Cancellative Abelian Monoids in Refutational Theorem Proving

Uwe Waldmann

Dissertation
zur Erlangung des Grades
des Doktors der Ingenieurwissenschaften (Dr.-Ing.)
der Technischen Fakultät
der Universität des Saarlandes

Saarbrücken
1997

Dekan:	Prof. Dr. Alexander Koch
Gutachter:	Prof. Dr. Harald Ganzinger
	Dr. Michaël Rusinowitch
Tag des Kolloquiums:	28. Juli 1997

Meinen Eltern,
für alles

Contents

Abstract vii
Zusammenfassung vii
Extended Abstract ix
Ausführliche Zusammenfassung xiii
Acknowledgments xvii

1 Introduction **1**

2 Superposition-Based Theorem Proving **5**
2.1 Logical Foundations 5
2.2 Rewrite Systems 11
2.3 Saturation and Redundancy 13
2.4 Resolution and Superposition 16
2.5 Theory Reasoning 23

3 Cancellative Superposition **31**
3.1 Preliminaries 31
3.2 Ideas and Concepts 36
3.3 The Inference System 45

4 Refutational Completeness **51**
4.1 Ideas and Concepts 51
4.2 Redundancy 59
4.3 Lifting 64
4.4 Rewriting on Equations 67
4.5 Model Construction 69
4.6 Confluence 72
4.7 Completeness 89

5 Refinements and Applications **95**
5.1 Simplification Techniques 95
5.2 Cancellative Superposition as a Decision Procedure 102
5.3 Eliminating Unshielded Variables 107
5.4 The Standard Case: $\Psi = \{1\}$ 111
5.5 The Torsion-Free Case: $\Psi = \mathbf{N}^{>0}$ 118
5.6 Divisible Torsion-Free Abelian Groups 124
5.7 Ordered Abelian Monoids 135

6 Conclusions **137**

A Cancellative Monoids **139**

Bibliography 145
List of Symbols 157
Index 161

Abstract

We present a constraint superposition calculus in which the axioms of cancellative abelian monoids and, optionally, further axioms (e. g., torsion-freeness) are integrated. Cancellative abelian monoids comprise abelian groups, but also such ubiquitous structures as the natural numbers or multisets. Our calculus requires 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 can even be avoided completely. Thanks to the equivalence of torsion-free cancellative and totally ordered abelian monoids, our calculus allows us to solve equational problems in totally ordered abelian monoids without requiring a detour via ordering literals.

Zusammenfassung

Wir stellen einen Constraint-Superpositionskalkül vor, in den die Axiome kürzbarer abelscher Monoide und weitere optionale Axiome (z. B. Torsionsfreiheit) eingebaut sind. Kürzbare abelsche Monoide umfassen abelsche Gruppen, aber auch so allgegenwärtige Strukturen wie die natürlichen Zahlen oder Multisets. Unser Kalkül erfordert weder erweiterte Klauseln noch explizite Inferenzen mit den Theorieaxiomen. Durch verschärfte Ordnungseinschränkungen und leistungsfähige Variableneliminationstechniken erzielen wir eine deutliche Einschränkung von Überlappungen mit Variablen; in dividierbaren torsionsfreien abelschen Gruppen werden Variablenüberlappungen sogar gänzlich überflüssig. Dank der Äquivalenz torsionsfreier kürzbarer und total geordneter abelscher Monoide bietet unser Kalkül die Möglichkeit, Gleichungsprobleme in total geordneten abelschen Monoiden ohne Umweg über Ordnungsliterale zu lösen.

Extended Abstract

In practical applications of theorem proving one is usually confronted with uninterpreted function and predicate symbols that are specific for a particular domain of application, as well as with standard algebraic structures or theories, such as the natural numbers, or abelian groups, or orderings. As a naïve handling of axioms like commutativity or associativity leads to an explosion of the search space, it is crucial to the performance of a prover that it incorporates specialized techniques to work efficiently within standard algebraic theories.

Typical examples of such techniques can be found in the superposition [12] and the AC-superposition calculus [9, 100]. The former may be considered as the result of incorporating the equality axioms into the resolution calculus; the latter extends superposition further by integrating associativity and commutativity, using AC-unification and extended clauses. In both cases, inferences with the theory axioms and certain inferences involving variables can be shown to be superfluous. Together with strengthened ordering restrictions and redundancy criteria, this leads to a significant reduction of the search space.

Although the AC-superposition calculus allows to avoid inferences with the associativity and commutativity axioms, it is by no means efficient. The extended clauses are responsible for a huge increase of the number of inferences, and minimal complete set of AC-unifiers may have doubly exponential size. Using constraints, the enumeration of unifiers can be avoided [69, 97]; it is still necessary, though, to solve the unifiability problem, which is NP-complete [53].

A radical improvement can be observed when we switch over from the abelian semigroup axioms AC (or the abelian monoid axioms ACU) to the axioms of abelian groups. We get a first impression when we compare the uniform word problem, which is EXPSPACE-complete for abelian semigroups or monoids [29, 64], but polynomial for abelian groups [52]. The operational difference between these theories in a rewrite or superposition-based calculus becomes apparent when we consider the two unit clauses $u_1 + \cdots + u_k \approx s$ and $v_1 + \cdots + v_l \approx t$. There is an AC-superposition inference (via extended clauses) between these clauses whenever *some* u_i is unifiable with *some* v_j. In the presence of the inverse axiom, the number of inferences is dramatically reduced: Only if the *maximal* u_i is unifiable with the *maximal* v_j, a superposition inference is required [63, 96].

The group axioms imply both the existence and the uniqueness of the difference of any two elements. Operationally, the second property is much more important than the first one. We can thus employ similar techniques for cancellative abelian monoids, that is, for abelian monoids in which the cancellation axiom $x + y \not\approx x + z \vee y \approx z$ holds. Cancellative abelian monoids are in some sense the most general algebraic structure allowing "abelian-group-like" reasoning. They comprise not only abelian groups, but also such ubiquitous structures as the natural numbers or multisets. We present a refined constraint superposition calculus for sets of clauses including the axioms of cancellative abelian monoids, that uses the cancellative superposition rule

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

Our refutational completeness proof for this calculus is based on the model construction technique of Bachmair and Ganzinger [12], using a novel kind of rewriting relation on equations, that allows to use an equation $mu + s \approx s'$ as a rewrite rule to transform $mu + t \approx t'$ into $s' + t \approx t' + s$.

As in the abelian group case described above, inferences are restricted to overlaps of maximal summands in maximal literals, and explicit inferences with the theory axioms and extended clauses are superfluous. In particular, superpositions with shielded variables are unnecessary. While inferences with unshielded variables can not generally be avoided, the number of literals with 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 $\Psi \subseteq \mathbf{N}^{>0}$). Torsion-freeness complicates certain cancellative superposition inferences; on the other hand, it makes it possible to eliminate all unshielded variables occurring only positively. Since an abelian monoid is cancellative and torsion-free if and only if it can be totally ordered, our calculus allows us to solve equational problems in ordered abelian monoids or groups without having to derive intermediate clauses containing ordering literals.

In divisible torsion-free abelian groups (e. g., the rational numbers or rational vector spaces), 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 unshielded variables can be eliminated completely. The variable elimination algorithm is not necessarily a simplification in the superposition calculus, though: 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. Furthermore, even AC unifications can be avoided, if clauses are fully abstracted eagerly.

Ausführliche Zusammenfassung

In praktischen Anwendungen wird ein Theorembeweiser üblicherweise sowohl mit uninterpretierten Funktions- und Prädikatsymbolen konfrontiert, die spezifisch für ein bestimmtes Anwendungsgebiet sind, als auch mit algebraischen Standardstrukturen und Theorien, wie etwa den natürlichen Zahlen, abelschen Gruppen oder Ordnungen. Eine naive Behandlung von Axiomen wie dem Assoziativ- oder Kommutativgesetz führt zu einer Explosion des Suchraumes. Es ist daher entscheidend für die Leistungsfähigkeit eines Beweisers, daß er über spezielle Mittel verfügt, um effizient in algebraischen Standardtheorien zu arbeiten.

Typische Beispiele solcher Techniken findet man im Superpositions- und im AC-Superpositionskalkül [9, 12, 100]. Ersterer kann als Resolutionskalkül mit eingebauten Gleichheitsaxiomen betrachtet werden, letzterer erweitert den Superpositionskalkül um Assoziativität und Kommutativität mittels erweiterter Regeln und AC-Unifikation. In beiden Fällen werden Inferenzen mit den Theorieaxiomen und gewisse Inferenzen mit Variablen überflüssig. Zusammen mit verstärkten Ordnungseinschränkungen und Redundanzkriterien bewirkt dies eine signifikante Suchraumverkleinerung.

Obgleich der AC-Superpositionskalkül Inferenzen mit dem Assoziativitäts- und Kommutativitätsaxiom unnötig macht, ist er doch keineswegs effizient. Die erweiterten Klauseln sind für eine Vielzahl von Inferenzen verantwortlich; überdies können minimale vollständige AC-Unifikatormengen doppelt exponentielle Größe besitzen. Durch Verwendung von Constraints ist es zwar möglich, das Aufzählen von Unifikatoren zu vermeiden [69, 97]; es bleibt aber erforderlich, das Unifizierbarkeitsproblem zu lösen, welches NP-vollständig ist [53].

Die Situation verbessert sich grundlegend, wenn wir von den Axiomen abelscher Halbgruppen (AC) oder abelscher Monoide (ACU) zu den Axiomen abelscher Gruppen übergehen. Einen ersten Eindruck erhalten wir beim Vergleich der uniformen Wortprobleme: Einem EXPSPACE-vollständigen uniformen Wortproblem für abelsche Halbgruppen oder Monoide [29, 64] steht ein polynomielles für abelsche Gruppen gegenüber [52]. Der operationale Unterschied zwischen diesen Theorien in einem auf Termersetzung oder Superposition basierenden Kalkül wird am Beispiel der beiden Einheitsklauseln $u_1 + \cdots + u_k \approx s$ und $v_1 + \cdots + v_l \approx t$ deutlich. Zwischen diesen kommt es zu einer AC-Superpositionsinferenz (mittels erweiterter Klauseln), sobald *irgendein* u_i mit *irgendeinem* v_j unifizierbar ist. Anders in Gegenwart des Inversenaxioms: Nun ist eine Superposition nur noch erforderlich, wenn das *maximale* u_i mit dem *maximalen* v_j unifizierbar ist [63, 96].

Aus den Gruppenaxiomen folgt sowohl die Existenz der Differenz von je zwei Elementen, als auch ihre Eindeutigkeit. Operational ist die zweite dieser beiden Eigenschaften ungleich wichtiger als die erste. Wir können daher ähnliche Techniken auch für kürzbare abelsche Monoide anwenden, d. h., für abelsche Monoide, in denen das Kürzungsaxiom $x + y \not\approx x + z \vee y \approx z$ gilt. Kürzbare abelsche Monoide sind in einem gewissen Sinn die allgemeinste algebraische Struktur, in der man ähnlich wie in abelschen Gruppen rechnen kann. Sie umfassen nicht nur abelsche Gruppen, sondern auch so allgegenwärtige Strukturen wie die natürlichen Zahlen oder Multisets. Wir präsentieren in dieser Arbeit eine Variante des Constraint-Superpositionskalküls mit der Hauptinferenzregel

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

Zum Beweis der Widerlegungsvollständigkeit dieses Kalküls benutzen wir die Modellkonstruktionstechnik von Bachmair und Ganzinger [12]. Dabei verwenden wir eine neuartige Ersetzungsrelation auf Gleichungen, die es erlaubt, mittels einer Regel $mu + s \approx s'$ die Gleichung $mu + t \approx t'$ in $s' + t \approx t' + s$ zu überführen.

Ebenso wie im oben beschriebenen Fall abelscher Gruppen erfordert unser Kalkül keine expliziten Inferenzen mit den Theorieklauseln oder erweiterte Klauseln; Inferenzen sind auf Überlappungen maximaler Summanden in maximalen Literalen beschränkt. Insbesondere sind damit Superpositionen mit geschützten Variablen unnötig. Auf Inferenzen mit ungeschützten Variablen kann im allgemeinen nicht verzichtet werden, jedoch existieren Simplifikationstechniken, um die Anzahl von Literalen mit ungeschützten Variablen zu vermindern. Zusätzlich bietet unser Kalkül die Möglichkeit, die Torsionsfreiheitsaxiome $\psi x \not\approx \psi y \vee x \approx y$ für alle $\psi \in \mathbf{N}^{>0}$ (und auch verallgemeinerte Formen der Torsionsfreiheit mit $\psi \in \Psi \subseteq \mathbf{N}^{>0}$) zu integrieren. Durch die Torsionsfreiheit werden zwar gewisse Superpositionsinferenzen komplizierter; jedoch wird es nun möglich, ungeschützte Variablen, die nur positiv vorkommen, gänzlich zu eliminieren. Da ein abelsches Monoid genau dann kürzbar und torsionsfrei ist, wenn es total geordnet werden kann, ermöglicht es unser Kalkül, Gleichungsprobleme in geordneten abelschen Monoiden zu lösen, ohne in Zwischenschritten Klauseln mit Ordnungsliteralen ableiten zu müssen.

Dividierbare torsionsfreie abelsche Gruppen (z. B. die rationalen Zahlen oder rationale Vektorräume) erweitern die Axiome abelscher Gruppen um die Torsionsfreiheitsaxiome, die Dividierbarkeitsaxiome $\forall x\ \exists y\colon\ ky \approx x$ für alle $k \in \mathbf{N}^{>0}$ und das Nichttrivialitätsaxiom $\exists y\colon\ y \not\approx 0$. In diesen Strukturen ist es möglich, ungeschützte Variablen vollständig zu eliminieren. Der Variableneliminationsalgorithmus ist allerdings nicht notwendigerweise eine Simplifikation im Sinne des Superpositionskalküls. Dies liegt daran, daß gewisse Grundinstanzen der transformierten Klauseln möglicherweise die für eine Simplifikation notwendigen Ordnungsbedingungen verletzen. Jedoch zeigt es sich, daß alle kritischen Instanzen durch Fallunterscheidung behandelt werden können. Der resultierende Kalkül erfordert weder Variablenüberlappungen noch explizite Inferenzen mit den Theorieaxiomen. Falls alle Eingabeklauseln frühzeitig voll abstrahiert werden, ist es überdies möglich, auch auf AC-Unifikation zu verzichten.

Acknowledgments

I am grateful to my thesis advisor Harald Ganzinger, who has supported me and my work for more than eleven years. The central idea underlying this thesis is due to him. I want to thank my colleagues at the Max-Planck-Institut für Informatik for a great working atmosphere, for their support and friendship. In particular I would like to thank Michaël Rusinowitch, who agreed to referee this thesis, an anonymous CADE-13 referee for pointing out to me a generalization of Thm. A.6, Leo Bachmair, Alexander Bockmayr, Ullrich Hustadt, Deepak Kapur, David Plaisted, Georg Struth, Jürgen Stuber, and Sergei Vorobyov for valuable discussions and literature pointers, and Jürgen Stuber and especially Georg Struth for proof reading parts of this thesis.

1 Introduction

Most applications of theorem proving make it necessary to deal with uninterpreted function and predicate symbols that are specific for a particular domain of application, 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 [87], 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 [25]). More often, however, the problem has been tackled from the other side, by integrating mathematical knowledge into a general-purpose theorem prover.

It is rarely possible 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 [17]) 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 [21]). Consequently, the integration has to be achieved on the level of the inference system and simplification techniques.

The superposition calculus (Bachmair and Ganzinger [12]), which can be seen as the result of building-in the equality axioms into the resolution calculus, illustrates the advantages of such an integration: Superposition makes resolution inferences with the equality axioms unnecessary. Compared with ordered resolution, the ordering restrictions and the redundancy criterion are strengthened; besides, superposition inferences at or below variables can be shown to be superfluous.

Similar techniques as for the equality axioms can be use for other theories. Bachmair and Ganzinger [9] and Wertz [100] have integrated the associativity and commutativity axioms into superposition using AC-unification and extended clauses, developed for the equational case by Peterson and Stickel [77]. 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 [69], Vigneron [97]); it is still necessary, though, to solve the unifiability problem, which is NP-complete (Kapur and Narendran [53]).

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 [19], Jouannaud and Marché [51]), which is only simply exponential (Kapur and Narendran [54]), and even unitary for the special case that sums of variables are to be unified. We observe a much more radical improvement when switching 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 $u_1 + \dots + u_k \approx s$ and $v_1 + \dots + v_l \approx t$. 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 a superposition 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 [63]) and in Stuber's extension of the superposition calculus for abelian groups [96].

The group axioms yield both the existence and the uniqueness of the difference of any two elements. Operationally, the second property is much more important than the first one. We can thus employ similar techniques for cancellative abelian monoids, that is, for abelian monoids in which the cancellation axiom $x + y \not\approx x + z \lor y \approx z$ holds. Cancellative abelian monoids are in some sense the most general algebraic structure where such an "abelian-group-like" reasoning is possible. They comprise not only abelian groups, but also such ubiquitous structures as the natural numbers or multisets. In this paper, we present a refined constraint 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. The restriction to overlaps of maximal summands in maximal sides of maximal literals implies in particular that there are no superpositions with shielded variables. While inferences with unshielded variables can not generally be avoided, the number of literals with 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 $\Psi \subseteq \mathbf{N}^{>0}$). Torsion-freeness complicates negative superposition inferences, but it also makes new variable elimination techniques applicable. Since an abelian monoid is cancellative and torsion-free if and only if it can be totally ordered, our calculus allows us to solve equational problems in ordered abelian monoids or groups without having to derive intermediate clauses containing ordering literals.

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. The variable elimination algorithm is not necessarily a simplification in the superposition calculus, though: 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. Furthermore, even AC unifications can be avoided, if clauses are fully abstracted eagerly.

The outline of this work is as follows: In Chapter 2, we provide the prerequisites of this work. We summarize the foundations of first-order logic, equational rewriting, and saturation-based theorem proving, and fix the necessary notations. Then we present the historical background of our calculus, from resolution to theory superposition. In Chapter 3, we present and explain the inference rules of the cancellative superposition calculus. The refutational completeness proof for this calculus, which follows in Chapter 4, is based on the model construction technique of Bachmair and Ganzinger [12], using a novel kind of rewriting on equations. Chapter 5 starts with a presentation of simplification techniques. An immediate consequence of some of these techniques is that the cancellative superposition calculus can be used as a decision procedure for various word problems. We discuss several refinements of the cancellative superposition calculus and its application to theorem proving in ordered abelian monoids. The conclusions follow in Chapter 6. In Appendix A, we provide some algebraic background information concerning cancellative (abelian) semigroups and monoids and their relationship to abelian groups.

Some of the results of this paper have been previously published in (Ganzinger and Waldmann [42, 43]). Section 2.3 is a modified version of (Bachmair, Ganzinger, and Waldmann [17], Sect. 3). The result of Section 5.6 has been discovered shortly after submission of this thesis; Section 5.6 has been added in the final version in agreement with the referees.

2 Superposition-Based Theorem Proving

2.1 Logical Foundations

We start this chapter by briefly summarizing the logical foundations of refutational first-order theorem proving. For a more detailed introduction the reader is referred to Fitting's book [39]. 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 predicate symbols.

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 $f : S_1 \ldots S_n \to S_0$,[1] every variable $x \in \mathcal{V}$ comes with a unique declaration $x : S_0$, where $S_0, \ldots, S_n \in \mathcal{S}$.

Definition 2.1 *The set of terms of sort S is the least set containing x whenever $x : S \in \mathcal{V}$, and containing $f(t_1, \ldots, t_n)$ whenever each t_i is a term of sort S_i and $f : S_1 \ldots S_n \to S \in \Sigma$.*

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 (i. e., that every sort is inhabited).

[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 list of the symbols used in this paper can be found on page 157.

DEFINITION 2.2 *An equation e is an ordered pair (t, t') of terms, usually written as $t \approx t'$, where t and t' have the same sort.*

The left-hand side and right-hand side of an equation 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 easily be coded using a function symbol p, so that $P(t_1, \ldots, t_n)$ is to be taken as an abbreviation for the equation $p(t_1, \ldots, t_n) \approx \mathrm{true}_p$, where $p(t_1, \ldots, t_n)$ and true_p have a new sort S_p.

DEFINITION 2.3 *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 $\mathrm{neg}(C)$. We use the symbol $\bot$ to denote the empty clause, i. e., the empty multiset of literals.

To assign a semantics to a set of formulae we need the concept of a model.

DEFINITION 2.4 *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 $f : S_1 \ldots S_n \to S_0 \in \Sigma$ a function $f^{\mathfrak{M}} : S_1^{\mathfrak{M}} \times \cdots \times S_n^{\mathfrak{M}} \to S_0^{\mathfrak{M}}$, 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.*

DEFINITION 2.5 *An $\mathfrak{M}$-assignment α is a mapping from the set of variables $\mathcal{V}$ to $\bigcup_{S \in \mathcal{S}} S^{\mathfrak{M}}$ such that $\alpha(x) \in S^{\mathfrak{M}}$ for every $x : S$.*

Every assignment α can be homomorphically extended to a mapping α^* from terms to $\bigcup_{S \in \mathcal{S}} S^{\mathfrak{M}}$ by defining recursively $\alpha^*(x) = \alpha(x)$ if $x \in \mathcal{V}$ and $\alpha^*(f(t_1, \ldots, t_n)) = f^{\mathfrak{M}}(\alpha^*(t_1), \ldots, \alpha^*(t_n))$ if $f \in \Sigma$. Usually, the assignment and its extension are denoted by the same symbol. For a ground term t, $\alpha(t)$ depends only on $\mathfrak{M}$, but not on α. To emphasize this point we often write $\mathfrak{M}(t)$ instead of $\alpha(t)$.

DEFINITION 2.6 *An interpretation $\mathfrak{M}$ is called term-generated, if for every element $\mathfrak{m}$ of some $S^{\mathfrak{M}}$ there is a ground term t of sort S such that $\mathfrak{m} = \mathfrak{M}(t)$.*

DEFINITION 2.7 *Let $\mathfrak{M}$ be an interpretation and α be an assignment. 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. If a literal or clause is not true, it is called false with respect to* $\mathfrak{M}$ *and* α.

It is clear that the empty clause $\bot$ is false with respect to all interpretations and assignments.

DEFINITION 2.8 *An interpretation* $\mathfrak{M}$ *is a model of a clause, if the clause is true with respect to* $\mathfrak{M}$ *and* α *for every* $\mathfrak{M}$*-assignment* α. *It is a model of a set of clauses, if it is a model of every clause in the set.*

If $\mathfrak{M}$ is a model of a set of clauses, we also say that it satisfies the set. A set of clauses is called satisfiable if it has a model, and unsatisfiable, otherwise.

First-order logic enjoys the compactness property. This means that every set of clauses (or even more generally, every set of first-order formulae) is satisfiable whenever each of its finite subsets is satisfiable:

THEOREM 2.9 *A set of clauses is unsatisfiable if and only if it contains a finite subset that 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 models: Let $\mathfrak{M}_0$ be any model of a set of clauses. Then we can construct a term-generated model $\mathfrak{M}$ by taking as $S^{\mathfrak{M}}$ the subset of elements of $S^{\mathfrak{M}_0}$ that are images of ground terms of sort S, and by restricting $f^{\mathfrak{M}}$ and $\approx^{\mathfrak{M}}$ accordingly.[2] Consequently, we obtain the following lemma.

LEMMA 2.10 *A set of clauses has a model if and only if it has a term-generated model.*

A substitution σ is a mapping from $\mathcal{V}$ to the set of terms over Σ and $\mathcal{V}$, such that x and $\sigma(x)$ have the same sort for every x. Substitutions can be considered as a particular kind of assignments, namely assignments into an interpretation $\mathfrak{M}$, where $S^{\mathfrak{M}}$ is the set of terms over Σ and $\mathcal{V}$ and $f^{\mathfrak{M}}$ is the function that maps $t_1, \dots, t_n$ to $f(t_1, \dots, t_n)$. As any other assignment, a substitution can be homomorphically extended to a function from terms to terms. Note that we have $\sigma(t) = \mathfrak{M}(t) = t$ for every ground term t. It is customary to use postfix notation for substitutions, i. e., to write $t\sigma$ rather than $\sigma(t)$; $\sigma\sigma'$ is the substitution that maps every x to $(x\sigma)\sigma'$.

[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.

The set $\mathrm{Dom}(\sigma) = \{\, x \in \mathcal{V} \mid x\sigma \neq x \,\}$ is called the domain of the substitution σ. A substitution with domain $\{x_1, \ldots, x_n\}$ that maps the variables $x_1, \ldots, x_n$ to the terms $t_1, \ldots, t_n$, respectively, is denoted by $\{x_1 \mapsto t_1, \ldots, x_n \mapsto t_n\}$. If σ_1 and σ_2 have disjoint domains, then we write $\sigma_1 \cup \sigma_2$ for the substitution that maps every $x \in \mathrm{Dom}(\sigma_i)$ to $x\sigma_i$ for $i \in \{1, 2\}$. A substitution σ is said to be idempotent, if $\sigma\sigma = \sigma$, i. e., if $\mathrm{Dom}(\sigma)$ and $\bigcup_{x \in \mathrm{Dom}(\sigma)} \mathrm{var}(x\sigma)$ are disjoint. 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 $x \in \mathcal{V}'$.

A term $\bar{s}$ is called an instance of a term s if $s\sigma = \bar{s}$ for some substitution σ. This terminology can be extended to equations by defining that $\bar{e}$ is an instance of e if $e = s \approx t$ and $\bar{e} = s\sigma \approx t\sigma$; in an analogous way it can be extended to other syntactic objects, such as clauses. A ground object that is an instance of some syntactic object Q is called a ground instance of Q.

A substitution σ is called a unifier of the terms $t_1, \ldots, t_n$, if $t_i\sigma = t_j\sigma$ for all $i, j \in \{1, \ldots, n\}$. A unifier σ of $t_1, \ldots, t_n$ is called a most general unifier, if for every unifier θ of $t_1, \ldots, t_n$ there exists a substitution ρ such that $\theta = \sigma\rho$ over $\mathrm{var}(\{t_1, \ldots, t_n\})$. Every set of terms $\{t_1, \ldots, t_n\}$ that has a unifier has an idempotent most general unifier.

If $\mathfrak{M}$ is a term-generated interpretation, then for every assignment α there is a substitution σ mapping variables to ground terms such that $\alpha(x) = \mathfrak{M}(x\sigma)$, and hence $\alpha(t) = \mathfrak{M}(t\sigma)$. This fact gives rise to the following lemma:

Lemma 2.11 *A term-generated interpretation is a model of a clause C if and only it is a model of all ground instances of C.*

As a consequence of Thm. 2.9, Lemma 2.10, and Lemma 2.11, we obtain the following corollary:

Corollary 2.12 *A set N of clauses is unsatisfiable if and only if there exists an unsatisfiable finite set of ground instances of clauses in N.*

As long as we restrict ourselves to term-generated models (which is possible thanks to Lemma 2.10), we may think of a non-ground clause as a finite representation of the set of all its ground instances. Generalization of this idea leads to the concept of a symbolic constraint: a formula that restricts the set of instances to be represented by a clause.

Definition 2.13 *A constraint is a first-order formula. A constrained clause $C\,[\![T]\!]$ is a pair consisting of a clause C and a constraint T.*

The first-order language in which the constraints are formulated comes with a fixed interpretation of the non-logical symbols. For instance, a typical constraint language might consist of all quantifier-free formulae built over true, false, $\wedge$, and the binary predicate symbols $\doteq$ and $\succ$, where $\doteq$ is interpreted as equality of terms and $\succ$ as some ordering on terms. A substitution θ satisfies a constraint T, if $T\theta$ evaluates to true in this fixed interpretation. A substitution that satisfies T is also called a solution of T. If for some substitution θ, $C\theta$ is ground and $T\theta = \text{true}$, then the clause $C\theta$ is called a ground instance of the constrained clause $C\,[\![T]\!]$. A term-generated interpretation is said to be a model of a constrained clause $C\,[\![T]\!]$, if it is a model of all ground instances of $C\,[\![T]\!]$. Sometimes we identify an unconstrained clause C with the constrained clause $C\,[\![\text{true}]\!]$. We say that a constraint T is satisfiable, if there exists a substitution θ such that $T\theta = \text{true}$. A constrained clause whose constraint is unsatisfiable is a tautology: is has no ground instances and is thus true in every interpretation.

We obtain a particularly important class of term-generated interpretations if we take the set of ground terms as domain:

Definition 2.14 *An interpretation $\mathfrak{M}$ is called a Herbrand interpretation, if for every $S \in \mathcal{S}$ and $f \in \Sigma$, the set $S^{\mathfrak{M}}$ is the set of* ground *terms of sort S and the function $f^{\mathfrak{M}}$ is the function that maps $t_1, \ldots, t_n$ to $f(t_1, \ldots, t_n)$.*

In a Herbrand interpretation, every ground term is interpreted by itself, that is, $\mathfrak{M}(t) = t$. As $S^{\mathfrak{M}}$ and $f^{\mathfrak{M}}$ are fixed, 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 the Herbrand interpretation $\mathfrak{M}$ with the set $E_{\mathfrak{M}}$. 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 $e \notin E_{\mathfrak{M}}$.

According to Lemma 2.10, we may confine ourselves to term-generated models for refutational theorem proving. The following theorem shows that this result can still be strengthened.

Theorem 2.15 *A set of (constrained) clauses has a term-generated model if and only if it has a Herbrand model.*

So far, we have considered equality as an arbitrary binary predicate symbol, to be interpreted by an arbitrary binary relation. However, when one uses the equality symbol $\approx$ in a logical language, one is usually interested in interpretations in which the relation represented by $\approx$ actually is equality. We refer to such interpretations as normal.

Definition 2.16 *An interpretation $\mathfrak{M}$ is called normal, if the relation $\approx^{\mathfrak{M}}$ is the equality relation on the domain of $\mathfrak{M}$, i. e., if $a \approx^{\mathfrak{M}} a'$ if and only if $a = a'$.*

It is easy to show that Lemma 2.10 holds also for normal models, that is, that a set of clauses has a normal model if and only if it has a term-generated normal model. Theorem 2.15, on the other hand, cannot be extended to normal interpretations: For any signature there is exactly one normal Herbrand interpretation, and this interpretation is trivial inasmuch as any two terms are different. 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.

DEFINITION 2.17 *The clauses*

$$x \approx x \qquad \text{(Reflexivity)}$$
$$x \not\approx y \lor y \approx x \qquad \text{(Symmetry)}$$
$$x \not\approx y \lor y \not\approx z \lor x \approx z \qquad \text{(Transitivity)}$$
$$x_1 \not\approx y_1 \lor \ldots \lor x_n \not\approx y_n \lor f(x_1, \ldots, x_n) \approx f(y_1, \ldots, y_n) \qquad \text{(Congruence)}$$

(for every n-ary function symbol $f \in \Sigma$) are called equality axioms.

DEFINITION 2.18 *An interpretation that is a model of the equality axioms is called an equality interpretation. If N is a set of clauses, then an equality interpretation that is a model of N is called an equality model of N.*

Term-generated normal interpretations and equality Herbrand interpretations are equivalent in the following sense:

LEMMA 2.19 *For every term-generated normal interpretation $\mathfrak{M}_1$ there exists an equality Herbrand interpretation $\mathfrak{M}_2$ (and vice versa), such that any ground literal is true in $\mathfrak{M}_1$ if and only if it is true in $\mathfrak{M}_2$.*

PROOF. If $\mathfrak{M}_1$ is given, let $\mathfrak{M}_2$ be the Herbrand interpretation where $t \approx^{\mathfrak{M}_2} t'$ if and only if $\mathfrak{M}_1(t) = \mathfrak{M}_1(t')$.

Conversely, if $\mathfrak{M}_2$ is given, define $\mathfrak{M}_1$ as the normal interpretation where $S^{\mathfrak{M}_1}$ is the set of all congruence classes of ground terms of sort S with respect to the relation $\approx^{\mathfrak{M}_2}$, and $f^{\mathfrak{M}_1}$ maps the congruence classes $[t_1], \ldots, [t_n]$ to $[f(t_1, \ldots, t_n)]$. □

As an interpretation is a model of a (constrained) clause if and only if it is a model of all its ground instances, this lemma can be generalized to clauses:

LEMMA 2.20 *For every term-generated normal interpretation $\mathfrak{M}_1$ there exists an equality Herbrand interpretation $\mathfrak{M}_2$ (and vice versa), such that for any (constrained) clause, $\mathfrak{M}_1$ is a model if and only if $\mathfrak{M}_2$ is a model.*

THEOREM 2.21 *A set of (constrained) clauses has a term-generated normal model if and only if it has an equality Herbrand model.*

DEFINITION 2.22 *Let N and N' be sets of constrained clauses. If every equality Herbrand model of N is a model of N', we say that N entails N' modulo equality and abbreviate this by $N \models_{\approx} N'$.*

By Lemma 2.20, we could equivalently define $\models_{\approx}$ using term-generated normal interpretations:

THEOREM 2.23 *If N and N' are sets of constrained clauses, then $N \models_{\approx} N'$ if and only if every term-generated normal of N is a model of 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 corresponding 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 Sections 5.3 and 5.7.

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 section serves mainly to fix the necessary notations; for more detailed information about rewrite systems we refer to Dershowitz and Jouannaud's survey [33].

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 $\text{pos}(t)$. 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.

DEFINITION 2.24 *We say that a binary relation $\rightarrow$ over terms is stable under contexts, if $t_1 \rightarrow t_2$ implies $s[t_1]_o \rightarrow s[t_2]_o$ for all terms t_1, t_2, and s, such that $s[t_1]_o$ and $s[t_2]_o$ are well-formed. It is called stable under substitutions, if $t_1 \rightarrow t_2$ implies $t_1\sigma \rightarrow t_2\sigma$ for all terms t_1, t_2 and all substitutions σ. It is called a rewrite relation, if it is stable under both contexts and substitutions.*

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.

DEFINITION 2.25 *A binary relation $\succ$ is called an ordering, if it transitive and irreflexive. 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 $\Pi^{\succ s}$ is an abbreviation for $\{\, t \in \Pi \mid t \succ s \,\}$; the sets $\Pi^{\succeq s}$, $\Pi^{\prec s}$, and $\Pi^{\preceq s}$ are defined analogously.

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

LEMMA 2.26 *Let $\succ$ and $\rightarrow$ be two binary relations over Π, such that $\succ$ is a noetherian ordering and $\rightarrow \subseteq \succ$. Let s and r be two elements of Π, such that r is irreducible with respect to $\rightarrow$ and define $\Pi^s_r = \{\, t \in \Pi \mid s \succeq t,\ t \rightarrow^* r \,\}$. 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^s_r and Π^s_r is closed under $\rightarrow$.*

PROOF. It is obviously sufficient to prove that for every $t_0 \in \Pi^s_r$ and $t'_1 \in \Pi$, $t_0 \rightarrow^* t'_1$ implies $t'_1 \in \Pi^s_r$. We show this by noetherian induction over the size of t_0. Let $t_0 \in \Pi^s_r$ 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 $t_0 \in \Pi^s_r$ 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 $t_2 \in \Pi^s_r$, hence by the induction hypothesis, $t_3 \in \Pi^s_r$ and thus $t_1 \in \Pi^s_r$. Since $t_0 \succ t_1$, we can use the induction hypothesis once more and obtain $t'_1 \in \Pi^s_r$ as required. □

DEFINITION 2.27 *A rewrite rule e is a pair of (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 $t \rightarrow t' \in R$.

