventually. As $k_i u + r_i \approx r_i'$ is $\delta\kappa$-irreducible with respect to R_C^Ψ, this can only happen by χ_i-fold γ-application of $mu + s = s'$, where $\psi k_i = \chi_i m$. These γ-steps are independent of any preceding rewrite steps. We can thus shift them to the front, obtaining a new derivation ①-②. As the remaining terms in the equation are smaller than u, the rewrite steps of ② can only use rules from R_C^Ψ. $\square$

We have mentioned above that the sets of secondary rules are usually infinite. In fact, if sufficiently many peaks can be joined, then *every* equation that is $\delta\kappa$-irreducible, different from $0 \approx 0$, and contained in the Ψ-truth set of R_C^Ψ (or $R_C^\Psi \cup E_C$) is a rewrite rule in R_C^Ψ (or $R_C^\Psi \cup E_C^\Psi$), as demonstrated by the next two lemmas.

LEMMA 4.18. *Let C be a clause in $\overline{N}$. If $e \in \mathrm{tr}_\Psi(R_C^\Psi)$ is $\delta\kappa$-irreducible with respect to R_C^Ψ, and $\to_{R_C^\Psi}$ is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e' \mid \mathrm{mt}(e) \succ \mathrm{mt}(e') \,\}$, then $e \in R_C^\Psi \cup \{0 \approx 0\}$. (Analogously for C replaced by ∞.)*

PROOF. We will prove the first part of the lemma, the proof of the second one being similar. Suppose that e is different from $0 \approx 0$ and let $v = \mathrm{mt}(e)$. By assumption, e is $\delta\kappa$-irreducible. If v did not have sort S_{CAM}, then it would also be γ-irreducible, so it could not be contained in $\mathrm{tr}_\Psi(R_C^\Psi)$. Hence we may suppose that e has the form $kv + t \approx t'$, where $v \succ t$ and $v \succ t'$. By definition of $\mathrm{tr}_\Psi(R_C^\Psi)$, there is a derivation

$$\psi' kv + \psi' t \approx \psi' t' \ \to^*_{R_C^\Psi} \ 0 \approx 0$$

for some $\psi' \in \Psi$. During this derivation all occurrences of v are deleted eventually. As e is $\delta\kappa$-irreducible, this can be done only by (possibly several) γ-rewriting steps, using a sequence of rules $e_i = k_i v + r_i \approx r_i'$ in R_C^Ψ for $i \in I$. By Lemma 4.15 all e_i are contained in the same E_D^Ψ for some $D \prec_{\mathrm{C}} C$. Since $\psi' kv$ is deleted completely, we have $\sum k_i = \psi' k$. (Here and in the rest of this proof, the summations range over all $i \in I$.) The remaining subterms in the equation are smaller than v. We may thus assume without loss of generality that the derivation has the form

$$\begin{array}{c}
\psi' kv + \psi' t \approx \psi' t' \\
① \Big\downarrow {}^{\gamma}_{+} \\
\sum r_i' + \psi' t \approx \psi' t' + \sum r_i \\
② \Big\downarrow {}_{*} \\
0 \approx 0
\end{array}$$

where the rewrite steps of ② use only rules from R_D^Ψ. Let $E_D = \{nv + s \approx s'\}$. According to Lemma 4.17, there exists a $\psi \in \Psi$ and for every $i \in I$ an $(R_D^\Psi \cup E_D)$-derivation

$$\begin{array}{c}
\psi k_i v + \psi r_i \approx \psi r_i' \\
③ \Big\downarrow {}^{\gamma}_{+} \\
\chi_i s' + \psi r_i \approx \psi r_i' + \chi_i s \\
④ \Big\downarrow {}_{*} \\
0 \approx 0
\end{array}$$

starting with χ_i-fold application of $nv + s \approx s'$ where $\psi k_i = \chi_i n$.

We construct a new derivation that combines ② and ④: By ψ-fold repetition of the steps of ② and by application of the steps of ④ for every $i \in I$, we obtain an R_D^Ψ-derivation ⑤ that starts from $\psi\,(\sum r_i' + \psi' t) + \sum(\chi_i s' + \psi r_i) \approx \psi\,(\psi' t' + \sum r_i) + \sum(\psi r_i' + \chi_i s)$.

$$
\begin{array}{c}
\psi \sum r_i' + \psi\psi' t + (\sum \chi_i)\, s' + \psi \sum r_i \approx \psi\psi' t' + \psi \sum r_i + \psi \sum r_i' + (\sum \chi_i)\, s \\
⑤ \Big\downarrow * \qquad \overset{⑥}{\searrow}{}^{*}_{\kappa} \\
\qquad\qquad \psi\psi' t + (\sum \chi_i)\, s' \approx \psi\psi' t' + (\sum \chi_i)\, s \\
\qquad \swarrow{}^{*} ⑦ \\
0 \approx 0
\end{array}
$$

Alternatively, we can cancel $\psi \sum r_i$ and $\psi \sum r_i'$ in the starting equation of ⑤, resulting in a derivation ⑥. By confluence, there is a derivation ⑦ which closes the diagram.

Noticing that $\psi\psi' k = \psi \sum k_i = n \sum \chi_i$, we see that we can rewrite $\psi\psi' kv + \psi\psi' t \approx \psi\psi' t'$ to the starting equation of ⑦ by $\sum \chi_i$-fold application of $nv + s \approx s' \in E_D$ ⑧.

$$
\begin{array}{c}
\psi\psi' kv + \psi\psi' t \approx \psi\psi' t' \\
⑧ \Big\downarrow {}^{\gamma}_{+} \\
\psi\psi' t + (\sum \chi_i)\, s' \approx \psi\psi' t' + (\sum \chi_i)\, s \\
⑦ \Big\downarrow {}_{*} \\
0 \approx 0
\end{array}
$$

As $\psi\psi' \in \Psi$ and $kv + t \approx t'$ is $\delta\kappa$-irreducible with respect to $R_D^\Psi \subseteq R_C^\Psi$, $kv + t \approx t'$ is contained in $E_D^\Psi \subseteq R_C^\Psi$ by Def. 4.11. □

Lemma 4.19. *If $C \in \overline{N}$, $e \in \mathrm{tr}_\Psi(R_C^\Psi \cup E_C)$ is $\delta\kappa$-irreducible with respect to $R_C^\Psi \cup E_C$, and $\rightarrow_{R_C^\Psi}$ is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e' \mid \mathrm{mt}(e) \succ \mathrm{mt}(e') \,\}$, then $e \in R_C^\Psi \cup E_C^\Psi \cup \{0 \approx 0\}$.*

Proof. If e is contained in $\mathrm{tr}_\Psi(R_C^\Psi)$, then $e \in R_C^\Psi \cup \{0 \approx 0\}$ by Lemma 4.18. Otherwise, let $E_C = \{nv + s \approx s'\}$ and $e = ku + t \approx t'$, such that $u = \mathrm{mt}(e)$. By definition of $\mathrm{tr}_\Psi(R_C^\Psi \cup E_C)$, there is a derivation

$$\psi ku + \psi t \approx \psi t' \;\rightarrow^*_{R_C^\Psi \cup E_C}\; 0 \approx 0$$

for some $\psi \in \Psi$. During this derivation all occurrences of u are deleted eventually. If u were larger than v, this would be impossible, as u is δ-irreducible with respect to $R_C^\Psi \cup E_C$. If u were smaller than v, then $nv + s \approx s'$ could not be used during this derivation, hence e would be contained in $\mathrm{tr}_\Psi(R_C^\Psi)$. Thus $u = v$, and by Def. 4.11, $e \in E_C^\Psi$. □

In the completeness proof for standard superposition, the rewrite systems that are generated are left-reduced: by construction there are no overlaps between left-hand sides of rules. In our case, overlaps between secondary rules derived from the same primary rule cannot be avoided. However, provided that sufficiently many peaks can be joined, the set of secondary rules has a closure property that allows us to handle such peaks. Intuitively, the following two lemmas show that the "difference" of two rewrite rules[20] from R_C^Ψ is either $0 \approx 0$ or also a rewrite rule from R_C^Ψ.

LEMMA 4.20. *Let $\{C, D, D_1\} \subseteq \overline{N}$, such that $C \succ_\mathrm{C} D \succeq_\mathrm{C} D_1$. Let $k_0 v + r_0 \approx r_0' \in E_D^\Psi$, $k_1 v + r_1 \approx r_1' \in E_{D_1}^\Psi$, where $k_0 > 0$ and $k_0 \geq k_1$.*[21] *Let w be the common part of r_0 and r_1, let w' be the common part of r_0' and r_1', and for $i \in \{0,1\}$, let $r_i = w + q_i$ and $r_i' = w' + q_i'$. If $\to_{R_C^\Psi}$ is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e' \mid v \succ \mathrm{mt}(e') \,\}$, then $(k_0 - k_1)v + (q_0 + q_1') \approx (q_0' + q_1) \in E_D^\Psi \cup R_D^\Psi \cup \{0 \approx 0\}$. (Analogously for C replaced by ∞.)*

PROOF. Let $E_D = \{nv + t \approx t'\}$. Then there are a $\psi \in \Psi$ and $(R_D^\Psi \cup E_D)$-derivations

$$\begin{tikzcd}
\psi k_0 v + \psi r_0 \approx \psi r_0' \arrow[d, "\gamma", "+"', "①"'] & \psi r_1' \approx \psi r_1 + \psi k_1 v \arrow[d, "\gamma"', "*", "③"] \\
\chi_0 t' + \psi r_0 \approx \psi r_0' + \chi_0 t \arrow[d, "*", "②"'] & \chi_1 t + \psi r_1' \approx \psi r_1 + \chi_1 t' \arrow[d, "*"', "④"] \\
0 \approx 0 & 0 \approx 0
\end{tikzcd}$$

Each derivation starts with χ_i-fold application of $nv + t \approx t'$, where $\psi k_i = \chi_i n$. If $D_1 = D$, this follows from Lemma 4.17; if $D_1 \prec_\mathrm{C} D$, it follows from Lemma 4.17 and Lemma 4.14.

Consider the two starting equations of ② and ④. 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 ⑤ of ② and ④.

$$\begin{tikzcd}
\chi_0 t' + \chi_1 t + \psi(r_0 + r_1') \approx \psi(r_0' + r_1) + \chi_0 t + \chi_1 t' \arrow[dd, "*", "⑤"'] \arrow[dr, "\kappa"', "*", "⑥"] & \\
& (\chi_0 - \chi_1)t' + \psi(q_0 + q_1') \approx \psi(q_0' + q_1) + (\chi_0 - \chi_1)t \arrow[dl, "*", "⑦"'] \\
0 \approx 0 &
\end{tikzcd}$$

Above, we have defined w as the common part of r_0 and r_1, w' as the common part of r_0' and r_1', and q_i, q_i' as the respective remainders. We can therefore construct an alternative derivation ⑥ by cancelling $\chi_1 t$, $\chi_1 t'$, and $\psi(w + w')$ in the starting equation of ⑤. By confluence, there is a derivation ⑦ which closes the diagram.

Since $\psi(k_0 - k_1) = n(\chi_0 - \chi_1)$, we can rewrite $\psi(k_0 - k_1)v + \psi(q_0 + q_1') \approx \psi(q_0' + q_1)$ to the starting equation of ⑦ by $(\chi_0 - \chi_1)$-fold application of $nv + t \approx t' \in E_D$ ⑧.

[20] The difference of two rewrite rules $s \approx s'$ and $t \approx t'$ is the result of κ-normalizing $s + t' \approx s' + t$.

[21] Deviating from our standard notational convention we allow $k_1 = 0$ (if and only if $D_1 \prec_\mathrm{C} D$) so that we can handle the cases $D_1 \prec_\mathrm{C} D$ and $D_1 = D$ simultaneously.

$$
\begin{array}{c}
\psi(k_0-k_1)v + \psi(q_0+q_1') \approx \psi(q_0'+q_1) \\
⑧\ \Big\downarrow {\scriptstyle \gamma}\ {\scriptstyle *} \\
(\chi_0-\chi_1)t' + \psi(q_0+q_1') \approx \psi(q_0'+q_1) + (\chi_0-\chi_1)t \\
⑦\ \Big\downarrow {\scriptstyle *} \\
0 \approx 0
\end{array}
$$

As $k_0v + r_0 \approx r_0'$ and $k_1v + r_1 \approx r_1'$ are $\delta\kappa$-irreducible with respect to $R_D^\Psi \cup E_D$, so is $e = (k_0-k_1)v + (q_0+q_1') \approx (q_0'+q_1)$. By Lemma 4.19, $e \in R_D^\Psi \cup E_D^\Psi \cup \{0 \approx 0\}$. □

LEMMA 4.21. *Let $\{C, D\} \subseteq \overline{N}$, such that $C \succ_{\mathrm{C}} D$. Let $v \approx r_0'$ and $v \approx r_1'$ be rules in E_D^Ψ, where v does not have sort S_{CAM}. If $\rightarrow_{R_C^\Psi}$ is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e' \mid v \succ \mathrm{mt}(e') \,\}$, then $r_0' = r_1'$. (Analogously for C replaced by ∞.)*

PROOF. Suppose that $E_D = \{v \approx t'\}$. As $v \approx r_0'$ and $v \approx r_1'$ are $\delta\kappa$-irreducible with respect to R_D^Ψ and v does not have sort S_{CAM}, there are $(R_D^\Psi \cup E_D)$-derivations

$$
\begin{array}{ccc}
v \approx r_0' & \qquad\qquad & v \approx r_1' \\
① \Big\downarrow {\scriptstyle \delta} & & {\scriptstyle \delta} \Big\downarrow ④ \\
t' \approx r_0' & & t' \approx r_1' \\
② \Big\downarrow {\scriptstyle \delta}\,{\scriptstyle *} & & {\scriptstyle *}\,{\scriptstyle \delta} \Big\downarrow ⑤ \\
r_0' \approx r_0' & & r_1' \approx r_1' \\
③ \Big\downarrow {\scriptstyle \kappa} & & {\scriptstyle \kappa} \Big\downarrow ⑥ \\
0 \approx 0 & & 0 \approx 0
\end{array}
$$

where ① and ④ use $v \approx t'$ and ② and ⑤ use rules from R_D^Ψ. As all δ-steps take place only on the left-hand sides of the equations, we can use the same rules as in ② and ⑤ to rewrite $t' \approx t'$ to $r_0' \approx r_1'$ ⑦.

$$
\begin{tikzcd}
t' \approx t' \arrow[dd, "⑧\ \kappa"'] \arrow[dr, "⑦\ \delta\ *"] & \\
& r_0' \approx r_1' \arrow[dl, "⑨\ *"] \\
0 \approx 0 &
\end{tikzcd}
$$

On the other hand, we can rewrite $t' \approx t'$ immediately to $0 \approx 0$ ⑧. By confluence, there is a derivation ⑨. As r'_0 and r'_1 are δ-irreducible and do not have sort S_{CAM}, ⑨ must consist of a single κ-step, hence $r'_0 = r'_1$. $\square$

THEOREM 4.22. *The relation* $\to_{R_C^\Psi}$ *is confluent on* $\mathrm{tr}(R_C^\Psi)$ *for every* $C \in \overline{N}$. *The relation* $\to_{R_\infty^\Psi}$ *is confluent on* $\mathrm{tr}(R_\infty^\Psi)$.

PROOF. Let us consider the relation $\to_{R_C^\Psi}$. (The case of $\to_{R_\infty^\Psi}$ is similar.) Traditionally the confluence of a noetherian relation is established in two steps. First, one proves by induction that the confluence of a noetherian relation follows from local confluence. Second, one shows that local confluence is implied by the convergence of certain critical pairs. In our case, the induction hypothesis is not only needed to show that local confluence implies confluence, but even to prove local confluence. Consequently, we have to embed the analysis of the critical pairs within the inductive confluence proof.

To show that $\to_{R_C^\Psi}$ is confluent on $\mathrm{tr}(R_C^\Psi)$, it suffices to show that it is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e \mid e_0 \succeq_{\text{L}} e \,\}$ for every $e_0 \in \mathrm{tr}(R_C^\Psi)$. We will do this by induction on the size of e_0 with respect to $\succ_{\text{L}}$. According to Lemma 2.1, we have to prove that for any peak

$$\begin{array}{ccc} & e & \\ {}^{R_C^\Psi}\swarrow & & \searrow^{R_C^\Psi} \\ e_1 & & e_2 \end{array}$$

such that $e_0 \succeq_{\text{L}} e$ and e_1 or e_2 can be reduced to $0 \approx 0$, there exists an e_3 such that

$$\begin{array}{ccc} e_1 & & e_2 \\ {}_{R_C^\Psi}\searrow^{*} & & {}^{*}\swarrow_{R_C^\Psi} \\ & e_3 & \end{array}$$

For $e_0 \succ_{\text{L}} e$, this follows directly from the induction hypothesis, so we assume $e_0 = e$.

Case 1: Trivial peaks.

As in traditional term rewriting, a peak converges if the two rewrite steps take place at disjoint redexes. Furthermore, local confluence is obvious, if one step is a κ-step, and the other one is a δ- or a κ-step. By Lemma 4.16, γ- and δ-steps can only take place at disjoint redexes. It remains thus to consider γ/γ-peaks, γ/κ-peaks, and δ/δ-peaks.

Case 2: γ/γ-peaks.

If two γ-steps take place at non-disjoint redexes, then both rewrite rules must be derived from the same $E_D = \{nv + t \approx t'\}$. Consider the two rules $k_0 v + r_0 \approx r'_0$ and $k_1 v + r_1 \approx r'_1$ from E_D^Ψ. Without loss of generality, let $k_0 \geq k_1$. If the two rules are applied to an equation $k_0 v + s \approx s'$ we obtain a peak

$$\begin{array}{ccc} & k_0 v + s \approx s' & \\ {}^{①}\swarrow_{\gamma} & & {}_{\gamma}\searrow^{②} \\ r'_0 + s \approx s' + r_0 & & (k_0 - k_1)v + r'_1 + s \approx s' + r_1 \end{array}$$

Let w be the common part of r_0 and r_1, let w' be the common part of r'_0 and r'_1, and for $i \in \{0,1\}$, let $r_i = w + q_i$ and $r'_i = w' + q'_i$. By the induction hypothesis, $\rightarrow_{R_C^\Psi}$ is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e \mid e' \succeq_{\mathrm{L}} e \,\}$ for every e' that is smaller than $k_0 v + s \approx s'$; so it is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e \mid v \succ \mathrm{mt}(e) \,\}$. We can thus apply Lemma 4.20, which yields that the equation $(k_0 - k_1)v + (q_0 + q'_1) \approx (q'_0 + q_1)$ is either $0 \approx 0$ or a rule in $E_D^\Psi \cup R_D^\Psi$. If it is $0 \approx 0$, there is nothing to show: the peak is trivial. Otherwise, we distinguish between two cases.

If $k_0 > k_1$, we can close the peak by γ-application of $(k_0 - k_1)v + (q_0 + q'_1) \approx (q'_0 + q_1)$ ③ and by cancellation of $q_1 + q'_1$ ④.

$$\begin{tikzcd}
k_0 v + s \approx s' \arrow[dr, "\textcircled{2}", "\gamma"'] \arrow[ddd, "\textcircled{1}"', "\gamma"] & \\
& (k_0 - k_1)v + q'_1 + w' + s \approx s' + q_1 + w \arrow[d, "\gamma"', "\textcircled{3}"] \\
& q'_0 + q_1 + q'_1 + w' + s \approx s' + q_1 + w + q_0 + q'_1 \arrow[dl, "\kappa"', "\textcircled{4}", "*" near end] \\
q'_0 + w' + s \approx s' + q_0 + w &
\end{tikzcd}$$

Otherwise, $k_0 = k_1$. Let $mu = \mathrm{mt}_\#(q_0 + q'_1 \approx q'_0 + q_1)$. Without loss of generality assume that mu occurs on the left-hand side of this equation (the other case is proved analogously), hence let $q_0 + q'_1 = mu + q_2$. Consider the equation $q'_1 + q'_0 + w' + s \approx s' + w + q_0 + q'_1$. We can construct two derivations starting from here: one by cancelling q'_1 ⑤, the other one by applying $mu + q_2 \approx q'_0 + q_1$ ⑥ and cancelling $q_2 + q'_0$ ⑦.

$$\begin{tikzcd}
q'_1 + q'_0 + w' + s \approx s' + w + q_0 + q'_1 \arrow[dr, "\textcircled{6}", "\gamma"'] \arrow[dd, "\textcircled{5}"', "\kappa", "*" near end] & \\
& q_2 + q'_1 + q'_0 + w' + s \approx s' + w + q_2 + q'_0 + q_1 \arrow[d, "\kappa"', "\textcircled{7}", "*"' near end] \\
q'_0 + w' + s \approx s' + w + q_0 & q'_1 + w' + s \approx s' + w + q_1
\end{tikzcd}$$

The derivations ⑤ and ⑥-⑦ lead to the same equations as ① and ②. By assumption, one of these two equations can be reduced to $0 \approx 0$. As $k_0 v + s \approx s'$ is larger than $q'_1 + q'_0 + w' + s \approx s' + q_0 + w + q'_1$, we can use the induction hypothesis to show that ⑤ and ⑥-⑦ can be joined. The joinability of ① and ② follows immediately.

Case 3: γ/κ-peaks.

Closing a peak between a κ-step and a γ-step is trivial if the latter takes place at some free function symbol. It suffices therefore to consider the application of a rewrite rule $kv + r \approx r' \in E_D^\Psi \subseteq R_C^\Psi$ with $k \geq 2$ at the top of an equation $kv + s \approx v + s'$. This yields a peak

$$\begin{tikzcd}[column sep=small]
 & kv + s \approx v + s' \arrow[dl, "①"', "\gamma"] \arrow[dr, "②", "\kappa"'] & \\
r' + s \approx v + s' + r & & (k-1)v + s \approx s'
\end{tikzcd}$$

where either $r' + s \approx v + s' + r$ or $(k-1)v + s \approx s'$ can be rewritten to $0 \approx 0$ by R_C^Ψ.