If E is a set of equations, then the rewrite relation $\leftrightarrow_E$ is the smallest rewrite relation containing $t \leftrightarrow_E t'$ and $t' \leftrightarrow_E t$ for every equation $t \approx t' \in E$. The reflexive-transitive closure $\leftrightarrow_E^*$ of E is also denoted by $=_E$.

2.3 Saturation and Redundancy

Most automated theorem provers for first-order logic are refutational provers. To show that a formula C' is a logical consequence of another formula C, they negate C' and try to show that $C \wedge \neg C'$ is contradictory. 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 some set of clauses. The prover is called refutationally complete, if it finds a refutation whenever the set of formulae is inconsistent.

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 conclusion and is denoted by $\mathrm{concl}(\iota)$. For instance, *Inf* might map N to the set of all ground resolution inferences

$$\frac{D' \vee e \qquad C' \vee \neg e}{D' \vee C'}$$

with premises $D' \vee e$ and $C' \vee \neg e$ in N. Typically, an inference system is sound with respect to a given semantical consequence relation $\models$, that is, $\{C_1, \cdots, C_k\} \models \{C_0\}$, for all inferences ι. The consequence relation $\models$ may for example be the relation $\models_\approx$ introduced in Section 2.1, or any other binary 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 "closed" (or "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 to define refutational completeness as a static property, rather than a dynamic one: An inference system is said to be refutationally complete, if saturated sets of formulae are unsatisfiable if and only if they contain a contradictory formula, say the empty clause $\bot$.

In practice, inference rules are equipped with strong local restrictions to keep the search space as small as possible. Nevertheless, the majority of the generated formulae are not actually needed for deriving a contradiction, and saturated sets tend to be very large, often infinite. Thus, techniques are employed to discard redundant formulae and a weaker notion of saturation is needed. For that purpose, we introduce a global concept of redundancy that applies to both formulae and inferences. 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 $\mathit{Red}^{\mathrm{I}}(N)$ are meant to specify formulae and inferences, respectively, deemed to be redundant in the context of a given set N. Under certain conditions, formulae in $\mathit{Red}^{\mathrm{C}}(N)$ may be removed from N, while inferences in $\mathit{Red}^{\mathrm{I}}(N)$ may be ignored. For instance, $\mathit{Red}^{\mathrm{C}}(N)$ may consist of all tautologies and formulae subsumed by N. (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.)

The following conditions characterize a reasonable notion of redundancy for refutational theorem proving:

Definition 2.28 *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) $N \setminus \mathit{Red}^{\mathrm{C}}(N) \models \mathit{Red}^{\mathrm{C}}(N)$.

(ii) If $N \subseteq N'$*, then* $\mathit{Red}^{\mathrm{C}}(N) \subseteq \mathit{Red}^{\mathrm{C}}(N')$ *and* $\mathit{Red}^{\mathrm{I}}(N) \subseteq \mathit{Red}^{\mathrm{I}}(N')$.

(iii) If $N' \subseteq \mathit{Red}^{\mathrm{C}}(N)$*, then* $\mathit{Red}^{\mathrm{C}}(N) \subseteq \mathit{Red}^{\mathrm{C}}(N \setminus N')$ *and* $\mathit{Red}^{\mathrm{I}}(N) \subseteq \mathit{Red}^{\mathrm{I}}(N \setminus N')$.

(iv) If $\iota \in \mathit{Inf}(N')$ *and* $\mathrm{concl}(\iota) \in N$*, then* $\iota \in \mathit{Red}^{\mathrm{I}}(N)$.

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

Condition (i) requires that redundant formulae logically follow from the non-redundant ones. Condition (ii) and condition (iii) indicate that redundant formulae and inferences must remain redundant if formulae are added or if redundant formulae are deleted. 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 2.29 *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 it satisfies the following properties for all sets of formulae* N *and* N'*:*

(i) If $N \vdash N'$*, then* $N \models N'$*.*

(ii) If $N \vdash N'$*, then* $N \setminus N' \subseteq \mathit{Red}^{\mathrm{C}}(N')$*.*

(iii) If $\iota \in \mathit{Inf}(N)$*, then* $N \vdash N \cup \{\mathrm{concl}(\iota)\}$*.*

Note that $N \vdash N'$ implies $N' \models N$ by condition (i) of Def. 2.28.

A derivation relation extends an inference system in such a way that we can not only compute inferences and add their conclusions but that we are also allowed to delete or to simplify formulae or to add lemmas. This is possible as long as all removed formulae are redundant and all new formulae are logical consequences of the old ones.

DEFINITION 2.30 *A triple* $(\mathit{Inf}, \mathit{Red}, \vdash)$ *consisting of an inference system* Inf*, a redundancy criterion* Red*, and a derivation relation* $\vdash$ *is called a theorem proving calculus.*

DEFINITION 2.31 *A set* N *of formulae is called saturated with respect to a theorem proving calculus* $(\mathit{Inf}, \mathit{Red}, \vdash)$*, if* $\mathit{Inf}(N) \subseteq \mathit{Red}^{\mathrm{I}}(N)$*.*

In other words, a set of formulae is saturated if all inferences from it are redundant.

A finite or infinite sequence $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$ is called an $(\mathit{Inf}, \mathit{Red}, \vdash)$-derivation, or simply derivation, if the theorem proving calculus is clear from the context. The set $N_\infty = \bigcup_i \bigcap_{j \geq i} N_j$ 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.

LEMMA 2.32 *For every derivation* $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$ *and every set* N_i *occurring in this derivation, we have* $\mathit{Red}^{\mathrm{C}}(N_i) \subseteq \mathit{Red}^{\mathrm{C}}(\bigcup_j N_j) = \mathit{Red}^{\mathrm{C}}(N_\infty)$ *and* $\mathit{Red}^{\mathrm{I}}(N_i) \subseteq \mathit{Red}^{\mathrm{I}}(\bigcup_j N_j) = \mathit{Red}^{\mathrm{I}}(N_\infty)$*.*

PROOF. Suppose that a formula C is contained in $\bigcup_j N_j \setminus N_\infty$. Since C has been deleted at some point of the derivation, there must exist some k such that $C \in N_k \setminus N_{k+1} \subseteq \mathit{Red}^{\mathrm{C}}(N_{k+1}) \subseteq \mathit{Red}^{\mathrm{C}}(\bigcup_j N_j)$. In other words, $\bigcup_j N_j \setminus N_\infty \subseteq \mathit{Red}^{\mathrm{C}}(\bigcup_j N_j)$. By property (iii) of redundancy criteria we get $\mathit{Red}^{\mathrm{C}}(\bigcup_j N_j) \subseteq \mathit{Red}^{\mathrm{C}}(\bigcup_j N_j \setminus (\bigcup_j N_j \setminus N_\infty)) = \mathit{Red}^{\mathrm{C}}(N_\infty)$ and similarly $\mathit{Red}^{\mathrm{I}}(\bigcup_j N_j) \subseteq \mathit{Red}^{\mathrm{I}}(\bigcup_j N_j \setminus (\bigcup_j N_j \setminus N_\infty)) = \mathit{Red}^{\mathrm{I}}(N_\infty)$. Furthermore, by property (ii) of redundancy criteria, $Red^{\mathrm{C}}(N_i) \subseteq Red^{\mathrm{C}}(\bigcup_j N_j)$, $Red^{\mathrm{C}}(N_\infty) \subseteq Red^{\mathrm{C}}(\bigcup_j N_j)$, $Red^{\mathrm{I}}(N_i) \subseteq Red^{\mathrm{I}}(\bigcup_j N_j)$, and $Red^{\mathrm{I}}(N_\infty) \subseteq Red^{\mathrm{I}}(\bigcup_j N_j)$. □

COROLLARY 2.33 *For every derivation $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$ and every set N_i occurring in this derivation, $N_i \subseteq N_\infty \cup Red^{\mathrm{C}}(N_\infty)$.*

PROOF. If $C \in N_i \setminus N_\infty$, then there is a $k \geq i$ such that $C \in N_k \setminus N_{k+1} \subseteq Red^{\mathrm{C}}(N_{k+1}) \subseteq Red^{\mathrm{C}}(N_\infty)$. □

DEFINITION 2.34 *A derivation $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$ is called fair, if*

$$Inf(N_\infty) \subseteq \bigcup_j Red^{\mathrm{I}}(N_j) \ .$$

LEMMA 2.35 *If a derivation is fair, then its limit is saturated.*

PROOF. If the derivation $N_0 \vdash N_1 \vdash \ldots$ is fair, then $Inf(N_\infty) \subseteq \bigcup_j Red^{\mathrm{I}}(N_j)$. By Lemma 2.32, $Red^{\mathrm{I}}(N_j) \subseteq Red^{\mathrm{I}}(N_\infty)$, hence $Inf(N_\infty) \subseteq Red^{\mathrm{I}}(N_\infty)$. □

DEFINITION 2.36 *A theorem proving calculus $(Inf, Red, \vdash)$ 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 (Inf, Red) instead of $(Inf, Red, \vdash)$.

2.4 Resolution and Superposition

Resolution. The forefather of our method is Robinson's resolution calculus [83]. It is based on a single inference rule,[3] ultimately a form of modus ponens, which can be written for ground clauses as follows:

$$\textit{Resolution (ground)} \qquad \frac{D' \vee e \qquad C' \vee \neg\, e}{D' \vee C'}$$

It allows to derive a new clause whenever the current clause set contains clauses with complementary literals e and $\neg\, e$.

In this ground form, the rule has already been used by Davis and Putnam [31].[4] The fundamental difference comes when we want to check the

[3] In (Robinson [83]), clauses are sets of literals, rather than multisets. In our framework, in which multisets are employed, an additional factoring rule is necessary.

[4] Davis [30] gives an survey of the history of logic and automated deduction from the seventeenth century to 1967. The time from 1965 to 1970 is covered in (Wos and Henschen [101]).

unsatisfiability of a non-ground clause set. Davis and Putnam did this by simply enumerating all possible ground instances of the given clauses and testing the first n instances (for $n = 1, 2, \dots$) for unsatisfiability. By Cor. 2.12, a set of clauses N is unsatisfiable if and only if a finite subset of instances of clauses in N is unsatisfiable. Since this finite set is eventually subsumed when n reaches a sufficiently large value, the program is guaranteed to stop whenever N is unsatisfiable.

Robinson's fundamental achievement was to extend the inference rule to non-ground clauses in such a way that the computation of appropriate instances became a by-product. By unifying the complementary literals, the variables of the clauses involved in the inference are instantiated as much as necessary for soundness, but no more:

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

where σ is a most general unifier of e_1 and e_2.

In contrast to earlier approaches it is no longer necessary to guess the right ground instances beforehand. Instead, the instantiations can be computed.

Resolution constituted such a significant improvement over previous calculi that some researchers even thought that the problems of automated theorem had essentially been solved. Soon it became obvious, however, that these high expectations were unwarrantable. Resolution was still by far too inefficient for solving practically interesting problems. This insight has been the motivation for developing numerous strategies and refinements of the resolution calculus. One of the first improvements is due to Robinson himself [82], who discovered that it is sufficient to consider only resolution inferences in which the first of the two premises contains only positive literals. By iterating inferences of this form,

$$\dfrac{D_3 \vee e_3 \qquad \dfrac{D_2 \vee e_2 \qquad \dfrac{D_1 \vee e_1 \qquad C' \vee \neg\, e_3 \vee \neg\, e_2 \vee \neg\, e_1}{D_1 \vee C' \vee \neg\, e_3 \vee \neg\, e_2}}{D_2 \vee D_1 \vee C' \vee \neg\, e_3}}{D_3 \vee D_2 \vee D_1 \vee C'}$$

we obtain a new derived inference rule, called *hyperresolution*, in which one clause with n negative literals (and possibly positive ones) reacts with n clauses with only positive literals and produces a new clause with only positive literals.

$$\frac{D_n \vee e_n \qquad \dots \qquad D_1 \vee e_1 \qquad C' \vee \neg\, e_n \vee \dots \vee \neg\, e_1}{D_n \vee \dots \vee D_1 \vee C'}$$

The non-ground version of the inference rule is obtained in a similar way as for resolution, the difference being that now σ must be a unifier of *every* pair of complementary literals.

Another important refinement of the resolution calculus results from the introduction of literal orderings (Slagle [92], Kowalski and Hayes [58]): In ordered resolution, only such inferences are necessary in which each of the two complementary literals is maximal in its premise. Both hyperresolution and ordered resolution lead to a significant reduction of the search space while preserving refutational completeness.

Paramodulation. Whereas techniques such as hyperresolution and ordered resolution are aimed to improve the efficiency of resolution in general, other extensions of resolution are focussed on specific application domains. Equational problems are one such domain where the weakness of resolution becomes particularly apparent. By Lemma 2.20, a theorem prover can reason in equational logic by including the equality axioms in the set of formulae. For a resolution-based prover, however, naïve use of the equality axioms turns out to be fatal: The symmetry axiom allows a resolution step with any other clause $C' \vee [\neg]\, t \approx t'$ that contains a (possibly negated) equality literal, producing the clause $C' \vee [\neg]\, t' \approx t$. A clause with n equality literals can thus be transformed into 2^n variants. This exponential increase is already bad enough, but it is harmless compared to the growth caused by the transitivity and congruence laws. Starting from a clause $C' \vee t \not\approx t'$, iterated resolution steps with the transitivity axiom may produce *infinitely* many clauses

$$
\begin{aligned}
&C' \vee t \not\approx y_1 \vee y_1 \not\approx t' \\
&C' \vee t \not\approx y_1 \vee y_1 \not\approx y_2 \vee y_2 \not\approx t' \\
&C' \vee t \not\approx y_1 \vee y_1 \not\approx y_2 \vee y_2 \not\approx y_3 \vee y_3 \not\approx t' \\
&\ldots
\end{aligned}
$$

This corresponds to the fact that an equational proof for $t \approx t'$ may involve arbitrarily many intermediate steps $t \approx t_1$, $t_1 \approx t_2$, $\ldots$, $t_n \approx t'$. Similarly, as an equation $t = t'$ allows to prove $s[t]_o \approx s[t']_o$ where t occurs arbitrarily deep in s, infinitely many clauses can result from iterated resolution steps with the congruence axiom. Few of these clauses are actually needed as intermediate steps to derive a contradiction; most only clutter the search space.

Several techniques have been proposed to dispense with explicit equality axioms and to integrate equality into the set of inference rules, for instance Brand's modification method [22] and Digricoli and Harrison's resolution by unification and equality [34]. The paramodulation calculus of Robinson and Wos [81] turned out to be the most influential. The paramodulation rule embodies the ideas of the resolution calculus and the operation of "replacing equals by equals" that is fundamental for term rewriting. Whenever a clause contains a positive literal $t \approx t'$, the paramodulation rule allows to rewrite a subterm t occurring in some literal $[\neg]\, e[t]_o$ of another clause to t'.

Paramodulation (ground) $$\frac{D' \vee t \approx t' \qquad C' \vee [\neg]\, e[t]_o}{D' \vee C' \vee [\neg]\, e[t']_o}$$

For non-ground clauses, this rule is modified in a similar way as the resolution rule above. Equality is replaced by unifiability, so that the resulting rule is essentially a combination of non-ground resolution and Knuth-Bendix completion [57].[5]

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

where σ is a most general unifier of t and w.

If we add the paramodulation rule to the resolution calculus, inferences with the equality axioms (except for resolution with the reflexivity axiom[6]) are unnecessary for refutational completeness. Paramodulation produces far less intermediate formulae than resolution with the equality axioms. This allows not only significantly shorter refutations, it also reduces the problem that the intermediate results spawn more and more useless consequence clauses.

Superposition. We have mentioned above that paramodulation combines resolution and Knuth-Bendix completion. Both resolution and completion are (or can be) subject to ordering restrictions: In the Knuth-Bendix completion procedure,[7] only overlaps at non-variable positions between the maximal sides of two rewrite rules produce a critical pair. Similarly, the resolution calculus remains a semidecision procedure if inferences are computed only if each of the two complementary literals is maximal in its premise. It is natural to

[5] Essentially the same rule (usually restricted to equational unit or Horn clauses) occurs in narrowing calculi (Fay [38]) used for theory unification.

[6] In their original proof, Robinson and Wos required additionally a so-called functionally reflexive axiom $f(x_1, \ldots, x_n) \approx f(x_1, \ldots, x_n)$ for any function symbol f. They conjectured that these axioms are unnecessary for completeness; the conjecture was verified by Brand [22] and Peterson [76].

[7] Or rather: in its unfailing variant (Bachmair [6]), which is a semidecision procedure for purely equational logic.

ask whether paramodulation may inherit the ordering restrictions of both its ancestors. More precisely: Let a paramodulation inference between clauses $D = D' \vee t \approx t'$ and $C = C' \vee [\neg]\, s[w]_o \approx s'$ be given as

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

where σ is the most general unifier of t and w. Does the calculus remain refutationally complete if we require, as in completion, that (i) w is not a variable, (ii) $t\sigma \not\preceq t'\sigma$, (iii) $(s[w]_o)\sigma \not\preceq s'\sigma$, and, as in ordered resolution, that (iv) $(t \approx t')\sigma$ is maximal in $D\sigma$, and (v) $(s[w]_o \approx s')\sigma$ is maximal in $C\sigma$?

A first result in this direction was obtained by Peterson [76], who showed the admissibility of restrictions (i) and (ii). It was extended to (i), (ii), (iii) for positive literals, and (v) by Rusinowitch [85], and to (i), (ii), (iv), and (v) by Hsiang and Rusinowitch [46]. The final answer was given by Bachmair and Ganzinger [7, 8, 12]: All five restrictions may be imposed on the paramodulation rule (which is named *superposition* then), however, an additional inference rule becomes necessary to cope with certain non-Horn clauses. If a clause C has the form

$$C' \vee t \approx t' \vee s \approx s'$$

where $s \approx s'$ is a maximal literal and s and t have the most general unifier σ and are maximal in their respective literals, then we need either an *equality factoring* rule that allows to derive

$$(C' \vee s' \not\approx t' \vee t \approx t')\sigma$$

from C, or a *merging paramodulation* rule, i. e., a rule allowing paramodulations of another clause into the non-maximal term s' of the maximal literal of C. The resulting inference system (on which our own system will be based) is the basis of the superposition calculus; it consists of the rules *superposition*, *equality resolution* (i. e., ordered resolution with the reflexivity axiom), and either *equality factoring* or *ordered factoring* and *merging paramodulation*.[8]

The superposition calculus does not only combine the ordering restrictions of resolution and completion; it can also be equipped with simplification techniques that subsume those of resolution and those of completion. The definition of redundancy introduced by Bachmair and Ganzinger [12] basically states that clauses are redundant if they follow from smaller clauses and equality, and that an inference is redundant if its conclusion follows from clauses that are smaller

[8]The version using the *merging paramodulation* rule appeared first in (Bachmair and Ganzinger [7, 8]). The *equality factoring* rule is due to Nieuwenhuis [66].

than the maximal premise and equality. Usual strategies for resolution-like calculi such as clause subsumption are encompassed by this definition, just as the simplification steps and critical-pair criteria [6] that can be found in completion procedures. Superposition can also be enhanced by selection functions, so that hyperresolution-like strategies become applicable.

Basic Calculi and Constraints. To explain what "basic" means in paramodulation-like calculi, we first have to define the skeleton of a clause. Let $N_0 \vdash N_1 \vdash \dots$ be a derivation. For each of the initially given clauses in N_0 we define the skeleton as the set of all its non-variable positions. If a clause in N_i $(i > 0)$ is the result of a paramodulation inference[9]

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

then we define its skeleton as the set of all non-variable positions of $D' \vee C' \vee [\neg]\, s[t']_o \approx s'$ that correspond to skeleton positions of one of the premises. In other words, a non-variable position in the conclusion is *not* part of the skeleton if it has been introduced either by the unifying substitution σ, or by the unifying substitution at a previous derivation step where one of its ancestors has been created. A paramodulation-like calculus is called basic, if we require that the position o in the paramodulation inference above must be contained in the skeleton of the second premise. The basic strategy was first developed by Degtyarev [32] for paramodulation and by Hullot [48] for equational narrowing. Bachmair, Ganzinger, Lynch, and Snyder [14] and Nieuwenhuis and Rubio [67] proved independently that basic restrictions are compatible with the superposition calculus. Although redundancy criteria for basic superposition are somewhat less powerful than their counterpart for standard superposition, the considerable reduction of the number of possible inferences leads to a significantly smaller search space.

Basic superposition calculi can be implemented in several ways. A clause

$$C\rho \;=\; f(\boxed{x}) \not\approx b \vee h(\boxed{g(x)}, \boxed{c}) \approx f(b)\,,$$

where the non-skeleton parts are indicated by boxes, can be coded as a *closure*, i. e., a pair $C \cdot \rho$ consisting of a clause $C = f(z_1) \not\approx b \vee h(z_2, z_3) \approx f(b)$ and a substitution $\rho = \{z_1 \mapsto x,\, z_2 \mapsto g(x),\, z_3 \mapsto c\}$. Using this representation, the superposition rule has the form

$$\frac{(D' \vee t \approx t') \cdot \rho_2 \qquad (C' \vee [\neg]\, s[w]_o \approx s') \cdot \rho_1}{(D' \vee C' \vee [\neg]\, s[t']_o \approx s') \cdot (\rho_2 \cup \rho_1)\sigma}$$

where σ is a most general unifier of $t\rho_2$ and $w\rho_1$.[10] The usual requirement that

[9] Analogously for other inferences such as factoring.

[10] We assume that $D' \vee t \approx t'$ and $C' \vee [\neg]\, s[w]_o \approx s'$ have no common variables.

w is not a variable ensures that the superposition takes place at a skeleton position of $(C' \vee [\neg]\, s[w]_o \approx s')\rho_1$.

Alternatively, we can code the clause $C\rho$ above by a constrained clause $C\,[\![T]\!]$, where the equality constraint T has the solution ρ. One possible candidate is of course the constraint $z_1 \doteq x \wedge z_2 \doteq g(x) \wedge z_3 \doteq c$, but any equality constraint with the same solution will do. In particular, when we compute the superposition of two clauses $D' \vee t \approx t'\,[\![T_2]\!]$ and $C' \vee [\neg]\, s[w]_o \approx s'\,[\![T_1]\!]$, there is no need to compute the most general unifier of t and w. Rather we can add the constraint $t \doteq w$ to the constraints of the premises:

$$\frac{D' \vee t \approx t'\,[\![T_2]\!] \qquad C' \vee [\neg]\, s[w]_o \approx s'\,[\![T_1]\!]}{D' \vee C' \vee [\neg]\, s[t']_o \approx s'\,[\![T_2 \wedge T_1 \wedge t \doteq w]\!]}$$

We may simplify the constraint immediately to some solved form by computing a unifier of t and w, but we may also delay the computation of unifiers until we have derived $\bot\,[\![T]\!]$, the empty clause with some constraint T. Only then it becomes necessary to check T for satisfiability, so that we can see whether $\bot\,[\![T]\!]$ actually has the instance $\bot$.

Equality constraints are not the only kind of constraints that are useful in superposition theorem proving. Syntactic ordering constraints, originally introduced by Kirchner, Kirchner, and Rusinowitch [55], allow to propagate ordering restrictions from one superposition inference to another one. To see the benefit of this technique consider the clause $D = f(x,y) \approx f(y,x)$. There is a superposition inference ι from D and $g(f(z,b)) \approx b$, namely

$$\frac{f(x,y) \approx f(y,x) \qquad g(f(z,b)) \approx b}{g(f(b,z)) \approx b}$$

There is another superposition inference ι' from D and the conclusion of ι, namely

$$\frac{f(x,y) \approx f(y,x) \qquad g(f(b,z)) \approx b}{g(f(z,b)) \approx b}$$

However, the ordering conditions of these two inferences are incompatible: We have to perform ι only if $f(z,b) \succ f(b,z)$, and ι' only if $f(b,z) \succ f(z,b)$. If the ordering condition is inherited by the conclusion of an inference as a constraint, the cycle is broken: The inference ι' becomes impossible, when its second premise is changed to $g(f(b,z)) \approx b\,[\![f(z,b) \succ f(b,z)]\!]$. Nieuwenhuis and Rubio [68, 70] have shown that the same techniques that are used in the completeness proof for basic superposition allow to show the completeness of superposition with ordering constraints, and that ordering and equality constraints can be combined in a single calculus.

2.5 Theory Reasoning

The Problem. Superposition is a refutationally complete calculus: Whenever a set of input clauses entails a contradiction, any fair superposition derivation will eventually produce the empty clause. Of course, this is a purely qualitative result. As first-order logic is only semi-decidable, the number of inference steps that are necessary to detect the contradiction is uncomputable and may be arbitrarily large.[11] Thus it is not surprising that there are many refutable inputs for which superposition needs more time to find a contradiction than a user is willing to wait. What is disappointing, however, is that superposition (just as other general-purpose theorem proving methods) often fails to find a contradiction, even if the input is almost trivial from a human's point of view.

Algebraic theories are typical examples for such a behaviour. Let us consider abelian groups. Whereas the word problem in abelian groups is efficiently solvable, it is very difficult for a usual general-purpose theorem prover to work in such a structure. The left and right-hand sides of the associativity and commutativity axioms contain few function symbols and relatively many variables. Such formulae are extremely prolific, allowing paramodulation inferences with every clause containing a subterm $t + t'$. Even worse, the left and right-hand side of the commutativity axiom are incomparable with respect to any reduction ordering, so that the paramodulation inferences above are not necessarily simplifications. Together, these effects lead to an explosion of the search space.

In general it can be said that automated theorem provers are competitive for two classes of problems: The first one consists of puzzles, mostly unstructured collections of formulae. The second one contains those problem formulations that are adapted to the way of reasoning of the prover, that are, for instance, sufficiently close to programs in a functional or logic programming language. In other words, automated theorem provers can beat humans, if either there is no structure, or if the structure of the problem can be exploited. If the problem is structured in some way, but the prover does not recognize or cannot make use of the structure, it will usually fail.

The Solution. What conclusion do we have to draw from the preceding paragraph? If we want a prover to work efficiently in a structured domain, we

[11] More precisely: For every computable total function $\phi : \mathbf{N} \to \mathbf{N}$ there exists an $m \in \mathbf{N}$ and an inconsistent set of clauses N, such that the number of occurrences of function symbols and variables in N does not exceed m and every refutation of N by superposition requires more than $\phi(m)$ derivation steps. The same is true for resolution and other saturation-based calculi, provided that for every finite set N_0 of clauses the set $\{ N_1 \mid N_0 \vdash N_1 \}$ is finite and computable.

have to enable it to exploit that structure: We have to build in domain-specific knowledge. This can be done in several ways. Sometimes, the deductive system of the prover can be left unchanged and mere preprocessing is sufficient to make the input better suited (or suited at all) for a theorem prover. Sometimes external (semi-)decision procedures can be linked to the prover in a black-box manner, so that the deductive system must be modified but the modification is independent of the particular theory. In other cases, even such a hierarchic approach is impossible and domain-specific changes of the inference rules or deduction strategies are necessary. We call the first kind of methods transformation techniques, the second one hierarchic techniques, and the third one integrating techniques.

Transformation Techniques. The first and still most important example of the transformation approach predates the first automated theorem provers by several decades: Skolemization [90, 91] serves to transform first-order formulae with arbitrary quantifications into formulae in which all variables are universally quantified. To obtain an even more restricted class of formulae skolemization can be combined with clause normal form transformation. Improved skolemization techniques have been presented for instance by Andrews [1] and Ohlbach and Weidenbach [74]. The differences between several clause normal form transformations and the consequences of the choice of such a transformations for a clausal theorem prover have been investigated by Baaz, Fermüller, and Leitsch [5], Boy de la Tour [20], Egly [36], and Egly and Rath [37].

Transformation methods allow us to extend the range of logics a prover can handle. Gabbay and Ohlbach [40] have shown that quantifier elimination techniques can be used to translate certain formulae of second-order logic into first-order logic. Translations from various modal logics to first-order predicate logic have been described by Auffray and Enjalbert [2], Nonnengart [71], Ohlbach [72], and Ohlbach, Schmidt, and Hustadt [73]. Other examples include Ganzinger's translation from order-sorted equational specifications into many-sorted logic [41], and the boolean ring method to deal with first-order predicate logic in an equational completion framework (Hsiang [45], Zhang [103]).

While transformation methods are the most important tool to make extensions of first-order logic digestible for a first-order theorem prover, applications for algebraic structures are less frequent. One such transformation that will be useful in our context is the replacement of the axioms of totally ordered abelian monoids by the axioms of torsion-free cancellative abelian monoids. By Thm. A.13, this replacement is possible if the ordering predicate appears only in the total ordering axioms and nowhere else. It allows us to prove equational problems in an equational framework without a detour via the ordering axioms.

Hierarchic Techniques. The transformation approach is obviously not a panacea. For many domains, no suitable preprocessing technique is available, so theory knowledge must be integrated *into* the deductive system. This can be done in two ways: Either an external (semi-)decision procedure for the theory is linked in a modular way as a subroutine to the prover, or theory knowledge is coded by means of new inference rules and deduction strategies. From an implementor's point of view, the first possibility is obviously more attractive. If the external procedure is added to the prover as a black box, then the modification of the deductive system can be independent of the particular theory. If suffices to change the prover in such a way that all formulae the external procedure can handle are forwarded to the external procedure.

In the context of logic programming, hierarchic techniques have been extremely successful (Jaffar and Lassez [50]). They work also for resolution theorem proving with theory unification (Bürckert [27, 28]) or certain instances of theory resolution (Stickel [95]). If we are dealing with equational theorem proving, however, their applicability is severely limited: Suppose that the external (semi-)decision procedure can handle formulae over some base vocabulary $\Sigma_{\mathrm{Base}} \subseteq \Sigma$. Then Bachmair, Ganzinger, and Waldmann's hierarchic superposition calculus requires that the function symbols from $\Sigma \setminus \Sigma_{\mathrm{Base}}$ are defined in a sufficiently complete manner on top of the base vocabulary Σ_{Base} [17]. Sufficient completeness is not as restrictive as the conditions required for combinations of decision procedures (Baader and Schulz [4], Nelson and Oppen [65], Ringeissen [80]), but nevertheless, in an algebraic context, it practically excludes uninterpreted function symbols. If we want to avoid sufficient completeness, we have to admit junk (Avenhaus and Becker [3]), and this is still less acceptable.

Even in situations where sufficient completeness is not a too restrictive requirement, a strictly hierarchic approach has a second drawback, which is due to insufficient communication between the external decision procedure and the general prover. This observation was made by Boyer and Moore [21] when they integrated a decision procedure for linear arithmetic into an inductive prover: Unless mathematical routines are tightly interwoven with the rest of the prover, they have few chances to contribute to the proof. If a decision procedure is really a "black box", it is mostly useless.

Integrating Techniques. If the domain-specific knowledge that we want to build into a prover can be handled neither by preprocessing nor by linking an external decision procedure, then we have no other choice than integrating it into the deductive system, usually by fine-tuning the set of inference rules.

We have already seen one example of this technique, namely the paramodulation calculus: the result of integrating the equality axioms into the resolution calculus. Paramodulation and its refinements illustrate what can be reached by incorporating theory axioms into a deductive system: Using paramodulation, all resolution inferences with the equality axioms become superfluous, with the only exception of resolution with the reflexivity law. The theory axioms can be integrated into the redundancy criterion, such that it is sufficient for redundancy if a clause follows from smaller clauses *and (arbitrarily large instances of) the equality axioms*. Compared to ordered resolution, the ordering restrictions can be strengthened: While ordered resolution requires only that inferences involve maximal literals, they must involve *maximal sides of maximal literals* in the superposition calculus. Finally, superposition inferences at or below variables can be shown to be unnecessary. This is particularly important, since unification with a variable succeeds always, so that without the variable restriction any clause could be superposed at a variable position.

The chaining rule for handling orderings is a second example for the integration of a theory into a general theorem prover. As we have already seen in the equality case, unrestricted resolution inferences with the transitivity axiom $\neg\, x < y \,\vee\, \neg\, y < z \,\vee\, x < z$ may produce arbitrarily many consequences. Selection functions (or hyperresolution) can be used to avoid this: If we select the two negative literals of the transitivity axiom, then the only possible sequences of inferences in which it can be used have the form

$$\cfrac{D' \vee t < t' \qquad \cfrac{C' \vee s < s' \qquad \neg\, x < y \vee \neg\, y < z \vee x < z}{C' \vee \neg\, x < s \vee x < s'}}{(D' \vee C' \vee t < s')\sigma}$$

where σ is a most general unifier of t' and s, and $t < t'$ and $s < s'$ are maximal in the respective clauses.

We can thus dispense with inferences with the transitivity axiom, if we use the chaining rule, introduced by Slagle [93], instead:

$$\frac{D' \vee t < t' \qquad C' \vee s < s'}{(D' \vee C' \vee t < s')\sigma}$$

where σ is a most general unifier of t' and s.

As in the equational case, we gain the possibility to furnish the chaining rule with additional restrictions: We may not only require that $t < t'$ and $s < s'$ are maximal literals, but also that t' and s are maximal terms in the premises. Besides we have the chance to exploit further information about the ordering. In particular, if $<$ is a dense total ordering without endpoints, then chaining inferences where either t' or s is a variable x can be completely avoided: If x is shielded, that is, if some term $w[x]$ occurs in the clause, then $w[x] \succ x$, so x cannot be maximal. If x is unshielded, then the literal in which x occurs can be eliminated using for instance the denseness axiom (Bachmair and Ganzinger [11, 13]).

If our set of theory axioms consists of unit equations, then one way to avoid inferences with them is to work with normal forms or equivalence classes. A typical example is the theory of associativity and commutativity, commonly abbreviated as AC. The integration of the AC axioms into the paramodulation calculus has been investigated already by Plotkin [78]; among his successors, we find Slagle [94], Rusinowitch and Vigneron [86], and Paul [75]. Wertz [100] and Bachmair and Ganzinger [9] have integrated the AC axioms into the superposition calculus, using the idea of extended clauses developed for the equational case by Peterson and Stickel [77]. All these approaches have in common that standard unification has to be replaced by unification modulo an equational theory.

The use of AC-unification algorithms has severe consequences for the efficiency of the prover: A minimal complete set of AC-unifiers may have doubly exponential size (Domenjoud [35], Kapur and Narendran [54]), so that computing all inferences between two clauses may not only require doubly exponential time, but may also produce doubly exponentially many conclusions. Under these circumstances, the possibility offered by the constraint framework to delay the unification becomes particularly valuable. As the enumeration of unifiers can be avoided, the number of generated conclusions is drastically reduced. Only if $\perp [\![T]\!]$ has been derived, it is necessary to solve T, that is, to decide the unifiability problem (Nieuwenhuis and Rubio [69], Vigneron [97]). This is still NP-complete, though (Kapur and Narendran [53]).

It appears that problems caused by the integration of algebraic structure can be mitigated by integrating *more* algebraic structure. If our theory contains also the identity axiom, then we can replace AC-unification by ACU-unification[12] (Boudet, Contejean, and Marché [19], Jouannaud and Marché [51]), which is only simply exponential (Kapur and Narendran [54]) and even unitary for the special case where sums of variables are to be unified.

A much more radical improvement can be observed when we switch over from the abelian semigroup axioms AC or the abelian monoid axioms ACU to the axioms of abelian groups. We get a first impression of the fundamental difference between these structures when we compare the uniform word problems:

[12]That is, unification modulo AC and identity. In the literature, the abbreviations ACU and AC1 are used interchangeably.

Whereas the word problem for abelian semigroups or monoids is EXPSPACE-complete (Cardoza, Lipton, and Meyer [29], Mayr and Meyer [64]), the word problem for abelian groups can be solved in polynomial time (Kandri-Rody, Kapur, and Narendran [52]).

To see the operational difference between abelian semigroups and abelian groups in a rewrite or superposition-based calculus consider the following example: Let $s \approx s'$ and $t \approx t'$ be two equations, where $s = u_1 + \cdots + u_k$ and $t = v_1 + \cdots + v_l$. We suppose without loss of generality that u_1 is maximal among the u_i and v_1 is maximal among the v_j. In AC-completion or AC-superposition, there is an inference between these two equations (via extended rules), whenever s and t are maximal in the respective equations and *some* u_i is unifiable with *some* v_j. The result is a new equation $(u_1 + \cdots + u_{i-1} + u_{i+1} + \cdots + u_k + t')\sigma \approx (v_1 + \cdots + v_{j-1} + v_{j+1} + \cdots + u_l + s')\sigma$. In the presence of the inverse axiom, however, the equations $s \approx s'$ and $t \approx t'$ can be equivalently written as $u_1 \approx s' + (-u_2) + \cdots + (-u_k)$ and $v_1 \approx t' + (-v_2) + \cdots + (-v_l)$. As the new equations do not have sums on their left-hand sides, extended rules become superfluous. Hence the two equations allow a superposition inference only if u_1 is unifiable with v_1. This technique, known as symmetrization (Le Chenadec [61]), can be found for instance in Zhang's work on distributivity and abelian groups in rewriting [102], in normalized rewriting (Marché [63]), in Stuber's extension of the superposition calculus for abelian groups [96], and in Gröbner bases and related theorem proving calculi [10, 15, 23, 24, 26]. Its superiority over pure AC-rewriting techniques is exemplified by Marché [63], who notes that his AG-normalized completion procedure, applied to a set of three equations taken from [60], produces only 11 critical pairs. On the other hand, for AC-completion of the same equations plus the abelian group axioms, REVEAL computes 183 and RRL even 837 critical pairs.

The group axioms yield both the existence and the uniqueness of the difference of any two elements. Operationally, the second property is much more important than the first one. It is the uniqueness, the fact that each term in the equation $\sum u_i \approx t'$ is determined by the others, that makes it sufficient to consider only overlaps of the maximal terms in a sum. This insight allows us to employ similar techniques for cancellative abelian monoids, that is, for abelian monoids in which the cancellation axiom $x + y \not\approx x + z \vee y \approx z$ holds, or in other words, for submonoids of abelian groups.[13] Cancellative abelian monoids are in some sense the most general algebraic structure where such an "abelian-group-like" reasoning is possible. They encompass not only abelian groups, but also such ubiquitous structures as the natural numbers or multisets.

[13] Some basic facts about cancellative monoids can be found in Appendix A.

In the rest of this paper we will present and explain a refined superposition calculus for sets of clauses including the axioms of cancellative abelian monoids. Our calculus requires neither explicit inferences with the theory axioms nor extended clauses. It generalizes the usual constraint superposition calculus and inherits its ordering restrictions, so that inferences involving non-maximal literals or non-maximal sides of equations are superfluous. Furthermore, similar to the calculi for abelian groups mentioned above, overlaps at non-maximal parts of a sum become unnecessary. As in the chaining calculus, this means in particular that there are no overlaps with shielded variables. While inferences with unshielded variables can not generally be avoided, we present several simplification techniques that allow us to reduce the number of literals with unshielded variables. Optionally, we can also integrate the torsion-freeness axioms $\psi x \not\approx \psi y \vee x \approx y$ for $\psi \in \mathbf{N}^{>0}$ into our calculus. Since an abelian monoid is cancellative and torsion-free if and only if it can be totally ordered, we get in this way the chance to solve equational problems in ordered abelian monoids or groups without having to derive intermediate clauses containing ordering literals.

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 $0 : \to S_{\text{CAM}}$ and $+ : S_{\text{CAM}} S_{\text{CAM}} \to S_{\text{CAM}}$. 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 of clauses and* AC, ACU, ACK, ACUK *for the respective subsets.*

By Lemma A.17, 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 *The symbol $=_{\text{ACU}}$ denotes the congruence generated by* ACU. *The* ACU*-congruence class of a term t is $[t]_{\text{ACU}} = \{\, t' \mid t =_{\text{ACU}} t' \,\}$.*

DEFINITION 3.3 *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 $t_1 \neq_{\text{ACU}} 0$, $t_2 \neq_{\text{ACU}} 0$.*

The set of all terms is the disjoint union of the three sets $\{\, t \mid \exists x\colon\ x$ is a variable, $t =_{\text{ACU}} x \,\}$, $\{\, t \mid \exists s\colon\ s$ is atomic, $t =_{\text{ACU}} s \,\}$, and $\{\, t \mid \exists s\colon\ s$ is a proper sum, $t =_{\text{ACU}} s \,\}$. We can therefore extend the terminology above to ACU-congruence classes and say that $[t]_{\text{ACU}}$ is a variable (an atomic term, a proper sum), if there is some $s \in [t]_{\text{ACU}}$ with this property.

We say that the term t occurs in the term s at the top, if there is a position $o \in \text{pos}(s)$ 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 a position $o \in \text{pos}(s)$ 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 $[t]_{\text{ACU}}$ occurs in $[s]_{\text{ACU}}$ at the top or (immediately) below a free function symbol, if there are some $t' \in [t]_{\text{ACU}}$ and $s' \in [s]_{\text{ACU}}$ with this property. For instance, $[2b]_{\text{ACU}}$ and $[f(2c) + b]_{\text{ACU}}$ occur at the top of $[c + 2b + 3f(2c)]_{\text{ACU}}$; $[c]_{\text{ACU}}$ occurs both at the top and below a free function symbol, but not immediately below a free function symbol; $[2c]_{\text{ACU}}$ occurs immediately below a free function symbol.

A substitution σ is called an ACU-unifier of the terms $t_1, \ldots, t_n$, if $t_i\sigma =_{\text{ACU}} t_j\sigma$ for all $i, j \in \{1, \ldots, n\}$. A set U of ACU-unifiers of $t_1, \ldots, t_n$ is called complete, if for every ACU-unifier θ of $t_1, \ldots, t_n$ there exists a $\sigma \in U$ and a substitution ρ such that $x\theta =_{\text{ACU}} x\sigma\rho$ for all $x \in \text{var}(\{t_1, \ldots, t_n\})$. ACU-unification is finitary: for every set of terms $\{t_1, \ldots, t_n\}$ there exists a (possibly empty) finite minimal complete set of idempotent ACU-unifiers.

DEFINITION 3.4 *A reduction ordering $\succ$ is called* ACU*-compatible, if $s' =_{\text{ACU}} s \succ t =_{\text{ACU}} t'$ 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 [51], we can obtain an ACU-compatible reduction ordering for ground terms from an AC-compatible ordering:

LEMMA 3.5 *Let $\succ_1$ be an* AC*-compatible reduction ordering, such that 0 is minimal with respect to $\succ_1$. Let $\succ$ be the ordering 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 [51], 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.[14] This is a serious problem, if one is interested in classical rewriting. It is not a hindrance, though, for calculi like superposition or unfailing completion, which are preferably implemented using constraints. In fact, Jouannaud and Marché's method can be considered as a variant of unfailing completion with constraints.

DEFINITION 3.6 *We say that an* ACU*-compatible ordering has the multiset property, if whenever a ground atomic term $u \neq 0$ is greater than t_i for every i in a finite index 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,[15] and satisfies $t \not\succ s[t]_o$ for every term $s[t]_o$.[16] 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 3.5. On the other hand, polynomial orderings (Ben Cherifa and Lescanne [18]) are unsuited, since they violate the multiset property.

CONVENTION 3.7 *For the remainder of this paper, we will work* only *with* ACU*-congruence classes, rather than with terms. To simplify notation, we will omit the $[\_]_{\mathrm{ACU}}$ and drop the subscript of $=_{\mathrm{ACU}}$. So* all *terms, equations, substitutions, inference rules, etc., are to be taken modulo* ACU*, that is, as representatives of their congruence classes. Furthermore, we will use the equality predicate as a symmetric operator, thus ignoring the difference between $t \approx t'$ and $t' \approx t$.*

DEFINITION 3.8 *Let t be a ground term, then the maximal atomic subterm of t (with or without multiplicity) is defined in the following way:*

- *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 $\mathrm{mt}_{\#}(t) = nu$.*

[14] Jouannaud and Marché's statement that "AC1-rewrite orderings cannot really exist" [51] should be taken with a grain of salt, however.

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

[16] In a many-sorted framework like ours this property does not follow automatically from $\succ$ being total and noetherian. As an example consider $\mathcal{S} = \{S, S'\}$ and $\Sigma = \{b : \to S,\ f : S \to S'\}$ with the ordering $b \succ f(b)$.

- *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* $\mathrm{mt}_{\#}(e) = \max\{\mathrm{mt}_{\#}(t), \mathrm{mt}_{\#}(t')\}$.

DEFINITION 3.9 *The symbol* $\mathrm{ms}(t)$ *denotes the multiset of all non-zero atomic terms occurring at the top of a ground term t, i. e.,*

- $\mathrm{ms}(t) = \{ v_j \mid j \in J \}$, *if* $t = \sum_{j \in J} v_j$ *and all v_j are non-zero atomic terms. (In particular* $\mathrm{ms}(0) = \emptyset$, *as J may be empty.)*
- $\mathrm{ms}(t) = \{t\}$, *if t does not have sort* S_{CAM}.

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

DEFINITION 3.10 *The ordering $\succ$ on terms is extended to an ordering $\succ_{\mathrm{L}}$ on literals as follows: Every ground literal $[\neg]\ s \approx t$ is mapped to the quadruple*

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

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

DEFINITION 3.11 *For a non-empty ground clause C, the maximal literal of C with respect to $\succ_{\mathrm{L}}$ is denoted by* $\mathrm{ml}(C)$.

Our constraint language consists of all quantifier-free formulae built over true, false, $\wedge$, and the binary predicates $\succ$, $\succ_{\mathrm{L}}$, $\succ_{\mathrm{C}}$ (to be interpreted as the respective orderings), and $\doteq$ (to be interpreted as ACU-equality).

If N and N' are sets of constrained clauses, we write $N \models_{\Psi} N'$ if $N \cup \mathrm{ACUKT}_{\Psi} \models_{\approx} N'$, that is, if every equality model of N and ACUKT_{Ψ} is a model of N'. In other words, $\models_\Psi$ denotes entailment modulo ACUKT$_\Psi$ and equality. If C is a constrained clause, $N \models_\Psi C$ is a shorthand for $N \models_\Psi \{C\}$. For a set of ground equations E, its ACUKT$_\Psi$-closure $\mathrm{cl}_\Psi(E)$ is the set of all ground equations e such that $E \models_\Psi \{e\}$.

In theorem proving without constraints, a prover has succeeded in proving an inconsistency once it has derived the single contradictory formula $\bot$. In constraint theorem proving, things are slightly more complicated. A constrained clause $\bot\,[\![T]\!]$ may have a model: if the constraint T is unsatisfiable, then $\bot\,[\![T]\!]$ is a tautology. Only if T is satisfiable, the constrained clause $\bot\,[\![T]\!]$ has the ground instance $\bot$ and is thus contradictory. Consequently, if we have derived $\bot$ with some constraint, the analysis of this constraint can no longer be delayed – we have to test whether it is satisfiable or not.

It is convenient to make a distinction between essential and non-essential constraints. Intuitively, a constraint is essential when it is necessary for the correctness of the calculus. A constraint is non-essential when we might just as well have derived the same formula without this constraint. Non-essential constraints can be considered as pure annotations that may be dropped ad libitum.

DEFINITION 3.12 *Let $N_0 \vdash N_1 \vdash N_2 \vdash \dots$ be a theorem proving derivation; let $C\,[\![T \wedge T']\!]$ be a constrained clause in N_k. Then the (sub-)constraint T' is called non-essential, if $N_0 \models_\Psi C\,[\![T]\!]$, and essential, otherwise.*

The inference rules of our calculus will have the property that all ordering constraints that they introduce during a derivation are non-essential. Hence when we encounter a clause $\bot\,[\![T \wedge T']\!]$ where T is an equality constraint and T' is an ordering constraint, we may replace this clause by $\bot\,[\![T]\!]$. It is therefore sufficient to test the equality constraint T for satisfiability, rather than the mixed equality/ordering constraint $T \wedge T'$.

Testing the equality constraint $s \doteq t$ for satisfiability amounts to checking s and t for ACU-unifiability. We say that σ is an ACU-unifier of the constraint $s_1 \doteq t_1 \wedge \cdots \wedge s_n \doteq t_n$ if it is an ACU-unifier of s_i and t_i for every $i \in \{1, \dots, n\}$. We can extend the terminology of unification to equality constraints, speaking of complete sets of unifiers of a constraint. Note, however, that in constraint calculi it is often unnecessary to compute complete sets of unifiers – as long as we are only interested in the satisfiability of a constraint, we may stop as soon as we have found a single solution.

To every idempotent substitution σ with finite domain we can associate an equality constraint $\mathrm{EQ}(\sigma)$. Let $\sigma = \{x_1 \mapsto t_1, \dots, x_n \mapsto t_n\}$, then $\mathrm{EQ}(\sigma)$ is the constraint $x_1 \doteq t_1 \wedge \cdots \wedge x_n \doteq t_n$. Every substitution $\sigma\rho$ is a solution of $\mathrm{EQ}(\sigma)$; conversely, if θ is a solution of $\mathrm{EQ}(\sigma)$, then $\theta = \sigma\theta$.

Definition 3.13 *Let x be a variable occurring in a literal or (sub-)clause. We say that x is shielded in the literal or (sub-)clause, if it occurs at least once below a free function symbol. Otherwise, x is called unshielded.*

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

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

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

Definition 3.14 *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 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 Section 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 $+$ is the only non-constant function symbol and that $\Psi = \{1\}$ (i. e., T_Ψ is void). In a cancellative abelian monoid, the congruence law and the cancellation law are in a certain sense complementary. The congruence law states that adding equal terms on both sides of an equation preserves truth, and conversely, that dropping equal terms on both sides of an equation preserves falsity. The cancellation law states that dropping equal terms on both sides of an equation preserves truth, and that adding equal terms on both sides of an equation preserves falsity. Hence, if we have an equation $u + t \approx t'$ where the atomic term u is larger than t and 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 (possibly negated) 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 $m \geq n \geq 1$.[17]

If the equation $mu + s \approx s'$ occurs negatively in the second clause, 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[18]
$$\frac{C' \vee \neg\, 0 \approx 0}{C'}$$

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

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

[18] 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.

maximal (or selected) in their clauses[19] and u is atomic and strictly larger than s, s', t, and t'.

EXAMPLE 3.15 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

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

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

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

[19] Except for the literal $mu + t \approx t'$ in the *cancellative equality factoring* rule.

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 $\psi m = \chi n$, $n \geq 1$, $\psi \in \Psi$, $\chi \in \mathbf{N}^{>0}$.

It is fairly easy to see that we may restrict to values of ψ and χ such that $\gcd(\psi, \chi) = 1$. With this additional condition, 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.

The only further change that becomes necessary applies to the *cancellative equality factoring* rule.

Canc. Eq. Factoring
$$\frac{C' \vee nu + t \approx n'u + t' \vee mu + s \approx s'}{C' \vee \neg\, \psi t + \chi s' \approx \psi t' + \chi s \vee nu + t \approx n'u + t'}$$

where $\psi(n - n') = \chi m$, $n - n' \geq 1$, $\psi \in \Psi$, $\chi \in \mathbf{N}^{>0}$.

Again, the additional restriction that $\gcd(\psi, \chi) = 1$ ensures that there exists 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 that contain variables and possibly non-trivial constraints. In the standard superposition calculus, for lifting one simply needs to replace equality in the ground inference by equality constraints (or by unification). As long as all variables in our clauses are shielded, the situation is similar here: In a clause $C = C' \vee [\neg]\, e_1$, the maximal equation e_1 need no longer have the form $mu + s \approx s'$, where u is the unique maximal atomic term. 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 constraints that state that all u_k (and the corresponding terms v_l from the other premise) are equal are added to the previously existing constraints of the premises.

For instance, the *positive cancellative superposition* rule has now the following form:

$$\textit{Pos. Canc. Superposition} \qquad \frac{D' \lor e_2 \; [\![T_2]\!] \qquad C' \lor e_1 \; [\![T_1]\!]}{D' \lor C' \lor e_0 \; [\![T_2 \land T_1 \land T_{\mathrm{E}} \land T_{\mathrm{O}}]\!]}$$

where

- $e_1 \;=\; \sum_{k\in K} m_k^* u_k + s \approx s'$.
- $e_2 \;=\; \sum_{l\in L} n_l^* v_l + t \approx t'$.
- $m \;=\; \sum_{k\in K} m_k^* \;\geq\; n \;=\; \sum_{l\in L} n_l^*$.
- u is one of the u_k or v_l $(k \in K,\, l \in L)$.
- $e_0 \;=\; (m-n)u + s + t' \approx t + s'$.
- $T_{\mathrm{E}} \;=\; \bigwedge_{k\in K} u_k \doteq u \,\land\, \bigwedge_{l\in L} v_l \doteq u$.
- $T_{\mathrm{O}} \;=\; u \succ s \,\land\, u \succ s' \,\land\, u \succ t \,\land\, u \succ t'$.

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, i. e., equality resolution, standard superposition, and standard equality factoring. But this is not sufficient, as shown by the following example.

EXAMPLE 3.16 Suppose that the ordering on constant symbols is given by $b \succ b' \succ c \succ d \succ d'$. In every ACUKT_Ψ-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. 3.15 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 the empty clause from the clauses (1), (2), (3), and (10).

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

Neither is it possible to derive the empty clause 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 [47, 84] have solved this problem by introducing the following inference rule:

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

In the example above, this rule allows to derive $\underline{4b} \approx 2b'$ from the first three clauses, 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 the term $4d$ is not maximal in (1), the rule can be only 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).[20] 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.

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

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

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 and has a non-trivial constraint, 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. 3.16 (assuming $2 \in \Psi$) can be refuted as follows:

EXAMPLE 3.17 Abstraction of (1) and (11) yields

$$y \not\approx 2b \vee \underline{f(y)} \not\approx \underline{f(b')} \; [\![2b \succ y]\!] \tag{12}$$

By (non-ground) cancellation of (12) with the equality constraint $f(y) \doteq f(b')$ we obtain

$$y \not\approx \underline{2b} \vee 0 \not\approx 0 \; [\![2b \succ y \wedge f(y) \doteq f(b')]\!] \tag{13}$$

At this point, y has become unshielded. We can either use the inference rules of the calculus for unshielded variables, to be presented in the sequel, or we can eliminate the unshielded variable by partially solving the constraint and applying the solution (the substitution $\{y \mapsto b'\}$) to the rest of the clause. In the latter case, we obtain

$$b' \not\approx \underline{2b} \vee 0 \not\approx 0 \; [\![2b \succ b']\!] \tag{14}$$

Cancellative superposition of (1) and (14) and simplification of the constraint yields

$$c + \underline{2b'} \not\approx 4d \vee 0 \not\approx 0 \; [\![\text{true}]\!] \tag{15}$$

which can be refuted in the same way as clause (5) in Ex. 3.15.

The relationship between the coefficients m and n in the *abstraction* rule above is not completely obvious. Intuitively, an *abstraction* inference between clauses $D = D' \vee nu + t \approx t'$ and $C = C' \vee [\neg]\, s[mu + q] \approx s'$ is necessary, 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. 3.16, 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 here 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 the clauses $D = \underline{b} + c \approx d$ 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 an *abstraction* inference is superfluous.

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 it occurs at least once below a free function symbol, i. e., 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 atomic terms. Now $\mu\bar{u}$ may be involved in a ground *positive* *cancellative superposition* or similar 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 the variable 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, it is now no longer possible to *count* the number of occurrences of the maximal terms in the respective clauses in order to compute the 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 $x_1\theta$.[21] 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 $\mathrm{var}(u_1) \cup \mathrm{var}(u_2) \cup \{\hat{x}_1, \hat{y}_1, \hat{y}_2, z\}$, thus $k\bar{u}$ 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 Chapter 5,

[21] 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.

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 $\Psi = \mathbf{N}^{>0}$.

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 Chapter 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 $\mathbf{N}^{>0}$.

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.

If a literal of a clause is selected, then an inference must not involve non-selected literals of this clause. We use the symbol $T_{\mathrm{O}}^{\mathrm{Lit}}$ to denote the conjunction of the following ordering constraints: (i) 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*). (ii) If an inference involves a selected literal, then it must be maximal among the selected literals of this clause. (iii) A positive literal that is involved in a *superposition* or *abstraction* inference must be strictly maximal in the respective clause. (iv) In all *superposition* and *abstraction* inferences, the left premise is smaller than the right premise. (v) 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 3.18 The inference system *CS-Inf*$_\Psi$ 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

$$\frac{C' \vee [\neg]\, e_1 \; [\![T_1]\!]}{C' \vee [\neg]\, e_0 \; [\![T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- $e_1 \;=\; \sum_{i\in I} m_i x_i + \sum_{k\in K} m_k^* u_k + s \approx \sum_{i\in I'} m_i' x_i + \sum_{k\in K'} m_k^{*\prime} u_k + s'.$
- $e_0 \;=\; z\sigma + \sum_{i\in I} m_i \check{x}_i + s \approx \sum_{i\in I'} m_i' \check{x}_i + s'.$
- $I \cup K \neq \emptyset$ and $I' \cup K' \neq \emptyset$.
- If $[\neg]\, e_1$ is a positive literal:
  $\{\, x_i \mid i \in I \,\} = \mathrm{elig}(C' \vee e_1) \cap \mathrm{var}(\mathrm{lhs}(e_1)) \setminus \mathrm{var}(\mathrm{neg}(C'))$,
  $\{\, x_i \mid i \in I' \,\} = \mathrm{elig}(C' \vee e_1) \cap \mathrm{var}(\mathrm{rhs}(e_1)) \setminus \mathrm{var}(\mathrm{neg}(C'))$.
  Otherwise:
  $\{\, x_i \mid i \in I \,\} = \mathrm{elig}(C' \vee \neg e_1) \cap \mathrm{var}(\mathrm{lhs}(e_1))$,
  $\{\, x_i \mid i \in I' \,\} = \mathrm{elig}(C' \vee \neg e_1) \cap \mathrm{var}(\mathrm{rhs}(e_1))$.
- If $K \cup K' \neq \emptyset$, u is one of the u_k ($k \in K \cup K'$), otherwise, u is a new variable.
- σ is a most general ACU-unifier of $\sum_{i\in I} m_i \hat{x}_i + \big(\sum_{k\in K} m_k^*\big) u$ and $z + \sum_{i\in I'} m_i' \hat{x}_i + \big(\sum_{k\in K'} m_k^{*\prime}\big) u$.
- $T_{\mathrm{E}} = \bigwedge_{i\in I\cup I'} x_i \doteq \hat{x}_i + \check{x}_i \;\wedge\; \bigwedge_{k\in K\cup K'} u_k \doteq u \;\wedge\; \mathrm{EQ}(\sigma)$.
- $T_{\mathrm{O}} = u \succ s \;\wedge\; u \succ s' \;\wedge\; \bigwedge_{i\in I\cup I'} u \succ \check{x}_i \;\wedge\; T_{\mathrm{O}}^{\mathrm{Lit}}$.

Equality Resolution

$$\frac{C' \vee \neg\, u \approx u' \; [\![T_1]\!]}{C' \; [\![T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- Either $T_{\mathrm{E}} = u \doteq u' \doteq 0$ or u and u' do not have sort S_{CAM} and $T_{\mathrm{E}} = u \doteq u'$.
- $T_{\mathrm{O}} = T_{\mathrm{O}}^{\mathrm{Lit}}$.

Standard Superposition

$$\frac{D' \lor t \approx t' \; [\![T_2]\!] \qquad C' \lor [\neg]\, s[w] \approx s' \; [\![T_1]\!]}{D' \lor C' \lor [\neg]\, s[t'] \approx s' \; [\![T_2 \land T_1 \land T_{\mathrm{E}} \land T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- w is not a variable.
- If s has sort S_{CAM}, then w occurs below a free function symbol in s.
- $T_{\mathrm{E}} = t \doteq w$.
- $T_{\mathrm{O}} = s[w] \succ s' \land t \succ t' \land T_{\mathrm{O}}^{\mathrm{Lit}}$.

Negative Cancellative Superposition

$$\frac{D' \lor e_2 \; [\![T_2]\!] \qquad C' \lor \neg\, e_1 \; [\![T_1]\!]}{D' \lor C' \lor \neg\, e_0 \; [\![T_2 \land T_1 \land T_{\mathrm{E}} \land T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- $e_1 = \sum_{i \in I} m_i x_i + \sum_{k \in K} m_k^* u_k + s \approx s'$.
- $e_2 = \sum_{j \in J} n_j y_j + \sum_{l \in L} n_l^* v_l + t \approx t'$.
- $e_0 = \sum_{i \in I} \psi m_i \check{x}_i + \psi s + \chi t' \approx \sum_{j \in J} \chi n_j \check{y}_j + \chi t + \psi s'$.
- $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
- $\{ x_i \mid i \in I \} = \mathrm{elig}(C' \lor \neg e_1) \cap \mathrm{var}(\mathrm{lhs}(e_1)) \setminus \mathrm{var}(\mathrm{rhs}(e_1))$,
  $\{ y_j \mid j \in J \} = \mathrm{elig}(D' \lor e_2) \cap \mathrm{var}(\mathrm{lhs}(e_2)) \setminus \mathrm{var}(\mathrm{rhs}(e_2)) \setminus \mathrm{var}(\mathrm{neg}(D'))$.
- $\mathrm{lhs}(e_1)$ is not a variable (i. e., either $\sum_{i \in I} m_i > 1$ or $\sum_{k \in K} m_k^* u_k + s \neq 0$).
- If $I = \{i_1\}$, $m_{i_1} = 1$, and $K = \emptyset$, then $\mathrm{lhs}(e_2)$ is not an atomic term. If additionally $\Psi = \{1\}$, then $J \neq \emptyset$ or $t \neq 0$.
- $\psi \in \Psi$ and $\chi \in \mathbf{N}^{>0}$, such that $\gcd(\psi, \chi) = 1$.
- 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.
- $T_{\mathrm{E}} = \bigwedge_{i \in I} x_i \doteq \hat{x}_i + \check{x}_i \land \bigwedge_{j \in J} y_j \doteq \hat{y}_j + \check{y}_j \land \bigwedge_{k \in K} u_k \doteq u \land \bigwedge_{l \in L} v_l \doteq u \land \sum_{i \in I} \psi m_i \hat{x}_i + \sum_{k \in K} \psi m_k^* u \doteq \sum_{j \in J} \chi n_j \hat{y}_j + \sum_{l \in L} \chi n_l^* u$.
- $T_{\mathrm{O}} = u \succ s \land u \succ s' \land u \succ t \land u \succ t' \land \bigwedge_{i \in I} u \succ \check{x}_i \land \bigwedge_{j \in J} u \succ \check{y}_j \land T_{\mathrm{O}}^{\mathrm{Lit}}$.

Positive Cancellative Superposition

$$\frac{D' \vee e_2 \; [\![T_2]\!] \qquad C' \vee e_1 \; [\![T_1]\!]}{D' \vee C' \vee e_0 \; [\![T_2 \wedge T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- $e_1 \;=\; \sum_{i \in I} m_i x_i + \sum_{k \in K} m_k^* u_k + s \approx s'$.
- $e_2 \;=\; \sum_{j \in J} n_j y_j + \sum_{l \in L} n_l^* v_l + t \approx t'$.
- $e_0 \;=\; z\sigma + \sum_{i \in I} m_i \check{x}_i + s + t' \approx \sum_{j \in J} n_j \check{y}_j + t + s'$.
- $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
- $\{\, x_i \mid i \in I \,\} = \mathrm{elig}(C' \vee e_1) \cap \mathrm{var}(\mathrm{lhs}(e_1)) \setminus \mathrm{var}(\mathrm{rhs}(e_1)) \setminus \mathrm{var}(\mathrm{neg}(C'))$, $\{\, y_j \mid j \in J \,\} = \mathrm{elig}(D' \vee e_2) \cap \mathrm{var}(\mathrm{lhs}(e_2)) \setminus \mathrm{var}(\mathrm{rhs}(e_2)) \setminus \mathrm{var}(\mathrm{neg}(D'))$.
- 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.
- σ is a most general ACU-unifier of $\sum_{i \in I} m_i \hat{x}_i + \big(\sum_{k \in K} m_k^*\big) u$ and $z + \sum_{j \in J} n_j \hat{y}_j + \big(\sum_{l \in L} n_l^*\big) u$.
- $T_{\mathrm{E}} = \bigwedge_{i \in I} x_i \doteq \hat{x}_i + \check{x}_i \wedge \bigwedge_{j \in J} y_j \doteq \hat{y}_j + \check{y}_j \wedge \bigwedge_{k \in K} u_k \doteq u \wedge \bigwedge_{l \in L} v_l \doteq u \wedge \mathrm{EQ}(\sigma)$.
- $T_{\mathrm{O}} = u \succ s \wedge u \succ s' \wedge u \succ t \wedge u \succ t' \wedge \bigwedge_{i \in I} u \succ \check{x}_i \wedge \bigwedge_{j \in J} u \succ \check{y}_j \wedge T_{\mathrm{O}}^{\mathrm{Lit}}$.

Abstraction

$$\frac{D' \vee e_2 \; [\![T_2]\!] \qquad C' \vee [\neg]\, s[w] \approx s' \; [\![T_1]\!]}{C' \vee \neg\, y \approx w \vee [\neg]\, s[y] \approx s' \; [\![T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- $w \;=\; \sum_{i \in I} m_i x_i + \sum_{k \in K} m_k^* u_k + q$.
- $e_2 \;=\; \sum_{j \in J} n_j y_j + \sum_{l \in L} n_l^* v_l + t \approx t'$.
- $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.

- None of the variables x_i occurs in the non-variable terms u_k or at the top of q.

- $\{ y_j \mid j \in J \} = \mathrm{elig}(D' \vee e_2) \cap \mathrm{var}(\mathrm{lhs}(e_2)) \setminus \mathrm{var}(\mathrm{rhs}(e_2)) \setminus \mathrm{var}(\mathrm{neg}(D'))$.

- w occurs in s immediately below some free function symbol.

- $\psi \in \Psi$ and $\chi \in \mathbf{N}^{>0}$, such that $\gcd(\psi, \chi) = 1$.

- 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.

- $T_{\mathrm{E}} = \bigwedge_{i \in I} x_i \doteq \hat{x}_i + \check{x}_i \wedge \bigwedge_{j \in J} y_j \doteq \hat{y}_j + \check{y}_j \wedge \bigwedge_{k \in K} u_k \doteq u \wedge \bigwedge_{l \in L} v_l \doteq u \wedge \sum_{i \in I} \psi m_i \hat{x}_i + \sum_{k \in K} \psi m_k^* u \doteq \psi z + \sum_{j \in J} \chi n_j \hat{y}_j + \sum_{l \in L} \chi n_l^* u$.

- 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.

- There exists a substitution ρ such that $T_{\mathrm{E}}\rho = \mathrm{true}$ and $\mathrm{lhs}(e_2)\rho$ is not a subterm of $w\rho$.

- $T_{\mathrm{O}} = w \succ y \wedge s[w] \succ s' \wedge u \succ t \wedge u \succ t' \wedge T_{\mathrm{O}}^{\mathrm{Lit}}$.

Standard Equality Factoring

$$\frac{C' \vee v \approx v' \vee u \approx u' \; [\![T_1]\!]}{C' \vee \neg\, u' \approx v' \vee v \approx v' \; [\![T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- u, u', v, and v' do not have sort S_{CAM}.

- $T_{\mathrm{E}} = u \doteq v$.

- $T_{\mathrm{O}} = u \succ u' \wedge u \succ v'$.

Cancellative Equality Factoring

$$\frac{C' \vee e_2 \vee e_1 \; [\![T_1]\!]}{C' \vee \neg\, e_0 \vee e_2 \; [\![T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

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

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

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

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

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

- $\psi \in \Psi$ and $\chi \in \mathbf{N}^{>0}$, such that $\gcd(\psi, \chi) = 1$.

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

- $T_{\mathrm{E}} \;=\; \bigwedge_{i\in I\cup J\cup J'} x_i \doteq \hat{x}_i + \check{x}_i \;\wedge\; \bigwedge_{k\in K} u_k \doteq u \;\wedge\; \bigwedge_{l\in L\cup L'} v_l \doteq u \;\wedge\; \sum_{i\in I} \chi m_i \hat{x}_i \;+\; \sum_{j\in J'} \psi n_j' \hat{x}_j \;+\; \big(\sum_{k\in K} \chi m_k^* \;+\; \sum_{l\in L'} \psi n_l^{*\prime}\big) u \;\doteq\; \sum_{j\in J} \psi n_j \hat{x}_j + \big(\sum_{l\in L} \psi n_l^*\big) u$.

- $T_{\mathrm{O}} = u \succ s \;\wedge\; u \succ s' \;\wedge\; u \succ t \;\wedge\; u \succ t' \;\wedge\; \bigwedge_{i\in I\cup J\cup J'} u \succ \check{x}_i \;\wedge\; T_{\mathrm{O}}^{\mathrm{Lit}}$.

THEOREM 3.19 *The inference rules of the cancellative superposition calculus are sound with respect to* ACUKT_Ψ, *i.e., for every inference rule*

$$\frac{C_k \;[\![T_k]\!] \;\;\ldots\;\; C_1 \;[\![T_1]\!]}{C_0 \;[\![T_0]\!]}$$

we have $\{C_k \,[\![T_k]\!], \ldots, C_1 \,[\![T_1]\!]\} \models_\Psi C_0 \,[\![T_0]\!]$.