Case 3.1: $r' + s \approx v + s' + r \to^* 0 \approx 0$.

At some step of the R_C^Ψ-derivation $r' + s \approx v + s' + r \to^* 0 \approx 0$ the term v must be eventually deleted. By Lemma 4.16 v is δ-irreducible, so this can happen only by a γ-step or a κ-step.

Case 3.1.1: v is deleted by a γ-step.

Suppose that the deletion happens by application of a rule $k_1 v + r_1 \approx r_1' \in E_D^\Psi$. Such a step requires the presence of $k_1 - 1$ further occurrences of v. As r and r' are smaller than v, these occurrences can only be derived from s or s'. We may thus assume without loss of generality that the derivation has the form ③-④-⑤:

$$\begin{tikzcd}[column sep=small]
 & kv + s \approx v + s' \arrow[dl, "①"', "\gamma"] \arrow[dr, "②", "\kappa"'] & \\
r' + s \approx v + s' + r \arrow[d, "③"', "*"] & & (k-1)v + s \approx s' \arrow[d, "⑦", "*"'] \\
r' + t \approx v + t' + (k_1 - 1)v + r \arrow[d, "④"', "\gamma"] & & (k-1)v + t \approx (k_1 - 1)v + t' \arrow[d, "⑧", "\kappa\,*"'] \\
r_1 + r' + t \approx t' + r_1' + r \arrow[d, "⑤"', "*"] \arrow[dr, "\kappa\,*", "⑥"'] & & (k - k_1)v + t \approx t' \arrow[dl, "\gamma"', "⑨"] \\
0 \approx 0 & q_1 + q' + t \approx t' + q_1' + q &
\end{tikzcd}$$

Let w be the common part of r and r_1, let w' be the common part of r' and r_1', and let $r = w + q$, $r_1 = w + q_1$, $r' = w' + q'$, and $r_1' = w' + q_1'$. We can thus use κ-steps ⑥ to cancel $w + w'$ in $r_1 + r' + t \approx t' + r_1' + r$, obtaining $q_1 + q' + t \approx t' + q_1' + q$. As the steps ③ take place only at s and s', we can simulate them by ⑦. Now we have to distinguish between two cases. If $k \neq k_1$, then we can first cancel the smaller of $(k_1 - 1)v$ or $(k-1)v$ ⑧. Let us assume that $k > k_1$, the case of $k_1 > k$ is proved similarly. By Lemma 4.20, $(k - k_1)v + (q + q_1') \approx (q' + q_1)$ is contained in E_D^Ψ; γ-application ⑨ of this rule closes the diagram.

If $k = k_1$, then by Lemma 4.20, $(q + q_1') \approx (q' + q_1)$ is either $0 \approx 0$ or contained in R_D^Ψ. If it is $0 \approx 0$, then the derivations ⑥ and ⑦ end at the same equation, so the peak is already joined. Otherwise, let $mu = \mathrm{mt}_\#(q + q_1' \approx q' + q_1)$. Without loss of generality assume that mu occurs on the left-hand side of this equation, that is, $q + q_1' = mu + q_2$ (the other case is similar). We can thus close the diagram by γ-application of $mu + q_2 \approx q' + q_1$ ⑩

followed by cancellation of $q_2 + q_1 + q'$ ⑪ on the one side, and by cancellation of $(k-1)v$ ⑫ on the other side.

$$
\begin{tikzcd}
 & kv + s \approx v + s' \arrow[dl, "①"', "\gamma"] \arrow[dr, "②", "\kappa"'] & \\
r' + s \approx v + s' + r \arrow[d, "③"', "*"] & & (k-1)v + s \approx s' \arrow[d, "*"', "⑦"] \\
r' + t \approx v + t' + (k-1)v + r \arrow[d, "④"', "\gamma"] & & (k-1)v + t \approx (k-1)v + t' \arrow[ddddd, "\kappa"', "⑫", "*"'] \\
r_1 + r' + t \approx t' + r_1' + r \arrow[dddd, "⑤"', "*"] \arrow[dr, "\kappa", "⑥"', "*"] & & \\
 & q_1 + q' + t \approx t' + mu + q_2 \arrow[d, "⑩"', "\gamma"] & \\
 & q_2 + q_1 + q' + t \approx t' + q' + q_1 + q_2 \arrow[ddr, "\kappa", "⑪"', "*"] & \\
 & & \\
0 \approx 0 & & t \approx t'
\end{tikzcd}
$$

Case 3.1.2: v is deleted by a κ-step.

The deletion of v by a κ-step requires the existence of another occurrence of v on the left-hand side. Again, this occurrence can only be derived from s or s'. We may thus assume that the derivation has the form ⑬-⑭-⑮:

$$
\begin{tikzcd}
 & kv + s \approx v + s' \arrow[dl, "①"', "\gamma"] \arrow[dr, "②", "\kappa"'] & \\
r' + s \approx v + s' + r \arrow[d, "⑬"', "*"] & & (k-1)v + s \approx s' \arrow[d, "*"', "⑯"] \\
r' + t + v \approx v + t' + r \arrow[d, "⑭"', "\kappa"] & & (k-1)v + t + v \approx t' \arrow[dll, "\gamma"', "⑰"] \\
r' + t \approx t' + r \arrow[d, "⑮"', "*"] & & \\
0 \approx 0 & &
\end{tikzcd}
$$

As the steps ⑬ take place only at s and s', we can simulate them by ⑯. Finally, we can close the diagram using γ-rewriting ⑰ by $kv + r \approx r'$.

Case 3.2: $(k-1)v + s \approx s' \to^* 0 \approx 0$.

If the R_C^Ψ-derivation of $(k-1)v + s \approx s'$ to $0 \approx 0$ consists only of κ-steps, then $(k-1)v + s$ is identical to s', so joining the peak is trivial. If the derivation contains at least one γ- or δ-step, then it has the form ⑱-⑲-⑳.

$$
\begin{tikzcd}
 & kv + s \approx v + s' \arrow[dl, "\gamma"', "\textcircled{1}"] \arrow[dd, "\gamma\cup\delta"', "\textcircled{21}"] \arrow[dr, "\kappa", "\textcircled{2}"'] & \\
r' + s \approx v + s' + r & & (k-1)v + s \approx s' \arrow[d, "\kappa\ *"', "\textcircled{18}"] \\
 & e_5 \arrow[dr, "\kappa\ +", "\textcircled{22}"'] & e_3 \arrow[d, "\gamma\cup\delta"', "\textcircled{19}"] \\
 & & e_4 \arrow[d, "*"', "\textcircled{20}"] \\
 & & 0 \approx 0
\end{tikzcd}
$$

Step ⑲ is independent of the preceding κ-steps, hence we can shift it to the front, obtaining a derivation ㉑-㉒-⑳. It remains to join the peak between ① and ㉑. This is done as in Case 2 if ① and ㉑ are γ-steps with overlapping redexes, it is trivial if ㉑ is a γ-step at a disjoint redex or a δ-step.

Case 4: δ/δ-peaks.

It remains to show that every δ/δ-peak converges. Suppose that the first rewrite step uses a rule $t_0 \approx r_0'$ from some E_D^Ψ, and that the second rewrite step uses a rule $t_1 \approx r_1'$ from some $E_{D_1}^\Psi$, where $D \succeq_{\mathrm{C}} D_1$. If the redexes are disjoint, there is nothing to show. As all rules in E_D^Ψ are δ-irreducible with respect to R_D^Ψ, the two rules cannot overlap below a free function symbol. We may thus suppose that the two rules rewrite the same redex or overlapping parts of a sum in the equation e_0. If t_0 and t_1 have sort S_{CAM}, let $v = \mathrm{mt}(t_0 \approx r_0')$ and let $t_i = k_i v + r_i$ for $i \in \{0, 1\}$. Deviating from our standard notational convention we allow $k_1 = 0$ (if and only if $D \succ_{\mathrm{C}} D_1$) so that we can handle the cases $D \succ_{\mathrm{C}} D_1$ and $D = D_1$ simultaneously. If $D = D_1$, we assume by symmetry that $k_0 \geq k_1$. Let w be the common part of r_0 and r_1, let w' be the common part of r_0' and r_1', and for $i \in \{0, 1\}$, let $r_i = w + q_i$ and $r_i' = w' + q_i'$. The peak has the form

$$
\begin{tikzcd}
 & e_0[k_0 v + w + q_0 + q_1] \arrow[dl, "\delta"', "\textcircled{1}"] \arrow[dr, "\delta", "\textcircled{2}"'] & \\
e_0[(k_0 - k_1)v + q_0 + q_1' + w'] & & e_0[q_0' + q_1 + w']
\end{tikzcd}
$$

By Lemma 4.20, $(k_0 - k_1)v + (q_0 + q_1') \approx (q_0' + q_1)$ is either $0 \approx 0$ or a rule in $E_D^\Psi \cup R_D^\Psi$. If it is $0 \approx 0$, then the peak is trivial; otherwise, we can join the peak between ① and ② as follows:

$$\begin{tikzcd}[column sep=small]
 & e_0[k_0 v + w + q_0 + q_1] \arrow[dl, "①"', "\delta"] \arrow[dr, "②", "\delta"'] & \\
e_0[(k_0 - k_1)v + q_0 + q_1' + w'] \arrow[rr, leftrightarrow, "\delta", "③"'] & & e_0[q_0' + q_1 + w']
\end{tikzcd}$$

where step ③ uses $(k_0 - k_1)v + (q_0 + q_1') \approx (q_0' + q_1)$.

It remains to show that the peak can be joined if t_0 and t_1 do not have sort S_{CAM}. This is proved similarly, using Lemma 4.21 rather than Lemma 4.20. $\square$

Corollary 4.23. *For every $C \in \overline{N}$, $\mathrm{tr}(R_C^\Psi)$ and $\mathrm{tr}(R_\infty^\Psi)$ satisfy the equality axioms.*

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

$$\begin{tikzcd}
r \approx s \arrow[d, "①"', "*"] & s \approx t \arrow[d, "②", "*"'] \\
0 \approx 0 & 0 \approx 0
\end{tikzcd}$$

If r, s and t have sort S_{CAM}, we can combine the derivations ① and ② and obtain a derivation ③:

$$\begin{tikzcd}
r + s \approx s + t \arrow[dd, "③"', "*"] \arrow[dr, "④", "\kappa\,*"'] & \\
 & r \approx t \arrow[dl, "⑤", "*"'] \\
0 \approx 0 &
\end{tikzcd}$$

On the other hand, we can use κ-steps ④ to cancel s on both sides of the equation. By Thm. 4.22, there is a derivation ⑤, hence $r \approx t \in \mathrm{tr}(R_C^\Psi)$.

If r, s and t do not have sort S_{CAM}, the derivations ① and ② must have the form ⑥-⑦ and ⑧-⑨:

$$
\begin{tikzcd}
r \approx s \arrow[d, "\delta", "*"', "⑥"'] & & s \approx t \arrow[d, "⑧", "\delta"', "*"'] \\
u \approx u \arrow[d, "\kappa", "⑦"'] & & v \approx v \arrow[d, "⑨", "\kappa"'] \\
0 \approx 0 & & 0 \approx 0
\end{tikzcd}
$$

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

$$
\begin{tikzcd}
s \approx s \arrow[dd, "\kappa", "⑫"'] \arrow[dr, "\delta", "⑩"', "*"] & & r \approx t \arrow[dl, "\delta"', "⑪", "*"'] \\
& u \approx v \arrow[dl, "⑬", "+"'] & \\
0 \approx 0 & &
\end{tikzcd}
$$

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

For the congruence axiom we have to show that $s \approx t \in \mathrm{tr}(R_C^\Psi)$ entails $r[s] \approx r[t] \in \mathrm{tr}(R_C^\Psi)$. If s does not have sort S_{CAM}, or if there is no free function symbol in r above s, this is trivial, so assume that s occurs in r below a free symbol. Consider derivation ①:

$$
\begin{tikzcd}
s \approx t \arrow[ddd, "*", "①"'] \arrow[dr, "⑫" description, phantom] \arrow[dr, "②", "\delta"', "*"] & \\
& w + w_0 \approx w' + w_0 \arrow[d, "③", "\kappa"', "*"'] \\
& w \approx w' \arrow[dl, "④", "*"'] \\
0 \approx 0 &
\end{tikzcd}
$$

We can $\delta\kappa$-normalize $s \approx t$, first by δ-rewriting s to $w + w_0$ and t to $w' + w_0$ ②, then by cancelling ③ the common part w_0. According to Thm. 4.22, there exists a derivation ④. The equation $w \approx w'$ is $\delta\kappa$-irreducible with respect to R_C^Ψ, hence it is contained in $R_C^\Psi \cup \{0 \approx 0\}$ by Lemma 4.18. Without loss of generality we assume $w \succeq w'$. This allows us to construct the following derivation:

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

where step 5 simulates 2 and step 6 uses $w \approx w'$ (if different from $0 \approx 0$). Summarizing we get $r[s] \approx r[t] \in \mathrm{tr}(R_C^\Psi)$. $\square$

COROLLARY 4.24. *For every $C \in \overline{N}$, $\mathrm{tr}(R_C^\Psi)$ and $\mathrm{tr}(R_\infty^\Psi)$ satisfy* ACUKT_Ψ.

PROOF. The proof of the cancellation axiom is analogous to the proof of the transitivity axiom; the Ψ-torsion-freeness axiom is proved in a similar way as the congruence axiom (Cor. 4.23). The associative, commutative, and identity axioms are obvious. $\square$

COROLLARY 4.25. *For every clause $C \in \overline{N}$, $\mathrm{tr}_\Psi(R_C^\Psi) = \mathrm{tr}(R_C^\Psi)$ and $\mathrm{tr}_\Psi(R_\infty^\Psi) = \mathrm{tr}(R_\infty^\Psi)$.*

COROLLARY 4.26. *Let e be a rewrite rule in $\mathrm{tr}(R_\infty^\Psi)$, such that $mv = \mathrm{mt}_\#(e)$ is γ-reducible with respect to E_D^Ψ. Then $E_D = \{nv + t \approx t'\}$ and there is a $\chi \in \mathbf{N}^{>0}$ and a $\psi \in \Psi$ such that $\psi m = \chi n$ and $\gcd(\psi, \chi) = 1$.*

PROOF. By Lemma 4.16, mv is δ-irreducible with respect to R_∞^Ψ. Hence $\delta\kappa$-normalization of $e = mv + s \approx s'$ yields an equation $mv + r \approx r'$, which is contained in $\mathrm{tr}(R_\infty^\Psi)$ since $\to_{R_\infty^\Psi}$ is confluent on $\mathrm{tr}(R_\infty^\Psi)$. By Thm. 4.22, Lemma 4.18, and Lemma 4.15, $mv + r \approx r'$ is a rule in E_D^Ψ. According to Lemma 4.17, there is a $\chi_0 \in \mathbf{N}^{>0}$ and a $\psi_0 \in \Psi$ such that $\psi_0 m = \chi_0 n$. Define $\chi = \chi_0/\gcd(\psi_0, \chi_0)$ and $\psi = \psi_0/\gcd(\psi_0, \chi_0)$, then $\chi \in \mathbf{N}^{>0}$, $\psi \in \Psi$, $\psi m = \chi n$, and $\gcd(\psi, \chi) = 1$. $\square$

COROLLARY 4.27. *Let $C = C' \vee e_2 \vee e_1$ be reductive for $e_1 = mu + s \approx s'$. Suppose that mu is irreducible with respect to R_C^Ψ and that e_2 is contained in $\mathrm{tr}_\Psi(R_C^\Psi \cup \{e_1\}) \setminus \mathrm{tr}(R_C^\Psi)$. Then e_2 has the form $nu + t \approx t'$ with $n > 0$, and there exists a $\psi \in \Psi$ and $\chi \in \mathbf{N}^{>0}$ such that $\gcd(\psi, \chi) = 1$, $\chi m = \psi n$, and $\chi s' + \psi t \approx \psi t' + \chi s \in \mathrm{tr}(R_C^\Psi)$.*

PROOF. Let $v = \mathrm{mt}(e_2)$ and $e_2 = nv + t \approx n'v + t'$ with $n \geq n' \geq 0$ and $n > 0$. As $e_1 \succ_{\mathrm{L}} e_2$, it is obvious that $u \succeq v$. Choose $\psi_0 \in \Psi$ such that $\psi_0 nv + \psi_0 t \approx \psi_0 n'v + \psi_0 t'$ has an $(R_C^\Psi \cup \{e_1\})$-derivation to $0 \approx 0$ that contains at least one rewriting step using e_1. If v were smaller than u, this would be impossible, hence $v = u$. Furthermore, n' must be 0, otherwise, due to the third component of the ordering quintuples, e_2 would be larger than e_1.

We have required mu to be irreducible with respect to R_C^Ψ. During the derivation of

$\psi_0 nu + \psi_0 t \approx \psi_0 t'$ to $0 \approx 0$, the occurrences of u can thus only be eliminated by χ_0-fold γ-application of e_1 ①, where $\psi_0 n = \chi_0 m$.

$$\begin{array}{c} \psi_0 nu + \psi_0 t \approx \psi_0 t' \\ ① \Big\downarrow {}^{\gamma}_{+} \\ \chi_0 s' + \psi_0 t \approx \psi_0 t' + \chi_0 s \\ ② \Big\downarrow {}_{*} \\ 0 \approx 0 \end{array}$$

All remaining terms in the equation are smaller than u, so the following rewrite steps can only use rules from R_C^Ψ ②.

We can now define $\chi = \chi_0 / \gcd(\psi_0, \chi_0)$ and $\psi = \psi_0 / \gcd(\psi_0, \chi_0)$. Then $\chi s' + \psi t \approx \psi t' + \chi s \in \mathrm{tr}(R_C^\Psi)$ as $\mathrm{tr}(R_C^\Psi)$ satisfies T_Ψ; besides $\gcd(\psi, \chi) = 1$ and $\chi m = \psi n$. □

The following corollary is proved analogously.

COROLLARY 4.28. *Let $C = C' \vee e_2 \vee e_1$ be reductive for $e_1 = u \approx s'$, where u does not have sort S_{CAM}. Suppose that u is irreducible with respect to R_C^Ψ and that e_2 is contained in $\mathrm{tr}_\Psi(R_C^\Psi \cup \{e_1\}) \setminus \mathrm{tr}(R_C^\Psi)$. Then e_2 has the form $u \approx t'$, and $s' \approx t' \in \mathrm{tr}(R_C^\Psi)$.*

4.5. REDUNDANCY

We have shown that $\mathrm{tr}(R_\infty^\Psi)$ is an equality model of the axioms of Ψ-torsion-free cancellative abelian monoids, but we still have to prove that $\mathrm{tr}(R_\infty^\Psi)$ is also a model of the clause set N, provided that N is saturated and does not contain $\bot$. In Subsection 2.3, we have given an abstract definition of saturation: The set N is saturated, if every inference from clauses in N is redundant. What is still missing, however, is the actual redundancy criterion complementing the inference system $CInf_\Psi$.

DEFINITION 4.29. *Let N be a set of clauses. We say that a clause C is $CRed_\Psi$-redundant with respect to N, if $\overline{N}^{\prec_{\mathrm{C}} C\theta} \models_\Psi C\theta$ for every ground instance $C\theta$.*

To obtain a definition of redundancy for inferences, we need the concept of a ground instance of an inference.

DEFINITION 4.30. *Let $C_0, C_1, \dots, C_k$ be clauses and let θ be a substitution such that $C_1\theta, \dots, C_k\theta$ are ground. If there are inferences*

$$\frac{C_k \ \dots \ C_1}{C_0}$$

and

$$\frac{C_k\theta \ \dots \ C_1\theta}{C_0\theta}$$

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

Note that for all inference rules except the *abstraction* rule the conclusion of a inference from ground premises is again ground.

Whenever we talk about instances of inferences, we assume that selected literals in the ground clauses and in the non-ground clauses correspond to each other.

DEFINITION 4.31. *Let N be a set of clauses. We say that a non-abstraction inference*

$$\frac{C_k \ \dots \ C_1}{C_0}$$

is $CRed_\Psi$-redundant with respect to N if $\overline{N}^{\prec_{\mathrm{C}} C_1\theta} \models_\Psi C_0\theta$ for every ground instance

$$\frac{C_k\theta \ \dots \ C_1\theta}{C_0\theta}$$

DEFINITION 4.32. *Let N be a set of clauses. An abstraction inference*

$$\frac{C_2 \qquad C_1}{C_0}$$

with $C_0 = C' \vee \neg\, y \approx w \vee [\neg]\, s[y] \approx s'$ is called $CRed_\Psi$-redundant with respect to N if for every ground instance

$$\frac{C_2\theta \qquad C_1\theta}{C_0\theta}$$

and every substitution ρ that maps y to a ground term $r \prec w\theta$, $\overline{N}^{\prec_{\mathrm{C}} C_1\theta} \models_\Psi C_0\theta\rho$.