PROOF. By routine computation. □

The inference rules remain sound if we ignore the ordering constraints of the premises and the conclusion. For every inference rule

$$\frac{C_k \;[\![T_k \wedge T_k']\!] \;\;\ldots\;\; C_1 \;[\![T_1 \wedge T_1']\!]}{C_0 \;[\![T_0 \wedge T_0']\!]}$$

where $T_0, \ldots, T_k$ are equality constraints and $T_0', \ldots, T_k'$ are ordering constraints, $\{C_k \,[\![T_k]\!], \ldots, C_1 \,[\![T_1]\!]\} \models_\Psi C_0 \,[\![T_0]\!]$ holds. In other words, if the ordering constraints of the premises are non-essential in a given theorem proving derivation, then so are the ordering constraints of the conclusion. For the *abstraction* rule, an even stronger result can be obtained: It remains sound, even if all its conditions are ignored.

4 Refutational Completeness

4.1 Ideas and Concepts

In the previous chapter 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 is consistent, i. e., has a model. If the formulae in question are constrained clauses, then contradictory formulae have the form $\bot \llbracket T \rrbracket$, where T is satisfiable. In other words, contradictory formulae are constrained clauses that possess the empty clause as a ground instance.

It is obvious that a set N of clauses does not have a model if the empty clause is among the ground instances of clauses in N. Our task is to show the reverse: Whenever N is saturated and does not contain a contradictory formula, then we will construct a model for N. The essential idea is due to Bachmair and Ganzinger [12]. Let N be saturated and let $\overline{N}$ be the set of all ground instances of clauses in N. 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 the clause 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 an atom 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. Here the interpretations are the sets of all equations $t \approx t'$, such that the terms t and t' can be rewritten to a common term t'', using the previously collected maximal literals as rewrite rules. A clause $C' \vee e$ may be productive only if the left-hand (i. e., 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 the left-hand side of 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 [12],[22] a ground clause C_0 is redundant with respect to a set of clauses N, if there are ground instances $D_1, \ldots, D_n$ of clauses in 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 with respect to a set of clauses N, if there are ground instances $D_1, \ldots, D_n$ of clauses in N such that the 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.

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

[22] Note that "redundancy" is called "compositeness" in [12]. In later papers the standard terminology has changed.

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 are copied 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.

EXAMPLE 4.1 Consider the set of ground clauses (1)–(5):

$$\underline{b'} \approx c' \tag{1}$$

$$\underline{b} \approx c \tag{2}$$

$$\underline{f(c')} \not\approx d \tag{3}$$

$$g(b, c') \neq g(c, b') \vee \underline{f(c)} \approx c' \tag{4}$$

$$\underline{f(b)} \approx b' \tag{5}$$

If the ordering on terms is the lexicographic path ordering induced by the precedence $f \succ g \succ b \succ b' \succ c \succ c' \succ d$, then these clauses are ordered by (1) $\prec_C$ (2) $\prec_C$ (3) $\prec_C$ (4) $\prec_C$ (5).

We start the model construction with the empty set of rewrite rules. Clause (1) is false in the empty interpretation, its only (thus maximal) literal is positive, and b' is irreducible with respect to $\emptyset$. Hence we add the rewrite rule $b' \rightarrow c'$.

Clause (2) is again false, b is irreducible with respect to $\{b' \rightarrow c'\}$, hence the clause is productive and we add the rewrite rule $b \rightarrow c$.

Clause (3) is true in the current interpretation, so the interpretation is left unchanged.

Clause (4) is false in the current interpretation, as both $g(b, c')$ and $g(c, b')$ can be rewritten to $g(c, c')$ using the set of rewrite rules $\{b' \rightarrow c',\ b \rightarrow c\}$, We add the maximal literal as a rewrite rule $f(c) \rightarrow c'$.

Finally we consider clause (5). As $f(b) \rightarrow f(c) \rightarrow c' \leftarrow b'$, this clause is true in the current interpretation.

It is easy to verify that the set of three rewrite rules $\{b' \rightarrow c',\ b \rightarrow c,\ f(c) \rightarrow c'\}$ generates a model of all clauses (1)–(5). In fact, this set of clauses is saturated: It allows only one inference, namely a *standard superposition* inference from (2) and (5); this inference is redundant because its conclusion $f(c) \approx b'$ follows from (1), (2), and (4), all these clauses being smaller than clause (5), the maximal premise.

Let us now see what happens if we replace clause (5) by

$$\underline{f(b)} \neq b' \tag{6}$$

The ordering on clauses is (1) $\prec_C$ (2) $\prec_C$ (3) $\prec_C$ (4) $\prec_C$ (6), so we start again by checking clauses (1) to (4), which produces the set of rewrite rules

$\{b' \to c',\, b \to c,\, f(c) \to c'\}$. When we inspect clause (6), it turns out that this clause is false in the current interpretation. However, it is now impossible to extend the rule set because the only literal of (6) is negative. Consequently, the set of clauses (1)–(4) and (6) can not be saturated. There must exist a non-redundant inference.

Which inference is this? Let us have a closer look at clause (6). It is false in the current interpretation because $f(b) \to f(c) \to c' \leftarrow b'$. The first rewrite step in this derivation uses the rewrite rule $b \to c$, that has been produced by clause (2). Indeed it is the *standard superposition* inference from clause (2) and clause (6), which violates saturation.

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 the left and right-hand side 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*,

$$\begin{array}{ccc} t & \qquad & t' \\ {\scriptstyle *}\Big\downarrow & & \Big\downarrow{\scriptstyle *} \\ t'' & & t'' \end{array}$$

Rather we use a rewrite relation on *equations*:

$$\begin{array}{c} t \approx t' \\ \Big\downarrow{\scriptstyle *} \\ t'' \approx t'' \end{array}$$

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 an equation $mu + t \approx t'$ 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.

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, additional problems arise.

EXAMPLE 4.2 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 $0 \approx 0$. This is possible by first applying the first rule, yielding $d \approx 2b' + c$, then by applying the second rule, yielding $d + c \approx d + c$, and finally by cancelling d and c, yielding $0 \approx 0$.

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 $0 \approx 0$.

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)

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

ordered by (1) $\prec_{\mathrm{C}}$ (2) $\prec_{\mathrm{C}}$ (3) $\prec_{\mathrm{C}}$ (4).

We start the model construction with empty sets of primary and secondary rules. Clause (1) is false in the empty interpretation, so $\underline{2b'} + c \approx d$ is turned into 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 $b \approx b'$ becomes a secondary rule.[23]

Since $\underline{b} \approx b'$ is now a secondary rule, both $\underline{3b} \approx 3b'$ 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 small 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).[24]

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 the usual analysis of various kinds of critical pairs). As an easy corollary we obtain that the interpretations satisfy the equality axioms and the theory axioms ACUKT_{Ψ}. Finally, we can demonstrate that the limit interpretation is a model of the ground instances in $\overline{N}$ whenever N is saturated and does not contain the empty clause. 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, even with constrained 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,

[23] 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.

[24] In some theory superposition calculi, for example in AC-superposition (Wertz [100], Bachmair and Ganzinger [9]), 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.

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: We say that a non-ground clause or inference is redundant, if all its ground instances have this property.

It is important to note that 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 example, if $C_2 = b \approx c$, $C_1 = f(x,x) \approx x$, $\theta = \{x \mapsto b\}$, and $b \succ c$, then there is a *standard superposition* inference from the ground instances $C_2\theta = b \approx c$ and $C_1\theta = f(b,b) \approx b$. However, there is no inference from C_2 and C_1 themselves, since the *standard superposition* rule prohibits overlaps at (or below) a variable position of C_1. Similar restrictions exist for other inferences, for instance for *negative cancellative superposition* inferences. The purpose of the so-called lifting lemmas is to show that all ground inferences that are actually needed in the refutational completeness proof are in fact ground instances of inferences from clauses in N.

Handling Constraints. While theory axioms add only little to the difficulty of the redundancy concept, things become significantly more complicated when we turn to constraint superposition. Let us return to the example mentioned in the previous paragraph. If $C_2 = b \approx c$ and $C_1 = f(x,x) \approx x$ are clauses in N, then there is a *standard superposition* inference $\bar\iota$ from the ground instances $C_2\theta = b \approx c$ and $C_1\theta = f(b,b) \approx b$, namely

$$\frac{b \approx c \qquad f(b,b) \approx b}{f(b,c) \approx b}$$

However, there exists no inference from C_2 and C_1 themselves. Saturation means closure up to redundancy of non-ground inferences, or equivalently, redundancy of ground instances. The inference $\bar\iota$ is not a ground instance from an inference from clauses in N, hence it is not covered by saturation.

If we are working with unconstrained clauses, it turns out that this is not a problem. Let the equality interpretation E be a model of all ground instances of clauses in N that are smaller than $C_1\theta$. In the completeness proof, we have to show that E is also a model of the conclusion of $\bar\iota$. We can do this without resorting to saturation: Let $\theta' = \{x \mapsto c\}$, then E is in particular a model of $C_2\theta = b \approx c$ and $C_1\theta' = f(c,c) \approx c$. As the conclusion of $\bar\iota$ follows from these two clauses and the equality axioms, it is also true in E.

If we work with constrained clauses, this argument fails. If we replace the unconstrained clause C_1 with $C_1 \llbracket T \rrbracket$ where T is the equality constraint $x \doteq b$, then the original substitution θ satisfies T, but the new substitution θ' does not. As $C_1\theta'$ is not a ground instance of $C_1 \llbracket T \rrbracket$, it need not be true in the interpretation E.

The problem can be solved as follows. Let x be a variable of C and let E be an interpretation. We say that x is E-variable minimal[25] in $C\theta$, if there exists no term r' such that $x\theta \succ r'$ and $x\theta \approx r' \in E$.[26] An instance $C\theta$ is called E-variable minimal, if all variables of C are variable minimal in $C\theta$. The rules of the model construction are modified in such a way that an instance $C\theta$ is allowed to be productive only if it is variable minimal with respect to the current interpretation. In this way, instances like $C_1\theta = f(b,b) \approx b$ above, which might allow non-liftable inferences, are ignored.

For unshielded variables that occur only positively, the definition of variable minimality has to be slightly weakened. With the definition given above, it might happen that an instance $\underline{f(b)} + c \approx d \vee \underline{f(b)} + c \approx d'$ of a clause $x \approx d \vee x \approx d'$ is variable minimal with respect to the current interpretation and productive, but adding its maximal literal $\underline{f(b)} + c \approx d$ to the current interpretation destroys variable minimality. For such variables x, we may only require that the non-maximal subterms of $x\theta$, here b and c, are minimal in E.[27] This requirement is still necessary, though. Without it, it would be impossible to show that the conclusions of inferences from variable minimal ground instances are again variable minimal.

The limit interpretation E obtained this way is a model of the instances of clauses in N that are E-variable minimal. If we want to show that it is a model of all ground instances of clauses in N, we need an additional property, called model generalizability. We will show that, if a theorem proving derivation starts from a set of unconstrained clauses, then every set of clauses occurring in the derivation is model generalizable.

The definition of redundancy for superposition with constraints is changed accordingly. In the standard superposition calculus, a non-ground clause is redundant with respect to N, if every ground instance follows from smaller ground instances of clauses in N. In the constraint case, we have to demand that every *E-variable minimal* ground instance follows from smaller *E-variable*

[25] The traditional name for this property is "irreducible", rather than "variable minimal". We use the latter to avoid a clash of nomenclature and reserve the adjective "irreducible" to refer to the rewrite relation on equations introduced earlier.

[26] This explanation is slightly simplified. The exact definition refers to the set of rewrite rules R that induces the interpretation E, rather than to E itself.

[27] For variables that do not have sort S_{CAM} there is a stronger condition that refers to the size of rules used to rewrite $x\theta$.

minimal ground instances of clauses in N, and possibly from sufficiently small equations in E. Of course, we have the problem here that E, that is the limit interpretation computed during the model construction, is not known. For this reason, we must require that the above condition holds not only for the limit interpretation E, but for all possible results of our model construction. We refer to such interpretations, or more precisely to the rules that generate them, as stratified.

4.2 Redundancy

DEFINITION 4.4 *A set R of ground equations is called stratified, if for all ground equations e_0 and e_1 with $\mathrm{mt}(e_0) \succeq \mathrm{mt}(e_1)$ we have $R \models_\Psi e_1$ implies $R^{\prec_{\mathrm{L}} \neg e_0} \models_\Psi e_1$.*

LEMMA 4.5 *Let R be a stratified set of ground equations. Then we have for all ground equations e_0 and e_1: If $\mathrm{mt}(e_0) \succ \mathrm{mt}(e_1)$, then $R \models_\Psi e_1$ implies $R^{\prec_{\mathrm{L}} e_0} \models_\Psi e_1$.*

PROOF. As R is stratified, we know that $R \models_\Psi e_1$ implies $R^{\prec_{\mathrm{L}} \neg e_1} \models_\Psi e_1$. If $\mathrm{mt}(e_0) \succ \mathrm{mt}(e_1)$, then $e_0 \succ_{\mathrm{L}} \neg\, e_1$, hence $R^{\prec_{\mathrm{L}} \neg e_1} \subseteq R^{\prec_{\mathrm{L}} e_0}$. □

DEFINITION 4.6 *Let $C\theta$ be a ground instance of a clause $C\,[\![T]\!]$. Then $x \in \mathrm{var}(C)$ is called a positive maximal variable of C and θ, if x is unshielded in C, occurs only in positive literals, and $x\theta \succeq \mathrm{mt}(e\theta)$ for all equations e that contain x.*

DEFINITION 4.7 *Let R be a stratified set of ground equations, let $C\theta$ be a ground instance of a clause $C\,[\![T]\!]$, and let x be a variable in $\mathrm{var}(C)$. Then x is called R-variable minimal in $C\theta$, if*

- *there is no term r' such that $x\theta \succ r'$ and $x\theta \approx r' \in \mathrm{cl}_\Psi(R)$; or*
- *x is a positive maximal variable of C and θ; x does not have sort S_{CAM}; there is no subterm r of $x\theta$ such that $r \approx r' \in \mathrm{cl}_\Psi(R)$ and $\mathrm{mt}(x\theta) \succ r \succ r'$; there is no literal $x \approx r''$ of C and $x\theta \approx r' \in R$ such that $x\theta \succ r'$ and $r''\theta \succ r'$; or*
- *x is a positive maximal variable of C and θ; x has sort S_{CAM}; there is no subterm r of $x\theta$ such that $r \approx r' \in \mathrm{cl}_\Psi(R)$ and $\mathrm{mt}(x\theta) \succ r \succ r'$.*

The clause $C\theta$ is called an R-variable minimal ground instance, if all variables in $\mathrm{var}(C)$ are R-variable minimal in $C\theta$.

The set of all R-variable minimal ground instances of $C\,[\![T]\!]$ is denoted by $\mathrm{vm}_R(C\,[\![T]\!])$. If N is a set of constrained clauses, $\mathrm{vm}_R(N)$ is the union of all $\mathrm{vm}_R(C\,[\![T]\!])$ for $C\,[\![T]\!] \in N$. In particular, if $R = \emptyset$, then every ground instance of a clause is R-variable minimal,[28] *hence $\mathrm{vm}_\emptyset(N)$ equals $\overline{N}$, the set of all ground instances of clauses in N.*

DEFINITION 4.8 *Let N be a set of constrained clauses. We say that a constrained clause $C\,[\![T]\!]$ is ACUKT_Ψ-redundant with respect to N, if for every stratified set of ground equations R and every R-variable minimal ground instance $C\theta$, $R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C\theta$.*

To obtain a definition of redundancy for inferences, we need the concept of a ground instance of an inference. Here a slight complication is due to the fact that the conclusion of an *abstraction* inference is not ground, even if the premises are.

DEFINITION 4.9 *Let $C_0\,[\![T_0]\!], C_1\,[\![T_1]\!], \ldots, C_k\,[\![T_k]\!]$ be constrained clauses and let θ be a substitution such that $C_i\theta$ is ground and $T_i\theta = \mathrm{true}$ for all $i \in \{0, \ldots, k\}$. If there are inferences (other than abstraction inferences)*

$$\frac{C_k\,[\![T_k]\!] \;\ldots\; C_1\,[\![T_1]\!]}{C_0\,[\![T_0]\!]}$$

and

$$\frac{C_k\theta \;\ldots\; C_1\theta}{C_0\theta}$$

then the latter is called a ground instance of the former. It is called R-variable minimal, if additionally $C_i\theta$ is an R-variable minimal ground instance of $C_i\,[\![T]\!]$ for all $i \in \{0, \ldots, k\}$.

DEFINITION 4.10 *Let $C_0\,[\![T_0]\!], C_1\,[\![T_1]\!], C_2\,[\![T_2]\!]$ be constrained clauses and let θ be a substitution such that $C_1\theta$ and $C_2\theta$ are ground and $T_1\theta = T_2\theta = \mathrm{true}$. If there are abstraction inferences*

$$\frac{C_2\,[\![T_2]\!] \qquad C_1\,[\![T_1]\!]}{C_0\,[\![T_0]\!]}$$

and

$$\frac{C_2\theta \qquad C_1\theta}{C_0\theta\,[\![T_0\theta]\!]}$$

[28] Note that $\emptyset$ is stratified and that $\mathrm{cl}_\Psi(\emptyset)$ is the set of all ground equations $t \approx t$ (modulo ACU).

then the latter is called a ground instance of the former. It is called R-variable minimal, if additionally $C_i\theta$ is an R-variable minimal ground instance of C_i $[\![T]\!]$ for $i \in \{1, 2\}$.

Whenever we talk about instances of inferences, we assume that in the ground clauses exactly those literals are selected that correspond to selected literals in the non-ground clauses.

DEFINITION 4.11 *Let N be a set of constrained clauses. A cancellation, equality resolution, or equality factoring inference*

$$\frac{C\ [\![T_1]\!]}{C_0\ [\![T_0]\!]}$$

is called ACUKT$_\Psi$*-redundant with respect to N if for every stratified set of equations R and every R-variable minimal ground instance*

$$\frac{C\theta}{C_0\theta}$$

$R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C_0\theta$.

DEFINITION 4.12 *Let N be a set of constrained clauses. A superposition inference*

$$\frac{D' \vee e_2\ [\![T_2]\!] \qquad C\ [\![T_1]\!]}{C_0\ [\![T_0]\!]}$$

is called ACUKT$_\Psi$*-redundant with respect to N if for every stratified set of equations R and every R-variable minimal ground instance*

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

where $e_2\theta \in R$, $R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C_0\theta$.

DEFINITION 4.13 *Let N be a set of constrained clauses. An abstraction inference*

$$\frac{D' \vee e_2\ [\![T_2]\!] \qquad C\ [\![T_1]\!]}{C_0\ [\![T_0]\!]}$$

is called ACUKT$_\Psi$*-redundant with respect to N if for every stratified set of equations R, every R-variable minimal ground instance*

$$\frac{D'\theta \vee e_2\theta \qquad C\theta}{C_0\theta\ [\![T_0\theta]\!]}$$

where $e_2\theta \in R$, and every substitution ρ such that $C_0\theta\rho$ is an R-variable minimal ground instance of C_0 $[\![T_0]\!]$, $R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C_0\theta\rho$.

Definition 4.14 *Let N be a set of constrained clauses. The set of all clauses that are* ACUKT_Ψ*-redundant with respect to N is denoted by* $\mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N)$*. The set of all inferences that are* ACUKT_Ψ*-redundant with respect to N is denoted by* $\mathit{CS\text{-}Red}^{\mathrm{I}}_{\Psi}(N)$*.*

Lemma 4.15 *The pair* $\mathit{CS\text{-}Red}_\Psi = (\mathit{CS\text{-}Red}^{\mathrm{I}}_{\Psi}, \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi})$ *is a redundancy criterion with respect to the inference system* $\mathit{CS\text{-}Inf}_\Psi$ *and the consequence relation* $\models_\Psi$*.*

Proof. We have to show that $\mathit{CS\text{-}Red}_\Psi$ satisfies the conditions (i)–(iv) of Def. 2.28. Condition (ii) is obvious. To show condition (i), let N be a set of constrained clauses, let $C\,[\![T]\!]$ be some clause in $\mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N)$ and let $C\theta$ be an instance of $C\,[\![T]\!]$. We have to prove that $N \setminus \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N) \models_\Psi C\theta$. Let $R = \emptyset$. We know that $\mathrm{vm}_\emptyset(N)$ is the set of all ground instances of clauses in N. As $C\,[\![T]\!] \in \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N)$, we have $\mathrm{vm}_\emptyset(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C\theta$. By the compactness of first-order logic there exists a finite subset of $\mathrm{vm}_\emptyset(N)^{\prec_{\mathrm{C}} C\theta}$ that entails $C\theta$. Let $\overline{N}_0$ be the minimal finite subset of $\mathrm{vm}_\emptyset(N)^{\prec_{\mathrm{C}} C\theta}$ (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 $\mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N)$, then there would exist $D_1, \ldots, D_n \in \mathrm{vm}_\emptyset(N)^{\prec_{\mathrm{C}} D}$ such that $\{D_1, \ldots, D_n\} \models_\Psi D$ and $\overline{N}_0 \cup \{D_1, \ldots, D_n\} \setminus \{D\} \models_\Psi C\theta$. This is impossible, however, as it contradicts the minimality of $\overline{N}_0$. Thus $N \setminus \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N) \models_\Psi \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N)$.

Let N and N' be sets of constrained clauses such that $N' \subseteq \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N)$, let $C\,[\![T]\!]$ be a clause in $\mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N)$. To prove condition (iii), we have to show that $C\,[\![T]\!]$ is also ACUKT_Ψ-redundant with respect to $N \setminus N'$. Let R be a stratified set of ground equations and let $C\theta$ be an R-variable minimal ground instance of $C\,[\![T]\!]$. As $C\,[\![T]\!] \in \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N)$, we know that $R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C\theta$. By the compactness of first-order logic there exists a finite subset of $R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)}$ and a finite subset of $\mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta}$ that entail $C\theta$. Let $\overline{N}_0$ be the minimal finite subset of $\mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta}$ (with respect to the multiset extension of $\succ_{\mathrm{C}}$) such that $R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)} \cup \overline{N}_0 \models_\Psi C\theta$. If some clause D in $\overline{N}_0$ were a ground instance of a clause in $\mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N)$, then there would exist $D_1, \ldots, D_n \in \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} D}$ such that $R^{\prec_{\mathrm{L}} \mathrm{ml}(D)} \cup \{D_1, \ldots, D_n\} \models_\Psi D$ and thus $R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)} \cup \overline{N}_0 \cup \{D_1, \ldots, D_n\} \setminus \{D\} \models_\Psi C\theta$. This is impossible, however, as it contradicts the minimality of $\overline{N}_0$. Therefore, $R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N \setminus \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N))^{\prec_{\mathrm{C}} C\theta} \models_\Psi C\theta$, which implies that $C\,[\![T]\!]$ is ACUKT_Ψ-redundant with respect to $N \setminus \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N)$. The second part of condition (iii) is proved analogously.

It remains to prove that condition (iv) is satisfied. Let ι be an arbitrary inference whose conclusion $C_0\,[\![T_0]\!]$ is contained in a set of clauses N. We have to show that $\iota \in \mathit{CS\text{-}Red}^{\mathrm{I}}_{\Psi}(N)$. Let R be a stratified set of ground equations.

If ι is not an *abstraction* inference, let $\iota\theta$ be an R-variable minimal ground instance of ι with maximal premise $C\theta$ and conclusion $C_0\theta$. As $C_0\theta \prec_{\mathrm{C}} C\theta$ and as $C_0\theta$ is an R-variable minimal ground instance of $C_0 [\![T_0]\!] \in N$, we have $R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C_0\theta$, which proves that ι is ACUKT$_\Psi$-redundant. The proof for *abstraction* inferences is similar, using the fact that if $\iota\theta$ is a ground instance of an *abstraction* inference with maximal premise $C\theta$ and conclusion $C_0\theta [\![T_0\theta]\!]$ and if $T_0\theta\rho = \mathrm{true}$, then $C_0\theta\rho \prec_{\mathrm{C}} C\theta$. □

DEFINITION 4.16 *Let $C [\![T]\!]$ be a constrained clause. A variable $x \in \mathrm{var}(C)$ is called lower bounded, if there are substitutions θ and ρ mapping all variables of $C [\![T]\!]$ to ground terms such that $x\theta \prec x\rho$, $y\theta = y\rho$ for every variable y different from x, and $T\rho = \mathrm{true}$, $T\theta \neq \mathrm{true}$.*

DEFINITION 4.17 *A constrained clause $C [\![T]\!]$ is called lower bounded, if some variable $x \in \mathrm{var}(C)$ is lower bounded in $C [\![T]\!]$.*

DEFINITION 4.18 *We say that a set of constrained clauses N is model generalizable, if for every stratified set R of equations, $\mathrm{cl}_\Psi(R)$ is a model of $\mathrm{vm}_\emptyset(N)$ whenever $\mathrm{cl}_\Psi(R)$ is a model of $\mathrm{vm}_R(N)$.*

LEMMA 4.19 *If N does not contain a clause that is lower bounded, then N is model generalizable.*

PROOF. Suppose that R is a stratified set of equations such that $\mathrm{cl}_\Psi(R)$ is a model of $\mathrm{vm}_R(N)$. Consider an arbitrary ground instance $C\rho$ of a clause $C [\![T]\!]$ in N. Let θ be the substitution that maps every variable $x \in \mathrm{var}(C)$ to the minimal term t such that $x\rho \approx t \in \mathrm{cl}_\Psi(R)$ and every variable $x \in \mathrm{var}(T) \setminus \mathrm{var}(C)$ to $x\rho$. Since $C [\![T]\!]$ is not lower bounded and $x\theta \preceq x\rho$ for every $x \in \mathrm{var}(C)$, we have $T\theta = T\rho = \mathrm{true}$, so $C\theta$ is a ground instance of $C [\![T]\!]$. Moreover, $C\theta$ is R-variable minimal, hence $\mathrm{cl}_\Psi(R)$ is a model of $C\theta$. As $\mathrm{cl}_\Psi(R)$ is also a model of $x\rho = x\theta$ for every $x \in \mathrm{var}(C)$, it must be a model of $C\rho$. □

LEMMA 4.20 *If $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$ is a (finite or infinite) derivation of the cancellative superposition calculus and N_0 is model generalizable, then the limit N_∞ of the derivation is model generalizable.*

PROOF. Let R be a stratified set of equations such that $\mathrm{cl}_\Psi(R)$ is a model of $\mathrm{vm}_R(N_\infty)$. Let $C [\![T]\!]$ be an arbitrary clause in N_0 with a ground instance $C\theta$ in $\mathrm{vm}_R(N_0)$. As $N_0 \subseteq N_\infty \cup \mathit{CS\text{-}Red}^{\mathrm{C}}_\Psi(N_\infty)$, we distinguish between two cases. If $C [\![T]\!] \in N_\infty$, then $C\theta$ in $\mathrm{vm}_R(N_\infty)$, so it is true in $\mathrm{cl}_\Psi(R)$ by assumption. Otherwise $C [\![T]\!] \in \mathit{CS\text{-}Red}^{\mathrm{C}}_\Psi(N_\infty)$, so $R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N_\infty)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C\theta$. As $\mathrm{cl}_\Psi(R)$ is a model of R and $\mathrm{vm}_R(N_\infty)$, $C\theta$ is true in $\mathrm{cl}_\Psi(R)$. This proves that

$\mathrm{cl}_\Psi(R)$ is a model of $\mathrm{vm}_R(N_0)$. Since N_0 is model generalizable, $\mathrm{cl}_\Psi(R)$ is a model of $\mathrm{vm}_\emptyset(N_0)$. Now $N_0 \models_\Psi N_\infty$, hence $\mathrm{cl}_\Psi(R)$ is a model of $\mathrm{vm}_\emptyset(N_\infty)$. □

4.3 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 4.21 *Let R be a stratified set of equations. Let $D\,[\![T_2]\!]$ and $C\,[\![T_1]\!]$ be two constrained clauses (without common variables) and let θ be a substitution such that $T_2\theta = T_1\theta =$ true and $D\theta$ and $C\theta$ are ground R-variable minimal instances. If there is a positive cancellative superposition inference from $D\theta$ and $C\theta$, then the inference is an R-variable minimal ground instance of a positive cancellative superposition inference from $D\,[\![T_2]\!]$ and $C\,[\![T_1]\!]$.*

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

$$\begin{aligned} x_i\theta &= \mu_i\bar{u} + \bar{s}_i \ \text{ for } i \in I, \text{ where } \mu_i \in \mathbf{N} \text{ and } \bar{u} \succ \bar{s}_i,\\ u_k\theta &= \bar{u} \ \text{ for } k \in K,\\ \bar{m} &= \textstyle\sum_{i\in I} m_i\mu_i + \sum_{k\in K} m_k^*,\\ \bar{s} &= \textstyle\sum_{i\in I} m_i\bar{s}_i + s\theta,\\ \bar{s}' &= s'\theta, \end{aligned}$$

and analogously

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

such that

$$\begin{aligned}
y_j\theta &= \nu_j \bar{u} + \bar{t}_j \;\text{ for } j \in J, \text{ where } \nu_j \in \mathbf{N} \text{ and } \bar{u} \succ \bar{t}_j,\\
v_l\theta &= \bar{u} \;\text{ for } l \in L,\\
\bar{n} &= \textstyle\sum_{j\in J} n_j\nu_j + \sum_{l\in L} n_l^*,\\
\bar{t} &= \textstyle\sum_{j\in J} n_j\bar{t}_j + t\theta,\\
\bar{t}' &= t'\theta.
\end{aligned}$$

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 a substitution θ' by

$$\theta' = \theta \cup \{\, \hat{x}_i \mapsto \mu_i\bar{u},\, \check{x}_i \mapsto \bar{s}_i,\, \hat{y}_j \mapsto \nu_j\bar{u},\, \check{y}_j \mapsto \bar{t}_j,\, z \mapsto (\bar{m}-\bar{n})\bar{u} \mid i \in I,\, j \in J \,\}.$$

Let u be any of the u_k or v_l ($k \in K$, $l \in L$), or a new variable, if $K \cup L = \emptyset$. Obviously θ' is an ACU-unifier of $\sum_{i\in I} m_i\hat{x}_i + \big(\sum_{k\in K} m_k^*\big)u$ and $z + \sum_{j\in J} n_j\hat{y}_j + \big(\sum_{l\in L} n_l^*\big)u$, therefore there is an idempotent most general ACU-unifier σ and a substitution ρ such that $\theta' = \sigma\rho$ over $\mathrm{Dom}(\theta')$. Define $\theta'' = \sigma\rho \cup \{u \mapsto \bar{u}\}$, if $K \cup L = \emptyset$, and $\theta'' = \sigma\rho$, otherwise. There is a *positive cancellative superposition* inference ι

$$\frac{D' \vee e_2 \;[\![T_2]\!] \qquad C' \vee e_1 \;[\![T_1]\!]}{D' \vee C' \vee e_0 \;[\![T_2 \wedge T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

where the equation e_0 and the constraints T_{E} and T_{O} are defined by

$$\begin{aligned}
e_0 &= z\sigma + \sum_{i\in I} m_i\check{x}_i + s + t' \approx \sum_{j\in J} n_j\check{y}_j + t + s'\\
T_{\mathrm{E}} &= \textstyle\bigwedge_{i\in I} x_i \doteq \hat{x}_i + \check{x}_i \;\wedge\; \bigwedge_{j\in J} y_j \doteq \hat{y}_j + \check{y}_j\\
&\qquad \wedge\; \textstyle\bigwedge_{k\in K} u_k \doteq u \;\wedge\; \bigwedge_{l\in L} v_l \doteq u \;\wedge\; \mathrm{EQ}(\sigma)\\
T_{\mathrm{O}} &= u \succ s \;\wedge\; u \succ s' \;\wedge\; u \succ t \;\wedge\; u \succ t'\\
&\qquad \wedge\; \textstyle\bigwedge_{i\in I} u \succ \check{x}_i \;\wedge\; \bigwedge_{j\in J} u \succ \check{y}_j \;\wedge\; T_{\mathrm{O}}^{\mathrm{Lit}}.
\end{aligned}$$

As $\theta'' = \theta$ over $\mathrm{var}(C\,[\![T_1]\!]) \cup \mathrm{var}(D\,[\![T_2]\!])$, we have $C\theta'' = C\theta$, $D\theta'' = D\theta$, $T_1\theta'' = T_1\theta = \mathrm{true}$, and $T_2\theta'' = T_2\theta = \mathrm{true}$; furthermore, it is easy to check that $T_{\mathrm{E}}\theta''$ and $T_{\mathrm{O}}\theta''$ are true. As $e_0\theta'' = (\bar{m}-\bar{n})\bar{u} + \bar{s} + \bar{t}' \approx \bar{t} + \bar{s}'$, we see that $\bar{\iota} = \iota\theta''$ is a ground instance of ι.

It remains to show that $\bar{\iota}$ is an R-variable minimal ground instance. The premises are R-variable minimal by assumption. To prove that the conclusion C_0 of $\bar{\iota}$ is R-variable minimal, too, let us first consider a variable $x \in \mathrm{var}(e_0)$.

If x occurs in s, s', t, or t', then is shielded or occurs also in some negative literal of C' or D'. Hence it is neither positive maximal in C_0 nor in C (or D). As x is R-variable minimal in the latter, it must be R-variable minimal in the former.

If x is one of the $\check{x}_i$ (or one of the $\check{y}_j$, this is proved analogously), then $\check{x}_i\theta''$ is a subterm of $x_i\theta''$, furthermore $\bar{u} \succ \bar{s}_i = \check{x}_i\theta''$. Therefore, $\check{x}_i$ is R-variable minimal, no matter whether x_i is positive maximal in C or not.

If $x \in \mathrm{var}(z\sigma)$, then x occurs also in $\mathrm{var}(\hat{x}_i\sigma)$ for some $i \in I$, or in $\mathrm{var}(u\sigma)$, if K is nonempty. This is obvious from the fact that σ is an ACU-unifier of $\sum_{i\in I} m_i\hat{x}_i + \left(\sum_{k\in K} m_k^*\right)u$ and $z + \sum_{j\in J} n_j\hat{y}_j + \left(\sum_{l\in L} n_l^*\right)u$. If K is nonempty and $x \in \mathrm{var}(u\sigma)$, choose some variable $y \in \mathrm{var}(u)$ such that $x \in \mathrm{var}(y\sigma)$, Then $x\theta'' = x\rho$ is a subterm of $y\sigma\rho = y\theta$, and as y is R-variable minimal in $C\theta = C\theta''$, so is x in $C_0\theta''$. On the other hand, if $x \in \mathrm{var}(\hat{x}_i\sigma)$ for some $i \in I$, then $x\theta'' = x\rho$ is a subterm of $\hat{x}_i\sigma\rho = \hat{x}\theta'$ and thus a subterm of $x_i\theta$. If $\mathrm{mt}(x\theta'') \prec \bar{u}$, then there is no subterm r of $x\theta''$ such that $r \succ r'$ and $r \approx r' \in E$. Otherwise, if $\mathrm{mt}(x\theta'') = \bar{u}$, then x is positive maximal and there is no subterm r of $x\theta''$ such that $\mathrm{mt}(x\theta'') = \bar{u} \succ r \succ r'$ and $r \approx r' \in E$. In both cases, x is R-variable minimal in $C_0\theta''$.

It remains to consider a variable $x \in \mathrm{var}(C_0) \setminus \mathrm{var}(e_0)$. In this case, x occurs in the same literals in C_0 and in one of the premises; hence it is positive maximal in the latter if and only if it is in the former. Consequently, x must be R-variable minimal. □

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

LEMMA 4.22 *Let R be a stratified set of equations. Let $C\,[\![T_1]\!]$ be a constrained clause and let θ be a substitution such that $T_1\theta = \mathrm{true}$ and $C\theta$ is a ground R-variable minimal instance. Then every cancellation, equality resolution, standard equality factoring, or cancellative equality factoring inference from $C\theta$ is an R-variable minimal ground instance of an inference from $C\,[\![T_1]\!]$.*

LEMMA 4.23 *Let R be a stratified set of equations. Let $D\,[\![T_2]\!]$ and $C\,[\![T_1]\!]$ be two constrained clauses (without common variables) and let θ be a substitution such that $T_2\theta = T_1\theta = \mathrm{true}$ and $D\theta$ and $C\theta$ are ground R-variable minimal instances. Let $D = D' \vee e_2$ such that $e_2\theta$ is strictly maximal in $D\theta$ and $e_2\theta \in R$; let $C = C' \vee [\neg]\, e_1$.*

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 an R-variable minimal

ground instance of a negative cancellative superposition inference from $D\,[\![T_2]\!]$ *and* $C\,[\![T_1]\!]$.

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), then the inference is an R-variable minimal ground instance of a standard superposition inference from* $D\,[\![T_2]\!]$ *and* $C\,[\![T_1]\!]$.

If there is an abstraction inference

$$\frac{D'\theta \vee e_2\theta \qquad C'\theta \vee [\neg]\, e_1\theta[\bar{m}\bar{v} + \bar{q}]}{C_0\ [\![T_0]\!]}$$

such that

- C_0 *equals* $C'\theta \vee \neg\, y \approx \bar{m}\bar{v} + \bar{q} \vee [\neg]\, e_1\theta[y]$,
- $\bar{m}\bar{v} + \bar{q} = w\theta$ *for some subterm* w *of* $\mathrm{lhs}(e_1)$,
- $\bar{m}\bar{v} + \bar{q}$ *is not a subterm of* $y'\theta$ *for any* $y' \in \mathrm{var}(\mathrm{lhs}(e_1))$,
- *the maximal atomic subterm of* $\mathrm{lhs}(e_2\theta)$ *equals* $\bar{v}$,
- *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 an R-variable minimal ground instance of an abstraction inference from $D\,[\![T_2]\!]$ *and* $C\,[\![T_1]\!]$.

4.4 Rewriting on Equations

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

DEFINITION 4.24 *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.

For simplicity, 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 $s' + t \approx t' + s$.[29]

DEFINITION 4.25 *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:*

- $mu + t \approx t' \;\to_{\gamma,R}\; s' + t \approx t' + s$,

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

- $t[s] \approx t' \;\to_{\delta,R}\; t[s'] \approx t'$,

  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.

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

  if u is atomic and different from 0.

The union of $\to_{\gamma,R}$, $\to_{\delta,R}$, and $\to_{\kappa}$ is denoted by $\to_R$.[30]

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 4.26 *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. □

LEMMA 4.27 *The relation $\to_R$ is contained in $\succ_{\mathrm{L}}$ and thus noetherian.*

[29] 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.

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

DEFINITION 4.28 *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 $s \approx s' \rightarrow_R^* 0 \approx 0$. 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 4.29 *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

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

During this derivation, all occurrences of u are deleted eventually. As s and s' are smaller than u, 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. □

4.5 Model Construction

DEFINITION 4.30 *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 4.31 *Let N be a set of (possibly non-ground) constrained clauses that does not contain the empty clause with a satisfiable constraint. Let $\overline{N}$ be the set of all ground instances of clauses in N. Using induction on the clause ordering we define sets of rules R_C, R_C^Ψ, E_C, and E_C^Ψ, for all clauses $C \in \overline{N}$. Let C be such a clause and assume that R_D, R_D^Ψ, E_D, and E_D^Ψ have already been defined for all $D \in \overline{N}$ such that $C \succ_{\mathrm{C}} D$. Then the set R_C of primary rules and the set R_C^Ψ of secondary rules are given by*

$$R_C = \bigcup_{D \prec_{\mathrm{C}} C} E_D \quad \text{and} \quad R_C^\Psi = \bigcup_{D \prec_{\mathrm{C}} C} E_D^\Psi\,.$$

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

The set E_C^Ψ is non-empty only if $E_C = \{e\}$. In this case, E_C^Ψ 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.

Finally, the sets R_∞ and R_∞^Ψ are defined by

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

Our goal is to show that $\mathrm{tr}(R_\infty^\Psi)$ is a 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 Section 4.6 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 Section 4.7 that $\mathrm{tr}(R_\infty^\Psi)$ is in fact a model of N, provided that N is saturated and does not contain the empty clause.

LEMMA 4.32 *For every $C \in \overline{N}$, $R_\infty^{\prec_\mathrm{L}\mathrm{ml}(C)} \subseteq R_C \subseteq R_\infty^{\preceq_\mathrm{L}\mathrm{ml}(C)}$.*

PROOF. To prove the first inclusion, take any $e \in R_\infty^{\prec_\mathrm{L}\mathrm{ml}(C)}$. Then $e \in E_D$ for some clause $D \in \overline{N}$. Assume that $D \succeq_\mathrm{C} C$, then, by the definition of the clause ordering, $e = \mathrm{ml}(D) \succeq_\mathrm{L} \mathrm{ml}(C)$. This is impossible, hence $D \prec C$ and $e \in E_D \subseteq R_C$. The second inclusion is obvious. □

LEMMA 4.33 *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.

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

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

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

where $\textcircled{2}$ uses $mu + s \approx s'$ and $\textcircled{3}$ simulates $\textcircled{1}$. 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 4.34 *For every $C \in \overline{N}$ we have $E_C \cup E_C^\Psi \subseteq \mathrm{tr}(R_C^\Psi \cup E_C^\Psi) \subseteq \mathrm{tr}(R_\infty^\Psi)$ and $R_C \cup R_C^\Psi \subseteq \mathrm{tr}(R_C^\Psi) \subseteq \mathrm{tr}(R_\infty^\Psi)$.*

LEMMA 4.35 *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 $E_D^\Psi \subseteq R_C^\Psi$, due to Lemma 4.33. This is impossible, however. □

LEMMA 4.36 *Let u be an atomic term. 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 exists a rule $ku + r \approx r' \in E_C^\Psi$, where $k \leq m$. 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 is strictly larger than $k'u$ 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 $t \approx t'$. □

4.6 Confluence

It is easy to see that the relations $\to_{R_C^\Psi}$ and $\to_{R_\infty^\Psi}$ are in general not confluent:

EXAMPLE 4.37 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 $\to_{R_C^\Psi}$ 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.

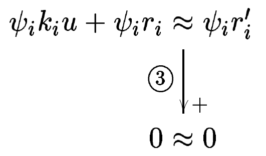
By definition, for every rewrite rule $ku + r \approx r'$ in some E_C^Ψ there is a $\psi \in \Psi$ and an $(R_C^\Psi \cup E_C)$-derivation from $\psi ku + \psi r \approx \psi r'$ to $0 \approx 0$. The purpose of the following lemma is twofold: First, it shows that we may enforce a particular structure upon this derivation. Second, 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 4.38 *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*

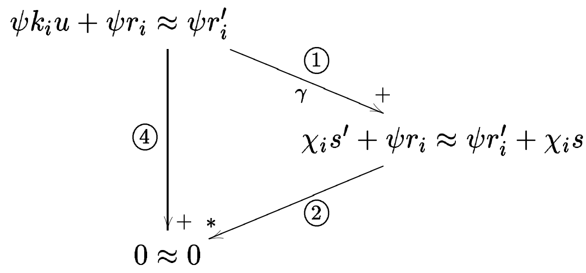
$$\begin{array}{c} \psi k_i u + \psi r_i \approx \psi r_i' \\ ①\ \big\downarrow {}^{\gamma}_{+} \\ \chi_i s' + \psi r_i \approx \psi r_i' + \chi_i s \\ ②\ \big\downarrow {}_{*} \\ 0 \approx 0 \end{array}$$

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. By definition of E_C^Ψ, for every $i \in I$ there exists a $\psi_i \in \Psi$ and an $(R_C^\Psi \cup E_C)$-derivation ③.



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.



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. $\square$

The next two lemmas state that every equation that is $\delta\kappa$-irreducible and contained in the Ψ-truth set of R_C^Ψ (or $R_C^\Psi \cup E_C$) is either $0 \approx 0$ or a rewrite rule in R_C^Ψ (or $R_C^\Psi \cup E_C^\Psi$), provided that sufficiently many peaks can be joined.

LEMMA 4.39 *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 $\rightarrow_{R_C^\Psi}$ is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e' \mid \mathrm{mt}(e) \succ \mathrm{mt}(e')\,\}$, then $e \in R_C^\Psi \cup \{0 \approx 0\}$. Similarly, if $e \in \mathrm{tr}_\Psi(R_\infty^\Psi)$ is $\delta\kappa$-irreducible with respect to R_∞^Ψ and $\rightarrow_{R_\infty^\Psi}$ is confluent on $\mathrm{tr}(R_\infty^\Psi) \cap \{\, e' \mid \mathrm{mt}(e) \succ \mathrm{mt}(e')\,\}$, then $e \in R_\infty^\Psi \cup \{0 \approx 0\}$.*

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 4.35 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

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

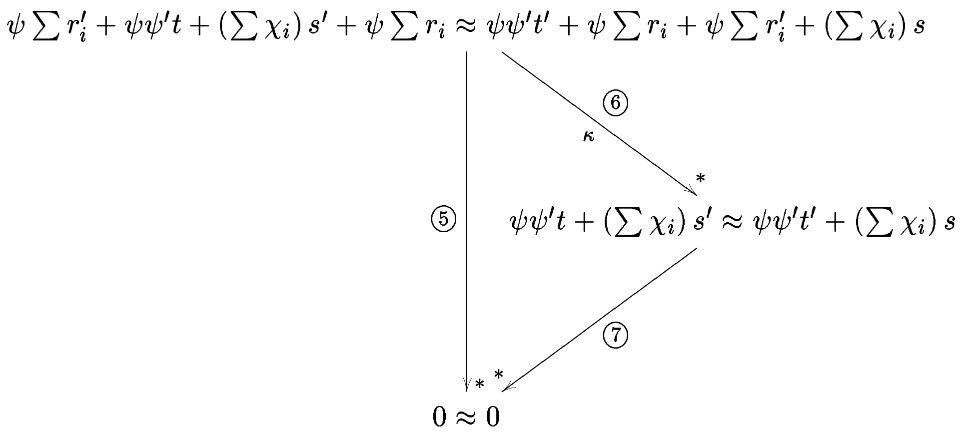
where the rewrite steps of ② use only rules from R_D^Ψ.

Let $E_D = \{nv + s \approx s'\}$. According to Lemma 4.38, there exists a $\psi \in \Psi$ and for every $i \in I$ an $(R_D^\Psi \cup E_D)$-derivation

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

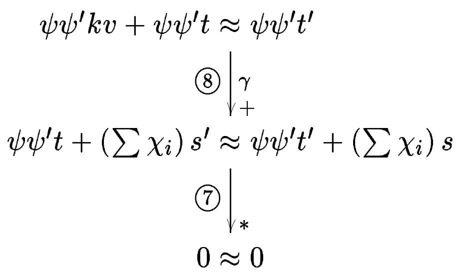
starting with χ_i-fold application of $nv + s \approx s'$ where $\psi k_i = \chi_i n$.

We will now 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 $\psi\,(\sum r_i' + \psi' t) + \sum(\chi_i s' + \psi r_i) \approx \psi\,(\psi' t' + \sum r_i) + \sum(\psi r_i' + \chi_i s)$.



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 it is possible to 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.



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. 4.31. □

Lemma 4.40 *If C is a clause 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 $\mathrm{tr}_\Psi(R^\Psi_C)$, then $e \in R^\Psi_C \cup \{0 \approx 0\}$ by Lemma 4.39. Otherwise, let $E_C = \{nv + s \approx s'\}$ and $e = ku + t \approx t'$, such that $u = \mathrm{mt}(e)$. By definition of $\mathrm{tr}_\Psi(R^\Psi_C \cup E_C)$, there is a derivation

$$\psi k u + \psi t \approx \psi t' \ \to^*_{R^\Psi_C \cup E_C} \ 0 \approx 0$$

for some $\psi \in \Psi$. During this derivation all occurrences of u are deleted eventually. If u were larger than v, then this would be impossible, as u is δ-irreducible with respect to $R^\Psi_C \cup E_C$. If u were smaller than v, then $nv + s \approx s'$ could not be used during this derivation, hence e would be contained in $\mathrm{tr}_\Psi(R^\Psi_C)$. Thus $u = v$, and by Def. 4.31, $e \in E^\Psi_C$. □

Intuitively, the following two lemmas show that the "difference" of two rewrite rules from R^Ψ_C is either $0 \approx 0$ or also a rewrite rule from R^Ψ_C, provided that sufficiently many peaks can be joined.

LEMMA 4.41 *Let $\{C, D, D_1\} \subseteq \overline{N}$, such that $C \succ_{\mathrm{C}} D \succeq_{\mathrm{C}} D_1$. Let $k_0 v + r_0 \approx r'_0 \in E^\Psi_D$, $k_1 v + r_1 \approx r'_1 \in E^\Psi_{D_1}$, where $k_0 > 0$ and $k_0 \geq k_1$.[31] 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 $\to_{R^\Psi_C}$ is confluent on $\mathrm{tr}(R^\Psi_C) \cap \{\, e' \mid v \succ \mathrm{mt}(e') \,\}$, then $(k_0 - k_1)v + (q_0 + q'_1) \approx (q'_0 + q_1) \in E^\Psi_D \cup R^\Psi_D \cup \{0 \approx 0\}$. (Analogously for C replaced by ∞.)*

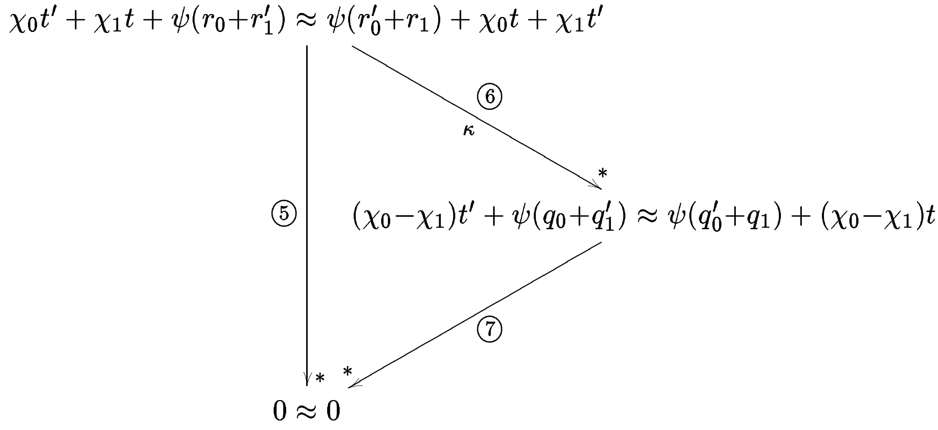
PROOF. Suppose that $E_D = \{nv + t \approx t'\}$. Then there are a $\psi \in \Psi$ and $(R^\Psi_D \cup E_D)$-derivations

$$\begin{array}{ccc}
\psi k_0 v + \psi r_0 \approx \psi r'_0 & \qquad & \psi r'_1 \approx \psi r_1 + \psi k_1 v \\
① \big\downarrow^{\gamma}_{+} & & {}^{\gamma}_{*}\big\downarrow ③ \\
\chi_0 t' + \psi r_0 \approx \psi r'_0 + \chi_0 t & & \chi_1 t + \psi r'_1 \approx \psi r_1 + \chi_1 t' \\
② \big\downarrow_{*} & & {}_{*}\big\downarrow ④ \\
0 \approx 0 & & 0 \approx 0
\end{array}$$

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 4.38; if $D_1 \prec_{\mathrm{C}} D$, it follows from Lemma 4.38 and Lemma 4.34.

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.

[31] 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.



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.

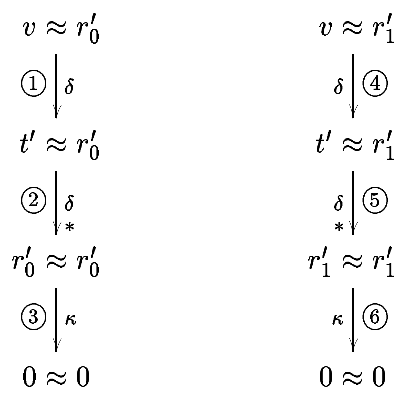
Since $\psi(k_0 - k_1) = n(\chi_0 - \chi_1)$, it is possible to rewrite $\psi(k_0{-}k_1)v + \psi(q_0{+}q_1') \approx \psi(q_0'{+}q_1)$ to the starting equation of ⑦ by $(\chi_0{-}\chi_1)$-fold application of $nv + t \approx t' \in E_D$ ⑧.

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

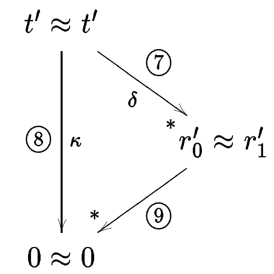
As $k_0 v + r_0 \approx r_0'$ and $k_1 v + r_1 \approx r_1'$ are $\delta\kappa$-irreducible with respect to $R_D^\Psi \cup E_D$, so is $(k_0{-}k_1)v + (q_0{+}q_1') \approx (q_0'{+}q_1)$. By Lemma 4.40, it is contained in $R_D^\Psi \cup E_D^\Psi \cup \{0 \approx 0\}$. □

LEMMA 4.42 *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 $\to_{R_C^\Psi}$ 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



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.

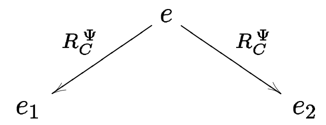


On the other hand, we can rewrite $t' \approx t'$ immediately to $0 \approx 0$ ⑧. By confluence, there is a derivation ⑨. As r'_0 and r'_1 are δ-irreducible and do not have sort S_{CAM}, ⑨ must consist of a single κ-step, hence $r'_0 = r'_1$. □

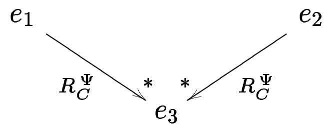
THEOREM 4.43 *The relation* $\to_{R_C^\Psi}$ *is confluent on* $\mathrm{tr}(R_C^\Psi)$ *for every* $C \in \overline{N}$. *The relation* $\to_{R_\infty^\Psi}$ *is confluent on* $\mathrm{tr}(R_\infty^\Psi)$.

PROOF. Let us consider the relation $\to_{R_C^\Psi}$. (The case of $\to_{R_\infty^\Psi}$ is similar.) The traditional way to establish the confluence of a noetherian relation proceeds 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. The situation is similar here, with one important exception: We need the induction hypothesis not only 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 the relation $\to_{R_C^\Psi}$ is confluent on $\mathrm{tr}(R_C^\Psi)$ it is sufficient to show that it is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e \mid e_0 \succeq_{\mathrm{L}} e \,\}$ 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 $\succ_{\mathrm{L}}$. According to Lemma 2.26, we have to prove that for any peak



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



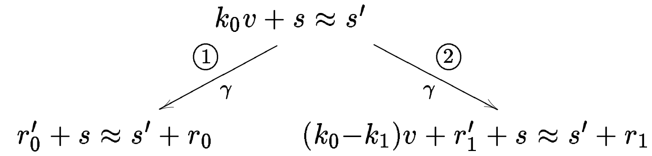
For $e_0 \succ_{\mathrm{L}} e$, this follows immediately from the induction hypothesis, so we may assume that $e_0 = e$.

Case 1: Trivial peaks.

As in the traditional term rewriting framework, every peak converges if the two rewrite steps take place at disjoint redexes. Furthermore, local confluence is obvious, if one step is a κ-step, and the other one is a δ- or a κ-step. Finally, Lemma 4.36 shows that 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.

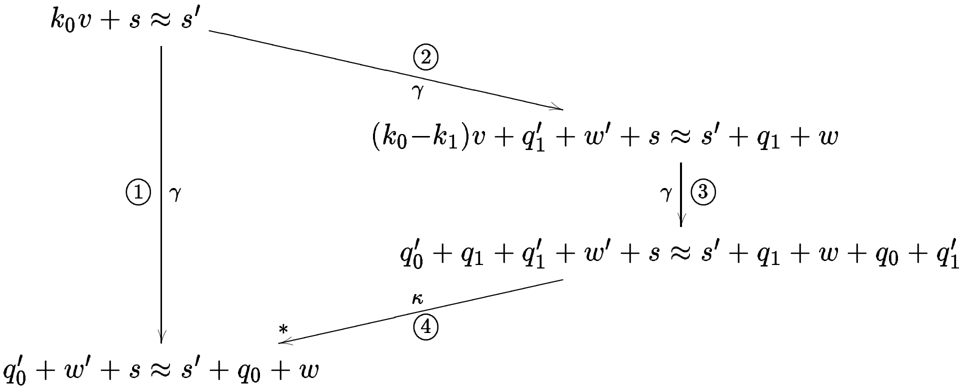
Case 2: γ/γ-peaks.

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

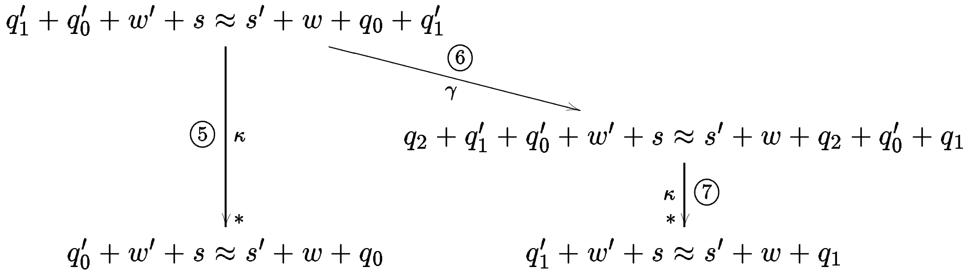


Let w be the common part of r_0 and r_1, let w' be the common part of r_0' and r_1', and for $i \in \{0, 1\}$, let $r_i = w + q_i$ and $r_i' = w' + q_i'$. By the induction hypothesis, $\rightarrow_{R_C^\Psi}$ is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e \mid e' \succeq_{\mathrm{L}} e \,\}$ for every e' that is smaller than $k_0 v + s \approx s'$; therefore, it is confluent on $\mathrm{tr}(R_C^\Psi) \cap \{\, e \mid v \succ \mathrm{mt}(e) \,\}$. We can thus apply Lemma 4.41, 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.



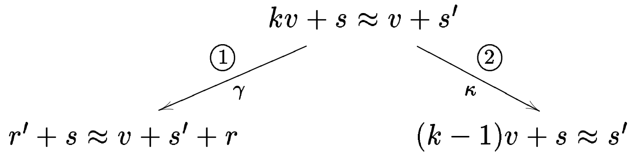
Otherwise, $k_0 = k_1$. Let $mu = \mathrm{mt}_{\#}(q_0 + q_1' \approx q_0' + q_1)$. Without loss of generality assume that mu occurs on the left-hand side of this equation (the other case is proved analogously), hence let $q_0 + q_1' = mu + q_2$. Consider the equation $q_1' + q_0' + w' + s \approx s' + w + q_0 + q_1'$. 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.



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 situation where a rewrite rule $kv + r \approx r' \in E_D^\Psi \subseteq R_C^\Psi$ with $k \geq 2$ is applied at the top of an equation $kv + s \approx v + s'$. This yields a peak



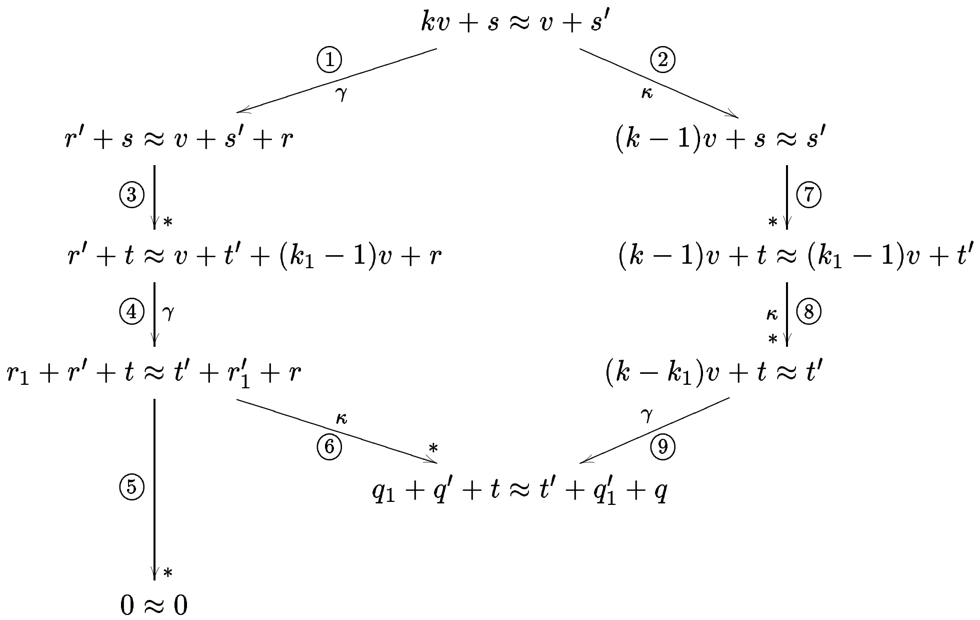
where either $r' + s \approx v + s' + r$ or $(k-1)v + s \approx s'$ can be rewritten to $0 \approx 0$ by R_C^Ψ.

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 4.36 v is δ-irreducible, so this can happen only by a γ-step or a κ-step.

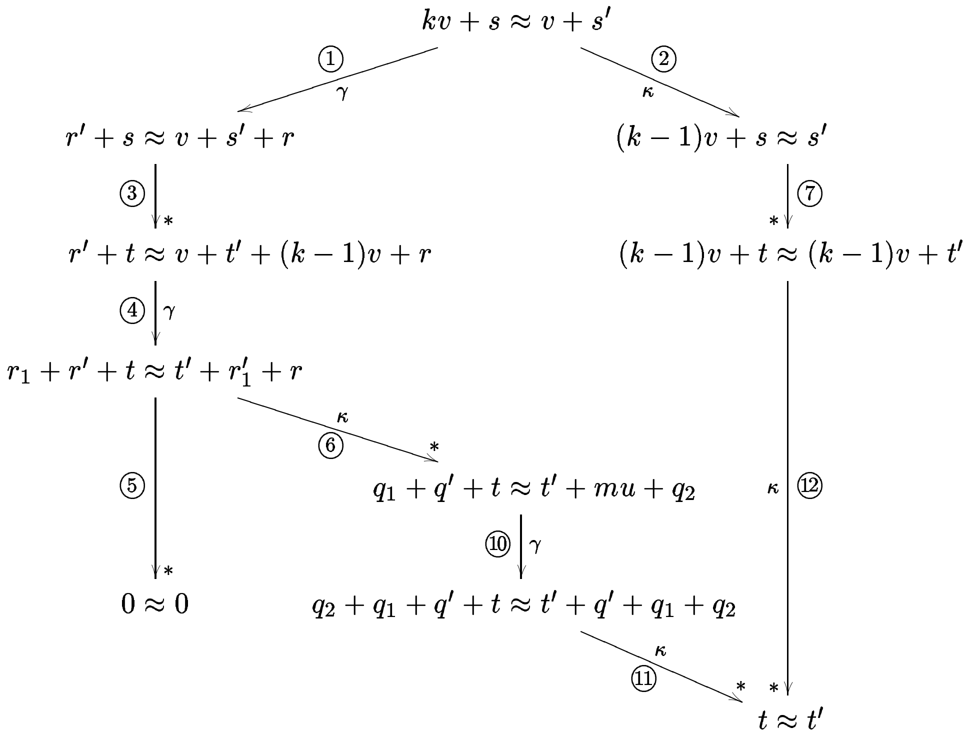
Case 3.1.1: v is deleted by a γ-step.

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:



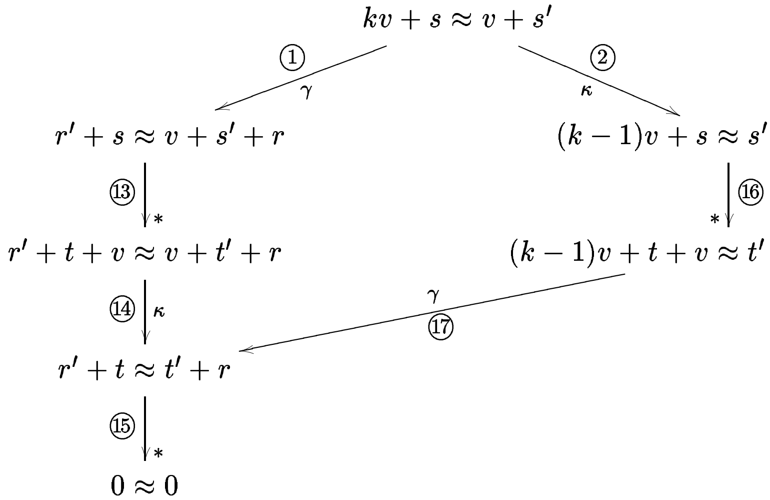
Let w be the common part of r and r_1, let w' be the common part of r' and r_1', and let $r = w + q$, $r_1 = w + q_1$, $r' = w' + q'$, and $r_1' = w' + q_1'$. We can thus use κ-steps ⑥ to cancel $w + w'$ in $r_1 + r' + t \approx t' + r_1' + r$, obtaining $q_1 + q' + t \approx t' + q_1' + q$. As the steps 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 4.41, $(k - k_1)v + (q + q_1') \approx (q' + q_1)$ is contained in E_D^Ψ; γ-application ⑨ of this rule closes the diagram.

If $k = k_1$, then by Lemma 4.41, $(q + q_1') \approx (q' + q_1)$ 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, i. e., $q + q_1' = mu + q_2$ (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.



Case 3.1.2: v is deleted by a κ-step.

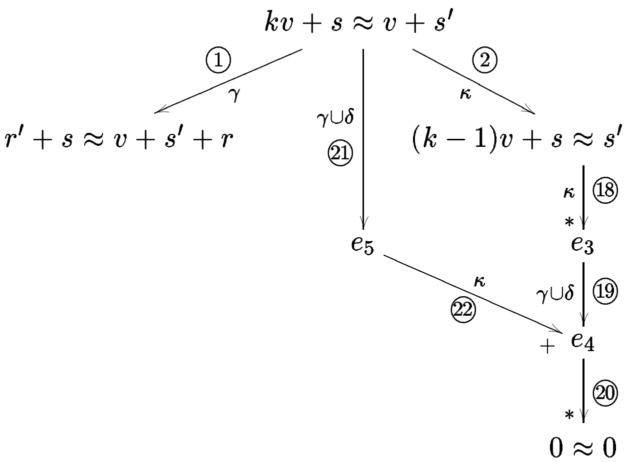
The deletion of v by a κ-step requires the existence of another occurrence of v on the left-hand side. Again, this occurrence can only be derived from s or s'. We may thus assume that the derivation has the form ⑬-⑭-⑮:



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.



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

$$\begin{array}{ccc} & e_0[k_0v + w + q_0 + q_1] & \\ \textcircled{1}\ \swarrow_{\delta} & & \searrow_{\delta}\ \textcircled{2} \\ e_0[(k_0{-}k_1)v + q_0 + q_1' + w'] & & e_0[q_0' + q_1 + w'] \end{array}$$

By Lemma 4.41, $(k_0 - k_1)v + (q_0 + q_1') \approx (q_0' + q_1)$ is either $0 \approx 0$ or a rule in $E_D^\Psi \cup R_D^\Psi$. If it is $0 \approx 0$ the peak is trivial; otherwise, we can join the peak between ① and ② as follows:

$$\begin{array}{ccc} & e_0[k_0v + w + q_0 + q_1] & \\ \textcircled{1}\ \swarrow_{\delta} & & \searrow_{\delta}\ \textcircled{2} \\ e_0[(k_0{-}k_1)v + q_0 + q_1' + w'] & \xleftrightarrow[\textcircled{3}]{\delta} & e_0[q_0' + q_1 + w'] \end{array}$$

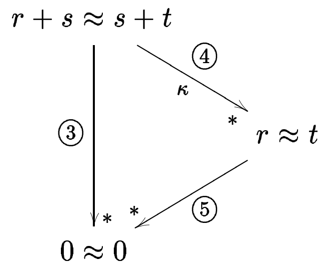
where step ③ uses $(k_0 - k_1)v + (q_0 + q_1') \approx (q_0' + q_1)$.

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 4.42 rather than Lemma 4.41. □

COROLLARY 4.44 *For every $C \in \overline{N}$, $\mathrm{tr}(R_C^\Psi)$ and $\mathrm{tr}(R_\infty^\Psi)$ are models of the equality axioms.*

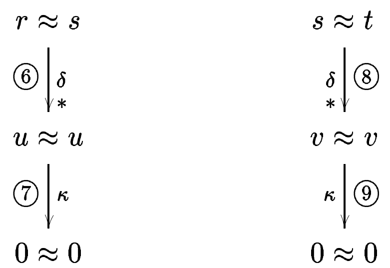
PROOF. We consider only $\mathrm{tr}(R_C^\Psi)$; the proof for $\mathrm{tr}(R_\infty^\Psi)$ is similar. It is obvious that $s \approx s \in \mathrm{tr}(R_C^\Psi)$ for every term s, and that $s \approx t \in \mathrm{tr}(R_C^\Psi)$ implies $t \approx s \in \mathrm{tr}(R_C^\Psi)$. For the transitivity axiom,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

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

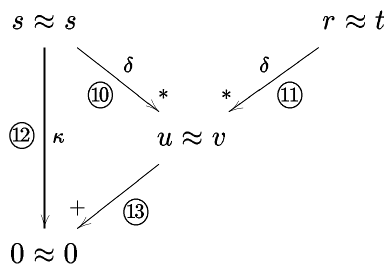


On the other hand, we can use κ-steps ④ to cancel s on both sides of the equation. By Thm. 4.43, there is a derivation ⑤, hence $r \approx t \in \mathrm{tr}(R_C^{\Psi})$.