DEFINITION 4.33. *Let N be a set of clauses. The set of all clauses that are $CRed_\Psi$-redundant with respect to N is denoted by $CRed^{\mathrm{C}}_\Psi(N)$. The set of all inferences that are $CRed_\Psi$-redundant with respect to N is denoted by $CRed^{\mathrm{I}}_\Psi(N)$.*

LEMMA 4.34. *The pair $CRed_\Psi = (CRed^{\mathrm{I}}_\Psi, CRed^{\mathrm{C}}_\Psi)$ is a redundancy criterion with respect to the inference system $CInf_\Psi$ and the consequence relation $\models_\Psi$.*

PROOF. We have to show that $CRed_\Psi$ satisfies the conditions (i)–(iv) of Def. 2.2. To show condition (i), let N be a set of clauses, let C be some clause in $CRed^{\mathrm{C}}_\Psi(N)$ and let $C\theta$ be a ground instance of C. We have to prove that $N \setminus CRed^{\mathrm{C}}_\Psi(N) \models_\Psi C\theta$. As $C \in CRed^{\mathrm{C}}_\Psi(N)$, we have $\overline{N}^{\prec_{\mathrm{C}} C\theta} \models_\Psi C\theta$. By the compactness of first-order logic there exists a finite subset of $\overline{N}^{\prec_{\mathrm{C}} C\theta}$ that entails $C\theta$. Let $\overline{N}_0$ be the minimal finite subset of $\overline{N}^{\prec_{\mathrm{C}} C\theta}$ (with respect to the multiset extension of $\succ_{\mathrm{C}}$) such that $\overline{N}_0 \models_\Psi C\theta$. If some clause D in $\overline{N}_0$ were a ground instance of a clause in $CRed^{\mathrm{C}}_\Psi(N)$, then there would exist $D_1, \dots, D_n \in \overline{N}^{\prec_{\mathrm{C}} D}$ such that $\{D_1, \dots, D_n\} \models_\Psi D$ and $\overline{N}_0 \cup \{D_1, \dots, D_n\} \setminus \{D\} \models_\Psi C\theta$. This is impossible, however, as it contradicts the minimality of $\overline{N}_0$. Thus $N \setminus CRed^{\mathrm{C}}_\Psi(N) \models_\Psi CRed^{\mathrm{C}}_\Psi(N)$.

Condition (ii) is obvious. Condition (iii) is proved in a similar way as condition (i). Condition (iv) follows from the fact that the conclusion of a ground inference is itself smaller than the maximal premise. □

4.6. LIFTING

Under which conditions is an inference from ground clauses $C_i\theta$ a ground instance of an inference from C_i ? This question will be answered by the so-called "lifting lemmas".

LEMMA 4.35. *Let D and C be two clauses (without common variables) and let θ be a substitution such that $D\theta$ and $C\theta$ are ground. If there is a positive cancellative superposition inference from $D\theta$ and $C\theta$ then the inference is a ground instance of a positive cancellative superposition inference from D and C.*

PROOF. We may assume that $C\theta$ and $D\theta$ have no selected literals and $C\theta \succ_{\mathrm{C}} D\theta$. Let $D = D' \vee e_2$ and $C = C' \vee e_1$, such that $e_2\theta$ and $e_1\theta$ are strictly maximal in $D\theta$ and $C\theta$. Suppose that $e_1\theta = \bar{m}\bar{u} + \bar{s} \approx \bar{s}'$, $e_2\theta = \bar{n}\bar{u} + \bar{t} \approx \bar{t}'$, where $\bar{u}$ is an atomic ground term, $\bar{u} \succ \bar{s}$, $\bar{u} \succ \bar{s}'$, $\bar{u} \succ \bar{t}$, $\bar{u} \succ \bar{t}'$, and $\bar{m} \geq \bar{n} \geq 1$. Then these clauses allow a *positive cancellative superposition* inference

$$\frac{D'\theta \vee \bar{n}\bar{u} + \bar{t} \approx \bar{t}' \qquad C'\theta \vee \bar{m}\bar{u} + \bar{s} \approx \bar{s}'}{D'\theta \vee C'\theta \vee (\bar{m}{-}\bar{n})\bar{u} + \bar{s} + \bar{t}' \approx \bar{t} + \bar{s}'}$$

that we denote by $\bar{\iota}$. By the ordering conditions above, $\bar{u} \succ y\theta$ for every variable y that is not eligible or occurs in the right-hand sides of e_1 or e_2 or in negative literals of C or D. Let $\{\, x_i \mid i \in I \,\} = \mathrm{elig}(C) \cap \mathrm{Var}(\mathrm{lhs}(e_1)) \setminus \mathrm{Var}(\mathrm{rhs}(e_1)) \setminus \mathrm{Var}(\mathrm{neg}(C))$ and $\{\, y_j \mid j \in J \,\} = \mathrm{elig}(D) \cap \mathrm{Var}(\mathrm{lhs}(e_2)) \setminus \mathrm{Var}(\mathrm{rhs}(e_2)) \setminus \mathrm{Var}(\mathrm{neg}(D))$. We may assume that

$$e_1 \;=\; \sum_{i\in I} m_i x_i + \sum_{k\in K} m_k^* u_k + s \approx s'$$

such that $x_i\theta = \mu_i\bar{u} + \bar{s}_i$ for $i \in I$, where $\mu_i \in \mathbf{N}$ and $\bar{u} \succ \bar{s}_i$, $u_k\theta = \bar{u}$ for $k \in K$, $\bar{m} = \sum_{i\in I} m_i\mu_i + \sum_{k\in K} m_k^*$, $\bar{s} = \sum_{i\in I} m_i\bar{s}_i + s\theta$, and $\bar{s}' = s'\theta$. Analogously

$$e_2 \;=\; \sum_{j\in J} n_j y_j + \sum_{l\in L} n_l^* v_l + t \approx t'$$

such that $y_j\theta = \nu_j\bar{u} + \bar{t}_j$ for $j \in J$, where $\nu_j \in \mathbf{N}$ and $\bar{u} \succ \bar{t}_j$, $v_l\theta = \bar{u}$ for $l \in L$, $\bar{n} = \sum_{j\in J} n_j\nu_j + \sum_{l\in L} n_l^*$, $\bar{t} = \sum_{j\in J} n_j\bar{t}_j + t\theta$, and $\bar{t}' = t'\theta$. Let z, $\hat{x}_i$, $\check{x}_i$, $\hat{y}_j$, and $\check{y}_j$ be new variables for $i \in I$ and $j \in J$. We define substitutions σ_1 and ρ_1 as follows: Let σ_1 map x_i to $\hat{x}_i + \check{x}_i$ and y_j to $\hat{y}_j + \check{y}_j$ for $i \in I$, $j \in J$. Let ρ_1 map $\hat{x}_i$ to $\mu_i\bar{u}$ and $\check{x}_i$ to $\bar{s}_i$ for $i \in I$, $\hat{y}_j$ to $\nu_j\bar{u}$ and $\check{y}_j$ to $\bar{t}_j$ for $j \in J$, z to $(\bar{m}{-}\bar{n})\bar{u}$, and every $y \in \mathrm{Var}(C) \cup \mathrm{Var}(D)$ to $y\theta$. It is easy to verify that θ equals $\sigma_1\rho_1$ over $\mathrm{Var}(C) \cup \mathrm{Var}(D)$.

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

Let u be any of the u_k or v_l ($k \in K$, $l \in L$), or a new variable, if $K \cup L = \emptyset$. Consider the terms $r = \sum_{i\in I} m_i\hat{x}_i + \big(\sum_{k\in K} m_k^*\big)u\sigma_2$ and $r' = z + \sum_{j\in J} n_j\hat{y}_j + \big(\sum_{l\in L} n_l^*\big)u\sigma_2$. As $r\rho_2 = r'\rho_2 = \bar{m}\bar{u}$, there exist a most general ACU-unifier σ_3 of r and r' and a substitution ρ_3 such that $\rho_2 = \sigma_3\rho_3$ over $\mathrm{Var}(C) \cup \mathrm{Var}(D) \cup \mathrm{Ran}(\sigma_1) \cup \mathrm{Ran}(\sigma_2) \cup \{z\}$. Let $\sigma = \sigma_1\sigma_2\sigma_3$.

We define

$$e_0 \;=\; z + \sum_{i\in I} m_i\check{x}_i + s + t' \approx \sum_{j\in J} n_j\check{y}_j + t + s'$$

then

$$\frac{D' \vee e_2 \qquad C' \vee e_1}{(D' \vee C' \vee e_0)\sigma}$$

is a *positive cancellative superposition* inference from D and C that we denote by ι. It is easy to see that the ordering conditions of the inference rule are satisfied.

It remains to prove that the *positive cancellative superposition* inference $\bar{\iota}$ given by

$$\frac{D'\theta \vee \bar{n}\bar{u} + \bar{t} \approx \bar{t}' \qquad C'\theta \vee \bar{m}\bar{u} + \bar{s} \approx \bar{s}'}{D'\theta \vee C'\theta \vee (\bar{m}{-}\bar{n})\bar{u} + \bar{s} + \bar{t}' \approx \bar{t} + \bar{s}'}$$

is a ground instance of ι. We will show that $\bar{\iota} = \iota\sigma\rho_3$: First, $\sigma\rho_3$ equals θ over $\mathrm{Var}(C) \cup \mathrm{Var}(D)$ and σ is idempotent, so $C'\sigma\sigma\rho_3 = C'\sigma\rho_3 = C'\theta$ and $e_1\sigma\rho_3 = \bar{m}\bar{u} + \bar{s} \approx \bar{s}'$, and analogously $D'\sigma\sigma\rho_3 = D'\sigma\rho_3 = D'\theta$ and $e_2\sigma\rho_3 = \bar{n}\bar{u} + \bar{t} \approx \bar{t}'$. Second, $\sigma\sigma\rho_3$ equals ρ_1 over $\mathrm{Var}(e_0)$. Therefore, $e_0\sigma\sigma\rho_3 = e_0\rho_1 = (\bar{m}{-}\bar{n})\bar{u} + \bar{s} + \bar{t}' \approx \bar{t} + \bar{s}'$. $\square$

The following two lemmas are proved in a similar way as the preceding one.

LEMMA 4.36. *Let C be a clause and let θ be a substitution such that $C\theta$ is ground. Then every cancellation, equality resolution, standard equality factoring, or cancellative equality factoring inference from $C\theta$ is a ground instance of an inference from C.*

LEMMA 4.37. *Let $D = D' \vee e_2$ and $C = C' \vee [\neg]\, e_1$ be two clauses (without common variables) and let θ be a substitution such that $D\theta$ and $C\theta$ are ground.*

If there is a negative cancellative superposition inference

$$\frac{D'\theta \vee e_2\theta \qquad C'\theta \vee \neg\, e_1\theta}{C_0}$$

(where the maximal atomic subterms of $\mathrm{lhs}(e_2\theta)$ *and* $\mathrm{lhs}(e_1\theta)$ *are overlapped), and* $\mathrm{lhs}(e_1)$ *is not a variable, then the inference is a ground instance of a negative cancellative superposition inference from D and C.*

If there is a standard superposition inference

$$\frac{D'\theta \vee e_2\theta \qquad C'\theta \vee [\neg]\, e_1\theta}{C_0}$$

(where $\mathrm{lhs}(e_2\theta)$ *and some subterm of* $\mathrm{lhs}(e_1\theta)$ *are overlapped), and* $\mathrm{lhs}(e_2\theta)$ *does not occur in* $\mathrm{lhs}(e_1\theta)$ *at or below a variable position of* $\mathrm{lhs}(e_1)$ *(that is, $x\theta = w[\mathrm{lhs}(e_2\theta)]$ for some $x \in \mathrm{Var}(\mathrm{lhs}(e_1))$), then the inference is a ground instance of a standard superposition inference from D and C.*

If there is an abstraction inference

$$\frac{D'\theta \vee e_2\theta \qquad C'\theta \vee [\neg]\, e_1\theta[\bar{m}\bar{v} + \bar{q}]}{C_0}$$

such that

(i) C_0 *equals* $C'\theta \vee \neg\, y \approx \bar{m}\bar{v} + \bar{q} \vee [\neg]\, e_1\theta[y]$,
(ii) $\bar{m}\bar{v} + \bar{q} = w\theta$ *for some subterm w of* $\mathrm{lhs}(e_1)$,
(iii) $\bar{m}\bar{v} + \bar{q}$ *is not a subterm of $y'\theta$ for any* $y' \in \mathrm{Var}(\mathrm{lhs}(e_1))$,
(iv) the maximal atomic subterm of $\mathrm{lhs}(e_2\theta)$ *equals* $\bar{v}$,
(v) if $w = x + q$ and $\bar{v}$ occurs in $x\theta$, then $q = q_1 + q_2$ and q_1 is a variable or a non-zero atomic term not containing x,

then the inference is a ground instance of an abstraction inference from D and C.

If a rewrite rule e is used in a derivation $e' \to^{+} 0 \approx 0$, then its maximal term $\mathrm{mt}(e)$ can not be larger than $\mathrm{mt}(e')$. The following two lemmas are consequences of this fact.

LEMMA 4.38. *Let $C\theta$ be a clause from $\overline{N}$. If $C\theta$ is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$, then it is also true in $\mathrm{tr}(R^{\Psi}_{\infty})$ and $\mathrm{tr}(R^{\Psi}_{D\theta})$ for any $D\theta \succ_{\mathrm{C}} C\theta$.*

LEMMA 4.39. *Let $C\theta = C'\theta \vee e\theta$ be a clause from $\overline{N}$ such that $E_{C\theta} = \{e\theta\}$. Then $C\theta$ is true and $C'\theta$ is false in $\mathrm{tr}(R^{\Psi}_{\infty})$ and $\mathrm{tr}(R^{\Psi}_{D\theta})$ for any $D\theta \succeq_{\mathrm{C}} C\theta$.*

LEMMA 4.40. *Let N be a set of clauses that is saturated up to $CRed_{\Psi}$-redundancy and does not contain the empty clause. Then we have for every ground clause $C\theta \in \overline{N}$:*
(i) If C has selected literals, then $C\theta$ is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$.
(ii) $E_{C\theta} = \emptyset$ if and only if $C\theta$ is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$.
(iii) $C\theta$ is true in $\mathrm{tr}(R^{\Psi}_{\infty})$ and in $\mathrm{tr}(R^{\Psi}_{D})$ for every $D \succ_{\mathrm{C}} C\theta$.

PROOF. We use induction on the clause ordering $\succ_{\mathrm{C}}$ and assume that (i)–(iii) are already satisfied for all clauses in $\overline{N}$ that are smaller than $C\theta$. Note that the "if" part of (ii) is obvious from the model construction and that condition (iii) follows immediately from (ii), Lemma 4.39, and Lemma 4.38.

Case 1: $x\theta$ equals some smaller term.

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

Case 2: C contains a selected or maximal negative literal.

Suppose that $C\theta$ does not fall into Case 1 and that $C\theta = C'\theta \vee \neg\, e_1\theta$, where $\neg\, e_1\theta$ is either maximal among the instances of selected literals in C (if C has selected literals), or maximal in $C\theta$ (otherwise). If $e_1\theta \notin \mathrm{tr}(R^{\Psi}_{C\theta})$, there is nothing to show, so assume that there is an $R^{\Psi}_{C\theta}$-derivation from $e_1\theta$ to $0 \approx 0$. Let $\bar{u} = \mathrm{mt}(e_1\theta)$.

Case 2.1: $\bar{u}$ occurs on both sides of $e_1\theta$.

If $e_1\theta$ equals $\bar{u} \approx \bar{u}$ where $\bar{u}$ either does not have sort S_{CAM} or equals 0, then there is an *equality resolution* inference

$$\frac{C'\theta \vee \neg\, \bar{u} \approx \bar{u}}{C'\theta}\,.$$

As shown in Lemma 4.36, this is an instance of an *equality resolution* inference from C. By saturation up to $CRed_{\Psi}$-redundancy, it is $CRed_{\Psi}$-redundant, hence $\overline{N}^{\prec_{\mathrm{C}} C\theta} \models_{\Psi} C'\theta$. By the induction hypothesis, all clauses in $\overline{N}^{\prec_{\mathrm{C}} C\theta}$ are true in $\mathrm{tr}(R^{\Psi}_{C\theta})$. Thus $C'\theta$ and $C\theta$ are true in $\mathrm{tr}(R^{\Psi}_{C\theta})$.

If $e_1\theta$ equals $\bar{m}\bar{u}+\bar{s}\approx\bar{m}'\bar{u}+\bar{s}'$ with $\bar{m}\ge\bar{m}'\ge 1$, then there is a *cancellation* inference

$$\frac{C'\theta \vee \neg\, \bar{m}\bar{u}+\bar{s}\approx\bar{m}'\bar{u}+\bar{s}'}{C'\theta \vee \neg\, (\bar{m}-\bar{m}')\bar{u}+\bar{s}\approx\bar{s}'}\;.$$

By Lemma 4.36, this is an instance of a *cancellation* inference from C. By saturation up to CRed_Ψ-redundancy, the inference is CRed_Ψ-redundant, hence $\overline{N}^{\prec_{\mathrm{C}} C\theta} \models_\Psi C'\theta \vee \neg\,(\bar{m}-\bar{m}')\bar{u}+\bar{s}\approx\bar{s}'$. By the induction hypothesis, all clauses in $\overline{N}^{\prec_{\mathrm{C}} C\theta}$ and thus $C'\theta \vee \neg\,(\bar{m}-\bar{m}')\bar{u}+\bar{s}\approx\bar{s}'$ and $C\theta$ are true in $\mathrm{tr}(R^\Psi_{C\theta})$.

Case 2.2: $\bar{u}$ occurs on only one side of $e_1\theta$.

If $\bar{u}$ occurs only on one side of $e_1\theta$, then $e_1\theta$ has either the form $\bar{m}\bar{u}+\bar{s}\approx\bar{s}'$ or $\bar{u}\approx\bar{u}'$ and $\bar{u}$ does not have sort S_{CAM}. We write $e_1\theta[\bar{u}]$ if the distinction between these cases is irrelevant.[22] By Lemma 4.9 we may assume that the derivation from $e_1\theta$ to $0\approx 0$ starts with a γ- or δ-step using a rule $e''\in E^\Psi_{D\theta}\subseteq R^\Psi_{C\theta}$ at (or inside) $\bar{m}\bar{u}$ or $\bar{u}$. (Without loss of generality we assume that C and D are variable disjoint; so we can use the same substitution θ.) Let $D\theta = D'\theta \vee e_2\theta$ with $E_{D\theta}=\{e_2\theta\}$. By part (i) and (ii) of the induction hypothesis and Lemma 4.39, D has no selected literals and $D'\theta$ is false in $\mathrm{tr}(R^\Psi_{C\theta})$.

Case 2.2.1: $\bar{m}\bar{u}$ is γ-reducible by e''.

If the reduction from $e_1\theta$ to $0\approx 0$ starts with a γ-application of e'' at $\bar{m}\bar{u}$, then, by Cor. 4.26, $e_2\theta$ is a rewrite rule $\bar{n}\bar{u}+\bar{t}\approx\bar{t}'$ and there are $\chi\in\mathbf{N}^{>0}$ and $\psi\in\Psi$ such that $\psi\bar{m}=\chi\bar{n}$ and $\gcd(\psi,\chi)=1$.

Consider the *negative cancellative superposition* inference

$$\frac{D'\theta \vee \bar{n}\bar{u}+\bar{t}\approx\bar{t}' \qquad C'\theta \vee \neg\, \bar{m}\bar{u}+\bar{s}\approx\bar{s}'}{D'\theta \vee C'\theta \vee \neg\, \psi\bar{s}+\chi\bar{t}'\approx\chi\bar{t}+\psi\bar{s}'}$$

As $\bar{m}\bar{u}+\bar{s}\approx\bar{s}'\in\mathrm{tr}(R^\Psi_{C\theta})$ and $\bar{m}\bar{u}+\bar{s}\succ\bar{s}'$, the left-hand side of e_1 cannot be a variable – otherwise $C\theta$ would be subject to Case 1 above. By Lemma 4.37 the inference is a ground instance of a *negative cancellative superposition* inference from D and C. As N is saturated, it is CRed_Ψ-redundant, thus its conclusion is true in $\mathrm{tr}(R^\Psi_{C\theta})$. Both $D'\theta$ and $\neg\,\psi\bar{s}+\chi\bar{t}'\approx\chi\bar{t}+\psi\bar{s}'$ are false in $\mathrm{tr}(R^\Psi_{C\theta})$, so $C'\theta$ and $C\theta$ must be true in $\mathrm{tr}(R^\Psi_{C\theta})$.

Case 2.2.2: $\bar{u}$ is δ-reducible by e''.

Otherwise, the reduction from $e_1\theta$ to $0\approx 0$ starts with a δ-application of e'' at or inside $\bar{u}$. We distinguish between two cases, depending on whether $\bar{u}$ is also δ-reducible by $e_2\theta$ or not.

Case 2.2.2.1: $\bar{u}$ is δ-reducible by both e'' and $e_2\theta$.

Suppose that $\bar{u}$ is also δ-reducible by $e_2\theta=\bar{t}\approx\bar{t}'$. Then $\bar{t}$ does not have sort S_{CAM} or $\bar{t}$ occurs in $\bar{u}$ below a free function symbol. Note that $\bar{t}$ cannot occur in $e_1\theta$ at or below a variable position of C, say $x\theta = w[\bar{t}]$, since otherwise $x\theta\approx w[\bar{t}']\in\mathrm{tr}(R^\Psi_{C\theta})$

[22] Recall that $\bar{m}\bar{u}$ is merely an abbreviation for the $\bar{m}$-fold sum $\bar{u}+\cdots+\bar{u}$. If $e_1\theta=\bar{m}\bar{u}+\bar{s}\approx\bar{s}'$, then the hole in $e_1\theta[\_]$ is the position of *one* of the $\bar{m}$ $\bar{u}$'s.

and $x\theta \succ w[\bar{t}']$, so $C\theta$ would be subject to Case 1 above. Consequently, the *standard superposition* inference

$$\frac{D'\theta \lor \bar{t} \approx \bar{t}' \qquad C'\theta \lor \neg\, e_1\theta[\bar{u}[\bar{t}]]}{D'\theta \lor C'\theta \lor \neg\, e_1\theta[\bar{u}[\bar{t}']]}$$

is a ground instance of a *standard superposition* inference from D and C. Again, by saturation, its conclusion is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$; and since $D'\theta$ and $\neg\, e_1\theta[\bar{u}[\bar{t}']]$ are false in $\mathrm{tr}(R^{\Psi}_{C\theta})$, both $C'\theta$ and $C\theta$ must be true.