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:

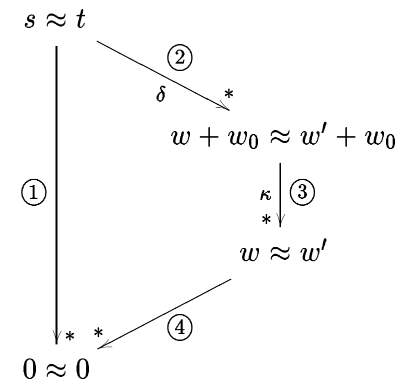


As the δ-steps in ⑥ and ⑧ rewrite each side of the equations separately, we can use the same rules to rewrite both $s \approx s$ ⑩ and $r \approx t$ ⑪ to $u \approx v$.

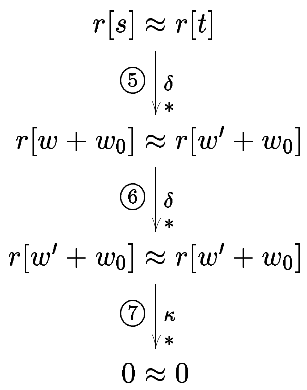


On the other hand, we can rewrite $s \approx s$ immediately to $0 \approx 0$ ⑫. By confluence, there is a derivation ⑬ and $r \approx t \in \mathrm{tr}(R_C^\Psi)$.

To prove 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 let us assume that s occurs in r below a free symbol. Consider the derivation ①:



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. 4.43, 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 4.39. Without loss of generality we assume $w \succeq w'$. This allows us to construct the following derivation:



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

Corollary 4.45 *For every $C \in \overline{N}$, $\mathrm{tr}(R_C^\Psi)$ and $\mathrm{tr}(R_\infty^\Psi)$ are models of the theory axioms* 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. 4.44). The associative, commutative, and identity axioms are obvious. $\Box$

COROLLARY 4.46 *For every clause* $C \in \overline{N}$, $\mathrm{tr}_\Psi(R_C^\Psi) = \mathrm{tr}(R_C^\Psi) = \mathrm{cl}_\Psi(R_C)$ *and* $\mathrm{tr}_\Psi(R_\infty^\Psi) = \mathrm{tr}(R_\infty^\Psi) = \mathrm{cl}_\Psi(R_\infty)$.

COROLLARY 4.47 *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 4.36, 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. 4.43, Lemma 4.39, and Lemma 4.35, $mv + r \approx r'$ is a rule in E_D^Ψ. According to Lemma 4.38, 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 $\gcd(\psi, \chi) = 1$. $\Box$

COROLLARY 4.48 *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* $nu + t \approx n'u + t'$ *with* $n > n'$ *and* $n > 0$, *and there exists a* $\psi \in \Psi$ *and* $\chi \in \mathbf{N}^{>0}$ *such that* $\gcd(\psi, \chi) = 1$, $\chi m + \psi n' = \psi n$, *and* $\psi t + \chi s' \approx \chi s + \psi t' \in \mathrm{tr}(R_C^\Psi)$.

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 $v = u$.

We have required mu to be irreducible with respect to R_C^Ψ. During the derivation of $\psi_0 nu + \psi_0 t \approx \psi_0 n'u + \psi_0 t'$ to $0 \approx 0$, the occurrences of u can thus only be eliminated by κ-steps or by γ-steps using e_1. Without loss of generality the derivation starts with k-fold κ-rewriting ① for some $k \in \mathbf{N}$. Then all remaining u's are removed by γ-steps; the rule e_1 is applied χ_0 times on the left-hand side and χ_0' times on the right-hand side ②, where $\psi_0 n - k = \chi_0 m$ and $\psi_0 n' - k = \chi_0' m$.

$$
\begin{array}{c}
\psi_0 nu + \psi_0 t \approx \psi_0 n'u + \psi_0 t' \\
\text{①}\Big\downarrow{}^{\kappa}_{*} \\
(\psi_0 n - k)u + \psi_0 t \approx (\psi_0 n' - k)u + \psi_0 t' \\
\text{②}\Big\downarrow{}^{\gamma}_{+} \\
\chi_0 s' + \psi_0 t + \chi'_0 s \approx \chi'_0 s' + \psi_0 t' + \chi_0 s \\
\text{③}\Big\downarrow{}^{\kappa}_{*} \\
\psi_0 t + \chi''_0 s' \approx \chi''_0 s + \psi_0 t' \\
\text{④}\Big\downarrow{}_{*} \\
0 \approx 0
\end{array}
$$

All remaining terms in the equation all smaller than u, hence the following rewrite steps can only use rules from R_C^Ψ. By confluence of $\rightarrow_{R_C^\Psi}$, we may assume that the derivation of $\chi_0 s' + \psi_0 t + \chi'_0 s \approx \chi'_0 s' + \psi_0 t' + \chi_0 s$ to $0 \approx 0$ starts with χ'_0-fold cancellation of $s + s'$ ③. Define $\chi''_0 = \chi_0 - \chi'_0$, then $\psi_0 t + \chi''_0 s' \approx \chi''_0 s + \psi_0 t' \in \mathrm{tr}(R_C^\Psi)$ and $\chi''_0 m + \psi_0 n' = \psi_0 n$ ④.

We still have to show that $n > n'$ and $\chi''_0 > 0$. Assume that $n = n'$. Then $\chi''_0 = 0$, and we could immediately rewrite $\psi_0 nu + \psi_0 t \approx \psi_0 n'u + \psi_0 t'$ to $\psi_0 t \approx \psi_0 t'$ using κ-steps, and then continue as in ④. Hence $nu + t \approx n'u + t'$ would be contained in R_C^Ψ, contradicting our assumptions.

We can now define $\chi = \chi''_0 / \gcd(\psi_0, \chi''_0)$ and $\psi = \psi_0 / \gcd(\psi_0, \chi''_0)$. Then $\psi t + \chi s' \approx \chi s + \psi t' \in \mathrm{tr}(R_C^\Psi)$ as $\mathrm{tr}(R_C^\Psi)$ satisfies T_Ψ; furthermore $\gcd(\psi, \chi) = 1$ and $\chi m + \psi n' = \psi n$. □

The following corollary is proved analogously.

COROLLARY 4.49 *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 $s' \approx t' \in \mathrm{tr}(R_C^\Psi)$.*

COROLLARY 4.50 *The set R_∞ is stratified.*

PROOF. We have to show that for all ground equations e_0 and e_1 with $\mathrm{mt}(e_0) \succeq \mathrm{mt}(e_1)$, $R_\infty \models_\Psi e_1$ implies $R_\infty^{\prec_{\mathrm{L}} \neg e_0} \models_\Psi e_1$. Without loss of generality let us assume that $R_\infty \neq R_\infty^{\prec_{\mathrm{L}} \neg e_0}$. Let C be the minimal clause in $\overline{N}$ such that $E_C \not\subseteq R_\infty^{\prec_{\mathrm{L}} \neg e_0}$. Then $R_C = R_\infty^{\prec_{\mathrm{L}} \neg e_0}$ and $E_C = \{e'\}$ with $e' \succeq_{\mathrm{L}} \neg e_0$, and therefore $\mathrm{mt}(e') \succ \mathrm{mt}(e_0) \succeq \mathrm{mt}(e_1)$. By assumption, $e_1 \in \mathrm{cl}_\Psi(R_\infty) = \mathrm{tr}(R_\infty^\Psi)$. As for all

$e \in R_\infty^\Psi \setminus R_C^\Psi$ we have $\mathrm{mt}(e) \succeq \mathrm{mt}(e') \succ \mathrm{mt}(e_1)$, the R_∞^Ψ-derivation from e_1 to $0 \approx 0$ can only use rules from R_C^Ψ, hence $e_1 \in \mathrm{tr}(R_C^\Psi) = \mathrm{cl}_\Psi(R_C) = \mathrm{cl}_\Psi(R_\infty^{\prec_L \neg e_0})$, as required. □

4.7 Completeness

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 three lemmas are consequences of this fact.

LEMMA 4.51 *Let $C\theta$ be a clause from $\overline{N}$. If $C\theta$ is true in $\mathrm{tr}(R_{C\theta}^\Psi)$, then it is also true in $\mathrm{tr}(R_\infty^\Psi)$ and $\mathrm{tr}(R_{D\theta}^\Psi)$ for any $D\theta \succ_C C\theta$.*

LEMMA 4.52 *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 $\mathrm{tr}(R_\infty^\Psi)$ and $\mathrm{tr}(R_{D\theta}^\Psi)$ for any $D\theta \succeq_C C\theta$.*

LEMMA 4.53 *Let $C\theta$ be an $R_{C\theta}$-variable minimal ground instance of a clause $C\,[\![T]\!] \in N$ such that $E_{C\theta} = \{e\theta\}$. Then $C\theta$ is an R_∞-variable minimal ground instance of $C\,[\![T]\!]$.*

LEMMA 4.54 *Let N be a set of constrained clauses that is saturated up to ACUKT_Ψ-redundancy and does not contain the empty clause. Then we have for every variable minimal ground clause $C\theta \in \mathrm{vm}_{R_\infty}(N)$ with $C\,[\![T_1]\!] \in N$ and $T_1\theta = \mathrm{true}$:*

(i) If C has selected literals, then $C\theta$ is true in $\mathrm{tr}(R_{C\theta}^\Psi)$.
(ii) $E_{C\theta} = \emptyset$ if and only if $C\theta$ is true in $\mathrm{tr}(R_{C\theta}^\Psi)$.
(iii) $C\theta$ is true in $\mathrm{tr}(R_\infty^\Psi)$ and in $\mathrm{tr}(R_D^\Psi)$ for every $D \succ_C C\theta$.

PROOF. We use induction on the clause ordering $\succ_C$ and assume that (i)–(iii) are already satisfied for all clauses in $\mathrm{vm}_{R_\infty}(N)$ 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 4.52, and Lemma 4.51.

Case 1: C contains a selected or maximal negative literal.

Suppose 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 $e_1\theta \notin \mathrm{tr}(R_{C\theta}^\Psi)$, there is nothing to show, so assume that there is an $R_{C\theta}^\Psi$-derivation from $e_1\theta$ to $0 \approx 0$. Let $\bar{u} = \mathrm{mt}(e_1\theta)$.

Case 1.1: $\bar{u}$ 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 4.22, this is an R_∞-variable minimal instance of an *equality resolution* inference from $C\,[\![T_1]\!]$. By saturation up to ACUKT_Ψ-redundancy, it is ACUKT_Ψ-redundant, hence $R_\infty^{\prec_{\text{L}}\text{ml}(C\theta)} \cup \text{vm}_{R_\infty}(N)^{\prec_{\text{C}} C\theta} \models_\Psi C'\theta$. By the induction hypothesis, all clauses in $\text{vm}_{R_\infty}(N)^{\prec_{\text{C}} C\theta}$ are true in $\text{tr}(R^\Psi_{C\theta})$; furthermore $R_\infty^{\prec_{\text{L}}\text{ml}(C\theta)} \subseteq R_{C\theta} \subseteq \text{tr}(R^\Psi_{C\theta})$. Thus $C'\theta$ and $C\theta$ are true in $\text{tr}(R^\Psi_{C\theta})$.

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

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

By Lemma 4.22, this is an R_∞-variable minimal instance of a *cancellation* inference from $C\,[\![T_1]\!]$. By saturation up to ACUKT_Ψ-redundancy, the inference is ACUKT_Ψ-redundant, hence $R_\infty^{\prec_{\text{L}}\text{ml}(C\theta)} \cup \text{vm}_{R_\infty}(N)^{\prec_{\text{C}} C\theta} \models_\Psi C'\theta \vee \neg\, (\bar{m}{-}\bar{m}')\bar{u} + \bar{s} \approx \bar{s}'$. By the induction hypothesis, all clauses in $\text{vm}_{R_\infty}(N)^{\prec_{\text{C}} C\theta}$ and thus $C'\theta \vee \neg\, (\bar{m}{-}\bar{m}')\bar{u} + \bar{s} \approx \bar{s}'$ and $C\theta$ are true in $\text{tr}(R^\Psi_{C\theta})$.

Case 1.2: $\bar{u}$ 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 two cases is irrelevant.[32] By Lemma 4.29 we may assume that the reduction from $e_1\theta$ to $0 \approx 0$ starts with a γ- or δ-application of 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 $C\,[\![T_1]\!]$ and $D\,[\![T_2]\!]$ are variable disjoint, so that we can use the same substitution θ.) Let $D\theta = D'\theta \vee e_2\theta$ with $T_2\theta = \text{true}$ and $E_{D\theta} = \{e_2\theta\}$. By part (i) and (ii) of the induction hypothesis, Lemma 4.52, and Lemma 4.53, D has no selected literals, $D'\theta$ is false in $\text{tr}(R^\Psi_{C\theta})$, and $D\theta$ is an R_∞-variable minimal ground instance of $D\,[\![T_2]\!]$.

Case 1.2.1: $\bar{m}\bar{u}$ 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. 4.47, $e_2\theta$ 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 $\gcd(\psi, \chi) = 1$.

[32] 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.

Consider the *negative cancellative superposition* inference

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

As $\bar{m}\bar{u} + \bar{s} \approx \bar{s}' \in \mathrm{tr}(R^{\Psi}_{C\theta}) \subseteq \mathrm{cl}_{\Psi}(R_\infty)$ and $\bar{m}\bar{u} + \bar{s} \succ \bar{s}'$, the left-hand side of e_1 cannot be a variable – it would not be R_∞-variable minimal otherwise. By Lemma 4.23 the inference is a R_∞-variable minimal ground instance of a *negative cancellative superposition* inference from $D\,[\![T_2]\!]$ and $C\,[\![T_1]\!]$. As N is saturated, it is ACUKT_{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 1.2.2: $\bar{u}$ 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 1.2.2.1: $\bar{u}$ 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. Consequently, there is a *standard superposition* inference

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

which is an R_∞-variable minimal ground instance of a *standard superposition* inference from $D\,[\![T_2]\!]$ and $C\,[\![T_1]\!]$. 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 1.2.2.2: $\bar{u}$ is δ-reducible by e'' but not by $e_2\theta$.

By the definition of $E^{\Psi}_{D\theta}$, 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

$$\frac{D'\theta \lor \bar{n}\bar{v} + \bar{t} \approx \bar{t}' \qquad C'\theta \lor \neg\, e_1\theta[\bar{u}[\bar{m}_0\bar{v} + \bar{r} + \bar{q}]]}{C_0\ [\![T_0]\!]}$$

where C_0 equals $C'\theta \vee \neg\, y \approx \bar{m}_0\bar{v} + \bar{r} + \bar{q} \vee \neg\, e_1\theta[\bar{u}[y]]$ and T_0 is the constraint $\psi\bar{m}_0\bar{v} \doteq \psi z + \chi\bar{n}\bar{v} \wedge \bar{m}_0\bar{v} + \bar{r} + \bar{q} \succ y$. Let $\bar{w}_0$ be the smallest term such that $\bar{w}_0 \approx \bar{m}_0\bar{v} + \bar{r} + \bar{q} \in \mathrm{tr}(R_C^\Psi)$. 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 $\rho = \{z \mapsto (\bar{m}_0 - \bar{m})\bar{v},\ y \mapsto \bar{w}_0\}$. It is easy to check that $T_0\rho = \mathrm{true}$ and $C_0\rho$ is ground.

If $\bar{m}\bar{v} + \bar{r}$ occurred in $e_1\theta$ at or below a variable position of C, then $C\theta$ could not be an R_∞-variable minimal instance, as $e'' = \bar{m}\bar{v} + \bar{r} \approx \bar{r}'$ is contained in $\mathrm{cl}_\Psi(R_\infty)$. 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 4.23, the inference is an R_∞-variable minimal ground instance of an *abstraction* inference from $D\,[\![T_2]\!]$ and $C\,[\![T_1]\!]$.

By saturation, the clause $C_0\rho$, that is

$$C'\theta \vee \neg\, \bar{w}_0 \approx \bar{m}_0\bar{v} + \bar{r} + \bar{q} \vee \neg\, e_1\theta[\bar{u}[\bar{w}_0]]$$

is true in $\mathrm{tr}(R_{C\theta}^\Psi)$; and since $\bar{w}_0 \approx \bar{m}_0\bar{v} + \bar{r} + \bar{q} \in \mathrm{tr}(R_C^\Psi)$, $C\theta$ must be true likewise.

Case 2: C does not contain a selected or maximal negative literal.

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

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

If $e_1\theta$ has the form $\bar{u} \approx \bar{u}$, then $C\theta$ is a tautology and thus true in $\mathrm{tr}(R_{C\theta}^\Psi)$. 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 1.1, we can show that $C\theta$ is true in $\mathrm{tr}(R_{C\theta}^\Psi)$.

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 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 2.2.1: $e_1\theta$ 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 an R_∞-variable minimal ground instance of an inference from $C\,[\![T_1]\!]$. 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 2.2.2: $e_1\theta$ 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 4.52, 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 1.2 we can thus show that $C\theta$ is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$.

Case 2.2.3: $e_1\theta$ 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 not larger 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 2.2.3.1: $\bar{u}$ has sort S_{CAM}.

If $\bar{u}$ has sort S_{CAM}, we know by Lemma 4.48 that $e_2\theta$ equals $\bar{n}\bar{u} + \bar{t} \approx \bar{n}'\bar{u} + \bar{t}'$ where $\chi\bar{m} + \psi\bar{n}' = \psi\bar{n}$ 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

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

which is an R_∞-variable minimal ground instance of a *cancellative equality factoring* inference from $C\,[\![T_1]\!]$. By saturation, its conclusion is true in $\mathrm{tr}(R^{\Psi}_{C\theta})$. As $\psi\bar{t} + \chi\bar{s}' \approx \chi\bar{s} + \psi\bar{t}' \in \mathrm{tr}(R^{\Psi}_{C\theta})$, $C''\theta \vee \bar{n}\bar{u} + \bar{t} \approx \bar{n}'\bar{u} + \bar{t}$ and thus $C\theta$ must be true in $\mathrm{tr}(R^{\Psi}_{C\theta})$. This contradicts our assumption above.

Case 2.2.3.2: $\bar{u}$ does not have sort S_{CAM}.

If $\bar{u}$ does not have sort S_{CAM}, we know by Lemma 4.49 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. □

COROLLARY 4.55 *Let N be a set of constrained clauses that is saturated up to ACUKT_{Ψ}-redundancy. Then $\mathrm{vm}_{R_\infty}(N) \cup \mathrm{ACUKT}_{\Psi}$ is equality unsatisfiable if and only if N contains the empty clause with a satisfiable constraint.*

PROOF. If N contains the empty clause with a satisfiable constraint, then the set $\mathrm{vm}_{R_\infty}(N)$ contains the empty clause and is unsatisfiable. Otherwise, $\mathrm{tr}(R^{\Psi}_{\infty})$ is a model of the equality axioms (by Cor. 4.44), of ACUKT_{Ψ} (by Cor. 4.45), and of $\mathrm{vm}_{R_\infty}(N)$ (by part (iii) of Lemma 4.54). □

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

THEOREM 4.56 *Let N be a set of constrained clauses that is model generalizable and saturated up to ACUKT_{Ψ}-redundancy. Then $N \cup \mathrm{ACUKT}_{\Psi}$ is equality unsatisfiable if and only if N contains the empty clause with a satisfiable constraint.*

PROOF. Suppose that N does not contain the empty clause with a satisfiable constraint. So the previous corollary shows that $\mathrm{vm}_{R_\infty}(N) \cup \mathrm{ACUKT}_{\Psi}$ has the equality model $\mathrm{tr}(R^{\Psi}_{\infty}) = \mathrm{cl}_{\Psi}(R_\infty)$. As N is model generalizable, $\mathrm{cl}_{\Psi}(R_\infty)$ is an equality model of $\mathrm{vm}_{\emptyset}(N) \cup \mathrm{ACUKT}_{\Psi}$. The reverse direction of the proof is trivial. □

THEOREM 4.57 *Let $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$ be a fair derivation of the cancellative superposition calculus, such that no clause from N_0 is lower bounded. 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 with a satisfiable constraint.*

PROOF. Suppose that N_∞ does not contain the empty clause with a satisfiable constraint. By fairness, N_∞ is saturated up to ACUKT_{Ψ}-redundancy. As no clause from N_0 is lower bounded, N_0 and thus N_∞ is model generalizable. Therefore the previous theorem shows that $N_\infty \cup \mathrm{ACUKT}_{\Psi}$ has an equality model. As $N_0 \subseteq N_\infty \cup \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N_\infty)$, this model is also an equality model of $N_0 \cup \mathrm{ACUKT}_{\Psi}$. The reverse direction of the proof is obvious since $N_0 \models_{\Psi} N_\infty$. □

5 Refinements and Applications

5.1 Simplification Techniques

Simplification. In any practicable saturation-based theorem prover methods to keep the current set of formulae as small as possible are indispensable. By Def. 2.29, formulae may be deleted during a theorem proving derivation only if they are redundant. If they do not become redundant accidentally, we still have the chance to *make* them redundant. This process is called simplification.

DEFINITION 5.1 *Let N be a set of clauses. We say that $M \subseteq N$ is simplified to a set M' of clauses, if $N \models_\Psi M'$ and if M is redundant with respect to $N \cup M'$.*

LEMMA 5.2 *If $M \subseteq N$ is simplified to M', then $N \vdash (N \cup M') \setminus M$ is an admissible derivation step according to Def. 2.29.*

PROOF. As $N \models_\Psi M'$, we have $N \models_\Psi (N \cup M') \setminus M$, so condition (i) of Def. 2.29 is satisfied. To prove condition (ii), we note that $N \setminus ((N \cup M') \setminus M) = M$, and $M \subseteq \mathit{CS\text{-}Red}_\Psi^{\mathrm{C}}(N \cup M')$ by assumption, and $\mathit{CS\text{-}Red}_\Psi^{\mathrm{C}}(N \cup M') \subseteq \mathit{CS\text{-}Red}_\Psi^{\mathrm{C}}((N \cup M') \setminus M)$ by part (iii) of Def. 2.28. □

In the sequel we discuss some techniques that are specific for the cancellative superposition calculus.

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

Demodulating a clause $C[s\sigma]$ means rewriting it to $C[s'\sigma]$ using another clause $s \approx s'$, where $s\sigma \succ s'\sigma$ and $C[s\sigma] \succ_{\mathrm{C}} s \approx s'$. In calculi without constraints, it is almost trivial to show that demodulation is a simplification. In the constraint case, however, it becomes more complicated. Even if the constraint of the demodulated clause entails the constraint of the demodulating clause, we have to take into account that redundancy is defined via variable minimal instances.

The following example is due to Nieuwenhuis and Rubio [67]. Let R be a stratified rewrite system such that $\mathrm{cl}_\Psi(R)$ contains an equation $b \approx t$ with $b \succ t$. Then the clause $D\,[\![T_2]\!] \;=\; g(x) \approx c\;[\![x \doteq b]\!]$ does not have any R-variable minimal ground instances. Consequently, we cannot use $D\,[\![T_2]\!]$ to demodulate $f(g(b)) \approx c$ to $f(c) \approx c$. If we want to simplify $f(g(b)) \approx c$, we first have to weaken $D\,[\![T_2]\!]$, that is, we must derive the clause $g(b) \approx c$. The situation is different if variables in the demodulating clause $D\,[\![T_2]\!]$ correspond to variables in the demodulated clause $C\,[\![T_1]\!]$. For example, we may use $g(x) \approx c\;[\![x \doteq b]\!]$ to simplify $f(g(x)) \approx c\;[\![x \doteq b]\!]$. Here, depending on R either both the ground instances $g(b) \approx c$ and $f(g(b)) \approx c$ are R-variable minimal, or none of them. In both cases, the redundancy criterion is satisfied – in the second case vacuously.

LEMMA 5.3 *Let R be a stratified set of ground equations, let $D\,[\![T]\!]$ be a clause, and let $\mathcal{V}'$ be a subset of $\mathrm{var}(D)$. Let $D\theta$ be a ground instance of $D\,[\![T]\!]$, such that no $x \in \mathcal{V}'$ is lower bounded in $D\,[\![T]\!]$, and such that all $x \in \mathrm{var}(D) \setminus \mathcal{V}'$ are R-variable minimal in $D\theta$. Let θ' be the substitution that maps x to $x\theta$ for $x \in \mathrm{var}(D\,[\![T]\!]) \setminus \mathcal{V}'$ and to the minimal t with $x\theta \approx t \in \mathrm{cl}_\Psi(R^{\prec_{\mathrm{L}}\mathrm{ml}(D\theta)})$ for $x \in \mathcal{V}'$. Then $D\theta'$ is an R-variable minimal ground instance of $D\,[\![T]\!]$.*

PROOF. As no variable in $\mathcal{V}'$ is lower bounded in $D\,[\![T]\!]$, $D\theta'$ is obviously an instance of $D\,[\![T]\!]$. It remains to show that all variables occurring in D are R-variable minimal in $D\theta'$. Let x be such a variable.

Suppose that $x \in \mathrm{var}(D) \setminus \mathcal{V}'$. Then $x\theta = x\theta'$, and x is R-variable minimal in $D\theta$. If x is a positive maximal variable of D and θ, then it is a positive maximal variable of D and θ', and the literals of D containing x are not larger in $D\theta'$ than in $D\theta$. From this, it follows that x is R-variable minimal in $D\theta'$.

Suppose now that $x \in \mathcal{V}'$. Then $x\theta'$ is the minimal t with $x\theta \approx t \in \mathrm{cl}_\Psi(R^{\prec_{\mathrm{L}}\mathrm{ml}(D\theta)})$. If there exists no s such that $x\theta' \succ s$ and $x\theta' \approx s \in \mathrm{cl}_\Psi(R)$, then x is R-variable minimal in $D\theta'$. Otherwise, we know that $x\theta' \approx s \in \mathrm{cl}_\Psi(R)$ and $x\theta' \approx s \notin \mathrm{cl}_\Psi(R^{\prec_{\mathrm{L}}\mathrm{ml}(D\theta)})$. We have assumed R to be stratified. By Def. 4.4 and Lemma 4.5, this implies that either $\mathrm{ml}(D\theta)$ is negative and $\mathrm{mt}(x\theta') = \mathrm{mt}(x\theta' \approx s) \succ \mathrm{mt}(\mathrm{ml}(D\theta))$, or $\mathrm{ml}(D\theta)$ is positive and $\mathrm{mt}(x\theta') = \mathrm{mt}(x\theta' \approx s) \succeq \mathrm{mt}(\mathrm{ml}(D\theta))$. Obviously, $\mathrm{mt}(x\theta')$ cannot be strictly larger than $\mathrm{mt}(\mathrm{ml}(D\theta))$, hence $\mathrm{ml}(D\theta)$ is positive and $\mathrm{mt}(x\theta') = \mathrm{mt}(\mathrm{ml}(D\theta))$. We conclude that x must be a positive maximal variable of D and θ'.

To prove that x is R-variable minimal in $D\theta'$, we still have to show that

(i) there is no subterm r of $x\theta'$ with $r \approx r' \in \mathrm{cl}_\Psi(R)$ and $\mathrm{mt}(x\theta') \succ r \succ r'$, and (ii) x has sort S_{CAM}, or there is no literal $x \approx r''$ in D and $x\theta' \approx r' \in R$ with $x\theta' \succ r'$ and $r''\theta' \succ r'$. We prove this by contradiction: If condition (i) were violated, then $r \approx r' \in \mathrm{cl}_\Psi(R^{\prec_{\mathrm{L}}\mathrm{ml}(D\theta)})$, as R is stratified. If condition (ii) were violated, then $x\theta' \approx r' \in R^{\prec_{\mathrm{L}}\mathrm{ml}(D\theta)}$. In both cases, $x\theta'$ could not be the minimal term t with $x\theta \approx t \in \mathrm{cl}_\Psi(R^{\prec_{\mathrm{L}}\mathrm{ml}(D\theta)})$. □

THEOREM 5.4 *Let $C\,[\![T_1]\!]$ and $D\,[\![T_2]\!]$ be two clauses in N and let σ be a substitution, such that*

- *$D = D' \vee t \approx t'$, $C = D'\sigma \vee C'[t\sigma]$,*
- *every solution of T_1 is a solution of $T_2\sigma \wedge C \succ_{\mathrm{C}} D\sigma \wedge t\sigma \succ t'\sigma$,*
- *for every variable x occurring in D, either x is not lower bounded in $D\,[\![T_2]\!]$, or $x\sigma$ is a variable of C, or $x\sigma = 0$,*
- *every variable of $t'\sigma$ occurs either negatively in C, or it is shielded in C, or it occurs in C and is not lower bounded in $C\,[\![T_1]\!]$, and*
- *if $t\sigma$ is a variable and does not have sort S_{CAM}, then it occurs either negatively in C, or it is shielded in C, or C contains a literal $t\sigma \approx s'$ and every solution of T_1 satisfies $s' \succeq t'\sigma$.*

Then $C\,[\![T_1]\!]$ can be simplified to $C_0\,[\![T_1]\!] = D'\sigma \vee C'[t'\sigma]\,[\![T_1]\!]$.

PROOF. Let R be an arbitrary stratified set of ground equations. We have to show that every R-variable minimal ground instance $C\theta$ follows from $R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)}$ and R-variable minimal ground instances of $D\,[\![T_2]\!]$ and $C_0\,[\![T_1]\!]$ that are smaller than $C\theta$. Define a substitution θ' that maps every variable $x \in \mathrm{var}(C)$ that is not lower bounded in $C\,[\![T_1]\!]$ to the minimal q with $x\theta \approx q \in \mathrm{cl}_\Psi(R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)})$, and every other variable x to $x\theta$. Obviously, $R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)} \cup \{C\theta'\} \models_\Psi C\theta$, furthermore $C\theta'$ is again an R-variable minimal ground instance of $C\,[\![T_1]\!]$ by Lemma 5.3.

We will first demonstrate that $C_0\theta'$ is an R-variable minimal ground instance of $C_0\,[\![T_1]\!]$, that is, that all variables of C_0 are R-variable minimal in $C_0\theta'$.

For a variable x not occurring in $t'\sigma$, this is obvious: Whenever x occurs in a literal $[\neg]\, e_0$ of C_0, then it occurs also in a literal $[\neg]\, e$ of C such that $[\neg]\, e_0\theta' \preceq_{\mathrm{L}} [\neg]\, e\theta'$; hence if x is a positive maximal variable of C and θ', then it is a positive maximal variable of C_0 and θ'.

If x does occur in $t'\sigma$, then two cases have to be distinguished: If there is no term r' such that $x\theta' \succ r'$ and $x\theta' \approx r' \in \mathrm{cl}_\Psi(R)$, then x is R-variable minimal in $C_0\theta'$. Otherwise the R-variable minimality of $C\theta'$ implies that x is a positive maximal variable of C and θ'. Since x occurs neither negatively nor shielded in C, it may not be lower bounded in $C\,[\![T_1]\!]$. By construction of θ', $x\theta'$ is the minimal q with $x\theta \approx q \in \mathrm{cl}_\Psi(R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)})$. On the other hand, $x\theta' \succ r'$ and $x\theta' \approx r' \in \mathrm{cl}_\Psi(R)$. Since R is stratified, we can conclude that either $\mathrm{ml}(C\theta)$ is negative and $\mathrm{mt}(x\theta' = r') \succ \mathrm{mt}(\mathrm{ml}(C\theta))$, or $\mathrm{ml}(C\theta)$ is positive and $\mathrm{mt}(x\theta' = r') \succeq \mathrm{mt}(\mathrm{ml}(C\theta))$. Now $\mathrm{mt}(\mathrm{ml}(C\theta)) \succeq \mathrm{mt}(x\theta) \succeq \mathrm{mt}(x\theta') = \mathrm{mt}(x\theta' = r')$, thus it follows that $\mathrm{ml}(C\theta)$ is positive, that $\mathrm{mt}(x\theta') = \mathrm{mt}(x\theta) = \mathrm{mt}(\mathrm{ml}(C\theta))$, and that x is a positive maximal variable of C and θ. Furthermore, we know that $t\sigma\theta$ occurs in $C\theta$ and $t\sigma\theta \succ t'\sigma\theta \succeq x\theta$. If x did not have sort S_{CAM}, this would be impossible, since it would imply $\mathrm{mt}(\mathrm{ml}(C\theta)) \succeq \mathrm{mt}(t\sigma\theta) \succ \mathrm{mt}(x\theta)$. Hence x has sort S_{CAM}. It is now easy to check that x is R-variable minimal in $C_0\theta$.

Let us now consider $D\,[\![T_2]\!]$. In a similar way as above, we can show that all variables in $D\sigma$ are R-variable minimal in $D\sigma\theta'$. From this we may conclude that all variables of D that σ maps to 0 or to a variable of C are R-variable minimal in $D\sigma\theta'$. Define a substitution θ'' that maps every variable x of D that is not lower bounded in $D\,[\![T_2]\!]$ to the minimal q' with $x\sigma\theta' \approx q' \in \mathrm{cl}_\Psi(R^{\prec_{\mathrm{L}}\mathrm{ml}(D\sigma\theta')})$, and every other variable x to $x\sigma\theta'$. By Lemma 5.3, $D\theta''$ is an R-variable minimal ground instance of $D\,[\![T_2]\!]$. Now $R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)} \cup \{C_0\theta', D\theta''\} \models_\Psi C\theta$, as required. □

We can extend demodulation from the traditional kind of rewriting to cancellative rewriting:

THEOREM 5.5 *Let $C\,[\![T_1]\!]$ and $D\,[\![T_2]\!]$ be two clauses in N and let σ be a substitution, such that*

- $D = D' \vee t + w \approx w'$, $C = D'\sigma \vee C' \vee [\neg]\, s + t\sigma \approx s'$,
- *every solution of T_1 is a solution of* $T_2\sigma \wedge C \succ_{\mathrm{C}} D\sigma \wedge s + t\sigma \approx s' \succ_{\mathrm{L}} s + w'\sigma \approx s' + w\sigma$,
- *for every variable x occurring in D, either x is not lower bounded in $D\,[\![T_2]\!]$, or $x\sigma$ is a variable of C, or $x\sigma = 0$,*
- *every variable of $w\sigma$ and $w'\sigma$ occurs either negatively in C, or it is shielded in C, or it occurs in C and is not lower bounded in $C\,[\![T_1]\!]$.*

Then $C\,[\![T_1]\!]$ can be simplified to $C_0\,[\![T_1]\!] = D'\sigma \vee C' \vee [\neg]\, s + w'\sigma \approx s' + w\sigma\,[\![T_1]\!]$.