Case 2.2.2.2: $\bar{u}$ is δ-reducible by e'' but not by $e_2\theta$.

By the definition of $E^{\Psi}_{D\theta}$, the rules e'' and $e_2\theta$ have the same maximal term. If $\bar{u}$ is δ-reducible by e'' but not by $e_2\theta$, then we may assume that $e_2\theta = \bar{n}\bar{v} + \bar{t} \approx \bar{t}'$ and $e'' = \bar{m}\bar{v} + \bar{r} \approx \bar{r}'$, such that there are $\chi \in \mathbf{N}^{>0}$ and $\psi \in \Psi$ with $\chi\bar{n} = \psi\bar{m}$ and $\gcd(\psi, \chi) = 1$. We may further assume that $e_1\theta = e_1\theta[\bar{u}[\bar{m}_0\bar{v} + \bar{r} + \bar{q}]]$, where $\bar{m}_0 \geq \bar{m}$ and $\bar{m}_0\bar{v} + \bar{r} + \bar{q}$ occurs in $\bar{u}$ immediately below a free function symbol. As $\bar{u}$ is δ-irreducible by $e_2\theta$, $\bar{n}\bar{v} + \bar{t}$ is not a subterm of $\bar{m}_0\bar{v} + \bar{r} + \bar{q}$. Consequently, there is an *abstraction* inference

$$\frac{D'\theta \lor \bar{n}\bar{v} + \bar{t} \approx \bar{t}' \qquad C'\theta \lor \neg\, e_1\theta[\bar{u}[\bar{m}_0\bar{v} + \bar{r} + \bar{q}]]}{C_0}$$

where C_0 equals $C'\theta \lor \neg\, y \approx \bar{m}_0\bar{v} + \bar{r} + \bar{q} \lor \neg\, e_1\theta[\bar{u}[y]]$. Let $\bar{w}_0$ be the smallest term such that $\bar{w}_0 \approx \bar{m}_0\bar{v} + \bar{r} + \bar{q} \in \mathrm{tr}(R^{\Psi}_{C})$. Obviously, $\bar{m}_0\bar{v} + \bar{r} + \bar{q} \succ (\bar{m}_0 - \bar{m})\bar{v} + \bar{r}' + \bar{q} \succeq \bar{w}_0$. We define a substitution $\rho = \{y \mapsto \bar{w}_0\}$.

If $\bar{m}\bar{v} + \bar{r}$ occurred in $e_1\theta$ at or below a variable position of C, then $C\theta$ would be subject to Case 1 above, as $e'' = \bar{m}\bar{v} + \bar{r} \approx \bar{r}'$ is contained in $\mathrm{tr}(R^{\Psi}_{C\theta})$ and $\bar{m}\bar{v} + \bar{r} \succ \bar{r}'$. Hence let $e_1 = e_1[u[w]]$, where $u[w]\theta = \bar{u}$ and $w\theta = \bar{m}_0\bar{v} + \bar{r} + \bar{q}$. Assume that w had the form $x + \sum_{j \in J} q_j$, where all q_j are non-zero atomic terms containing x and $\bar{v}$ occurs in $x\theta$. Then $x\theta$ could be written as $\bar{m}_0\bar{v} + \bar{r} + \bar{r}''$, since $\bar{v} \succ \bar{r}$. This is impossible, though, as $\bar{m}\bar{v} + \bar{r}$ must not occur at or below a variable position. Therefore, by Lemma 4.37, the inference is a ground instance of an *abstraction* inference from D and C.

By saturation, the clause $C_0\rho$, that is $C'\theta \lor \neg\, \bar{w}_0 \approx \bar{m}_0\bar{v} + \bar{r} + \bar{q} \lor \neg\, e_1\theta[\bar{u}[\bar{w}_0]]$ is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$; and since $\bar{w}_0 \approx \bar{m}_0\bar{v} + \bar{r} + \bar{q} \in \mathrm{tr}(R^{\Psi}_{C})$, $C\theta$ must be true likewise.

Case 3: C does not contain a selected or maximal negative literal.

Suppose that $C\theta$ does not fall into Cases 1 or 2. Then C can be written as $C' \lor e_1$, where $e_1\theta$ is a maximal literal of $C\theta$. If $E_{C\theta} = \{e_1\theta\}$ or $C'\theta$ is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$, then there is nothing to show, so assume that $E_{C\theta} = \emptyset$ and that $C'\theta$ is false in $\mathrm{tr}(R^{\Psi}_{C\theta})$. Let $\bar{u} = \mathrm{mt}(e_1\theta)$.

Case 3.1: $\bar{u}$ occurs on both sides of $e_1\theta$.

If $e_1\theta$ has the form $\bar{u} \approx \bar{u}$, then $C\theta$ is a tautology and thus true in $\mathrm{tr}(R^{\Psi}_{C\theta})$. If $e_1\theta$ equals $\bar{m}\bar{u} + \bar{s} \approx \bar{m}'\bar{u} + \bar{s}'$ with $\bar{m} \geq \bar{m}' \geq 1$, then there is a *cancellation* inference from $C\theta$. As in Case 2.1, we can show that $C\theta$ is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$.

Case 3.2: $\bar{u}$ occurs on only one side of $e_1\theta$.

If $\bar{u}$ occurs only on one side of $e_1\theta$, then either $e_1\theta = \bar{m}\bar{u} + \bar{s} \approx \bar{s}'$, or $e_1\theta = \bar{u} \approx \bar{s}'$ and $\bar{u}$ does not have sort S_{CAM}.

Case 3.2.1: $e_1\theta$ is maximal in $C\theta$, but not strictly maximal.

If $e_1\theta$ is maximal in $C\theta$, but not strictly maximal, then $C\theta$ can be written as $C''\theta \vee e_2\theta \vee e_1\theta$, where $e_1\theta = e_2\theta$. In this case, there is either a *cancellative equality factoring* inference

$$\frac{C''\theta \vee \bar{m}\bar{u} + \bar{s} \approx \bar{s}' \vee \bar{m}\bar{u} + \bar{s} \approx \bar{s}'}{C''\theta \vee \neg\, \bar{s} + \bar{s}' \approx \bar{s} + \bar{s}' \vee \bar{m}\bar{u} + \bar{s} \approx \bar{s}'}$$

(if $\bar{u}$ has sort S_{CAM}), or a *standard equality factoring* inference

$$\frac{C''\theta \vee \bar{u} \approx \bar{s}' \vee \bar{u} \approx \bar{s}'}{C''\theta \vee \neg\, \bar{s}' \approx \bar{s}' \vee \bar{u} \approx \bar{s}'}$$

(if $\bar{u}$ does not have sort S_{CAM}). This inference is a ground instance of an inference from C. By saturation, its conclusion is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$. As $\bar{s} + \bar{s}' \approx \bar{s} + \bar{s}'$ or $\bar{s}' \approx \bar{s}'$ are contained in $\mathrm{tr}(R^{\Psi}_{C\theta})$, $C\theta$ must be true in $\mathrm{tr}(R^{\Psi}_{C\theta})$.

Case 3.2.2: $e_1\theta$ is strictly maximal in $C\theta$ and $\mathrm{mt}_{\#}(e_1\theta)$ is reducible.

Suppose that $e_1\theta$ is strictly maximal in $C\theta$ and $\mathrm{mt}_{\#}(e_1\theta)$ is reducible by some rule $e'' \in E^{\Psi}_{D\theta} \subseteq R^{\Psi}_{C\theta}$. Let $D\theta = D'\theta \vee e_2\theta$ and $E_{D\theta} = \{e_2\theta\}$. By part (i) and (ii) of the induction hypothesis and Lemma 4.39, D has no selected literals and $D'\theta$ is false in $\mathrm{tr}(R^{\Psi}_{C\theta})$. Depending on whether $\mathrm{mt}_{\#}(e_1\theta)$ is γ- or δ-reducible by e'' and whether $\mathrm{mt}(e_1\theta)$ is reducible or irreducible by $e_2\theta$, there is either a *positive cancellative superposition* inference, or a *standard superposition* inference, or an *abstraction* inference from $D\theta$ and $C\theta$. Using essentially the same techniques as in Case 2.2 we can thus show that $C\theta$ is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$.

Case 3.2.3: $e_1\theta$ is strictly maximal in $C\theta$ and $\mathrm{mt}_{\#}(e_1\theta)$ is irreducible.

Suppose that $e_1\theta$ is strictly maximal in $C\theta$ and $\mathrm{mt}_{\#}(e_1\theta)$ is irreducible by $R^{\Psi}_{C\theta}$. Then either $C\theta$ is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$, or $C'\theta$ is true in $\mathrm{tr}_{\Psi}(R^{\Psi}_{C\theta} \cup \{e_1\theta\})$, or $E_{C\theta} = \{e_1\theta\}$. In the first and the third case, there is nothing to show. Let us therefore assume that $C\theta$ is false in $\mathrm{tr}(R^{\Psi}_{C\theta})$ and $C'\theta$ is true in $\mathrm{tr}_{\Psi}(R^{\Psi}_{C\theta} \cup \{e_1\theta\})$. Then $C'\theta = C''\theta \vee e_2\theta$, where the literal $e_2\theta$ is smaller than $e_1\theta$ and is contained in $\mathrm{tr}_{\Psi}(R^{\Psi}_{C\theta} \cup \{e_1\theta\}) \setminus \mathrm{tr}(R^{\Psi}_{C\theta})$.

Case 3.2.3.1: $\bar{u}$ has sort S_{CAM}.

If $\bar{u}$ has sort S_{CAM}, we know by Lemma 4.27 that $e_2\theta$ equals $\bar{n}\bar{u} + \bar{t} \approx \bar{t}'$ where $\chi\bar{m} = \psi\bar{n}$ for some $\psi \in \Psi$ and $\chi \in \mathbf{N}^{>0}$ with $\gcd(\psi, \chi) = 1$, and that $\psi\bar{t} + \chi\bar{s}' \approx \chi\bar{s} + \psi\bar{t}' \in \mathrm{tr}(R^{\Psi}_{C\theta})$. Consequently, there is a *cancellative equality factoring* inference

$$\frac{C''\theta \vee \bar{n}\bar{u} + \bar{t} \approx \bar{t}' \vee \bar{m}\bar{u} + \bar{s} \approx \bar{s}'}{C''\theta \vee \neg\, \psi\bar{t} + \chi\bar{s}' \approx \chi\bar{s} + \psi\bar{t}' \vee \bar{n}\bar{u} + \bar{t} \approx \bar{t}'}$$

which is a ground instance of a *cancellative equality factoring* inference from C. By saturation, its conclusion is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$. As $\psi\bar{t} + \chi\bar{s}' \approx \chi\bar{s} + \psi\bar{t}' \in \mathrm{tr}(R^{\Psi}_{C\theta})$, $C''\theta \vee \bar{n}\bar{u} + \bar{t} \approx \bar{t}'$ and thus $C\theta$ must be true in $\mathrm{tr}(R^{\Psi}_{C\theta})$. This contradicts our assumption above.

Case 3.2.3.2: $\bar{u}$ does not have sort S_{CAM}.

If $\bar{u}$ does not have sort S_{CAM}, we know by Lemma 4.28 that $e_2\theta = \bar{u} \approx \bar{t}'$ and $\bar{s}' \approx \bar{t}' \in \mathrm{tr}(R^{\Psi}_{C\theta})$. Hence there is a *standard equality factoring* inference

$$\frac{C''\theta \vee \bar{u} \approx \bar{t}' \vee \bar{u} \approx \bar{s}'}{C''\theta \vee \neg\, \bar{s}' \approx \bar{t}' \vee \bar{u} \approx \bar{t}'}$$

whose conclusion is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$. Again, $C\theta$ must be true in $\mathrm{tr}(R^{\Psi}_{C\theta})$, contradicting our assumption. This concludes the proof of the lemma. $\square$

We can now prove the two central theorems of this paper.

THEOREM 4.41. *Let N be a set of clauses that is saturated up to $CRed_{\Psi}$-redundancy. Then $N \cup \mathrm{ACUKT}_{\Psi}$ is equality unsatisfiable if and only if N contains the empty clause.*

PROOF. If N contains the empty clause, then it is unsatisfiable. Otherwise, $\mathrm{tr}(R^{\Psi}_{\infty})$ is a model of the equality axioms (by Cor. 4.23), of ACUKT_{Ψ} (by Cor. 4.24), and of $\overline{N}$ (by part (iii) of Lemma 4.40). $\square$

THEOREM 4.42. *Let $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$ be a fair derivation of the cancellative superposition calculus. Let N_∞ be the limit of the derivation. Then $N_0 \cup \mathrm{ACUKT}_{\Psi}$ is equality unsatisfiable if and only if N_∞ contains the empty clause.*

PROOF. Suppose that N_∞ does not contain the empty clause. By fairness, N_∞ is saturated up to $CRed_{\Psi}$-redundancy, hence $N_\infty \cup \mathrm{ACUKT}_{\Psi}$ has an equality model. As $N_0 \subseteq N_\infty \cup CRed^{\mathrm{C}}_{\Psi}(N_\infty)$, this model is also an equality model of $N_0 \cup \mathrm{ACUKT}_{\Psi}$. The reverse direction of the proof is obvious since $N_0 \models_{\Psi} N_\infty$. $\square$

5. Refinements and Applications

5.1. Simplification Techniques

In any practicable saturation-based theorem prover methods to keep the number and the size of formulae as small as possible are indispensable. Deletion of a formula during a theorem proving derivation is allowed only if this formula has become redundant. However, if formulae do not become redundant accidentally, we still have the chance to simplify them, that is, to *make* them redundant by adding some other clauses (see Def. 2.4).

The easiest simplification rules are those which transform a clause into an ACUKT_{Ψ}-equivalent smaller one. For example, independently of N every clause $C \vee [\neg]\, s + t \approx s + t'$ with $s \succ 0$ can be simplified to $C \vee [\neg]\, t \approx t'$; every clause $C \vee [\neg]\, \psi t \approx \psi t'$ with $\psi \in \Psi \setminus \{1\}$ can be simplified to $C \vee [\neg]\, t \approx t'$.

Demodulation is a classical simplification technique. Demodulating a clause $C[t\sigma]$ means rewriting it to $C[t'\sigma]$ using another clause $t \approx t'$, where $t\sigma \succ t'\sigma$ and $C[t\sigma] \succ_{\mathrm{C}} t \approx t'$. More generally, if C and D are two clauses in N such that $D = D' \vee t \approx t'$, $C = D'\sigma \vee C'[t\sigma]$, $C \succ_{\mathrm{C}} D\sigma$ and $t\sigma \succ t'\sigma$, then C can be simplified to $D'\sigma \vee C'[t'\sigma]$.

We can extend demodulation from the traditional kind of rewriting to cancellative rewriting: If C and D are two clauses in N such that $D = D' \vee t + w \approx w'$, $C = D'\sigma \vee C' \vee [\neg]\, s + t\sigma \approx s'$, $C \succ_{\mathrm{C}} D\sigma$, and $s + t\sigma \approx s' \succ_{\mathrm{L}} s + w'\sigma \approx s' + w\sigma$, then C can be simplified to $D'\sigma \vee C' \vee [\neg]\, s + w'\sigma \approx s' + w\sigma$.

For example, if D is the inverse axiom $(-x) + x \approx 0$, then every clause $C' \vee [\neg]\, (-s) + t \approx t'$ that is larger than $(-s) + s \approx 0$ can be simplified to $C' \vee [\neg]\, t \approx t' + s$. Similarly, for every *cancellation*, *equality resolution*, or *superposition* inference from ground unit clauses, the maximal premise follows from the smaller premise (if any) and the conclusion; hence any such inference constitutes a simplification of the maximal premise. In particular, when we start with a finite set of ground unit clauses over $+$ and constants, then every inference step leads to a set of ground unit clauses over $+$ and constants that is smaller in the multiset extension of the clause ordering $\succ_{\mathrm{C}}$, hence the cancellative superposition calculus must terminate:

LEMMA 5.1. *If membership in Ψ is decidable, then the cancellative superposition calculus is a decision procedure for the satisfiability of finite sets of ground unit clauses over $+$ and constants with respect to* ACUKT_Ψ.

In other words, we can decide the word problem for Ψ-torsion-free cancellative abelian monoids (and the equivalent problem for Ψ-torsion-free abelian groups) whenever membership in Ψ is decidable. Using a refinement of *positive cancellative superposition* similar to *gcd superposition* in (Stuber 1996), this can even be done in quadratic time.

It is crucial for Lemma 5.1 that all free function symbols are constants. Lemma 5.1 is thus weaker than Theorem 5.7 of Marché (1996). If the set of clauses contains non-constant free function symbols, then not only *cancellation*, *cancellative superposition*, and *equality resolution* inferences, but also *standard superposition* and *abstraction* inferences have to be computed. While *standard superposition* is harmless, *abstraction* is not: As the conclusion of an *abstraction* inference from ground unit premises is neither ground nor a unit clause, the subsequent inferences are no longer guaranteed to be simplifications. To use cancellative superposition as a decision procedure for arbitrary ground unit clauses, a significantly more elaborate strategy would be necessary; and it is not known whether such a strategy exists.

5.2. ELIMINATING UNSHIELDED VARIABLES

Let us now return to non-ground cancellative superposition. As we have seen, 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, however. As an example, consider the clauses $\underline{b} + c \approx d$ and $\underline{x} + c \not\approx d$ with the ordering $b \succ c \succ d$. Since unification is not an effective filter, clauses with eligible variables are extremely prolific. In this subsection, we will concentrate on simplification techniques that reduce the number of clauses with eligible variables. As we will show, certain clauses with eligible variables can be removed from the clause set.

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 $\mathfrak{M}$ be a term-generated normal model of $\{C\} \cup \mathrm{ACK}$. Then $\mathfrak{M}$ is a model of C', or $S^{\mathfrak{M}}_{\mathrm{CAM}}$ is a group under the operation $+^{\mathfrak{M}}$.

PROOF. Let $\mathfrak{M}$ be a term-generated normal model of $\{C\} \cup \mathrm{ACK}$. Then $S^{\mathfrak{M}}_{\mathrm{CAM}}$ is a cancellative abelian semigroup under $+^{\mathfrak{M}}$. Suppose that $\mathfrak{M}$ is not a model of C'. Let θ be a substitution that maps all variables in C except x to constants, such that the ground clause $C'\theta$ is false in $\mathfrak{M}$. Consequently, $\mathfrak{M}$ satisfies

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

Hence for every $\mathfrak{m} \in S^{\mathfrak{M}}_{\mathrm{CAM}}$ there is an $i \in I$ such that $\mathfrak{m} +^{\mathfrak{M}} (n_i - 1)\mathfrak{m} +^{\mathfrak{M}} \alpha(s_i\theta) = \alpha(t_i\theta)$, where α is the assignment mapping x to $\mathfrak{m}$. As $t_i\theta$ is ground, $\mathfrak{t}_i = \alpha(t_i\theta)$ is independent of α. Now $\{\, \mathfrak{t}_i \mid i \in I \,\}$ is finite, hence by Thm. A.3, $S^{\mathfrak{M}}_{\mathrm{CAM}}$ is a group. □

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. Bachmair, Ganzinger, and Waldmann 1993). 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, then the first branch can be closed immediately. In this case C implies the identity and inverse axioms, and, although it is not required by fairness, it may be wise to add them to find an easier proof. Second, if some subset N' of N implies that, for every model $\mathfrak{M}$, $S^{\mathfrak{M}}_{\mathrm{CAM}}$ 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.

A similar lemma holds for torsion-free cancellative abelian monoids:

LEMMA 5.3. *Let $\Psi = \mathbf{N}^{>0}$ and 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 s_i or t_i. Let $\mathfrak{M}$ be a term-generated normal model of $\{C\} \cup \mathrm{ACUKT}_\Psi$. Then $\mathfrak{M}$ is a model of C', or $S^{\mathfrak{M}}_{\mathrm{CAM}}$ is a singleton.

PROOF. This is proved analogously to Lemma 5.2 using Cor. A.6. □

In the presence of a non-triviality axiom[23]

$$a \not\approx 0 \qquad \text{(Non-triviality)}$$

(which excludes singleton models), Lemma 5.3 allows us to eliminate every literal with unshielded variables that occur only positively in a clause.[24] In Subsection 5.4 we will exploit this possibility to construct a finite branching variant of the inference system CInf_Ψ for $\Psi = \mathbf{N}^{>0}$.

If a non-triviality axiom is not initially given, then Lemma 5.3 suggests proving theorems in torsion-free cancellative abelian monoids in two steps: Before checking whether some set $N \cup \mathrm{ACUKT}_\Psi$ of clauses has an equality model, we inspect the extended set $N' = N \cup \{x = 0\}$. If $N' \cup \mathrm{ACUKT}_\Psi$ has a model, then so has $N \cup \mathrm{ACUKT}_\Psi$. Otherwise, $N \cup \mathrm{ACUKT}_\Psi$ has no term-generated normal model in which $S^{\mathfrak{M}}_{\mathrm{CAM}}$ is a singleton, so we may consider $N'' = N \cup \{a \not\approx 0\}$ and continue as above.

Unshielded variables occurring negatively are somewhat harder to handle than positive ones. If an unshielded variable occurs in only one negative literal and in no positive literals, we can eliminate it, provided that the coefficient of the variable is 1 and that the inverse axiom

$$(-x) + x \approx 0 \qquad \text{(Inverse)}$$

has been derived: In a group, every clause of the form $C \vee \neg\, x + s \approx s'$, where x does not occur in C, can be simplified to C. If the coefficient is different from 1, the inverse

[23] In non-skolemized form: $\exists y$: $y \not\approx 0$.
[24] We assume that variables occurring on both sides of an equation are cancelled immediately.

axiom alone does not yield enough information to eliminate the literal. In this case, we additionally need division operators *divided-by*$_n$ and divisibility axioms[25]

$$n\ \textit{divided-by}_n(x) \approx x \qquad \text{(Divisibility)}$$

(for $n \in \mathbf{N}^{>0}$). This will be further investigated in Section 6.

So far we have presented ways to deal with unshielded variables that occur either only positively (in "non-groups" and torsion-free cancellative abelian monoids) or negatively and in only one literal (in groups). There is a variant of "rewriting with equations of conditions" which can sometimes be applied if an unshielded variable occurs in more than one literal and at least once negatively. Consider a clause C_0 of the form

$$C' \vee \neg\, nx + s \approx s' \vee [\neg]\, mx + t \approx t'\,.$$

If $m \geq n \geq 1$, then C_0 is equivalent to

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

If $n/\gcd(m,n) \in \Psi$, then it is also equivalent to

$$C' \vee \neg\, nx + s \approx s' \vee [\neg]\, \psi t + \chi s' \approx \psi t' + \chi s\,,$$

where $\psi = n/\gcd(m,n)$ and $\chi = m/\gcd(m,n)$. Exploiting the first and then the second equivalence repeatedly until they are no longer applicable we obtain a clause C_1 in which x occurs only in one negative literal $kx + w \approx w'$ and, if $\Psi \neq \mathbf{N}^{>0}$, possibly in some positive literals (with coefficients smaller than k). Unfortunately, a ground instance $C_1\theta$ of the resulting clause is not necessarily smaller than the corresponding instance $C_0\theta$. This transformation is therefore not a simplification. A method to integrate this kind of variable elimination into cancellative superposition will be presented in Section 6.

5.3. The Standard Case: $\Psi = \{1\}$

It is easy to see that the inference system *CInf*$_\Psi$ is more or less unusable as soon as clauses with unshielded variables have to be handled. Even if we restrict to the special case $\Psi = \{1\}$, the *negative cancellative superposition* and *cancellative equality factoring* rules may produce infinitely many inferences for a given premise or pair of premises. However, if $\Psi = \{1\}$, the system *CInf*$_{\{1\}}$ can be refined to a finitely branching system *WCInf*$_{\{1\}}$. In *WCInf*$_{\{1\}}$, the inference rules *negative cancellative superposition*, *abstraction*, and *cancellative equality factoring* are replaced by new rules whose names are distinguished from the old versions by the prefix *weak*. As saturation with respect to the *weak* rules implies saturation with respect to the original rules, the completeness proof of the previous section carries over to the new calculus.

The main idea behind the *weak negative cancellative superposition* rule is very simple. If we have ground clauses $D\theta = D'\theta \vee \bar{n}\bar{u} + \bar{t} \approx \bar{t}'$ and $C\theta = C'\theta \vee \neg\, \bar{m}\bar{u} + \bar{s} \approx \bar{s}'$, where $\bar{m} = \chi\bar{n}$ and $\chi \in \mathbf{N}^{>0}$, then a *negative cancellative superposition* inference ι produces the clause $C_0 = D'\theta \vee C'\theta \vee \neg\, \bar{s} + \chi\bar{t}' \approx \chi\bar{t} + \bar{s}'$. The premise $C\theta$ contains χ copies of $\bar{n}\bar{u}$; these are replaced by χ copies of $\bar{t}'$, and χ copies of $\bar{t}$ are added to the right-hand side. If we replace only one copy of $\bar{n}\bar{u}$, rather than all χ copies, we obtain the clause $C_0' = D'\theta \vee C'\theta \vee \neg\, (\bar{m}{-}\bar{n})\bar{u} + \bar{s} + \bar{t}' \approx \bar{t} + \bar{s}'$. Like C_0, the clause C_0' is smaller than

[25] In non-skolemized form: $\forall n \in \mathbf{N}^{>0}\ \forall x\ \exists y\colon ny \approx x$.

the premise $C\theta$; furthermore, together with $D\theta$, it implies C_0, making the inference ι redundant. The new inference

$$\frac{D\theta \qquad C\theta}{C_0'}$$

is independent of χ, therefore it can be lifted to non-ground clauses in the same way as the *positive cancellative superposition* rule.

To obtain a finitely branching variant of the *cancellative equality factoring* rule, we consider the size of the literals in a ground *cancellative equality factoring* inference

$$\frac{C'\theta \ \vee \ \bar{n}\bar{u} + \bar{t} \approx \bar{t}' \ \vee \ \bar{m}\bar{u} + \bar{s} \approx \bar{s}'}{C'\theta \ \vee \ \neg \ \bar{t} + \chi\bar{s}' \approx \chi\bar{s} + \bar{t}' \ \vee \ \bar{n}\bar{u} + \bar{t} \approx \bar{t}'}$$

where $\chi\bar{m} = \bar{n}$ and $\chi \in \mathbf{N}^{>0}$. Obviously, $\bar{n} \geq \bar{m}$. On the other hand, the literal $\bar{m}\bar{u} + \bar{s} \approx \bar{s}'$ must be maximal, implying $\bar{m} \geq \bar{n}$. It follows that $\bar{m} = \bar{n}$ and $\chi = 1$. With these additional restrictions, the lifted rule becomes finitary branching.

As we have already mentioned, the *abstraction* rule differs from the other inference rules in that the validity of its conclusion depends only on the validity of its second premise and that its conditions serve only to restrict the number of necessary inferences. The conditions are not required for correctness – if checking them turns out to be too expensive, they may be weakened or even ignored. For instance, we note that an *abstraction* inference is usually possible if the set of variables $\mathrm{lhs}(e_2) \setminus \mathrm{Var}(\mathrm{rhs}(e_2)) \setminus \mathrm{Var}(\mathrm{neg}(D'))$ contains an eligible variable, or if w has the form $mx + q$ and either $m > 1$, or $m = 1$ and $q = q_1 + q_2$ where q_1 is a variable or a non-zero atomic term not containing x as a subterm. It remains only to consider the situation where $e_2 = \sum_{l \in L} n_l^* v_l + t \approx t'$ and $w = \sum_{k \in K} m_k^* u_k + q$ with atomic terms u_k and v_l. In this case, an inference is possible as soon as all u_k and v_l are ACU-unifiable and $\sum_{k \in K} m_k^* \geq \sum_{l \in L} n_l^*$. We combine these conditions in the *weak abstraction* inference rule.

Inference System 5.4. The inference system $\mathit{WCInf}_{\{1\}}$ consists of the inference rules *cancellation*, *equality resolution*, *standard superposition*, *positive cancellative superposition*, and *standard equality factoring* of Inference System 3.16, and of the inference rules *weak negative cancellative superposition*, *weak abstraction*, and *weak cancellative equality factoring*, as described below.

Weak Negative Cancellative Superposition

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

if the following conditions are satisfied:

(i) $e_1 \ = \ \sum_{i \in I} m_i x_i + \sum_{k \in K} m_k^* u_k + s \approx s'$.
(ii) $e_2 \ = \ \sum_{j \in J} n_j y_j + \sum_{l \in L} n_l^* v_l + t \approx t'$.
(iii) $e_0 \ = \ z + \sum_{i \in I} m_i \check{x}_i + s + t' \approx \sum_{j \in J} n_j \check{y}_j + t + s'$.
(iv) $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
(v) $\{ x_i \mid i \in I \} = \mathrm{elig}(C' \vee \neg e_1) \cap \mathrm{Var}(\mathrm{lhs}(e_1)) \setminus \mathrm{Var}(\mathrm{rhs}(e_1))$,
$\{ y_j \mid j \in J \} = \mathrm{elig}(D' \vee e_2) \cap \mathrm{Var}(\mathrm{lhs}(e_2)) \setminus \mathrm{Var}(\mathrm{rhs}(e_2)) \setminus \mathrm{Var}(\mathrm{neg}(D'))$.
(vi) $\mathrm{lhs}(e_1)$ is not a variable (i. e., either $\sum_{i \in I} m_i > 1$ or $\sum_{k \in K} m_k^* u_k + s \neq 0$).
(vii) If $I = \{i_1\}$, $m_{i_1} = 1$, and $K = \emptyset$, then $J \neq \emptyset$ or $t \neq 0$.
(viii) If $K \cup L \neq \emptyset$, u is one of the u_k or v_l ($k \in K$, $l \in L$), otherwise, u is a new variable.

(ix) σ_1 maps x_i to $\hat{x}_i + \check{x}_i$ and y_j to $\hat{y}_j + \check{y}_j$ for $i \in I$, $j \in J$; σ_2 is a most general ACU-unifier of all u_k and v_l $(k \in K,\, l \in L)$; σ_3 is a most general ACU-unifier of $\sum_{i\in I} m_i\hat{x}_i + \bigl(\sum_{k\in K} m_k^*\bigr)u\sigma_2$ and $z + \sum_{j\in J} n_j\hat{y}_j + \bigl(\sum_{l\in L} n_l^*\bigr)u\sigma_2$ such that $\bigl(\sum_{j\in J} n_j\hat{y}_j + \bigl(\sum_{l\in L} n_l^*\bigr)u\sigma_2\bigr)\sigma_3$ is not identical to 0; and $\sigma = \sigma_1\sigma_2\sigma_3$.

(x) $u\sigma \not\preceq s\sigma$, $u\sigma \not\preceq s'\sigma$, $u\sigma \not\preceq t\sigma$, $u\sigma \not\preceq t'\sigma$, $u\sigma \not\preceq \check{x}_i\sigma$ for $i \in I$, $u\sigma \not\preceq \check{y}_j\sigma$ for $j \in J$.

Weak Abstraction

$$\frac{D' \vee e_2 \qquad C' \vee [\neg]\, s[w] \approx s'}{C' \vee \neg\, y \approx w \vee [\neg]\, s[y] \approx s'}$$

if the following conditions are satisfied:

(i) w occurs in s immediately below some free function symbol and has sort S_{CAM}.

(ii) Either:

(a) $\mathrm{elig}(D' \vee e_2) \cap \mathrm{Var}(\mathrm{lhs}(e_2)) \setminus \mathrm{Var}(\mathrm{rhs}(e_2)) \setminus \mathrm{Var}(\mathrm{neg}(D')) \neq \emptyset$.

(b) w is not a variable.

or:

(a) $e_2 \;=\; \sum_{l\in L} n_l^* v_l + t \approx t'$.

(b) $L \neq \emptyset$.

(c) Either:

(α) $w \;=\; mx + q$.

(β) Either $m > 1$, or $m = 1$ and $q = q_1 + q_2$ where q_1 is a variable or a non-zero atomic term not containing x as a subterm.

(γ) $\mathrm{lhs}(e_2)$ is neither an atomic term nor a subterm of w.

(δ) $v_l \not\preceq t$, $v_l \not\preceq t'$ for some $l \in L$.

or:

(α) $w \;=\; \sum_{k\in K} m_k^* u_k + q$.

(β) ρ is a most general ACU-unifier of all u_k and v_l $(k \in K,\, l \in L)$.

(γ) $\sum_{k\in K} m_k^* \geq \sum_{l\in L} n_l^*$.

(δ) $\mathrm{lhs}(e_2)\rho$ is not a subterm of $w\rho$.

(ϵ) $v_l\rho \not\preceq t\rho$, $v_l\rho \not\preceq t'\rho$ for all $l \in L$.

Weak Cancellative Equality Factoring

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

if the following conditions are satisfied:

(i) $e_1 \;=\; \sum_{i\in I} m_i x_i + \sum_{k\in K} m_k^* u_k + s \approx s'$.

(ii) $e_2 \;=\; \sum_{j\in J} n_j x_j + \sum_{l\in L} n_l^* v_l + t \approx t'$.

(iii) $e_0 \;=\; \sum_{j\in J} n_j \check{x}_j + t + s' \approx \sum_{i\in I} m_i \check{x}_i + s + t'$.

(iv) $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.

(v) $\{ x_i \mid i \in I \} = \mathrm{elig}(C' \vee e_2 \vee e_1) \cap \mathrm{Var}(\mathrm{lhs}(e_1)) \setminus \mathrm{Var}(\mathrm{rhs}(e_1)) \setminus \mathrm{Var}(\mathrm{rhs}(e_2)) \setminus \mathrm{Var}(\mathrm{neg}(C'))$,
$\{ x_j \mid j \in J \} = \mathrm{elig}(C' \vee e_2 \vee e_1) \cap \mathrm{Var}(\mathrm{lhs}(e_2)) \setminus \mathrm{Var}(\mathrm{rhs}(e_1)) \setminus \mathrm{Var}(\mathrm{rhs}(e_2)) \setminus \mathrm{Var}(\mathrm{neg}(C'))$.

(vi) If $K \cup L \neq \emptyset$, u is one of the u_k or v_l ($k \in K$, $l \in L$), otherwise, u is a new variable.

(vii) σ_1 maps x_i to $\hat{x}_i + \check{x}_i$ for $i \in I \cup J$; σ_2 is a most general ACU-unifier of all u_k and v_l ($k \in K, l \in L$); σ_3 is a most general ACU-unifier of $\sum_{i\in I} m_i \hat{x}_i + \big(\sum_{k\in K} m_k^*\big) u\sigma_2$ and $\sum_{j\in J} n_j \hat{x}_j + \big(\sum_{l\in L} n_l^*\big) u\sigma_2$; and $\sigma = \sigma_1\sigma_2\sigma_3$.

(viii) $u\sigma \not\preceq s\sigma$, $u\sigma \not\preceq s'\sigma$, $u\sigma \not\preceq t\sigma$, $u\sigma \not\preceq t'\sigma$, $u\sigma \not\preceq \check{x}_i\sigma$ for $i \in I \cup J$.

Instances of *weak negative cancellative superposition*, *weak abstraction*, and *weak cancellative equality factoring* inferences are defined as previously in Def. 4.30. With these implements, we obtain again lifting lemmas analogously to Lemmas 4.36 and 4.37. The definitions of redundancy (Def. 4.31 and 4.32) are extended verbatim to the new kinds of inferences. It is easy to show that Lemma 4.34 holds also for Inference System 5.4, that is, that $\mathit{CRed}_{\{1\}}$ is also a redundancy criterion with respect to $\mathit{WCInf}_{\{1\}}$. We can then link saturation with respect to $\mathit{WCInf}_{\{1\}}$ to saturation with respect to $\mathit{CInf}_{\{1\}}$ as described at the beginning of this section:

LEMMA 5.5. *Let N be a set of clauses. If $\Psi = \{1\}$ and if every weak negative cancellative superposition inference from clauses in N is redundant with respect to N, then every negative cancellative superposition inference from clauses in N is redundant with respect to N. (Analogously for weak abstraction inferences and weak cancellative equality factoring inferences.)*

THEOREM 5.6. *If a set of clauses if saturated with respect to $\mathit{WCInf}_{\{1\}}$ and $\mathit{CRed}_{\{1\}}$, then it is also saturated with respect to $\mathit{CInf}_{\{1\}}$ and $\mathit{CRed}_{\{1\}}$.*

This means that we can obtain a $\mathit{CInf}_{\{1\}}$-saturated set of clauses not only as the limit of a fair $\mathit{CInf}_{\{1\}}$-derivation, but also as the limit of a fair $\mathit{WCInf}_{\{1\}}$-derivation. Theorem 4.42 can therefore be extended from $\mathit{CInf}_{\{1\}}$-derivations to $\mathit{WCInf}_{\{1\}}$-derivations.

5.4. The Torsion-Free Case: $\Psi = \mathbf{N}^{>0}$

In the previous subsection, we have shown how the infinitely branching inference system $\mathit{CInf}_{\{1\}}$ can be transformed into a finitely branching system $\mathit{WCInf}_{\{1\}}$. A similar refinement is possible if $\Psi = \mathbf{N}^{>0}$. For lack of space, we will only sketch the construction of this inference system; the reader is referred to (Waldmann 1997) for the details.

We have shown in Subsection 5.2 that in the $\mathbf{N}^{>0}$-torsion-free case all unshielded variables that occur only positively can be eliminated. The new inference system $\mathit{WCInf}_{\mathbf{N}^{>0}}$ is equivalent to $\mathit{CInf}_{\mathbf{N}^{>0}}$, provided that the elimination of unshielded variables is performed eagerly. Again, the inference rules *negative cancellative superposition*, *abstraction*, and *cancellative equality factoring* are replaced by new rules, whose redundancy implies the redundancy of the old rules.

Let $\Psi = \mathbf{N}^{>0}$ and let us consider a *negative cancellative superposition* inference ι from ground clauses $D\theta = D'\theta \vee \bar{n}\bar{u} + \bar{t} \approx \bar{t}'$ and $C\theta = C'\theta \vee \neg\, \bar{m}\bar{u} + \bar{s} \approx \bar{s}'$, producing the clause $C_0 = D'\theta \vee C'\theta \vee \neg\, \psi\bar{s} + \chi\bar{t}' \approx \chi\bar{t} + \psi\bar{s}'$, where $\psi\bar{m} = \chi\bar{n}$, $\psi \in \mathbf{N}^{>0}$, and

$\chi \in \mathbf{N}^{>0}$. Provided that $\bar{m} \geq \bar{n}$, we can use the same trick as in the case $\Psi = \{1\}$: By replacing $\bar{n}\bar{u}$ by $\bar{t}'$ in $\bar{m}\bar{u} + \bar{s} \approx \bar{s}'$ and by adding $\bar{t}$ to the right-hand side, we obtain a clause $C_0' = D'\theta \vee C'\theta \vee \neg\, (\bar{m}{-}\bar{n})\bar{u} + \bar{s} + \bar{t}' \approx \bar{t} + \bar{s}'$. This clause is smaller than the premise $C\theta$; together with $D\theta$, it implies C_0. Hence, if $\bar{m} \geq \bar{n}$, then the inference ι is redundant whenever the inference

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

is redundant. We call inferences of this kind *weak negative cancellative superposition (I)* inferences; they can be lifted in the same way as in Subsection 5.3.

It remains to consider ground inferences

$$\frac{D'\theta \vee \bar{n}\bar{u} + \bar{t} \approx \bar{t}' \qquad C'\theta \vee \neg\, \bar{m}\bar{u} + \bar{s} \approx \bar{s}'}{D'\theta \vee C'\theta \vee \neg\, \psi\bar{s} + \chi\bar{t}' \approx \chi\bar{t} + \psi\bar{s}'}$$

where $\psi\bar{m} = \chi\bar{n}$, $\psi \in \mathbf{N}^{>0}$, and $\chi \in \mathbf{N}^{>0}$, and additionally $0 < \bar{m} < \bar{n}$. We call such inferences *weak negative cancellative superposition (II)* inferences. Lifting them in a finitely branching manner becomes possible by virtue of the fact that we restrict to clauses C and D without unshielded variables occurring only positively. If $\bar{n}\bar{u} + \bar{t} \approx \bar{t}'$ is the ground instance of an equation $\sum_{l \in L} n_l^* v_l + t \approx t'$, then $\bar{u}$ does not result from instantiations of variables. Every $\bar{u}$ is the instance of some v_l, hence $\bar{n} = \sum_{l \in L} n_l^*$. In other words, even for non-ground clauses, the number $\bar{n}$ of the occurrences of the maximal atomic subterm in the ground instance is known. As $0 < \bar{m} < \bar{n}$, this leaves only finitely many possibilities for $\bar{m}$. Since there exists at most one pair (ψ, χ) for given $\bar{n}$ and $\bar{m}$, it is no problem anymore to lift the inference in a finitely branching way.

The *weak abstraction* rule is constructed according to the same principles as the corresponding rule of Inference System 5.4. Some differences are, on the one hand, due to the fact that unshielded variables occurring only positively can be excluded. On the other hand, for $\Psi = \mathbf{N}^{>0}$, a clause $D\theta = D'\theta \vee \bar{n}\bar{u} + \bar{t} \approx \bar{t}'$ makes the abstraction of a term $\bar{m}\bar{u} + \bar{q}$ necessary, even if $\bar{n} > \bar{m}$.

Making the *cancellative equality factoring* rule of Inference System 3.16 finitely branching turns out to be almost trivial: In the absence of unshielded variables occurring only positively, the conditions of this rule can be satisfied only if $\chi = n^* / \gcd(m^*, n^*)$ and $\psi = m^* / \gcd(m^*, n^*)$, where $m^* = \sum_{k \in K} m_k^*$ and $n^* = \sum_{l \in L} n_l^*$.

THEOREM 5.7. *If a set of clauses if saturated with respect to $\mathit{WCInf}_{\mathbf{N}^{>0}}$ and $\mathit{CRed}_{\mathbf{N}^{>0}}$ and none of the clauses contains unshielded variables that occur only positively, then it is also saturated with respect to $\mathit{CInf}_{\mathbf{N}^{>0}}$ and $\mathit{CRed}_{\mathbf{N}^{>0}}$.*

By Thm. A.7, an abelian monoid is cancellative and torsion-free if and only if it can be extended to a totally ordered abelian monoid. As a consequence of this fact, the calculi $\mathit{WCInf}_{\mathbf{N}^{>0}}$ (or $\mathit{CInf}_{\mathbf{N}^{>0}}$) can also be used to prove theorems in totally ordered abelian monoids, as long as the ordering predicate occurs only in the ordering axioms but nowhere else in the theorem (Waldmann 1997). Otherwise, a combination of cancellative superposition and chaining (Slagle 1972, Bachmair and Ganzinger 1994b) is necessary (cf. Waldmann 2001).