PROOF. Analogously to the proof of Thm. 5.4. □

For example, if D is the inverse axiom $(-x) + x \approx 0$, then every clause $C' \vee [\neg]\, (-s) + t \approx t' \; [\![T_1]\!]$ that is larger than $(-s) + s \approx 0$ can be simplified to $C' \vee [\neg]\, t \approx t' + s \; [\![T_1]\!]$. Similarly, if a (cancellative) superposition inference from unconstrained ground unit clauses produces a conclusion with a true constraint, then its larger premise follows from the smaller premise and the conclusion; hence any such inference constitutes a simplification of the larger premise.

Quasisimplification. When a set of clauses is simplified, then the ground instances of the removed clauses follow from smaller ground instances of other clauses. What happens, if we relax "smaller" to "smaller or equal"? Consider the following example.

EXAMPLE 5.6 Consider the set of clauses (1)–(5):

$$f(x) \not\approx b \vee g(x) \approx b \quad (1)$$
$$f(d) \approx b \quad (2)$$
$$f(x) \not\approx b \vee f(h(x)) \approx b \quad (3)$$
$$f(c) \approx b \quad (4)$$
$$g(c) \not\approx b \quad (5)$$

If we define $\succ$ as the lexicographic path ordering over the precedence $f \succ g \succ b \succ c \succ h \succ d$, then d is the minimal ground term, hence every ground instance of clause (1) is an instance of either clause (6) or (7):

$$f(d) \not\approx b \vee g(d) \approx b \quad (6)$$
$$f(x) \not\approx b \vee g(x) \approx b \; [\![x \succ d]\!] \quad (7)$$

As one of these two clauses (namely clause (6)) can be further simplified using clause (2) to

$$g(d) \approx b \quad (8)$$

it is tempting to replace clause (1) by clause (7) and (8). But this is not a simplification – for a good reason: The clauses (2) and (3) entail $f(h^n(d)) \approx b$ for every $n \in \mathbf{N}$. As $h(d)$ is the minimal ground term larger than d, we might again replace clause (7) by

$$f(h(d)) \not\approx b \vee g(h(d)) \approx b \quad (9)$$
$$f(x) \not\approx b \vee g(x) \approx b \; [\![x \succ h(d)]\!] \quad (10)$$

followed by a simplification of clause (9) to

$$g(h(d)) \approx b \tag{11}$$

And so on. Eventually, we derive every clause of the form

$$f(x) \not\approx b \vee g(x) \approx b \; [\![x \succ h^n(d)]\!],$$

and delete it immediately afterwards. None of these clauses is persistent, hence fairness does *not* require to perform the superpositions with clauses (4) and (5), which would lead to a contradiction. In other words, although the set of clauses (1)–(5) is contradictory, there exists a fair derivation that does not produce the empty clause: Refutational completeness is destroyed.

In spite of this problem, such operations on clause sets, which we will call quasisimplifications in the sequel, have their uses in automated theorem proving.

DEFINITION 5.7 *Let M, M', and N be sets of constrained clauses. We say that M is quasisimplified to M' with respect to N, if $N \cup M \models_\Psi M'$ and for every stratified set of ground equations R and every R-variable minimal ground instance $C\theta$ of $C\,[\![T]\!] \in M$, $R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N \cup M')^{\preceq_{\mathrm{C}} C\theta} \models_\Psi C\theta$. We denote this by $M \leadsto_N M'$.*

LEMMA 5.8 *If M is quasisimplified to M' with respect to N, then*

(i) $N \cup M' \models_\Psi M$.

(ii) $\mathit{CS}\text{-}\mathit{Red}^{\mathrm{C}}_{\Psi}(N \cup M) \subseteq \mathit{CS}\text{-}\mathit{Red}^{\mathrm{C}}_{\Psi}(N \cup M')$.

(iii) $\mathit{CS}\text{-}\mathit{Red}^{\mathrm{I}}_{\Psi}(N \cup M) \subseteq \mathit{CS}\text{-}\mathit{Red}^{\mathrm{I}}_{\Psi}(N \cup M')$.

(iv) If $N \cup M$ is model generalizable, then $N \cup M'$ is model generalizable.

PROOF. Part (i) is follows immediately, if we set $R = \emptyset$ in Def. 5.7.

To prove part (ii), let $C\,[\![T]\!]$ be a clause from $\mathit{CS}\text{-}\mathit{Red}^{\mathrm{C}}_{\Psi}(N \cup M)$, let R be a stratified set of ground equations, and let $C\theta$ be an R-variable minimal ground instance of $C\,[\![T]\!]$. As $C\,[\![T]\!]$ is redundant, $R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(M)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C\theta$. Now for every $D\rho \in \mathrm{vm}_R(M)^{\prec_{\mathrm{C}} C\theta}$, we have $R^{\prec_{\mathrm{L}} \mathrm{ml}(D\rho)} \cup \mathrm{vm}_R(N \cup M')^{\preceq_{\mathrm{C}} D\rho} \models_\Psi D\rho$. Since $R^{\prec_{\mathrm{L}} \mathrm{ml}(D\rho)} \subseteq R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)}$ and $\mathrm{vm}_R(N \cup M')^{\preceq_{\mathrm{C}} D\rho} \subseteq \mathrm{vm}_R(N \cup M')^{\prec C\theta}$, this implies $R^{\prec_{\mathrm{L}} \mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N \cup M')^{\prec C\theta} \models_\Psi C\theta$, hence $C\,[\![T]\!]$ is redundant with respect to $N \cup M'$.

Part (iii) is proved analogously to part (ii).

It remains to show part (iv). Let R be a stratified set of ground equations such that $\mathrm{cl}_\Psi(R)$ is a model of $\mathrm{vm}_R(N \cup M')$. Let $C\,[\![T]\!]$ be a clause in $N \cup$

M with a ground instance $C\theta \in \mathrm{vm}_R(N \cup M)$. If $C\,[\![T]\!] \in N$, then $C\theta \in \mathrm{vm}_R(N \cup M')$, so it is true in $\mathrm{cl}_\Psi(R)$. Otherwise $C\,[\![T]\!] \in M$, then $R^{\prec_{\mathrm{Lml}(C\theta)}} \cup \mathrm{vm}_R(N \cup M')^{\preceq_{\mathrm{C}} C\theta} \models_\Psi C\theta$. As $\mathrm{cl}_\Psi(R)$ is a model of both R and $\mathrm{vm}_R(N \cup M')$, $C\theta$ is true in $\mathrm{cl}_\Psi(R)$. This proves that $\mathrm{cl}_\Psi(R)$ is a model of $\mathrm{vm}_R(N \cup M)$. Since $N \cup M$ is model generalizable, $\mathrm{cl}_\Psi(R)$ is a model of $\mathrm{vm}_\emptyset(N \cup M)$, and as $N \cup M \models_\Psi N \cup M'$, it is also a model of $\mathrm{vm}_\emptyset(N \cup M')$. □

The preceding lemma implies that we can mix derivation steps of the cancellative superposition calculus with quasisimplification steps, provided that the latter occur only finitely often in a derivation. Consider a fair mixed derivation $N_0, N_1, N_2, \ldots$, where $N_{i-1} \vdash N_i$ or $N_{i-1} \rightsquigarrow_\emptyset N_i$ for every $i > 0$. If the number of quasisimplification steps is guaranteed to be finite, then there is an index n such that the subderivation starting with N_n is a pure derivation of the cancellative superposition calculus. By Lemma 2.32 and part (ii) and (iii) of Lemma 5.8, clauses and inferences that have become redundant in some set N_i of the derivation remain redundant in all following sets. Every inference that is redundant with respect to some N_i with $i < n$ is therefore redundant with respect to N_n. Consequently, the subderivation starting with N_n is fair and its limit is saturated. Furthermore, due to Lemma 4.20 and part (iv) of Lemma 5.8, N_n is model generalizable whenever N_0 has this property.

While unrestricted quasisimplifications of the whole clause set may endanger refutational completeness, it is always permissible to quasisimplify those clauses that are added during an inference step with respect to the remaining ones:

LEMMA 5.9 *If $N \vdash N'$ is an admissible derivation step and $N' \setminus N \rightsquigarrow_{N \cap N'} M'$, then $N \vdash (N \cap N') \cup M'$ is an admissible derivation step.*

PROOF. We must show that $N \setminus ((N \cap N') \cup M')$ is a subset of $\mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}((N \cap N') \cup M')$. It is easy to check that $N \setminus ((N \cap N') \cup M')$ equals $N \setminus (N' \cup M')$, which is a subset of $N \setminus N'$. As $N \vdash N'$ is an admissible derivation step, we have $N \setminus N' \subseteq \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N')$. Splitting N' into $(N \cap N') \cup (N' \setminus N)$, we see that $\mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N')$ equals $\mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}((N \cap N') \cup (N' \setminus N))$, hence by part (ii) of Lemma 5.8, $\mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}(N') \subseteq \mathit{CS\text{-}Red}^{\mathrm{C}}_{\Psi}((N \cap N') \cup M')$. Combining these inclusions, we get the desired result. □

To ensure fairness, every inference from persistent clauses has to be made redundant at some point of the derivation. If we cannot prove that an inference ι is already redundant, we have to make it redundant, for instance by adding its conclusion to the current set of clauses. The following lemma shows that we may alternatively quasisimplify $\mathrm{concl}(\iota)$ to M' with respect to the current set N of clauses before adding it. The inference ι becomes redundant, whether we add the conclusion itself or its quasisimplification.

LEMMA 5.10 *Let ι be an inference and let N be a set of constrained clauses. If $\{\mathrm{concl}(\iota)\} \leadsto_N M'$, then $\iota \in \mathit{CS\text{-}Red}^{\mathrm{I}}_{\Psi}(N \cup M')$.*

PROOF. By condition (iv) of Def. 2.28, we have $\iota \in \mathit{CS\text{-}Red}^{\mathrm{I}}_{\Psi}(N \cup \{\mathrm{concl}(\iota)\})$. Now part (iii) of Lemma 5.8 implies $\iota \in \mathit{CS\text{-}Red}^{\mathrm{I}}_{\Psi}(N \cup M')$. □

An important kind of quasisimplifications is case splitting. If $T_1 \vee T_2 = \mathrm{true}$, we may replace a clause $C\,[\![T]\!]$ by the two clauses $C\,[\![T \wedge T_1]\!]$ and $C\,[\![T \wedge T_2]\!]$. This operation is particularly useful if one of the two new clauses can be simplified further.

Partial solution of a constraint is another quasisimplification technique. If T_1 is an equality constraint and U a complete set of ACU-unifiers of T_1, the clause $C\,[\![T_1 \wedge T_2]\!]$ can be quasisimplified to the set of clauses $C\rho\,[\![T_2\rho]\!]$ for all $\rho \in U$. Since we assume that all ordering constraints are non-essential anyway, this technique allows us to eliminate constraints completely, if needed, or to transform a clause into another one in which for example all unshielded variables are unconstrained.

The last example that we want to mention here is subsumption. If $C\,[\![T]\!]$ is a constrained clause and σ a substitution such that no variable of $\mathrm{Dom}(\sigma)$ is lower bounded in $C\,[\![T]\!]$, then, by Lemma 5.3, $C' \vee C\sigma\,[\![T\sigma]\!]$ can be quasisimplified to $\emptyset$ with respect to $C\,[\![T]\!]$. In fact, if C' is non-empty, this is not only a quasisimplification, but a simplification.[33]

5.2 Cancellative Superposition as a Decision Procedure

The Ground Unit Case. We have seen in the previous section that some cancellative superposition inferences are actually simplifications. It is an easy consequence of this observation that cancellative superposition is a decision procedure for certain sets of clauses.

CONVENTION 5.11 *Whenever we talk about ground clauses and ground inferences in Section 5.2, we consider only clauses with a true constraint and inferences whose premises and conclusion have a true constraint.*

[33]As long as $x\sigma$ contains a free function symbol for some $x \in \mathrm{Dom}(\sigma) \cap \mathrm{var}(C)$, it can be turned into a simplification even if C' is empty. However, for this purpose, we must base the definition of redundancy on a more complicated ordering that compares not only ground instances but also the non-ground clauses from which the ground instances have been produced. Details of this technique can be found in (Bachmair and Ganzinger [12]).

LEMMA 5.12 *Let ι be a CS-Inf$_\Psi$-inference from ground unit clauses over $+$ and constants. Then the conclusion of ι is either the empty clause, or again a ground unit clause over $+$ and constants. Furthermore, the largest premise of ι follows from the conclusion and the smaller premise (if any), hence ι simplifies its largest premise.*

PROOF. The only inferences that are possible from ground unit clauses over $+$ and constants are *cancellation*, *cancellative superposition*, and *equality resolution* inferences. The conclusions of the first two are again unit clauses, *equality resolution* produces the empty clause. None of these inference rules introduces variables or new free function symbols. Routine computation shows that in all cases the largest premise is simplified. □

We recall that for ground clauses there is always at most one pair (ψ, χ) such that the constraint T_E of the *negative cancellative superposition* inference rule is satisfiable.

LEMMA 5.13 *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.

PROOF. Let N be a finite set of ground unit clauses over $+$ and constants. Starting from $N_0 = N$, we construct iteratively a sequence $N_0, N_1, \ldots$ of sets of clauses in the following way:

Suppose that N_j has already been defined. If there are *CS-Inf*$_\Psi$-inferences from N_j, let ι_j be such an inference (otherwise stop). Let C_j be the conclusion of ι_j, let C'_j be its maximal premise. Then define $N_{j+1} = N_j \cup \{C_j\} \setminus \{C'_j\}$.

Note that in every iteration C_j is again a ground unit clause over $+$ and constants and that C'_j becomes redundant by adding C_j. Since the conclusion of a ground inference is always smaller than its maximal premise, the sequence of sets $N_0, N_1, \ldots$ is strictly decreasing with respect to the multiset extension of the clause ordering $\succ_C$. As this ordering is noetherian, the derivation $N_0 \vdash N_1 \vdash \cdots$ must terminate; by fairness, its last element N_k is saturated. The sets $N_0, N_1, \ldots, N_k$ are unsatisfiable with respect to ACUKT_Ψ if and only if $\bot \in N_k$. □

Lemma 5.13 is limited to unit clauses. This is not a severe restriction, however: If $C = C_1 \vee C_2$ is a non-unit ground clause, then $N \cup \{C\}$ is satisfiable if and only if $N \cup \{C_1\}$ or $N \cup \{C_2\}$ is satisfiable. It is therefore easy to reduce the decision problem for non-unit ground clauses to the decision problem for unit ground clauses by splitting every non-unit clause and checking all possible combinations of unit subclauses.

It is crucial for Lemma 5.13 that all free function symbols are constants. Lemma 5.13 is thus weaker than Marché's Theorem 5.7 [63]. If the set of clauses contains non-constant free function symbols, then *standard superposition* and *abstraction* inferences have to be taken into account. While the former are harmless, the latter are 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.

Word Problems. As a special case of Lemma 5.13 we can decide the word problem for Ψ-torsion-free cancellative abelian monoids whenever membership in Ψ is decidable. That means that we can check whether some equation

$$e \;=\; m_1 b_1 + \cdots + m_i b_i \approx n_1 c_1 + \cdots + n_j c_j$$

holds in every Ψ-torsion-free cancellative abelian monoid in which the finite set of equations

$$e_k \;=\; m_1^k b_1^k + \cdots + m_i^k b_i^k \approx n_1^k c_1^k + \cdots + n_j^k c_j^k$$

for $k \in K$ hold. All we have to do is to negate the first equation and check the set of unit clauses $\{\neg\, e\} \cup \{\, e_k \mid k \in K\,\}$ for unsatisfiability with respect to ACUKT_Ψ.

Can we also solve the word problem for Ψ-torsion-free abelian groups? Consider the equation

$$e' \;=\; m_1 b_1 + \cdots + m_i b_i + n_1(-c_1) + \cdots + n_j(-c_j) \approx 0$$

and the finite set of equations

$$e'_k \;=\; m_1^k b_1^k + \cdots + m_i^k b_i^k + n_1^k(-c_1^k) + \cdots + n_j^k(-c_j^k) \approx 0$$

for $k \in K$. We want to know whether e' holds in every Ψ-torsion-free group in which the equations e'_k hold, or equivalently, whether

$$\{\neg\, e'\} \cup \{\, e'_k \mid k \in K\,\} \cup \{\,(-x) + x \approx 0\,\} \cup \mathrm{ACUKT}_\Psi$$

is unsatisfiable. It turns out that the word problems for Ψ-torsion-free cancellative abelian monoids and for Ψ-torsion-free abelian groups are essentially the same:

LEMMA 5.14 *Let the equations e, e_k, e', e'_k be defined as above, let* $N = \{\neg\, e\} \cup \{\, e_k \mid k \in K\,\}$ *and* $N' = \{\neg\, e'\} \cup \{\, e'_k \mid k \in K\,\}$. *Then* $N \cup \mathrm{ACUKT}_\Psi$ *is satisfiable if and only if* $N' \cup \{\, (-x) + x \approx 0\,\} \cup \mathrm{ACUKT}_\Psi$ *is satisfiable.*

PROOF. By the inverse axiom $(-x) + x \approx 0$, the clause $\neg\, e'$ is equivalent to $\neg\, e$ and each e'_k is equivalent to e_k. Hence $N' \cup \{\, (-x) + x \approx 0\,\} \cup \mathrm{ACUKT}_\Psi$ is satisfiable if and only if $N \cup \{\, (-x) + x \approx 0\,\} \cup \mathrm{ACUKT}_\Psi$ is satisfiable. We check satisfiability of this set of clauses using the cancellative superposition calculus. Obviously, the inverse axiom cannot take part in any *cancellation*, *equality resolution*, or *equality factoring* inference. As the subterm $-x$ is strictly maximal in the inverse axiom, and the negation function occurs in no other axiom, the inverse axiom does not take part in any *cancellative superposition* inference. Furthermore, *abstraction* and *standard superposition* inferences are excluded since the term occurring below the negation function is a variable. Hence there are no inferences at all with the inverse axiom. This implies that there is a *CS-Inf*$_\Psi$-derivation from $N \cup \{\, (-x) + x \approx 0\,\}$ producing $\bot$ if and only if there is such a derivation starting with N. In other words, $N \cup \{\, (-x) + x \approx 0\,\} \cup \mathrm{ACUKT}_\Psi$ is satisfiable if and only if $N \cup \mathrm{ACUKT}_\Psi$ is. □

GCD Superposition. The following example illustrates a useful optimization for unit ground clauses that we have not yet mentioned.

EXAMPLE 5.15 Consider the two clauses

$$\underline{24b} + c \approx d \tag{1}$$

$$\underline{9b} + c' \approx d' \tag{2}$$

By *cancellative superposition* of (2) and (1) we obtain

$$\underline{15b} + c + d' \approx d + c' \tag{3}$$

and may delete (1). By *cancellative superposition* of (2) and (3) we obtain

$$\underline{6b} + c + 2d' \approx d + 2c' \tag{4}$$

and may delete (3). Now *cancellative superposition* of (4) and (2) yields

$$\underline{3b} + 3c' + d \approx 3d' + c \tag{5}$$

followed by deletion of (2). From (5) and (4) we derive

$$\underline{3b} + 2c + 5d' \approx 2d + 5c' \tag{6}$$

and may delete (4). Finally, *cancellative superposition* of (5) and (6) produces

$$\underline{3c} + 8d' \approx 3d + 8c' \tag{7}$$

and allows us to delete (6).

It is evident that the sequence of steps we have performed is nothing else than Euclid's algorithm. Starting from two clauses (1) and (2) with the same maximal term b, we get in the end clause (7), in which b has been erased completely, and clause (5), in which the coefficient of b is the greatest common divisor of the coefficients of b in (1) and (2). All other clauses, including (1) and (2), have become redundant eventually.

A refinement of *positive cancellative superposition* for ground unit clauses allows us to derive the clauses (5) and (7) of the example above directly, bypassing the intermediate clauses (3), (4), and (6). This operation is called *gcd superposition* in (Stuber [96]):

Suppose we have ground clauses $C_1 = mu + s \approx s'$ and $C_2 = nu + t \approx t'$ such that u is larger than s, s', t, and t', $C_1 \succeq_{\mathrm{C}} C_2$, and thus $m \geq n$. If m is divisible by n, we can obviously simplify C_1 to $s + (m/n)t' \approx s' + (m/n)t$.

Otherwise, we can make use of an algorithm of Knuth [56] that allows to compute natural numbers λ, μ, and ν such that either (i) $\mu m - \nu n = \lambda = \gcd(m, n)$ or (ii) $\nu n - \mu m = \lambda = \gcd(m, n)$. We will demonstrate that both C_1 and C_2 may be simplified to two new clauses D_1 and D_2, where

$$D_1 \;=\; \lambda u + \mu s + \nu t' \approx \mu s' + \nu t$$

in case (i), or

$$D_1 \;=\; \lambda u + \mu s' + \nu t \approx \mu s + \nu t'$$

in case (ii), and

$$D_2 \;=\; (n/\lambda)s + (m/\lambda)t' \approx (n/\lambda)s' + (m/\lambda)t\,.$$

Let us assume case (i); case (ii) is similar. First we show that both D_1 and D_2 follow from C_1 and C_2. Observe that we obtain D_1 if we multiply the left and right-hand side of C_2 by ν, flip sides, multiply C_1 by μ, add left and right-hand sides, respectively, and cancel νnu. Similarly, if we multiply C_2 by m/λ, flip sides, multiply C_1 by n/λ, add left and right-hand sides, respectively, and cancel $(mn/\lambda)u$, we obtain D_2.

To prove that D_1 and D_2 make C_1 and C_2 redundant we have to show that both D_1 and D_2 are smaller than C_1 and C_2, which is clear since $\lambda < n \leq m$, and we have to show that D_1 and D_2 entail C_1 and C_2. To show that

$\{D_1, D_2\} \models_\Psi C_1$, we note that we obtain C_1 if we multiply D_2 by ν, flip sides, multiply D_1 by m/λ, add left and right-hand sides, respectively, and cancel $(\nu m/\lambda)(t + t') + (\nu n/\lambda)(s + s')$. In the same way, we can multiply D_2 by μ, flip sides, multiply D_1 by n/λ, add left and right-hand sides, respectively, cancel $(\mu n/\lambda)(s + s') + (\nu n/\lambda)(t + t')$, and get C_2.

Similarly to the Gaussian elimination algorithm, every set of n positive ground unit clauses over $+$ and constants can be saturated using only *gcd superposition* and simplification by *cancellation*. At most $n(n-1)/2$ gcd superpositions are necessary. The resulting set of clauses is in triangular form: no two clauses have the same maximal term.[34]

The *gcd superposition* rule can be extended to non-ground non-unit clauses $C_1' \vee mu + s \approx s' \llbracket T_1 \rrbracket$ and $C_2' \vee nu + t \approx t' \llbracket T_2 \rrbracket$. Provided that $mu + s \approx s'$ and $nu + t \approx t'$ are strictly the largest literals in the respective clauses, u is larger than s, s', t, and t', and $C_1' = C_2'$ and $T_1 = T_2$, it is again a simplification of both the original clauses.

5.3 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 section, we will concentrate on simplification and quasisimplification techniques that help to reduce the number of clauses with eligible variables. As we will show, certain clauses with eligible variables can be removed from the clause set. Others can not be removed, but at least, they can be equipped with more restrictive constraints.

LEMMA 5.16 *Let $C \llbracket T \rrbracket$ be a clause*

$$\bigvee_{i \in I} n_i x + s_i \approx t_i \; \llbracket T \rrbracket$$

where x occurs neither in one of the t_i nor in T, and T is satisfiable. Then in every term-generated normal model $\mathfrak{M}$ of $\{C \llbracket T \rrbracket\} \cup \mathrm{ACK}$, the set $S^{\mathfrak{M}}_{\mathrm{CAM}}$ is a group under the operation $+^{\mathfrak{M}}$.

[34]This algorithm is very similar to Kandri-Rody, Kapur, and Narendran's procedure to solve the word problem in finitely presented abelian groups [52]. Essentially the same construction can also be found in algorithms to compute canonical structures of abelian groups, for instance in (Iliopoulos [49]).

PROOF. Let θ be a substitution that satisfies $T\theta = \text{true}$ and maps all variables in C except x to constants. Let $\mathfrak{M}$ be a term-generated normal model of $\{C [\![T]\!]\} \cup \text{ACK}$, with $S^{\mathfrak{M}}_{\text{CAM}}$ being the set corresponding to the sort S_{CAM}. Since $\mathfrak{M}$ satisfies ACK, $S^{\mathfrak{M}}_{\text{CAM}}$ is a cancellative abelian semigroup under $+^{\mathfrak{M}}$. Furthermore, $\mathfrak{M}$ satisfies $C\theta$, hence for every $\mathfrak{m} \in S^{\mathfrak{M}}_{\text{CAM}}$ 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, $\alpha(t_i\theta) = \mathfrak{M}(t_i\theta)$ is independent of α. Now $\{\, \mathfrak{M}(t_i\theta) \mid i \in I \,\}$ is finite, hence by Thm. A.6, $S^{\mathfrak{M}}_{\text{CAM}}$ is a group. □

LEMMA 5.17 *Let $C [\![T]\!]$ be a clause*

$$C' \vee \bigvee_{i \in I} n_i x + s_i \approx t_i \; [\![T]\!]$$

where x occurs neither in C', nor in the t_i, nor in T. Then for every term-generated normal model $\mathfrak{M}$ of $\{C [\![T]\!]\} \cup \text{ACK}$, either $\mathfrak{M}$ is a model of $C' [\![T]\!]$, or $S^{\mathfrak{M}}_{\text{CAM}}$ is a group under the operation $+^{\mathfrak{M}}$.

PROOF. If T is unsatisfiable, then this is obvious since neither $C [\![T]\!]$ nor $C' [\![T]\!]$ have any instances. Otherwise suppose that there is a term-generated normal model $\mathfrak{M}$ of $\{C [\![T]\!]\} \cup \text{ACK}$ that is not a model of $C' [\![T]\!]$. Let θ be a substitution that satisfies $T\theta = \text{true}$ and maps all variables in C except x to constants, such that the ground clause $C'\theta$ is false in $\mathfrak{M}$. Consequently,

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

must be true in $\mathfrak{M}$. By Lemma 5.16, $S^{\mathfrak{M}}_{\text{CAM}}$ 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. [16]). 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 and T is satisfiable, then the first branch can be closed immediately. In this case C implies the identity and inverse axioms, and, although it not required by fairness, it may be wise to add them to find an easier proof. Second, if some subset N' of N implies that, for every model $\mathfrak{M}$, $S^{\mathfrak{M}}_{\text{CAM}}$ 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, $C [\![T]\!]$ can be simplified to $C' [\![T]\!]$. 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 5.18 *Let* $\Psi = \mathbf{N}^{>0}$ *and let* $C\,[\![T]\!]$ *be a clause*

$$\bigvee_{i\in I} n_i x + s_i \approx t_i \;[\![T]\!]$$

where x *occurs neither in one of the* s_i *or* t_i *nor in* T, *and* T *is satisfiable. Then in every term-generated normal model* $\mathfrak{M}$ *of* $\{C\,[\![T]\!]\} \cup \mathrm{ACUKT}_\Psi$, *the set* $S^{\mathfrak{M}}_{\mathrm{CAM}}$ *is a singleton.*

PROOF. Let θ be a substitution that satisfies $T\theta = \mathrm{true}$ and maps all variables in C except x to constants. Let $\mathfrak{M}$ be a term-generated normal model of $\{C\,[\![T]\!]\} \cup \mathrm{ACUKT}_\Psi$, with $S^{\mathfrak{M}}_{\mathrm{CAM}}$ being the set corresponding to the sort S_{CAM}. By Lemma 5.16, $S^{\mathfrak{M}}_{\mathrm{CAM}}$ is an abelian group with addition $+^{\mathfrak{M}}$, and since $\mathfrak{M}$ satisfies ACUKT_Ψ, it is torsion-free. Let $-^{\mathfrak{M}}$ denote the implicitly given inverse operation of the group $S^{\mathfrak{M}}_{\mathrm{CAM}}$. The model $\mathfrak{M}$ satisfies $C\theta$, hence for every $\mathfrak{m} \in S^{\mathfrak{M}}_{\mathrm{CAM}}$ there is an $i \in I$ such that $n_i\mathfrak{m} +^{\mathfrak{M}} \alpha(s_i\theta) = \alpha(t_i\theta)$, where α is the assignment mapping x to $\mathfrak{m}$. As $s_i\theta$ and $t_i\theta$ are ground, $\alpha(s_i\theta) = \mathfrak{M}(s_i\theta)$ and $\alpha(t_i\theta) = \mathfrak{M}(t_i\theta)$ are independent of α, hence for every $\mathfrak{m} \in S^{\mathfrak{M}}_{\mathrm{CAM}}$ there is an $i \in I$ such that $n_i\mathfrak{m} = \mathfrak{M}(t_i\theta) +^{\mathfrak{M}} (-^{\mathfrak{M}}\,\mathfrak{M}(s_i\theta))$. Now $\{\, \mathfrak{M}(t_i\theta) +^{\mathfrak{M}} (-^{\mathfrak{M}}\,\mathfrak{M}(s_i\theta)) \mid i \in I \,\}$ is finite, hence by Cor. A.11, $S^{\mathfrak{M}}_{\mathrm{CAM}}$ is a singleton. □

LEMMA 5.19 *Let* $\Psi = \mathbf{N}^{>0}$ *and let* $C\,[\![T]\!]$ *be a clause*

$$C' \vee \bigvee_{i\in I} n_i x + s_i \approx t_i \;[\![T]\!]$$

where x *occurs neither in* C', *nor in the* s_i *or* t_i, *nor in* T. *Then for every term-generated normal model* $\mathfrak{M}$ *of* $\{C\,[\![T]\!]\} \cup \mathrm{ACUKT}_\Psi$, *either* $\mathfrak{M}$ *is a model of* $C'\,[\![T]\!]$, *or* $S^{\mathfrak{M}}_{\mathrm{CAM}}$ *is a singleton.*

PROOF. This is proved analogously to Lemma 5.17. □

In the presence of a non-triviality axiom[35]

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

(which excludes singleton models), Lemma 5.19 allows us to eliminate every literal with unshielded variables that are unconstrained and occur only positively in a clause.[36] If a variable is unshielded but constrained, we can either

[35] In non-skolemized form: $\exists y\colon y \not\approx 0$.

[36] We assume that variables occurring on both sides of an equation are cancelled immediately.

drop the constraint (if it is an ordering constraint, thus non-essential), or solve the constraint partially as mentioned earlier (if it is an equality constraint). This is done in the beginning for all initially given clauses, and later once for every clause that is added as the result of an inference. In Section 5.5 we will exploit this possibility to construct a finite branching variant of the inference system *CS-Inf*$_\Psi$ for $\Psi = \mathbf{N}^{>0}$.

If a non-triviality axiom is not initially given, then Lemma 5.19 suggests to prove 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 neither in the constraint nor in positive literals, and only in one negative literal, 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

$$C \vee \neg\, x + s \approx s' \; [\![T]\!],$$

where x does not occur in C and T, can be simplified to $C\; [\![T]\!]$. If the coefficient is different from 1, the inverse axiom alone does not yield enough information to eliminate the literal. In this case, we need additionally division operators *divided-by*$_n$ and divisibility axioms[37]

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

(for $n \in \mathbf{N}^{>0}$). This will be further investigated in Section 5.6.

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 $C_0\; [\![T]\!]$ of the form

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

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

where every variable of C_0 occurs either negatively in C_0, or it is shielded in C_0, or is not lower bounded in $C_0\,[\![T]\!]$. If $m \geq n \geq 1$, then $C_0\,[\![T]\!]$ is equivalent to

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

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

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

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 more applicable we obtain a clause $C_1\,[\![T]\!]$ 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.

One way to cope with this problem is to use case splitting, that is, to replace $C_0\,[\![T]\!]$ by the two new clauses $C_0\,[\![T \wedge C_1 \succeq_{\mathrm{C}} C_0]\!]$ and $C_0\,[\![T \wedge C_1 \prec_{\mathrm{C}} C_0]\!]$, where only the latter can be simplified to $C_1\,[\![T \wedge C_1 \prec_{\mathrm{C}} C_0]\!]$. It should be noted that $C_1 \prec_{\mathrm{C}} C_0$ is satisfied in particular by all substitutions θ for which $\mathrm{mt}(x\theta) \succ \max\{\mathrm{mt}(s\theta), \mathrm{mt}(s'\theta), \mathrm{mt}(t\theta), \mathrm{mt}(t'\theta)\}$. Thus even though the simplification is only partial, it renders all superposition inferences with C_0 redundant that involve only x but no subterm of s, s', t, or t'. An alternative method to integrate this kind of variable elimination into cancellative superposition will be presented in Section 5.6.

5.4 The Standard Case: $\Psi = \{1\}$

It is easy to see that the inference system *CS-Inf*$_\Psi$ 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 *CS-Inf*$_{\{1\}}$ can be refined to a finitely branching system *WCS-Inf*$_{\{1\}}$. In *WCS-Inf*$_{\{1\}}$, 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 chapter 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 $C_0' = D'\theta \vee C'\theta \vee \neg\, (\bar{m}{-}\bar{n})\bar{u} + \bar{s} + \bar{t}' \approx \bar{t} + \bar{s}'$. Just as C_0, the clause C_0' is smaller than 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 see how a finitely branching variant of the *cancellative equality factoring* rule can be obtained, we consider the size of the literals in a ground *cancellative equality factoring* inference

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

where $\chi\bar{m} + \bar{n}' = \bar{n}$ and $\chi \in \mathbf{N}^{>0}$. As $\chi\bar{m} + \bar{n}' = \bar{n}$, we have $\bar{n} \geq \bar{m}$. 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 $\bar{m} = \bar{n}$, $\bar{n}' = 0$ and $\chi = 1$. With these additional restrictions, the lifted rule becomes finitary branching; furthermore, as $\bar{n}' = 0$, unshielded variables occurring in $\mathrm{rhs}(e_2)$ in the *cancellative equality factoring* rule can be ignored.

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. For a given second premise $C' \vee [\neg]\ e_1[w]\,[\![T_1]\!]$ and term w to be abstracted out, several distinct conclusions can be derived, but these conclusions differ only in constraints which are irrelevant for correctness anyway. It is plausible here to content oneself with an inference rule that produces only a single conclusion with a slightly weaker constraint, subsuming all conclusions of the original rule. To obtain such a rule, we note that the constraint T_{E} of the *abstraction* rule is always satisfiable whenever the first premise $D' \vee e_2$ contains an unshielded variable in $\mathrm{lhs}(e_2)$. Similarly, it is not worth spending time in equality constraint solving, 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 to consider the case that $e_2 = \sum_{l \in L} n_l^* v_l + t \approx t'$ and $w = \sum_{k \in K} m_k^* u_k + q$ with atomic terms u_k and v_l. In this case the equality constraint T_{E} boils down to checking that all u_k and v_l are ACU-unifiable and $\sum_{k \in K} m_k^* \geq \sum_{l \in L} n_l^*$.