5.5. Full Abstraction

The *abstraction* inference rule is problematic from an efficiency point of view. In comparison with the second premise, the conclusion of an inference

$$\frac{D \qquad C[t]}{y \not\approx t \vee C[y]}$$

contains more variables and more incomparable terms, which both add to the number of inferences in which this clause can participate.[26] Furthermore, even though $y \not\approx t \vee C[y]$ is logically equivalent to $C[t]$, an *abstraction* inference is in general not a simplification, hence both the second premise and the conclusion must be kept and may be used in further inferences.

In our calculi, working with abstracted clauses is necessary for completeness. It is possible, however, to avoid working with both abstracted *and unabstracted* clauses:

A clause C is called fully abstracted, if no non-variable term of sort S_{CAM} occurs below a free function symbol in C. Every clause C can be transformed into an equivalent fully abstracted clause $\text{abs}(C)$ by iterated rewriting

$$C[f(\dots,t,\dots)] \quad\to\quad x \not\approx t \vee C[f(\dots,x,\dots)]\,,$$

where x is a new variable and t is a non-variable term of sort S_{CAM} occurring immediately below the free function symbol f in C.

We can replace every clause C in the input of the inference system by the logically equivalent clause $\text{abs}(C)$ before we start the saturation process. With a little care, we can also ensure that the clauses derived during the saturation process are again fully abstracted. We only have to avoid selecting literals in such a way that a shielded variable becomes eligible.[27]

In the worst case, full abstraction may intensify the efficiency problems of the the *abstraction* inference rule. On the other hand, the cancellative superposition calculus requires many *abstraction* inferences anyway, in particular in the presence of clauses with unshielded variables in positive literals. In this case, almost every clause in which a non-variable term of sort S_{CAM} occurs below a free function symbol can participate as a second premise in an *abstraction* inference. Furthermore, if all clauses are fully abstracted, then the terms that must be compared during the saturation do not contain the operator $+$. This greatly enlarges the assortment of orderings with which the calculus can be parameterized: We are no longer restricted to the small number of known ACU-orderings, but may use an arbitrary reduction ordering over terms not containing $+$ that is total on ground terms and for which 0 is minimal: As every ordering of this kind can be extended to an ordering that is ACU-compatible and has the multiset property (Waldmann 1998a), the completeness proof is still justified. In particular, full abstraction allows us to use classes of orderings that are more efficient in practice than LPO or RPO, for instance the Knuth-Bendix ordering.

Surprisingly, there are even examples where full abstraction *reduces* the number of inferences, compared with, say, Godoy and Nieuwenhuis's (2000) paramodulation calculus modulo abelian groups (where the presence of the inverse axiom makes it possible to

[26] Note that *equality resolution* inferences with the new variable y are prohibited by the ordering restrictions, though.

[27] In other words: a variable is shielded is the full clause, but unshielded in the subclause consisting of all the selected literals.

avoid abstraction). Consider the two clauses $2g(y) = b$ and $f(2x + b) \not\approx f(2x)$. In Godoy and Nieuwenhuis's calculus there is an *AG-superposition* inference of the first into the second clause using variable splitting. The conclusion is $f(2x_2 + 2b) \not\approx f(2x)$ with the constraint $x = x_1 + x_2 \wedge 2x_1 = 2g(y)$. Since this conclusion contains again a variable with multiplicity 2 below a free function symbol, we can continue with a further *AG-superposition* inference of the first clause into the new clause, yielding $f(2x_4 + 3b) \not\approx f(2x)$ with the constraint $x_2 = x_3 + x_4 \wedge 2x_3 = 2g(y') \wedge x = x_1 + x_2 \wedge 2x_1 = 2g(y)$, and so on. Consequently, Godoy and Nieuwenhuis's calculus does not terminate for the two input clauses given above. On the other hand, full abstraction of the second clause yields $z_1 \not\approx 2x + b \vee z_2 \not\approx 2x \vee f(z_1) \not\approx f(z_2)$. If we select the last literal, then only one inference is possible with this clause, namely a *cancellation* inference yielding $z_1 \not\approx 2x + b \vee z_1 \not\approx 2x \vee 0 \not\approx 0$. This clause can be simplified immediately to $b \not\approx 0$ using the techniques of Subsection 5.2. From $b \not\approx 0$ and $2g(y) = b$ no further inferences are possible.

6. Divisible Torsion-Free Abelian Groups

As shown in Subsection 5.2, certain unshielded variables can be eliminated in the presence of the torsion-freeness axioms, the non-triviality axiom, the inverse axiom, or the divisibility axioms. We will now investigate the effect of the combination of these axioms.

Let us denote the union of the set of divisibility axioms, the inverse axiom, and the non-triviality axiom by DivInvNt. Algebraic structures that satisfy the axioms $\mathrm{ACUKT}_{\mathbf{N}^{>0}} \cup \mathrm{DivInvNt}$ are called divisible torsion-free abelian groups. Typical examples are the rational numbers and rational vector spaces. It is well-known that the theory of divisible torsion-free abelian groups allows quantifier elimination: every formula over 0, $+$, and $\approx$ can be transformed into an equivalent quantifier-free formula. In particular, every closed formula over this vocabulary is either true in all divisible torsion-free abelian groups or false in all divisible torsion-free abelian groups. In the presence of free function symbols (and possibly other sorts), there is of course no way to eliminate all variables from a clause, but we can at least give an effective method to eliminate all *unshielded* variables. Using this elimination algorithm, we will then construct a new inference system that operates on clauses without unshielded variables and again produces such clauses.

The integration of the variable elimination algorithm demands a restriction of the cancellative superposition calculus: We have to dispense with selection functions. The reason for this is rather technical in nature and will become clear later.

We write $\models_{\mathrm{DTAG}}$ for the entailment relation modulo $\mathrm{ACUKT}_{\mathbf{N}^{>0}} \cup \mathrm{DivInvNt}$ and equality. In other words, $N \models_{\mathrm{DTAG}} N'$ if and only if $N \cup \mathrm{DivInvNt} \models_{\mathbf{N}^{>0}} N'$. Throughout this section we consider only sets of clauses comprising the set of additional theory axioms DivInvNt, so that deduction using $\models_{\mathrm{DTAG}}$ is legal. However, our redundancy concept is still based on $\models_{\mathbf{N}^{>0}}$. This has several important consequences: First of all, it means that the axioms DivInvNt themselves are in general not redundant, and even some inferences with them may be necessary. Secondly, if we want to prove the redundancy of some other clause using the theory axioms $\mathrm{ACUKT}_{\mathbf{N}^{>0}} \cup \mathrm{DivInvNt}$, then we may use arbitrarily large instances of the $\mathrm{ACUKT}_{\mathbf{N}^{>0}}$ axioms, whereas the instances of the DivInvNt axioms must obey the size restriction spelled out in the definition of redundancy.

Let x be a variable of sort S_{CAM}. We define a binary relation $\to_x$ over clauses by

CancelVar	$C' \vee [\neg]\, mx + s \approx m'x + s' \quad \to_x \quad C' \vee [\neg]\, (m{-}m')x + s \approx s'$ if $m \geq m' \geq 1$.
ElimNeg	$C' \vee \neg\, mx + s \approx s' \quad \to_x \quad C'$ if $m \geq 1$ and x does not occur in C', s, s'.
ElimPos	$C' \vee m_1 x + s_1 \approx s'_1 \vee \ldots \vee m_k x + s_k \approx s'_k \quad \to_x \quad C'$ if $m_i \geq 1$ and x does not occur in C', s_i, s'_i, for $1 \leq i \leq k$.
Coalesce	$C' \vee \neg\, mx + s \approx s' \vee [\neg]\, nx + t \approx t'$ $\to_x \quad C' \vee \neg\, mx + s \approx s' \vee [\neg]\, \psi t + \chi s' \approx \psi t' + \chi s$ if $m \geq 1$, $n \geq 1$, $\psi = m/\gcd(m,n)$, $\chi = n/\gcd(m,n)$, and x does not occur at the top of s, s', t, t'.

LEMMA 6.1. *If $C_0 \to_x C_1$, then $\{C_1\} \models_{\mathbf{N}^{>0}} C_0$ and $\{C_0\} \models_{\text{DTAG}} C_1$. If $C_0\theta$ is a ground instance of C_0, then $\{C_1\theta\} \models_{\mathbf{N}^{>0}} C_0\theta$.*

PROOF. If $C_0 \to_x C_1$ by *CancelVar*, the equivalence of C_0 and C_1 modulo $\text{ACUKT}_{\mathbf{N}^{>0}}$ follows from cancellation; for *Coalesce*, from cancellation and torsion-freeness. The soundness of *ElimNeg* follows from the divisibility and the inverse axiom; for *ElimPos* it is implied by torsion-freeness and non-triviality (see Lemma 5.3). □

LEMMA 6.2. *If $C_0 \to_x C_1$, then every variable or atomic term occurring (negatively) in C_1 also occurs (negatively) in C_0.*

COROLLARY 6.3. *If $C_0 \to_x C_1$, x occurs in both C_0 and C_1, and x is unshielded in C_0, then x is unshielded in C_1.*

LEMMA 6.4. *The relation $\to_x$ is noetherian.*

LEMMA 6.5. *If $x \in \text{Var}(C)$ is unshielded in C, then C is reducible with respect to $\to_x$.*

COROLLARY 6.6. *If C_1 is a normal form of C_0 with respect to $\to_x$, and x is unshielded in C_0, then x does not occur in C_1.*

The relation $\to_{\text{elim}}$ is defined in such a way that $C_0 \to_{\text{elim}} C_1$ if and only if C_0 contains an unshielded variable x and C_1 is a normal form of C_0 with respect to $\to_x$.

Combining Lemma 6.2 and Cor. 6.6, we see that $C_0 \to_{\text{elim}} C_1$ implies $\text{Var}(C_1) \subset \text{Var}(C_0)$. As the number of variables in a clause is finite, we obtain the following corollary:

COROLLARY 6.7. *The relation $\to_{\text{elim}}$ is noetherian.*

Every $\to_{\text{elim}}$-rewrite step eliminates at least one unshielded variable (non-deterministically). For a clause C, $\text{elim}(C)$ denotes some (arbitrary but fixed) normal form of C with respect to the relation $\to_{\text{elim}}$.

COROLLARY 6.8. *For any C, $\text{elim}(C)$ contains no unshielded variables.*

COROLLARY 6.9. *For every clause C, every variable or atomic term occurring (negatively) in* $\mathrm{elim}(C)$ *occurs also (negatively) in C. For every ground instance $C\theta$, every atomic term occurring (negatively) in* $\mathrm{elim}(C)\theta$ *occurs also (negatively) in $C\theta$.*

COROLLARY 6.10. *For every clause C,* $\{C\} \models_{\mathrm{DTAG}} \mathrm{elim}(C)$ *and* $\{\mathrm{elim}(C)\} \models_{\mathbf{N}^{>0}} C$. *For every ground instance $C\theta$,* $\{\mathrm{elim}(C)\theta\} \models_{\mathbf{N}^{>0}} C\theta$.

Using the technique sketched so far, every clause C_0 can be transformed into a clause $\mathrm{elim}(C_0)$ that does not contain unshielded variables, follows from C_0 and the divisible torsion-free abelian group axioms, and implies C_0 modulo $\mathrm{ACUKT}_{\mathbf{N}^{>0}}$. Obviously, we can perform this transformation for all initially given clauses *before* we start the saturation process. However, the set of clauses without unshielded variables is not closed under the inference system $\mathit{CInf}_{\mathbf{N}^{>0}}$: inferences from clauses without unshielded variables may produce clauses with unshielded variables. To eliminate these clauses *during* the saturation process, logical equivalence is not sufficient: We have to require either that the transformed clause $\mathrm{elim}(C_0)$ makes the original clause C_0 redundant, or at least that it makes the inference producing C_0 redundant.

The second condition it slightly easier to satisfy: Let ι be an inference with maximal premise C and conclusion C_0. For the redundancy of C_0 it is necessary that each of its ground inferences $C_0\theta$ follows from $\overline{N}^{\prec_{\mathrm{C}} C_0\theta}$. For the redundancy of ι, it is sufficient that for each ground instance of ι, $C_0\theta$ follows from $\overline{N}^{\prec_{\mathrm{C}} C\theta}$. But even the latter property is not guaranteed for our variable elimination algorithm, as the following example demonstrates:

EXAMPLE 6.11. Let the ordering on constants be given by $b \succ c$ and consider the clause

$$C \;=\; 3x \not\approx c \,\vee\, x + f(z) \approx 0 \,\vee\, f(x) + b \approx f(y)\,.$$

A *cancellation* inference ι from C yields

$$C_0 \;=\; 3x \not\approx c \,\vee\, x + f(z) \approx 0 \,\vee\, b \approx 0\,.$$

The conclusion C_0 contains the unshielded variable x. Eliminating x from C_0, we obtain

$$\mathrm{elim}(C_0) \;=\; c + 3f(z) \approx 0 \,\vee\, b \approx 0\,.$$

Now let $\theta = \{x \mapsto b,\, z \mapsto b\}$, then

$$\mathrm{elim}(C_0)\theta \;=\; c + 3f(b) \approx 0 \,\vee\, b \approx 0$$

is not only strictly larger than

$$C_0\theta \;=\; 3b \not\approx c \,\vee\, b + f(b) \approx 0 \,\vee\, b \approx 0\,,$$

but even strictly larger than

$$C\theta \;=\; 3b \not\approx c \,\vee\, b + f(b) \approx 0 \,\vee\, f(b) + b \approx f(b)\,.$$

Hence the clause $\mathrm{elim}(C_0)$ makes neither C_0 nor ι redundant.

To integrate the variable elimination algorithm into the cancellative superposition calculus, it has to be supplemented by a case analysis technique.

Let $C_1, \dots, C_k$ be clauses without unshielded variables and let ι be a $\mathit{CInf}_{\mathbf{N}^{>0}}$-inference

$$\frac{C_k \;\dots\; C_1}{C_0\sigma}$$

We call the unifying substitution σ that is computed during ι and applied to the conclusion the pivotal substitution of ι. (For *abstraction* inferences and all ground inferences, the pivotal substitution is the identity mapping.) If C_1 has the form $C_1' \vee [\neg]\, e$, where the replacement or abstraction or cancellation takes place at $[\neg]\, e$, then we call $[\neg]\, e\sigma$ the pivotal literal of ι. Finally, if u_0 is the atomic term that is cancelled out in ι, or in which some subterm is replaced or abstracted out,[28] then we call $u_0\sigma$ the pivotal term of ι.

Two properties of pivotal terms are important for us: First, whenever an inference ι from clauses *without* unshielded variables produces a conclusion *with* unshielded variables, then all these unshielded variables occur in the pivotal term of ι. Second, no atomic term in the conclusion of ι can be larger than the pivotal term of ι.

Lemma 6.12. *In every ground inference, the pivotal term is maximal among the atomic terms occurring in the premises, and the pivotal literal is maximal among the literals of the premises.*

Notice that Lemma 6.12 does not hold in the presence of selection functions. Nor does it hold for the *merging paramodulation* rule of Bachmair and Ganzinger (1994c).

Lemma 6.13. *Let ι be an inference from clauses without unshielded variables, let $\iota\theta$ be a ground instance of ι. Then the pivotal term of $\iota\theta$ is a ground instance of the pivotal term of ι.*

Whenever we talk about a ground instance $\iota\theta$ of an inference ι, we assume without loss of generality that θ is defined on all variables of the pivotal term u_0 of ι, and that the pivotal term of $\iota\theta$ is $u_0\theta$.

Lemma 6.14. *Let ι be an inference from clauses without unshielded variables; let C_0 be the conclusion and σ be the pivotal substitution of ι. Let C be some premise of ι (if ι is an abstraction inference: the second premise). If t is an atomic term that occurs in $C\sigma$, but not in C_0, then t is a subterm of the pivotal term of ι.*

Corollary 6.15. *Let ι be an inference from clauses without unshielded variables. Then every variable that is unshielded in the conclusion of ι occurs in the pivotal term of ι.*

Lemma 6.16. *Let ι be a non-abstraction inference with maximal premise C and conclusion C_0; let $D_0 = \mathrm{elim}(C_0)$. Let $\iota\theta$ be a ground instance of ι, and let $[\neg]\, e\theta$ be the pivotal literal of $\iota\theta$. If $C\theta \preceq_{\mathrm{C}} D_0\theta$, then the multiset difference $D_0\theta \setminus C_0\theta$ contains a literal $[\neg]\, e_1\theta$, such that $[\neg]\, e_1\theta$ has the same polarity as $[\neg]\, e\theta$ and the pivotal term of $\iota\theta$ occurs in $[\neg]\, e_1\theta$.*

[28] More precisely, u_0 is the maximal atomic subterm of s containing t (or w) in *standard superposition* or *abstraction* inferences, and the term u in all other inferences.

PROOF. In every ground inference, the conclusion consists of the literals of the last premise, minus the pivotal literal, plus possibly other literals that are smaller than the pivotal literal. As the clause ordering is the multiset extension of the literal ordering, the conclusion is thus smaller than the last premise.

Let us consider the ground inference $\iota\theta$. Its conclusion $C_0\theta$ is smaller than $C\theta$. If all literals in $D_0\theta \setminus C_0\theta$ are smaller than $[\neg]\, e\theta$, we can conclude that $C\theta \succ_{\mathrm{C}} D_0\theta$.

Conversely, if $C\theta \preceq_{\mathrm{C}} D_0\theta$, then some literal $[\neg]\, e_1\theta$ in $D_0\theta \setminus C_0\theta$ is greater or equal than $[\neg]\, e\theta$. As the literal ordering depends first on the maximal atomic terms in the literals to be compared, $[\neg]\, e_1\theta \succeq_{\mathrm{L}} [\neg]\, e\theta$ implies $\mathrm{mt}(e_1\theta) \succeq \mathrm{mt}(e\theta)$. On the other hand, every atomic term in $D_0\theta$ occurs also in $C_0\theta$, and is thus not greater than the pivotal term of $\iota\theta$, that is $\mathrm{mt}(e\theta)$. Consequently, $\mathrm{mt}(e_1\theta) = \mathrm{mt}(e\theta)$.

If the maximal atomic terms in the literals to be compared are equal, then the literal ordering compares the polarity of the literals, so $[\neg]\, e_1\theta \succeq_{\mathrm{L}} [\neg]\, e\theta$ and $\mathrm{mt}(e_1\theta) = \mathrm{mt}(e\theta)$ implies that either both $[\neg]\, e_1\theta$ and $[\neg]\, e\theta$ are negative, or both are positive, or $[\neg]\, e_1\theta$ is negative and $[\neg]\, e\theta$ is positive. We will show that the last of these cases is impossible: Assume $[\neg]\, e_1\theta$ is negative. Then $\mathrm{mt}(e_1\theta) = \mathrm{mt}(e\theta)$ occurs in a negative literal of $D_0\theta = \mathrm{elim}(C_0)\theta$, and hence also in a negative literal of $C_0\theta$. But if $[\neg]\, e\theta$, that is, the pivotal literal of $\iota\theta$ is positive, then $\mathrm{mt}(e\theta)$ does not occur negatively in $C\theta$. So $C\theta$ would be smaller than $C_0\theta$, which is impossible. Hence, $[\neg]\, e_1\theta$ and $[\neg]\, e\theta$ have the same polarity. ☐

LEMMA 6.17. *Let ι be a non-abstraction inference from clauses without unshielded variables with maximal premise C, conclusion C_0, pivotal literal $[\neg]\, e$, and pivotal term u; let $D_0 = \mathrm{elim}(C_0)$. Let $\iota\theta$ be a ground instance of ι. If $C\theta \preceq_{\mathrm{C}} D_0\theta$, then the multiset difference $D_0 \setminus C_0$ contains a literal $[\neg]\, e_1$, such that:*

(i) $[\neg]\, e_1$ *has the same polarity as* $[\neg]\, e$,
(ii) there is an atomic term u_1 occurring at the top of e_1,
(iii) for every minimal complete set U of ACU*-unifiers of u and u_1, there is a $\tau \in U$ such that $C_0\theta$ is a ground instance of $C_0\tau$.*

Furthermore, for every $\tau \in U$, $C_0\tau$ has no unshielded variables.

PROOF. By Lemma 6.16, the multiset difference $D_0\theta \setminus C_0\theta$ contains a literal that has the same polarity as $[\neg]\, e$ and contains the pivotal term $\bar{u} = u\theta$ of $\iota\theta$. It is easy to see that this literal is an instance $[\neg]\, e_1\theta$ of some literal $[\neg]\, e_1$ in the multiset difference $D_0 \setminus C_0$.

Assume that the term $\bar{u}$ in $e_1\theta$ results from instantiating a variable x occurring at the top of e_1. We will show that this is impossible: Let $x\theta = m\bar{u} + \bar{s}$. As D_0 has no unshielded variables, some atomic term $v[x]$ must occur in D_0 and, by Cor. 6.9, in C_0. Consequently, the term $v[x]\theta = v\theta[m\bar{u} + \bar{s}]$ occurs in $C_0\theta$. But $v\theta[m\bar{u} + \bar{s}]$ is larger than the pivotal term $\bar{u}$ of $\iota\theta$, contradicting the fact that $C_0\theta \prec_{\mathrm{C}} C\theta$.

As $\bar{u}$ in $e_1\theta$ cannot result from instantiating a variable x of e_1, some atomic term u_1 with $u_1\theta = \bar{u}$ must occur at the top of e_1. On the other hand, $u\theta = \bar{u}$, so θ is an ACU-unifier of u and u_1. Then every minimal complete set U of ACU-unifiers of u and u_1 contains a τ such that $C_0\theta$ is a ground instance of $C_0\tau$.

It remains to prove that $C_0\tau$ has no unshielded variables for any $\tau \in U$. For every variable $y \in \mathrm{Var}(C_0\tau)$ there is a variable $x \in \mathrm{Var}(C_0)$ such that $y \in \mathrm{Var}(x\tau)$. Assume that y is unshielded in $C_0\tau$. Then x must be unshielded in C_0. By Cor. 6.15, x occurs in the pivotal term u. As $u\tau = u_1\tau$, $x\tau$ is a subterm of $u_1\tau$. The atomic term u_1 occurs in

$D_0 = \mathrm{elim}(C_0)$ and, by Cor. 6.9, also in C_0. Consequently, the atomic term $u_1\tau$ occurs in $C_0\tau$. Hence every variable in $\mathrm{Var}(x\tau)$ is shielded in $C_0\tau$, contradicting our assumption. $\square$

A similar lemma can be proved for *abstraction* inferences.

LEMMA 6.18. *Let ι be an abstraction inference with maximal premise C and conclusion $C_0 = y \not\approx w \vee C_0'$; let $D_0 = \mathrm{elim}(C_0)$. Let $\iota\theta$ be a ground instance of ι, and let $[\neg]\, e\theta$ be the pivotal literal of $\iota\theta$. Let ρ be a substitution that maps y to a ground term smaller than $w\theta$. If $C\theta \preceq_{\mathrm{C}} D_0\theta\rho$, then the multiset difference $D_0\theta\rho \setminus C_0\theta\rho$ contains a literal $[\neg]\, e_1\theta\rho$, such that $[\neg]\, e_1\theta\rho$ has the same polarity as $[\neg]\, e\theta$ and the pivotal term of $\iota\theta$ occurs in $[\neg]\, e_1\theta\rho$.*

LEMMA 6.19. *Let ι be an abstraction inference from clauses without unshielded variables with maximal premise C, conclusion $C_0 = y \not\approx w \vee C_0'$, pivotal literal $[\neg]\, e$, and pivotal term u; let $D_0 = \mathrm{elim}(C_0)$. Let $\iota\theta$ be a ground instance of ι. Let ρ be a substitution that maps y to a ground term smaller than $w\theta$. If $C\theta \preceq_{\mathrm{C}} D_0\theta\rho$, then the multiset difference $D_0 \setminus C_0$ contains a literal $[\neg]\, e_1$, such that:*

(i) $[\neg]\, e_1$ *has the same polarity as* $[\neg]\, e$,
(ii) *there is an atomic term u_1 occurring at the top of e_1,*
(iii) *for every minimal complete set U of* ACU*-unifiers of u and u_1, there is a $\tau \in U$ such that $C_0\theta$ is a ground instance of $C_0\tau$.*

Furthermore, for every $\tau \in U$, $C_0\tau$ has no unshielded variables.

6.3. Integration of the Elimination Algorithm

Using the results above, we can now transform the inference system $\mathit{CInf}_{\mathbf{N}^{>0}}$ into a new inference system that operates on clauses without unshielded variables and produces again such clauses. The new system *DInf* is given by two meta-inference rules:

Eliminating Inference

$$\frac{C_n \quad \ldots \quad C_1}{\mathrm{elim}(C_0)}$$

if the following condition is satisfied:

(i) $\dfrac{C_n \quad \ldots \quad C_1}{C_0}$ is a $\mathit{CInf}_{\mathbf{N}^{>0}}$-inference.

Instantiating Inference

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

if the following conditions are satisfied:

(i) $\dfrac{C_n \quad \ldots \quad C_1}{C_0}$ is a $\mathit{CInf}_{\mathbf{N}^{>0}}$-inference with pivotal literal $[\neg]\, e$ and pivotal term u.

(ii) The multiset difference $\mathrm{elim}(C_0) \setminus C_0$ contains a literal $[\neg]\, e_1$ with the same polarity as $[\neg]\, e$.

(iii) An atomic term u_1 occurs at the top of e_1.
(iv) τ is contained in a minimal complete set of ACU-unifiers of u and u_1.

Recall that the inference system $\mathit{CInf}_{\mathbf{N}^{>0}}$ is finitely branching for premises without unshielded variables.

Like the inference systems introduced previously, the inference system *DInf* has to be complemented by a redundancy criterion. For the redundancy of clauses we can use the same definition that we have used so far. The redundancy of *DInf*-inferences, however, has to be defined in a different way. As most *DInf*-inferences do not have ground instances at all, the traditional method to define the redundancy of inferences in terms of their ground instances does not work here. Rather, our definition of redundancy of *DInf*-inferences is based on the redundancy of $\mathit{CInf}_{\mathbf{N}^{>0}}$-inferences.

We call the $\mathit{CInf}_{\mathbf{N}^{>0}}$-inference that is used in the definition of a *DInf*-inference ι a parent inference of ι. Note that parent inferences are not defined uniquely; for instance, both the *cancellation* $\mathit{CInf}_{\mathbf{N}^{>0}}$-inferences

$$\frac{b \not\approx 0 \vee x + \underline{f(x)} + g(y) \approx y + \underline{f(x)} + g(y)}{b \not\approx 0 \vee x + g(y) \approx y + g(y)}$$

and

$$\frac{b \not\approx 0 \vee x + f(x) + \underline{g(y)} \approx y + f(x) + \underline{g(y)}}{b \not\approx 0 \vee x + f(x) \approx y + f(x)}$$

are parent inferences of the eliminating *DInf*-inference

$$\frac{b \not\approx 0 \vee x + f(x) + g(y) \approx y + f(x) + g(y)}{b \not\approx 0}$$

since elim maps their conclusions to the same clause $b \not\approx 0$.

DEFINITION 6.20. *An eliminating DInf-inference*

$$\frac{C_n \quad \ldots \quad C_1}{D_0}$$

is called DRed-redundant with respect to N, if for all its parent inferences

$$\iota \;=\; \frac{C_n \quad \ldots \quad C_1}{C_0}$$

with $D_0 = \mathrm{elim}(C_0)$, all ground instances $\iota\theta$ with $D_0\theta \prec_{\mathrm{C}} C_1\theta$ are $\mathit{CRed}_{\mathbf{N}^{>0}}$-redundant with respect to N.[29]

An instantiating DInf-inference

$$\frac{C_n \quad \ldots \quad C_1}{D_0}$$

is called DRed-redundant with respect to N, if for all its parent inferences

$$\iota \;=\; \frac{C_n \quad \ldots \quad C_1}{C_0}$$

[29] If ι is an *abstraction* inference with $C_0 = C' \vee \neg\, y \approx w \vee [\neg]\, s[y] \approx s'$, the condition is replaced by "if for every substitution ρ that maps y to a ground term $r \prec w\theta$ and all ground instances $\iota\theta$ with $\mathrm{elim}(C_0)\theta\rho \prec_{\mathrm{C}} C_1\theta$, $\overline{N}^{\prec_{\mathrm{C}} C_1\theta} \models_{\mathbf{N}^{>0}} C_0\theta\rho$."

with $D_0 = C_0\tau$ *all ground instances* $\iota\tau\theta$ *with* $\mathrm{elim}(C_0)\tau\theta \succeq_{\mathrm{C}} C_1\tau\theta$ *are* $\mathit{CRed}_{\mathbf{N}^{>0}}$*-redundant with respect to* N.[30]

The set of all inferences that are DRed-redundant with respect to a set N *of clauses is denoted by* $\mathit{DRed}^{\mathrm{I}}(N)$.

LEMMA 6.21. *The pair* $\mathit{DRed} = (\mathit{DRed}^{\mathrm{I}}, \mathit{CRed}^{\mathrm{C}}_{\mathbf{N}^{>0}})$ *is a redundancy criterion with respect to the inference system DInf and the consequence relation* $\models_{\mathrm{DTAG}}$.

PROOF. We have to show that *DRed* satisfies conditions (i)–(iv) of Def. 2.2. Conditions (i)–(iii) follow directly from the fact that $\mathit{CRed}_{\mathbf{N}^{>0}}$ is a redundancy criterion with respect to $\mathit{CInf}_{\mathbf{N}^{>0}}$ and $\models_{\mathbf{N}^{>0}}$. Condition (iv) is a consequence of Lemmas 6.17 and 6.19. □

LEMMA 6.22. *Let* N *be a set of clauses without unshielded variables. If every inference in* $\mathit{DInf}(N)$ *is DRed-redundant with respect to* N*, then every inference in* $\mathit{CInf}_{\mathbf{N}^{>0}}(N)$ *is* $\mathit{CRed}_{\mathbf{N}^{>0}}$*-redundant with respect to* N.

PROOF. By the definition of *DInf*, Lemmas 6.17 and 6.19, and Cor. 6.10. □

EXAMPLE 6.23. Let us consider once more the *cancellation* inference ι from Ex. 6.11:

$$\frac{3x \not\approx c \vee x + f(z) \approx 0 \vee f(x) + b \approx f(y)}{3x \not\approx c \vee x + f(z) \approx 0 \vee b \approx 0}$$

We denote the premise of ι by C and the conclusion by C_0. The conclusion contains one unshielded variable, namely x, which occurs in the pivotal term $f(x)$. Eliminating x from C_0, we obtain

$$\mathrm{elim}(C_0) \;=\; c + 3f(z) \approx 0 \vee b \approx 0\,.$$

The multiset difference $\mathrm{elim}(C_0) \setminus C_0$ equals $\{c + 3f(z) \approx 0\}$; the pivotal term $f(x)$ and $f(z)$ are ACU-unifiable. The singleton set containing the substitution $\tau = \{x \mapsto z\}$ is a minimal complete set of ACU-unifiers. Applying τ to C_0 we obtain the clause

$$C_0\tau \;=\; 3z \not\approx c \vee z + f(z) \approx 0 \vee b \approx 0\,.$$

The clause $\mathrm{elim}(C_0)$ makes all ground instances $\iota\theta$ $\mathit{CRed}_{\mathbf{N}^{>0}}$-redundant that satisfy $C\theta \succ_{\mathrm{C}} \mathrm{elim}(C_0)\theta$, that is, in particular, every $\iota\theta$ with $x\theta \succ z\theta$. The only remaining ground instances are those where $x\theta = z\theta$; these are made $\mathit{CRed}_{\mathbf{N}^{>0}}$-redundant by $C_0\tau$.

THEOREM 6.24. *If a set of clauses is saturated with respect to DInf and none of the clauses contains unshielded variables, then it is also saturated with respect to* $\mathit{CInf}_{\mathbf{N}^{>0}}$.

While deletion of redundant clauses works as usual in the *DInf* calculus, we have to take some care if we want to simplify clauses: Like the inference rules of *DInf*, the simplification techniques that we employ must not introduce unshielded variables. As

[30] If ι is an *abstraction* inference with $C_0 = C' \vee \neg\, y \approx w \vee [\neg]\, s[y] \approx s'$, the condition is replaced by "if for every substitution ρ that maps y to a ground term $r \prec w\theta$ and all ground instances $\iota\tau\theta$ with $\mathrm{elim}(C_0)\tau\theta\rho \prec_{\mathrm{C}} C_1\tau\theta$, $\overline{N}^{\prec_{\mathrm{C}} C_1\tau\theta} \models_{\mathbf{N}^{>0}} C_0\tau\theta\rho$."

demonstrated by the following example, this requirement can make even seemingly innocuous techniques from Sect. 5.1 such as syntactic cancellation inapplicable.[31]

EXAMPLE 6.25. Consider the clause

$$\begin{aligned} C_1 \;=\; & 3x \not\approx a \vee 5z \not\approx c \vee x + z + f(x', y') \approx 0 \vee b + g(y, z) \approx g(y', z) \\ & \qquad\qquad\qquad\qquad\qquad\qquad \vee\; d + f(x, y) \approx f(x'', y). \end{aligned}$$

Cancellation of $f(x, y)$ and $f(x'', y)$ in the last literal yields

$$C_0 \;=\; 3x \not\approx a \vee 5z \not\approx c \vee x + z + f(x', y') \approx 0 \vee b + g(y, z) \approx g(y', z) \vee d \approx 0.$$

This $\mathit{CInf}_{\mathbf{N}^{>0}}$-inference ι is the parent inference of an eliminating inference; elimination of the unshielded variable x yields

$$\mathrm{elim}(C_0) \;=\; 5z \not\approx c \vee a + 3z + 3f(x', y') \approx 0 \vee b + g(y, z) \approx g(y', z) \vee d \approx 0.$$

As $\mathrm{elim}(C_0) \setminus C_0$ contains the literal $a + 3z + 3f(x', y') \approx 0$ and $f(x, y)$ is unifiable with $f(x', y')$, ι is also the parent inference of an instantiating inference with the conclusion

$$C_0\theta \;=\; 3x' \not\approx a \vee 5z \not\approx c \vee x' + z + f(x', y') \approx 0 \vee b + g(y', z) \approx g(y', z) \vee d \approx 0,$$

with $\theta = \{x \mapsto x',\, y \mapsto y'\}$. Now $g(y', z)$ occurs on both sides of the fourth literal of $C_0\theta$, and it seems to be natural to simplify $C_0\theta$ by syntactic cancellation of $g(y', z)$. However, the resulting clause

$$D \;=\; 3x' \not\approx a \vee 5z \not\approx c \vee x' + z + f(x', y') \approx 0 \vee b \approx 0 \vee d \approx 0$$

contains again an unshielded variable, namely z. One might be tempted to fix this by applying the usual variable elimination algorithm once more to D:

$$\mathrm{elim}(D) \;=\; 3x' \not\approx a \vee 5x' + c + 5f(x', y') \approx 0 \vee b \approx 0 \vee d \approx 0.$$

However, if we consider ground instances $C_1\theta$ with $f(x, y)\theta = f(x'', y)\theta = f(x', y')\theta$, then $\mathrm{elim}(D)\theta$ is larger than $C_1\theta$ and even larger than $\mathrm{elim}(C_0)\theta$. And worse, the trick used above, namely to unify the cancelled term $f(x, y)$ with the remaining term $f(x', y')$, can not be iterated: there is no non-variable term in D that can be unified with $g(y', z)$. The only way out is to keep $C_0\theta$ as it is, without cancelling $g(y', z)$.

We have mentioned in the beginning of this section that we consider only sets of clauses comprising the set of additional theory axioms DivInvNt. The axioms $\mathrm{ACUKT}_{\mathbf{N}^{>0}}$ are integrated into the inference rules *and the redundancy criterion* of the cancellative superposition calculus and no inferences with these axioms are required. What about inferences with DivInvNt? If there are no clauses with unshielded variables, then a non-*abstraction* inference with, say, the inverse axiom is only possible if the maximal atomic term $-x$ of the inverse axiom overlaps with a maximal atomic term in another clause, that is, if the negation function occurs in another clause. Similarly, for a non-*abstraction* inference with one of the divisibility axioms $k\ \mathit{divided\text{-}by}_k(x) \approx x$ it is necessary that some other clause contains the function symbol $\mathit{divided\text{-}by}_k$. The only inferences that are possible if the negation function or the symbol $\mathit{divided\text{-}by}_k$ does not occur otherwise are *abstraction* inferences where the theory axiom is the first premise. Note that in this case

[31] Note that avoiding syntactic cancellation is impossible in the calculus of Godoy and Nieuwenhuis (2000), which works on clauses in normal form with respect to abelian groups, thus making syntactic cancellation (even of non-maximal terms) mandatory.

the conclusion does not depend on the first premise. So, although there are infinitely many divisibility axioms, it suffices to compute *one* such inference. In fact, as we will show in the sequel, by performing abstraction eagerly, *abstraction* inferences and inferences with DivInvNt during the saturation process can be avoided completely.

6.4. Full Abstraction

Let us define a new inference system $D^{abs}Inf$ that contains exactly the inference rules of *DInf* except the *abstraction* rule. As *abstraction* inferences from fully abstracted clauses are impossible, the following theorem is an obvious consequence of Thm. 6.24.

THEOREM 6.26. *If a set of fully abstracted clauses is saturated with respect to $D^{abs}Inf$ and none of the clauses contains unshielded variables, then it is also saturated with respect to $CInf_{\mathbf{N}^{>0}}$.*

The following two lemmas show that, for effective saturation of a set of clauses with respect to $D^{abs}Inf$, it is sufficient to perform full abstraction once in the beginning.

LEMMA 6.27. *Let C be a fully abstracted clause. Then* $\mathrm{elim}(C)$ *is fully abstracted.*

LEMMA 6.28. *Let ι be a $D^{abs}Inf$ (or more generally, DInf) inference from fully abstracted clauses without unshielded variables. Then the conclusion of ι is a fully abstracted clause without unshielded variables.*

If we replace every clause C in the input of the inference system by the logically equivalent clause $\mathrm{elim}(\mathrm{abs}(C))$ before we start the saturation process, then all clauses produced by $D^{abs}Inf$-inferences are again fully abstracted and contain no unshielded variables. For such inputs, $D^{abs}Inf$-superposition is therefore refutationally complete.

In fact $D^{abs}Inf$-superposition can be used as a decision procedure. Define the class $\mathcal{D}_{\mathrm{c}}^{\mathrm{elim}}$ as follows: A clause C is contained in $\mathcal{D}_{\mathrm{c}}^{\mathrm{elim}}$ if it has no unshielded variables and if there is a finite sequence of distinct variables $x_1, \ldots, x_n$ such that, for every literal $[\neg]\ s \approx s'$ in C, both s and s' are sums $\sum n_k t_k$, and each t_k is either a variable x_i of sort S_{CAM} or an atomic term $f(x_1, \ldots, x_i)$ for some $i \leq n$. Then there is a strategy for $D^{abs}Inf$-superposition that is guaranteed to terminate on every finite subset of $\mathcal{D}_{\mathrm{c}}^{\mathrm{elim}}$. As every closed formula with arbitrary quantifiers over the vocabulary $+$, 0, $-$, *divided-by*$_k$, $\approx$, and variables of sort S_{CAM} can be translated into such a finite subset, $D^{abs}Inf$-superposition decides the theory of divisible torsion-free abelian groups (Waldmann 1999).

We have described the general advantages and disadvantages of full abstraction in Sect. 5.5. In the case of $D^{abs}Inf$-superposition, there is one more advantage: If all clauses are fully abstracted and do not contain unshielded variables, then the terms that must be unified during the saturation do not contain the operator $+$. For such terms, ACU-unification and syntactic unification are equivalent. Thus we may reformulate $D^{abs}Inf$ in terms of syntactic unification. In an implementation of the calculus, this means that efficient indexing techniques for non-AC calculi become available again. Besides, if all clauses are fully abstracted, then the negation function or the symbols *divided-by*$_k$ can occur only at the top of a clause. In this case, it is easy to eliminate them initially from all non-theory clauses, so that there is no need for further inferences with the clauses DivInvNt during the saturation.

7. Conclusions

We have presented calculi for first-order equational theorem proving in the presence of the axioms of cancellative abelian monoids and, optionally, the torsion-freeness axioms. These calculi are refutationally complete without requiring extended clauses or explicit inferences with the theory clauses. Compared to the conventional superposition calculus, on which they are based, the ordering restrictions are strengthened in such a way that we may not only restrict to inferences that involve maximal sides of maximal literals, but even to inferences that involve maximal summands occurring in maximal sides.

In traditional AC-superposition, extended rules show a rather prolific behaviour, since they produce an inference between two clauses whenever two summands in the maximal sides of the respective maximal literals are unifiable. This is already bad enough if all summands are ground, and it has truly fatal consequences for the search space, if one of the summands is a variable. In our approach, cancellative superposition makes extended rules superfluous, and the ordering restrictions mentioned above allow to exclude overlaps with shielded variables altogether. The degree to which removal of unshielded variables is possible depends on additional algebraic structure: Certain elimination techniques for unshielded variables are applicable in the presence of the non-group axiom, the inverse axiom, the non-triviality and torsion-freeness axioms, or the divisibility axioms.

In divisible torsion-free abelian groups, unshielded variables can be eliminated completely. We have presented two calculi that integrate this variable elimination algorithm into the cancellative superposition calculus, rendering all variable overlaps superfluous. The two calculi differ in the way they handle abstraction. The first one contains an explicit *abstraction* inference rule with the usual ordering restrictions. By contrast, for the second one, it is required that all input clauses are fully abstracted in advance. Full abstraction is detrimental to the search space, as it increases the number of inferences in which a clause can participate. On the other hand, it allows us to dispense with ACU-unification and ACU-orderings. Both these operations are costly, and being able to avoid them greatly simplifies the integration of the calculus into existing theorem provers, whose performance depends crucially on efficient indexing data structures. It remains to be investigated whether full abstraction is generally advantageous in practice.

The integration of the torsion-freeness axioms makes the cancellative superposition calculus also applicable to theorem proving in totally ordered abelian monoids, as long as the ordering predicate occurs only in the ordering axioms but nowhere else in the theorem. To relax this restriction, cancellative superposition must be combined with an ordered chaining calculus for total orderings (Bachmair and Ganzinger 1994b). A superposition and chaining calculus for totally ordered divisible abelian groups has been described in (Waldmann 2001).

At the time of writing this paper, we cannot yet report about practical experiences with our calculus. We are working on an implementation of the inference system $D^{abs}\mathit{Inf}$ within the SPASS system (Weidenbach, Gaede, and Rock 1996).

A. Cancellative Monoids

We summarize some basic facts about cancellative semigroups and monoids. Most of this material is part of the algebraic folklore; it can be found for instance in the books of Lang (1984) and Gilmer (1984). Theorems A.3 and A.5 are new (to the best of my knowledge).

A semigroup is an algebraic structure consisting of a non-empty set G and a binary function $+ : G \times G \to G$ that is associative: $x + (y + z) = (x + y) + z$ for all $x, y, z \in G$. The semigroup is said to be abelian (or commutative), if $x + y = y + x$ for all $x, y \in G$.

An element 0 of a semigroup $(G, +)$ is called a left identity element if $0 + x = x$ for all $x \in G$ (analogously for right identity). It is called an identity element if it is both a left and right identity element. A monoid $(G, +, 0)$ is a semigroup $(G, +)$ with an identity element 0.

Let $(G, +)$ be a semigroup. If $x \in G$ and $m \in \mathbf{N}^{>0}$, then mx is an abbreviation for the m-fold sum $x + \cdots + x$. In a monoid, $0x$ is defined as 0.

A semigroup $(G, +)$ or monoid $(G, +, 0)$ is called left-cancellative, if for all $x, y, z \in G$, $x + y = x + z$ implies $y = z$. (analogously for right-cancellative). It is called cancellative, if it is left- and right-cancellative.

A monoid $(G, +, 0)$ is called a group if for every $x \in G$ there is a left and right inverse $-x \in G$ such that $(-x) + x = x + (-x) = 0$. In fact, it is already sufficient to require a right identity element and a right inverse:

Lemma A.1. *Let G be a semigroup. Then G is a group if and only if G has a right identity element 0_{R} and for every $x \in G$ a right inverse $(-x)$ with respect to 0_{R}, that is, $x + 0_{\mathrm{R}} = x$ and $x + (-x) = 0_{\mathrm{R}}$.*

Cancellative abelian semigroups and abelian groups are closely related: If $(G, +, 0)$ is an abelian group and $G' \subseteq G$ is closed under addition, then $(G', +)$ is a cancellative abelian semigroup; if furthermore $0 \in G'$, then $(G', +, 0)$ is a cancellative abelian monoid. Conversely, every cancellative abelian semigroup $(G', +)$ or monoid $(G', +, 0)$ is isomorphic to a subset of some abelian group G. This so-called Grothendieck group can be constructed as follows: Consider the set $G' \times G'$ with componentwise addition. We define an equivalence relation $\cong$ on $G' \times G'$ by $(x, y) \cong (x', y') \iff x + y' = x' + y$. Then G is the quotient of $G' \times G'$ under $\cong$; the identity element is $[(x, x)]_{\cong}$ for an arbitrary $x \in G'$, and the inverse of $[(x, y)]_{\cong}$ is $[(y, x)]_{\cong}$. The embedding homomorphism of G' into G given by $x \mapsto [(2x, x)]_{\cong}$ has the universal property with respect to homomorphisms of G' into abelian groups. In particular, G and G' are isomorphic, if G' itself is already a group.

Lemma A.2. *Let $(G, +)$ be a left-cancellative semigroup. Let H be a finite subset of G, such that for every x in G there is a y in G with $x + y \in H$. Then G has a left identity element 0_{L} and right inverses with respect to 0_{L}, that is, for all x in G, $0_{\mathrm{L}} + x = x$ and $x + (-x) = 0_{\mathrm{L}}$.*

Proof. Let $\mathbf{H}$ be the set of all finite subsets of G with the same property as H above. As $H \in \mathbf{H}$, $\mathbf{H}$ is non-empty. Let H' be an element of $\mathbf{H}$ with minimal cardinality.

For all $x \in H'$ and $y \in G$, $x + y \in H'$ implies $x + y = x$. We show this by contradiction: Assume there are $b \in H'$ and $c \in G$ such that $b + c \in H'$ and $b + c \neq b$. For every x in G such that there is a y in G with $x + y = b$, there is a y' (namely $y + c$) in G with $x + y' = b + c$. Hence $H' \setminus \{b\}$ is an element of $\mathbf{H}$, contradicting the minimality of H'.

We can furthermore show that H' is a singleton $\{b\}$. Assume that $b, c \in H'$. As $H' \in \mathbf{H}$, there is a $d \in G$ such that $b + b + d \in H'$. By the previous paragraph, $b + b + d = b$. Hence $b + b + d + c = b + c$, and by left-cancellation $b + d + c = c$. Therefore, $c = b$.

As $H' \in \mathbf{H}$, there is a $c \in G$ such that $b + b + c = b$. Let x be an arbitrary element of G. Then $b + b + c + x = b + x$, and by left-cancellation $b + c + x = x$; Define $0_{\mathrm{L}} = b + c$.

As $0_{\mathrm{L}} + x = x$, 0_{L} is a left identity element.

It remains to show that G has right inverses. Let x be an arbitrary element of G. There is a $y \in G$ such that $x + y = b$, hence $x + y + c = b + c = 0_{\mathrm{L}}$. Define $(-x) = y + c$, then $x + (-x) = 0_{\mathrm{L}}$ as required. □

The conditions of the previous lemma are not sufficient to deduce that G is a group. As a counterexample consider the so-called right-zero semigroup with two elements, that is the semigroup $(\{0,1\}, \oplus)$ where $\oplus$ is defined by $x \oplus y = y$ and $H = \{0,1\}$. We have to require one more property: either that G has a right identity, or that G is right-cancellative, or that G is abelian. (The first two properties are actually equivalent.)

THEOREM A.3. *Let $(G, +)$ be a left-cancellative semigroup. Let H be a finite subset of G, such that for every x in G there is a y in G with $x + y \in H$. If additionally (i) G has a right identity, or (ii) G is right-cancellative, or (iii) G is abelian, then G is a group.*[32]

PROOF. (i) According to the preceding lemma G has a left identity element 0_{L} and right inverses with respect to 0_{L}. If G has also a right identity 0_{R}, then $0_{\mathrm{L}} = 0_{\mathrm{L}} + 0_{\mathrm{R}} = 0_{\mathrm{R}}$. Therefore, every right inverse with respect to 0_{L} is also a right inverse with respect to 0_{R}. By Lemma A.1, G is a group.

(ii) As 0_{L} is a left identity, $0_{\mathrm{L}} = 0_{\mathrm{L}} + 0_{\mathrm{L}}$. Hence, for every $d \in G$, $d + 0_{\mathrm{L}} = d + 0_{\mathrm{L}} + 0_{\mathrm{L}}$. If G is right cancellative, $d = d + 0_{\mathrm{L}}$, so 0_{L} is a right identity. By (i), G is a group.

(iii) If G is abelian, the left identity 0_{L} is also a right identity. By (i), G is a group. □

Let $(G, +, 0)$ be an abelian group. An element $x \in G$ is called a torsion element, if $kx = 0$ for some positive integer k. If G' is the set of all torsion elements of G, then $(G', 0, +)$ is a subgroup of G, the so-called torsion subgroup of G. We say that G is torsion-free, if $G' = \{0\}$, that is, if for all $k \in \mathbf{N}^{>0}$ and $x \in G$, $kx = 0$ implies $x = 0$.

A cancellative abelian semigroup $(G, +)$ is called torsion-free, if its Grothendieck group is torsion-free.

LEMMA A.4. *A cancellative abelian semigroup $(G, +)$ is torsion-free, if and only if for all $k \in \mathbf{N}^{>0}$ and $x, y \in G$, $kx = ky$ implies $x = y$.*

If should be noted that the conditions (i) "$kx = ky$ implies $x = y$ for all $k \in \mathbf{N}^{>0}$", and (ii) "$kx = 0$ implies $x = 0$ for all $k \in \mathbf{N}^{>0}$" are not equivalent for cancellative abelian semigroups, or even for cancellative abelian monoids. As an example consider the monoid $(\mathbf{N} \times (\mathbf{Z}/2\mathbf{Z})) \setminus \{(0,1)\}$ with addition defined componentwise. Here $kx = (0,0)$ implies $x = (0,0)$ for every $k \in \mathbf{N}^{>0}$. On the other hand, $2 \cdot (1,0) = (2,0) = 2 \cdot (1,1)$. The Grothendieck group of this monoid is (isomorphic to) $\mathbf{Z} \times (\mathbf{Z}/2\mathbf{Z})$.

THEOREM A.5. *Let $(G, +)$ be a cancellative abelian semigroup. Let H be a finite subset of G, such that for every x in G there is a $k \in \mathbf{N}^{>0}$ with $kx \in H$. Then G is an abelian group and every $x \in G$ is a torsion element.*

[32] Part (iii) of this theorem appeared in a draft of (Ganzinger and Waldmann 1996b) submitted for CADE-13 and in (Ganzinger and Waldmann 1996a), with a different proof. An anonymous CADE-13 referee pointed out to me that the theorem holds also in the non-abelian case (without giving me the proof). Independently, the extension to non-abelian cancellative semigroups, that is, part (ii) of this theorem was found by Pröhle (1998).

PROOF. For every $x \in G$, there is a $k \in \mathbf{N}^{>0}$ such that $x + (k-1)x \in H$. Hence G is an abelian group by Thm. A.3.

To show that every $x \in G$ is a torsion element, let $\mathbf{H}$ be the set of all finite subsets of G with the same property as H above. As $H \in \mathbf{H}$, $\mathbf{H}$ is non-empty. Let H' be an element of $\mathbf{H}$ with minimal cardinality. Obviously, $0 \in H'$.

We will first prove that for all $y \in H'$ and $k \in \mathbf{N}^{>0}$, $ky \in H'$ implies $ky = y$. Assume that $b \in H'$, $k \in \mathbf{N}^{>0}$, such that $kb \in H'$ and $kb \neq b$. For every x in G such that there is an n in $\mathbf{N}^{>0}$ with $nx = b$, we have $knx = kb$. So $H' \setminus \{b\}$ is an element of $\mathbf{H}$, contradicting the minimality of H'.

Let $y \in H'$ be arbitrarily chosen. By assumption, there is an $n \in \mathbf{N}^{>0}$ such that $n \cdot 2y \in H'$. By the previous paragraph, $2ny = y$, thus $(2n-1)y = 0$. As $2n - 1 \in \mathbf{N}^{>0}$ and $0 \in H'$, we have $y = 0$, and therefore $H' = \{0\}$. Since for each $x \in G$ there is a $k \in \mathbf{N}^{>0}$ with $kx \in H'$, each $x \in G$ is a torsion element. $\square$

COROLLARY A.6. *Let $(G, +)$ be a torsion-free cancellative abelian semigroup. Let H be a finite subset of G, such that for every x in G there is a $k \in \mathbf{N}^{>0}$ with $kx \in H$. Then G is a singleton.*

A totally ordered abelian semigroup $(G, +, <)$ is an abelian semigroup $(G, +)$ together with a binary relation $<$ that is (i) transitive: $x < y$ and $y < z$ implies $x < z$ for all $x, y, z \in G$, (ii) irreflexive: there is no $x \in G$ with $x < x$, (iii) total: for all $x, y \in G$, either $x < y$ or $y < x$ or $x = y$, and (iv) compatible: for all $x, y, z \in G$, if $x < y$, then $x + z < y + z$. Totally ordered abelian monoids and groups are defined analogously.

The following theorem due to Levi (1913) gives the connection between totally ordered and torsion-free cancellative abelian semigroups. A proof can be found in (Gilmer 1984).

THEOREM A.7. *For an abelian semigroup $(G, +)$ (or monoid, or group), the following two properties are equivalent:*

(i) $(G, +)$ is cancellative and torsion-free.
(ii) There exists a binary relation $<$ over G such that $(G, +, <)$ is a totally ordered abelian semigroup.

To be able to present the general and the torsion-free case in a uniform way, we generalize torsion-freeness to Ψ-torsion-freeness.

Let $\Psi \subseteq \mathbf{N}^{>0}$. We say that a cancellative abelian semigroup is Ψ-torsion-free, if for all $\psi \in \Psi$ and $x, y \in G$, $\psi x = \psi y$ implies $x = y$. It is obvious that Ψ-torsion-freeness is equivalent to torsion-freeness for $\Psi = \mathbf{N}^{>0}$, and that it is void for $\Psi = \{1\}$.

LEMMA A.8. *Let $(G, +)$ be an abelian semigroup, let Ψ be the set of* all *$\psi \in \mathbf{N}^{>0}$ such that $\psi x = \psi y$ implies $x = y$ for all $x, y \in G$. Then Ψ contains 1 and is closed under multiplication and factors, that is, for all $\psi, \psi' \in \mathbf{N}^{>0}$*

$$\psi \in \Psi \wedge \psi' \in \Psi \quad \textit{if and only if} \quad \psi\psi' \in \Psi\,.$$

PROOF. If $\psi \in \Psi$ and $\psi' \in \Psi$, then $\psi\psi' x = \psi\psi' y$ implies $\psi' x = \psi' y$ and $x = y$ for all $x, y \in G$; hence $\psi\psi' \in \Psi$. Conversely, if $\psi' \notin \Psi$, then there exist $x, y \in G$ such that $x \neq y$ and $\psi' x = \psi' y$, hence $\psi\psi' x = \psi\psi' y$ and $\psi\psi' \notin \Psi$. $\square$

References

Bachmair, L., Ganzinger, H. (1994a). Associative-commutative superposition. In Dershowitz, N., Lindenstrauss, N., editors, *Conditional and Typed Rewriting Systems, 4th International Workshop, CTRS-94*, LNCS 968, pages 1–14, Jerusalem, Israel. Springer-Verlag.

Bachmair, L., Ganzinger, H. (1994b). Ordered chaining for total orderings. In Bundy, A., editor, *Twelfth International Conference on Automated Deduction*, LNAI 814, pages 435–450, Nancy, France. Springer-Verlag.

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

Bachmair, L., Ganzinger, H., Waldmann, U. (1993). Superposition with simplification as a decision procedure for the monadic class with equality. In Gottlob, G., Leitsch, A., Mundici, D., editors, *Computational Logic and Proof Theory, Third Kurt Gödel Colloquium*, LNCS 713, pages 83–96, Brno, Czech Republic. Springer-Verlag.

Bachmair, L., Ganzinger, H., Waldmann, U. (1994). Refutational theorem proving for hierarchic first-order theories. *Applicable Algebra in Engineering, Communication and Computing*, **5**(3/4):193–212.

Ben Cherifa, A., Lescanne, P. (1987). Termination of rewriting systems by polynomial interpretations and its implementation. *Science of Computer Programming*, **9**(2):137–159.

Boudet, A., Contejean, E., Marché, C. (1996). AC-complete unification and its application to theorem proving. In Ganzinger, H., editor, *Rewriting Techniques and Applications, 7th International Conference, RTA-96*, LNCS 1103, pages 18–32, New Brunswick, NJ, USA. Springer-Verlag.

Boyer, R. S., Moore, J. S. (1988). Integrating decision procedures into heuristic theorem provers: A case study of linear arithmetic. In Hayes, J. E., Michie, D., Richards, J., editors, *Machine Intelligence 11: Logic and the acquisition of knowledge*, chapter 5, pages 83–124. Oxford University Press.

Buchberger, B. (1996). Symbolic computation: Computer algebra and logic. In Baader, F., Schulz, K. U., editors, *Frontiers of Combining Systems, First International Workshop*, Applied Logic Series, Vol. 3, pages 193–219, Munich, Germany. Kluwer Academic Publishers.

Dershowitz, N., Jouannaud, J.-P. (1990). Rewrite systems. In van Leeuwen, J., editor, *Handbook of Theoretical Computer Science*, volume B: Formal Models and Semantics, chapter 6, pages 243–320. Elsevier Science Publishers B.V., Amsterdam, The Netherlands.

Fitting, M. (1990). *First-Order Logic and Automated Theorem Proving.* Springer-Verlag, New York, NY, USA.

Ganzinger, H., Waldmann, U. (1996a). Theorem proving in cancellative abelian monoids. Technical Report MPI-I-96-2-001, Max-Planck-Institut für Informatik, Saarbrücken, Germany.

Ganzinger, H., Waldmann, U. (1996b). Theorem proving in cancellative abelian monoids (extended abstract). In McRobbie, M. A., Slaney, J. K., editors, *Automated Deduction – CADE-13, 13th International Conference on Automated Deduction*, LNAI 1104, pages 388–402, New Brunswick, NJ, USA. Springer-Verlag.

Gilmer, R. (1984). *Commutative Semigroup Rings.* Chicago Lectures in Mathematics. The University of Chicago Press, Chicago, IL, USA.

Godoy, G., Nieuwenhuis, R. (2000). Paramodulation with built-in abelian groups. In *Fifteenth Annual IEEE Symposium on Logic in Computer Science*, pages 413–424, Santa Barbara, CA, USA. IEEE Computer Society, Los Alamitos, CA, USA.

Hsiang, J., Rusinowitch, M. (1991). Proving refutational completeness of theorem-proving strategies: The transfinite semantic tree method. *Journal of the ACM*, **38**(3):559–587.

Hsiang, J., Rusinowitch, M., Sakai, K. (1987). Complete inference rules for the cancellation laws (extended abstract). In McDermott, J., editor, *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, volume 2, pages 990–992, Milan, Italy. Morgan Kaufmann Publishers.

Jouannaud, J.-P., Marché, C. (1992). Termination and completion modulo associativity, commutativity and identity. *Theoretical Computer Science*, **104**(1):29–51.

Kapur, D., Narendran, P. (1992a). Complexity of unification problems with associative-commutative operators. *Journal of Automated Reasoning*, **9**(2):261–288.

Kapur, D., Narendran, P. (1992b). Double-exponential complexity of computing a complete set of AC-unifiers. In *Seventh Annual IEEE Symposium on Logic in Computer Science*, pages 11–21, Santa Cruz, CA, USA. IEEE Computer Society Press.

Lang, S. (1984). *Algebra.* Addison-Wesley, Redwood City, CA, USA, second edition.

Levi, F. (1913). Arithmetische Gesetze im Gebiete discreter Gruppen. *Rend. Circ. Mat. Palermo*, **35**:225–236. Cited in Gilmer (1984).

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

Nieuwenhuis, R., Rubio, A. (1994). AC-superposition with constraints: no AC-unifiers needed. In Bundy, A., editor, *Twelfth International Conference on Automated Deduction*, LNAI 814, pages 545–559, Nancy, France. Springer-Verlag.

Pais, J., Peterson, G. E. (1991). Using forcing to prove completeness of resolution and paramodulation. *Journal of Symbolic Computation*, **11**:3–19.

Paul, E. (1992). A general refutational completeness result for an inference procedure based on associative-commutative unification. *Journal of Symbolic Computation*, **14**(6):577–618.

Peterson, G. E. (1983). A technique for establishing completeness results in theorem proving with equality. *SIAM Journal on Computing*, **12**(1):82–100.

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

Plotkin, G. D. (1972). Building-in equational theories. In Meltzer, B., Michie, D., editors, *Machine Intelligence 7*, chapter 4, pages 73–90. American Elsevier, New York, NY, USA.

Pr̃ohle, P. (1998). Which of the cancellative semigroups are groups? *Semigroup Forum*, **57**(3):438–439.

Rusinowitch, M. (1989). *Démonstration Automatique: Techniques de Réécriture*, chapitre 7: Ensembles complets de règles d'inférence pour les axiomes de régularité, pages 111–127. InterEditions, Paris, France.

Rusinowitch, M. (1991). Theorem-proving with resolution and superposition. *Journal of Symbolic Computation*, **11**(1&2):21–49.

Rusinowitch, M., Vigneron, L. (1995). Automated deduction with associative-commutative operators. *Applicable Algebra in Engineering, Communication and Computing*, **6**(1):23–56.

Shostak, R. E. (1979). A practical decision procedure for arithmetic with function symbols. *Journal of the ACM*, **26**(2):351–360.

Slagle, J. R. (1972). Automatic theorem proving with built-in theories including equality, partial ordering, and sets. *Journal of the ACM*, **19**(1):120–135.

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

Stuber, J. (1996). Superposition theorem proving for abelian groups represented as integer modules. In Ganzinger, H., editor, *Rewriting Techniques and Applications, 7th International Conference, RTA-96*, LNCS 1103, pages 33–47, New Brunswick, NJ, USA. Springer-Verlag.

Vigneron, L. (1994). Associative-commutative deduction with constraints. In Bundy, A., editor, *Twelfth International Conference on Automated Deduction*, LNAI 814, pages 530–544, Nancy, France. Springer-Verlag.

Waldmann, U. (1997). *Cancellative Abelian Monoids in Refutational Theorem Proving*. Dissertation, Universität des Saarlandes, Saarbrücken, Germany. `http://www.mpi-sb.mpg.de/~uwe/paper/PhD.ps.gz`.

Waldmann, U. (1998a). Extending reduction orderings to ACU-compatible reduction orderings. *Information Processing Letters*, **67**(1):43–49.

Waldmann, U. (1998b). Superposition for divisible torsion-free abelian groups. In Kirchner, C., Kirchner, H., editors, *Automated Deduction – CADE-15, 15th International Conference on Automated Deduction*, LNAI 1421, pages 144–159, Lindau, Germany. Springer-Verlag.

Waldmann, U. (1999). Cancellative superposition decides the theory of divisible torsion-free abelian groups. In Ganzinger, H., McAllester, D., Voronkov, A., editors, *Logic for Programming and Automated Reasoning, 6th International Conference, LPAR'99*, LNAI 1705, pages 131–147, Tbilisi, Georgia. Springer-Verlag.

Waldmann, U. (2001). Superposition and chaining for totally ordered divisible abelian groups (extended abstract). In Goré, R., Leitsch, A., Nipkow, T., editors, *Automated Reasoning, First International Joint Conference, IJCAR 2001*, LNAI 2083, pages 226–241, Siena, Italy. Springer-Verlag.

Weidenbach, C., Gaede, B., Rock, G. (1996). SPASS & FLOTTER, version 0.42. In McRobbie, M. A., Slaney, J. K., editors, *Automated Deduction – CADE-13, 13th International Conference on Automated Deduction*, LNAI 1104, pages 141–145, New Brunswick, NJ, USA. Springer-Verlag.

Wertz, U. (1992). First-order theorem proving modulo equations. Technical Report MPI-I-92-216, Max-Planck-Institut für Informatik, Saarbrücken, Germany.

Zhang, H., Kapur, D. (1988). First-order theorem proving using conditional rewrite rules. In Lusk, E., Overbeek, R., editors, *9th International Conference on Automated Deduction*, LNCS 310, pages 1–20, Argonne, Illinois, USA. Springer-Verlag.