INFERENCE SYSTEM 5.20 The inference system *WCS-Inf*$_{\{1\}}$ consists of the inference rules *cancellation*, *equality resolution*, *standard superposition*, *positive cancellative superposition*, and *standard equality factoring* of Inference System 3.18, and of the inference rules *weak negative cancellative superposition*, *weak abstraction*, and *weak cancellative equality factoring*, as described below.

Weak Negative Cancellative Superposition

$$\frac{D' \vee e_2 \; [\![T_2]\!] \qquad C' \vee \neg\, e_1 \; [\![T_1]\!]}{D' \vee C' \vee \neg\, e_0 \; [\![T_2 \wedge T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- $e_1 \;=\; \sum_{i \in I} m_i x_i + \sum_{k \in K} m_k^* u_k + s \approx s'$.
- $e_2 \;=\; \sum_{j \in J} n_j y_j + \sum_{l \in L} n_l^* v_l + t \approx t'$.
- $e_0 \;=\; z\sigma + \sum_{i \in I} m_i \check{x}_i + s + t' \approx \sum_{j \in J} n_j \check{y}_j + t + s'$.
- $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
- $\{\, x_i \mid i \in I \,\} = \mathrm{elig}(C' \vee \neg e_1) \cap \mathrm{var}(\mathrm{lhs}(e_1)) \setminus \mathrm{var}(\mathrm{rhs}(e_1))$,
  $\{\, y_j \mid j \in J \,\} = \mathrm{elig}(D' \vee e_2) \cap \mathrm{var}(\mathrm{lhs}(e_2)) \setminus \mathrm{var}(\mathrm{rhs}(e_2)) \setminus \mathrm{var}(\mathrm{neg}(D'))$.
- $\mathrm{lhs}(e_1)$ is not a variable (i. e., either $\sum_{i \in I} m_i > 1$ or $\sum_{k \in K} m_k^* u_k + s \neq 0$).
- If $I = \{i_1\}$, $m_{i_1} = 1$, and $K = \emptyset$, then $J \neq \emptyset$ or $t \neq 0$.
- 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.
- σ is a most general ACU-unifier of $\sum_{i \in I} m_i \hat{x}_i + \big(\sum_{k \in K} m_k^*\big) u$ and $z + \sum_{j \in J} n_j \hat{y}_j + \big(\sum_{l \in L} n_l^*\big) u$.
- $T_{\mathrm{E}} = \bigwedge_{i \in I} x_i \doteq \hat{x}_i + \check{x}_i \;\wedge\; \bigwedge_{j \in J} y_j \doteq \hat{y}_j + \check{y}_j \;\wedge\; \bigwedge_{k \in K} u_k \doteq u \;\wedge\; \bigwedge_{l \in L} v_l \doteq u \;\wedge$ $\mathrm{EQ}(\sigma)$.
- $T_{\mathrm{O}} = u \succ s \;\wedge\; u \succ s' \;\wedge\; u \succ t \;\wedge\; u \succ t' \;\wedge\; \bigwedge_{i \in I} u \succ \check{x}_i \;\wedge\; \bigwedge_{j \in J} u \succ \check{y}_j \;\wedge\; T_{\mathrm{O}}^{\mathrm{Lit}}$.

Weak Abstraction

$$\frac{D' \lor e_2 \; [\![T_2]\!] \qquad C' \lor [\neg]\, s[w] \approx s' \; [\![T_1]\!]}{C' \lor \neg\, y \approx w \lor [\neg]\, s[y] \approx s' \; [\![T_1 \land T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- w occurs in s immediately below some free function symbol and has sort S_{CAM}.
- Either:
  - $\mathrm{elig}(D' \lor e_2) \cap \mathrm{var}(\mathrm{lhs}(e_2)) \setminus \mathrm{var}(\mathrm{rhs}(e_2)) \setminus \mathrm{var}(\mathrm{neg}(D')) \neq \emptyset$.
  - w is not a variable.

  or:

  - $e_2 \;=\; \sum_{l \in L} n_l^* v_l + t \approx t'$.
  - $L \neq \emptyset$.
  - Either:
    * $w \;=\; mx + q$.
    * 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.
    * $\mathrm{lhs}(e_2)$ is neither an atomic term nor a subterm of w.
    * There is a ground instance $(D' \lor e_2)\theta$ of the first premise such that $v_l\theta \succ t\theta$, $v_l\theta \succ t'\theta$ for all $l \in L$.

    or:

    * $w \;=\; \sum_{k \in K} m_k^* u_k + q$.
    * ρ is a most general ACU-unifier of all u_k and v_l $(k \in K, l \in L)$.
    * $\sum_{k \in K} m_k^* \geq \sum_{l \in L} n_l^*$.
    * $\mathrm{lhs}(e_2)\rho$ is not a subterm of $w\rho$.
    * There is a ground instance $(D' \lor e_2)\rho\theta$ of the first premise such that $v_l\rho\theta \succ t\rho\theta$, $v_l\rho\theta \succ t'\rho\theta$ for all $l \in L$.
- $T_{\mathrm{O}} = w \succ y \;\land\; s[w] \succ s' \;\land\; T_{\mathrm{O}}^{\mathrm{Lit}}$.

Weak Cancellative Equality Factoring

$$\frac{C' \vee e_2 \vee e_1 \, [\![T_1]\!]}{C' \vee \neg\, e_0 \vee e_2 \, [\![T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- $e_1 \;=\; \sum_{i \in I} m_i x_i + \sum_{k \in K} m_k^* u_k + s \approx s'$.
- $e_2 \;=\; \sum_{j \in J} n_j x_j + \sum_{l \in L} n_l^* v_l + t \approx t'$.
- $e_0 \;=\; \sum_{j \in J} n_j \check{x}_j + t + s' \approx \sum_{i \in I} m_i \check{x}_i + s + t'$.
- $I \cup K \neq \emptyset$ and $J \cup L \neq \emptyset$.
- $\{ x_i \mid i \in I \} = \mathrm{elig}(C' \vee e_2 \vee e_1) \cap \mathrm{var}(\mathrm{lhs}(e_1)) \setminus \mathrm{var}(\mathrm{rhs}(e_1)) \setminus \mathrm{var}(\mathrm{rhs}(e_2)) \setminus \mathrm{var}(\mathrm{neg}(C'))$,
  $\{ x_j \mid j \in J \} = \mathrm{elig}(C' \vee e_2 \vee e_1) \cap \mathrm{var}(\mathrm{lhs}(e_2)) \setminus \mathrm{var}(\mathrm{rhs}(e_1)) \setminus \mathrm{var}(\mathrm{rhs}(e_2)) \setminus \mathrm{var}(\mathrm{neg}(C'))$.
- 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.
- $T_{\mathrm{E}} = \bigwedge_{i \in I \cup J} x_i \doteq \hat{x}_i + \check{x}_i \wedge \bigwedge_{k \in K} u_k \doteq u \wedge \bigwedge_{l \in L} v_l \doteq u \wedge \sum_{i \in I} m_i \hat{x}_i + (\sum_{k \in K} m_k^*) u \doteq \sum_{j \in J} n_j \hat{x}_j + (\sum_{l \in L} n_l^*) u$.
- $T_{\mathrm{O}} = u \succ s \wedge u \succ s' \wedge u \succ t \wedge u \succ t' \wedge \bigwedge_{i \in I \cup J} u \succ \check{x}_i \wedge T_{\mathrm{O}}^{\mathrm{Lit}}$.

Instances of *weak negative cancellative superposition*, *weak abstraction*, and *weak cancellative equality factoring* inferences are defined as previously in Def. 4.9 and 4.10. With these implements, we obtain again lifting lemmas:

LEMMA 5.21 *Let R be a stratified set of ground equations. Let $D\,[\![T_2]\!]$ and $C\,[\![T_1]\!]$ be two constrained clauses (without common variables) and let θ be a substitution such that $T_2\theta = T_1\theta = \mathrm{true}$ and $D\theta$ and $C\theta$ are ground R-variable minimal instances. Let $D = D' \vee e_2$ such that $e_2\theta$ is strictly maximal in $D\theta$ and $e_2\theta \in R$; let $C = C' \vee [\neg]\, e_1$.*

If there is a weak 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 $\text{lhs}(e_2\theta)$ *and* $\text{lhs}(e_1\theta)$ *are overlapped), and* $\text{lhs}(e_1)$ *is not a variable, then the inference is an* R*-variable minimal ground instance of a* weak negative cancellative superposition *inference from* $D\,[\![T_2]\!]$ *and* $C\,[\![T_1]\!]$.

If there is a weak abstraction *inference*

$$\frac{D'\theta \vee e_2\theta \qquad C'\theta \vee [\neg]\, e_1\theta[\bar{m}\bar{v} + \bar{q}]}{C_0\ [\![T_0]\!]}$$

such that

- C_0 *equals* $C'\theta \vee \neg\, y \approx \bar{m}\bar{v} + \bar{q} \vee [\neg]\, e_1\theta[y]$,
- $\bar{m}\bar{v} + \bar{q} = w\theta$ *for some subterm* w *of* $\text{lhs}(e_1)$,
- $\bar{m}\bar{v} + \bar{q}$ *is not a subterm of* $y'\theta$ *for any* $y' \in \text{var}(\text{lhs}(e_1))$,
- *the maximal atomic subterm of* $\text{lhs}(e_2\theta)$ *equals* $\bar{v}$,
- *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 an R*-variable minimal ground instance of a* weak abstraction *inference from* $D\,[\![T_2]\!]$ *and* $C\,[\![T_1]\!]$.

LEMMA 5.22 *Let* R *be a stratified set of ground equations. Let* $C\,[\![T_1]\!]$ *be a constrained clause and let* θ *be a substitution such that* $T_1\theta = \text{true}$ *and* $C\theta$ *is a ground* R*-variable minimal instance. Then every* weak cancellative equality factoring *inference from* $C\theta$ *is an* R*-variable minimal ground instance of a* weak cancellative equality factoring *inference from* $C\,[\![T_1]\!]$.

The definitions of redundancy (Def. 4.12 and 4.13) are extended verbatim to the new kinds of inferences. It is easy to show that Lemma 4.15 holds also for Inference System 5.20, i. e., that $\mathit{CS\text{-}Red}_{\{1\}}$ is also a redundancy criterion with respect to $\mathit{WCS\text{-}Inf}_{\{1\}}$.

LEMMA 5.23 *Let* N *be a set of constrained clauses. 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.

PROOF. Let $D\,[\![T_2]\!] = D' \vee e_2\ [\![T_2]\!]$ and $C\,[\![T_1]\!] = C' \vee \neg\, e_1\ [\![T_1]\!]$ be two clauses in N and let ι be a *negative cancellative superposition* inference

$$\frac{D\,[\![T_2]\!] \qquad C\,[\![T_1]\!]}{C_0\,[\![T_0]\!]}$$

By definition of the inference rule we know that $\mathrm{lhs}(e_1)$ is not a variable. Let R be a stratified set of ground equations and let $\iota\theta$ be an R-variable minimal ground instance

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

such that $\bar{m} = \chi\bar{n}$ and $e_2\theta \in R$. Then there exists a ground *weak negative cancellative superposition* inference $\bar{\iota}'$ from $D\theta$ and $C\theta$, namely

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

By Lemma 5.21, $\bar{\iota}'$ is an R-variable minimal ground instance of a *weak negative cancellative superposition* inference ι' of the form

$$\frac{D\,[\![T_2]\!] \qquad C\,[\![T_1]\!]}{C_0'\,[\![T_0']\!]}$$

As ι' is redundant with respect to N, $R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C_0'\theta$. Furthermore, $\bar{n}\bar{u} + \bar{t} \approx \bar{t}'$ and $\neg\, (\bar{m}-\bar{n})\bar{u} + \bar{s} + \bar{t}' \approx \bar{t} + \bar{s}'$ entail $\neg\, \bar{s} + \chi\bar{t}' \approx \chi\bar{t} + \bar{s}'$, hence $\{D\theta, C_0'\theta\} \models_\Psi C_0\theta$. Since $D\theta \in \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta}$, we obtain $R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C_0\theta$. This proves that ι is redundant with respect to N. □

The following two lemmas are proved analogously.

LEMMA 5.24 *Let N be a set of constrained clauses. If every weak abstraction inference from clauses in N is redundant with respect to N, then every abstraction inference from clauses in N is redundant with respect to N.*

LEMMA 5.25 *Let N be a set of constrained clauses. If every weak cancellative equality factoring inference from clauses in N is redundant with respect to N, then every cancellative equality factoring inference from clauses in N is redundant with respect to N.*

The following theorem is now obvious.

THEOREM 5.26 *If a set of clauses if saturated with respect to $WCS\text{-}Inf_{\{1\}}$ and $CS\text{-}Red_{\{1\}}$, then it is also saturated with respect to $CS\text{-}Inf_{\{1\}}$ and $CS\text{-}Red_{\{1\}}$.*

This means that we can obtain a $CS\text{-}Inf_{\{1\}}$-saturated set of clauses not only as the limit of a fair $CS\text{-}Inf_{\{1\}}$-derivation, but also as the limit of a fair $WCS\text{-}Inf_{\{1\}}$-derivation. Theorem 4.57 can therefore be extended from $CS\text{-}Inf_{\{1\}}$-derivations to $WCS\text{-}Inf_{\{1\}}$-derivations.

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

In Section 5.3 we have shown that in the $\mathbf{N}^{>0}$-torsion-free case all unshielded variables that occur only positively can be eliminated. We will now construct a finitely branching inference system that is equivalent to *CS-Inf*$_{\mathbf{N}^{>0}}$, provided that the elimination of unshielded variables is performed eagerly. Again, we replace the inference rules *negative cancellative superposition* and *abstraction* by new rules, whose redundancy implies the redundancy of the old rules.

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 $\psi\bar{m} = \chi\bar{n}$, $\psi \in \Psi$, 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 Section 5.4.

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 $\psi\bar{m} = \chi\bar{n}$, $\psi \in \Psi$, 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 to the non-ground level 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 $\bar{m}$ as well. 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 that we present below is constructed according to the same principles as the corresponding rule of Inference System 5.20. The 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 3.18 finitely branching turns out to be almost trivial: In the absence of unshielded variables occurring only positively, the constraint T_{E} of this rule can have a solution only if $\chi = n^*/\gcd(m^*, n^*)$ and $\psi = m^*/\gcd(m^*, n^*)$, where $m^* = \sum_{k \in K} m^*_k$ and $n^* = \sum_{l \in L} n^*_l - \sum_{l \in L'} n^{*\prime}_l$.

INFERENCE SYSTEM 5.27 The inference system $\mathit{WCS\text{-}Inf}_{\mathbf{N}^{>0}}$ consists of the inference rules *cancellation*, *equality resolution*, *standard superposition*, *positive cancellative superposition*, *standard equality factoring*, and *cancellative equality factoring* of Inference System 3.18, and of the inference rules *weak negative cancellative superposition (I)*, *weak negative cancellative superposition (II)*, and *weak abstraction*, as described below. (Below we also present the inference rule *cancellative equality factoring*. This is not a new rule but only the obvious specialization of the corresponding rule of Inference System 3.18 to clauses without unshielded variables occurring only positively.)

Weak Negative Cancellative Superposition (I)

$$\frac{D' \vee e_2 \; [\![T_2]\!] \qquad C' \vee \neg\, e_1 \; [\![T_1]\!]}{D' \vee C' \vee \neg\, e_0 \; [\![T_2 \wedge T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- $e_1 \;=\; \sum_{i \in I} m_i x_i + \sum_{k \in K} m^*_k u_k + s \approx s'$.
- $e_2 \;=\; \sum_{l \in L} n^*_l v_l + t \approx t'$.
- $e_0 \;=\; z\sigma + \sum_{i \in I} m_i \check{x}_i + s + t' \approx t + s'$.
- $I \cup K \neq \emptyset$ and $L \neq \emptyset$.
- $\{\, x_i \mid i \in I \,\} = \mathrm{elig}(C' \vee \neg e_1) \cap \mathrm{var}(\mathrm{lhs}(e_1)) \setminus \mathrm{var}(\mathrm{rhs}(e_1))$.
- $\mathrm{lhs}(e_1)$ is not a variable (i. e., either $\sum_{i \in I} m_i > 1$ or $\sum_{k \in K} m^*_k u_k + s \neq 0$).
- If $I = \{i_1\}$, $m_{i_1} = 1$, and $K = \emptyset$, then $t \neq 0$.
- u is one of the u_k or v_l ($k \in K$, $l \in L$).

- σ is a most general ACU-unifier of $\sum_{i \in I} m_i \hat{x}_i + \big(\sum_{k \in K} m_k^*\big) u$ and $z + \big(\sum_{l \in L} n_l^*\big) u$.

- $T_{\mathrm{E}} = \bigwedge_{i \in I} x_i \doteq \hat{x}_i + \check{x}_i \,\wedge\, \bigwedge_{k \in K} u_k \doteq u \,\wedge\, \bigwedge_{l \in L} v_l \doteq u \,\wedge\, \mathrm{EQ}(\sigma)$.

- $T_{\mathrm{O}} = u \succ s \,\wedge\, u \succ s' \,\wedge\, u \succ t \,\wedge\, u \succ t' \,\wedge\, \bigwedge_{i \in I} u \succ \check{x}_i \,\wedge\, T_{\mathrm{O}}^{\mathrm{Lit}}$.

Weak Negative Cancellative Superposition (II)

$$\frac{D' \,\vee\, e_2 \; [\![T_2]\!] \qquad C' \,\vee\, \neg\, e_1 \; [\![T_1]\!]}{D' \,\vee\, C' \,\vee\, \neg\, e_0 \; [\![T_2 \wedge T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

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

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

- $e_0 \;=\; \displaystyle\sum_{i \in I} \psi m_i \check{x}_i + \psi s + \chi t' \approx \chi t + \psi s'$.

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

- $\{\, x_i \mid i \in I \,\} = \mathrm{elig}(C' \vee \neg e_1) \cap \mathrm{var}(\mathrm{lhs}(e_1)) \setminus \mathrm{var}(\mathrm{rhs}(e_1))$.

- $\mathrm{lhs}(e_1)$ is not a variable (i. e., either $\sum_{i \in I} m_i > 1$ or $\sum_{k \in K} m_k^* u_k + s \neq 0$).

- If $I = \{i_1\}$, $m_{i_1} = 1$, and $K = \emptyset$, then $\mathrm{lhs}(e_2)$ is not an atomic term.

- $0 < \bar{m} < n^* = \sum_{l \in L} n_l^*$.

- $\chi = \bar{m} / \gcd(\bar{m}, n^*)$, $\psi = n^* / \gcd(\bar{m}, n^*)$.

- u is one of the u_k or v_l ($k \in K$, $l \in L$).

- $T_{\mathrm{E}} \;=\; \bigwedge_{i \in I} x_i \doteq \hat{x}_i + \check{x}_i \,\wedge\, \bigwedge_{k \in K} u_k \doteq u \,\wedge\, \bigwedge_{l \in L} v_l \doteq u \,\wedge\, \sum_{i \in I} m_i \hat{x}_i + \sum_{k \in K} m_k^* u \doteq \bar{m} u$.

- $T_{\mathrm{O}} = u \succ s \,\wedge\, u \succ s' \,\wedge\, u \succ t \,\wedge\, u \succ t' \,\wedge\, \bigwedge_{i \in I} u \succ \check{x}_i \,\wedge\, T_{\mathrm{O}}^{\mathrm{Lit}}$.

Weak Abstraction

$$\frac{D' \ \vee\ e_2\ [\![T_2]\!] \qquad C' \ \vee\ [\neg]\ s[w] \approx s'\ [\![T_1]\!]}{C' \ \vee\ \neg\ y \approx w \ \vee\ [\neg]\ s[y] \approx s'\ [\![T_1 \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- $w \ = \ mu + q$.
- $e_2 \ = \ nv + t \approx t'$.
- w occurs in s immediately below some free function symbol.
- If u is a variable:
  - 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 u as a subterm.
  - $\mathrm{lhs}(e_2)$ is neither an atomic term nor a subterm of w.
  - There is a ground instance $(D' \ \vee\ e_2)\theta$ of the first premise such that $v\theta$ is a maximal atomic subterm of $(nv + t)\theta$ and $v\theta \succ t'\theta$.
- If u is not a variable:
  - u and v are atomic terms and have a most general ACU-unifier ρ.
  - $\mathrm{lhs}(e_2)\rho$ is not a subterm of $w\rho$.
  - There is a ground instance $(D' \ \vee\ e_2)\rho\theta$ of the first premise such that $v\rho\theta$ is a maximal atomic subterm of $(nv + t)\rho\theta$ and $v\rho\theta \succ t'\rho\theta$.
- $T_{\mathrm{O}} = w \succ y \ \wedge\ s[w] \succ s' \ \wedge\ T_{\mathrm{O}}^{\mathrm{Lit}}$.

Cancellative Equality Factoring

$$\frac{C' \ \vee\ e_2 \ \vee\ e_1\ [\![T_1]\!]}{C' \ \vee\ \neg\ e_0 \ \vee\ e_2\ [\![T_1 \wedge T_{\mathrm{E}} \wedge T_{\mathrm{O}}]\!]}$$

if the following conditions are satisfied:

- $e_1 \ = \ \sum_{k\in K} m_k^* u_k + s \approx s'$.
- $e_2 \ = \ \sum_{l\in L} n_l^* v_l + t \approx \sum_{l\in L'} n_l^{*\prime} v_l + t'$.
- $e_0 \ = \ \psi t + \chi s' \approx \chi s + \psi t'$.

- $m^* = \sum_{k \in K} m_k^* > 0$, $n^* = \sum_{l \in L} n_l^* - \sum_{l \in L'} n_l^{*\prime} > 0$.
- $\chi = n^* / \gcd(m^*, n^*)$, $\psi = m^* / \gcd(m^*, n^*)$.
- u is one of the u_k or v_l ($k \in K$, $l \in L \cup L'$).
- $T_\mathrm{E} = \bigwedge_{k \in K} u_k \doteq u \;\wedge\; \bigwedge_{l \in L \cup L'} v_l \doteq u$.
- $T_\mathrm{O} = u \succ s \;\wedge\; u \succ s' \;\wedge\; u \succ t \;\wedge\; u \succ t' \;\wedge\; T_\mathrm{O}^\mathrm{Lit}$.

Instances of *weak negative cancellative superposition (I)/(II)* and *weak abstraction* inferences are defined as previously in Def. 4.9 and 4.10. Again, this allows us to prove a lifting lemma:

LEMMA 5.28 *Let R be a stratified set of ground equations. Let $D\,[\![T_2]\!]$ and $C\,[\![T_1]\!]$ be two constrained clauses (without common variables), such that neither of them contains an unshielded variable that occurs only positively. Let θ be a substitution such that $T_2\theta = T_1\theta = \mathrm{true}$ and $D\theta$ and $C\theta$ are ground R-variable minimal instances. Let $D = D' \vee e_2$ such that $e_2\theta$ is strictly maximal in $D\theta$ and $e_2\theta \in R$; let $C = C' \vee [\neg]\, e_1$.*

If there is a weak negative cancellative superposition (I) 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 an R-variable minimal ground instance of a weak negative cancellative superposition (I) inference from $D\,[\![T_2]\!]$ and $C\,[\![T_1]\!]$.*

If there is a weak negative cancellative superposition (II) 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 an R-variable minimal ground instance of a weak negative cancellative superposition (II) inference from $D\,[\![T_2]\!]$ and $C\,[\![T_1]\!]$.*

If there is a weak abstraction inference

$$\frac{D'\theta \vee e_2\theta \qquad C'\theta \vee [\neg]\, e_1\theta[\bar{m}\bar{v} + \bar{q}]}{C_0\ [\![T_0]\!]}$$

such that

- *C_0 equals $C'\theta \vee \neg\, y \approx \bar{m}\bar{v} + \bar{q} \vee [\neg]\, e_1\theta[y]$,*

- $\bar{m}\bar{v} + \bar{q} = w\theta$ *for some subterm* w *of* $\mathrm{lhs}(e_1)$,
- $\bar{m}\bar{v} + \bar{q}$ *is not a subterm of* $y'\theta$ *for any* $y' \in \mathrm{var}(\mathrm{lhs}(e_1))$,
- *the maximal atomic subterm of* $\mathrm{lhs}(e_2\theta)$ *equals* $\bar{v}$,
- *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 an R*-variable minimal ground instance of a* weak abstraction inference *from* $D\,[\![T_2]\!]$ *and* $C\,[\![T_1]\!]$.

The definitions of redundancy (Def. 4.12 and 4.13) are extended verbatim to the new kinds of inferences. It is easy to show that Lemma 4.15 holds also for Inference System 5.27, i. e., that $\mathit{CS}\text{-}\mathit{Red}_{\mathbf{N}^{>0}}$ is also a redundancy criterion with respect to $\mathit{WCS}\text{-}\mathit{Inf}_{\mathbf{N}^{>0}}$.

LEMMA 5.29 *Let* N *be a set of constrained clauses. If every* weak negative cancellative superposition (I)/(II) *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 *whenever the premises do not contain unshielded variables that occur only positively.*

PROOF. Let $D\,[\![T_2]\!] = D' \vee e_2\,[\![T_2]\!]$ and $C\,[\![T_1]\!] = C' \vee \neg\, e_1\,[\![T_1]\!]$ be two clauses in N that do not contain unshielded variables occurring only positively. Let ι be a *negative cancellative superposition* inference

$$\frac{D\,[\![T_2]\!] \qquad C\,[\![T_1]\!]}{C_0\,[\![T_0]\!]}$$

By definition of the inference rule we know that $\mathrm{lhs}(e_1)$ is not a variable. Let R be a stratified set of ground equations and let $\iota\theta$ be an R-variable minimal ground instance

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

such that $\psi\bar{m} = \chi\bar{n}$ and $e_2\theta \in R$.

Now have to distinguish between two cases. If $\bar{m} \geq \bar{n}$, then there exists a ground *weak negative cancellative superposition (I)* inference $\bar{\iota}'$ from $D\theta$ and $C\theta$, namely

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

By Lemma 5.28, $\bar{\iota}'$ is an R-variable minimal ground instance of a *weak negative cancellative superposition (I)* inference ι' of the form

$$\frac{D\,[\![T_2]\!] \qquad C\,[\![T_1]\!]}{C_0'\,[\![T_0']\!]}$$

As ι' is redundant with respect to N, $R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C_0'\theta$. Furthermore, $\bar{n}\bar{u} + \bar{t} \approx \bar{t}'$ and $\neg\,(\bar{m}{-}\bar{n})\bar{u} + \bar{s} + \bar{t}' \approx \bar{t} + \bar{s}'$ entail $\neg\,\psi\bar{s} + \chi\bar{t}' \approx \chi\bar{t} + \psi\bar{s}'$, hence $\{D\theta, C_0'\theta\} \models_\Psi C_0\theta$. Since $D\theta \in \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta}$, we obtain $R^{\prec_{\mathrm{L}}\mathrm{ml}(C\theta)} \cup \mathrm{vm}_R(N)^{\prec_{\mathrm{C}} C\theta} \models_\Psi C_0\theta$. This proves that ι is redundant with respect to N.

Otherwise $\bar{m} < \bar{n}$. In this case, $\iota\theta$ is a *weak negative cancellative superposition (II)* from $D\theta$ and $C\theta$. By Lemma 5.28, $\iota\theta$ is an R-variable minimal ground instance of a *weak negative cancellative superposition (II)* inference ι'. Again, as ι' is redundant with respect to N, we can show that ι is redundant with respect to N. □

Analogously we can prove the following lemma.

LEMMA 5.30 *Let N be a set of constrained clauses. If every weak abstraction inference from clauses in N is redundant with respect to N, then every abstraction inference from clauses in N is redundant with respect to N whenever the premises do not contain unshielded variables that occur only positively.*

Theorem 5.31 is now obvious.

THEOREM 5.31 *If a set of clauses if saturated with respect to WCS-Inf$_{\mathbf{N}^{>0}}$ and CS-Red$_{\mathbf{N}^{>0}}$ and none of the clauses contains unshielded variables that occur only positively, then it is also saturated with respect to CS-Inf$_{\mathbf{N}^{>0}}$ and CS-Red$_{\mathbf{N}^{>0}}$.*

5.6 Divisible Torsion-Free Abelian Groups

Variable Elimination. In Section 5.3, we have shown that 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, that is, that for every formula over 0, +, and $\approx$ there exists 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 is closed under clauses without unshielded variables.

The integration of the variable elimination algorithm demands some restrictions of the cancellative superposition calculus. We have mentioned already in Section 5.3 that variable elimination works only for unconstrained variables, so that if a variable is unshielded but constrained, the constraint either has to be dropped (if it is an ordering constraint, thus non-essential), or partially solved (if it is an equality constraint). To simplify the presentation, we restrict ourselves in this section to completely unconstrained clauses: We assume that *all* satisfiable constraints are immediately dropped or solved when a constrained clause is generated, so that an inference

$$\frac{C_k \ [\![\text{true}]\!] \ \ldots \ C_1 \ [\![\text{true}]\!]}{C_0 \ [\![T_0 \wedge T_0']\!]}$$

where T_0 is an equality constraint and T_0' is an ordering constraint, and $T_0 \wedge T_0'$ is satisfiable, is effectively replaced by inferences

$$\frac{C_k \ \ldots \ C_1}{C_0\sigma}$$

where σ ranges over a complete set of ACU-unifiers of T_0. In other words, we assume that the ground inference system is lifted to non-ground clauses by unification. Furthermore, we have to dispense with selection functions. The reason for this restriction is rather technical in nature and will become clear later.

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

CancelVar $\quad C' \vee [\neg]\, mx + s \approx m'x + s' \quad \to_x \quad C' \vee [\neg]\, (m{-}m')x + s \approx s'$

if $m \geq m' \geq 1$.

ElimNeg $\quad C' \vee \neg\, mx + s \approx s' \quad \to_x \quad C'$

if $m \geq 1$ and x does not occur in C', s, s'.

ElimPos $C' \vee m_1 x + s_1 \approx s'_1 \vee \ldots \vee m_k x + s_k \approx s'_k \quad \rightarrow_x \; C'$

if $m_i \geq 1$ and x does not occur in C', s_i, s'_i, for $1 \leq i \leq k$.

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

$\rightarrow_x \; C' \vee \neg\, mx + s \approx s' \vee [\neg]\, \psi t + \chi s' \approx \psi t' + \chi s$

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 5.32 *If $C_0 \rightarrow_x C_1$, then $\{C_1\} \models_{\mathbf{N}^{>0}} C_0$ and $\{C_0\} \cup \mathrm{DivInvNt} \models_{\mathbf{N}^{>0}} C_1$. If $C_0\theta$ is a ground instance of C_0, then $\{C_1\theta\} \models_{\mathbf{N}^{>0}} C_0\theta$.*

PROOF. If $C_0 \rightarrow_x C_1$ by *CancelVar*, the equivalence of C_0 and C_1 modulo $\mathrm{ACUKT}_{\mathbf{N}^{>0}}$ follows from cancellation; for *Coalesce*, from cancellation and torsion-freeness. The soundness of *ElimNeg* follows from the divisibility and and inverse axiom, for *ElimPos* it is implied by torsion-freeness and non-triviality (see Lemma 5.19). □

LEMMA 5.33 *If $C_0 \rightarrow_x C_1$, then every variable or atomic term occurring (negatively) in C_1 occurs also (negatively) in C_0.*

COROLLARY 5.34 *If $C_0 \rightarrow_x C_1$, x occurs in both C_0 and C_1, and x is unshielded in C_0, then x is unshielded in C_1.*

LEMMA 5.35 *The relation $\rightarrow_x$ is noetherian.*

LEMMA 5.36 *If x occurs unshielded in C, then C is reducible with respect to $\rightarrow_x$.*

COROLLARY 5.37 *If C_1 is a normal form of C_0 with respect to $\rightarrow_x$, and x occurs unshielded in C_0, then x does not occur in C_1.*

The binary relation $\rightarrow_{\mathrm{elim}}$ over clauses is defined in such a way that $C_0 \rightarrow_{\mathrm{elim}} C_1$ if and only if C_0 contains an unshielded variable x and C_1 is a normal form of C_0 with respect to $\rightarrow_x$.

Combining Lemma 5.33 and Cor. 5.37, we see that $C_0 \rightarrow_{\mathrm{elim}} C_1$ implies $\mathrm{var}(C_1) \subset \mathrm{var}(C_0)$. As the number of variables in a clause is finite, we obtain the following corollary:

COROLLARY 5.38 *The relation $\rightarrow_{\mathrm{elim}}$ is noetherian.*

For a clause C, $\mathrm{elim}(C)$ denotes some (arbitrary but fixed) normal form of C with respect to the relation $\rightarrow_{\mathrm{elim}}$.

Corollary 5.39 *For every clause C,* $\mathrm{elim}(C)$ *contains no unshielded variables.*

Corollary 5.40 *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 5.41 *For every clause C,* $\{C\} \cup \mathrm{DivInvNt} \models_{\mathbf{N}^{>0}} \mathrm{elim}(C)$ *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 $\mathit{CS\text{-}Inf}_{\mathbf{N}^{>0}}$, i. e., 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 ground instances of clauses in N that are smaller than $C_0\theta$. For the redundancy of ι, it is sufficient that for each ground instance of ι, $C_0\theta$ follows from ground instances of clauses in N that are smaller than $C\theta$. As demonstrated by the following example, however, even the latter property is not guaranteed for our variable elimination algorithm.

Example 5.42 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 the inference ι, which produces C_0, redundant.

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

Pivotal Terms. Let $C_1, \dots, C_k$ be clauses without unshielded variables and let ι be an 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 $[\neg]\, e$ is the last literal of the last premise of ι, 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,[38] 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 5.43 *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 5.43 does not hold in the presence of selection functions. Neither does it hold for the *merging paramodulation* rule of Bachmair and Ganzinger [12].

[38] 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.

LEMMA 5.44 *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 5.45 *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 5.46 *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 5.47 *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 $[\neg]\, e_1\theta$.*

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 $[\neg]\, e\theta$, then we can conclude analogously 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 $[\neg]\, e\theta$. Under which circumstances is this possible? The literal ordering depends first on the maximal atomic terms in the literals to be compared, hence $[\neg]\, e_1\theta \succeq_{\mathrm{L}} [\neg]\, e\theta$ 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 three cases is impossible: Suppose that $[\neg]\, e_1\theta$ 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 $C_0\theta$. If $[\neg]\, e\theta$, that is, the pivotal literal of $\iota\theta$ is positive, however, 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 5.48 *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:*

- *$[\neg]\, e_1$ has the same polarity as $[\neg]\, e$,*
- *there is an atomic term u_1 occurring at the top of e_1,*
- *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 the previous lemma, we know that the multiset difference $D_0\theta \setminus C_0\theta$ contains a literal that has the same polarity as $[\neg]\, e$ and that 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. 5.40, 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, there must be an atomic term u_1 occurring at the top of e_1, such that $u_1\theta = \bar{u}$. On the other hand, we know that $u\theta = \bar{u}$. As θ is an ACU-unifier of u and u_1, it is clear that every minimal complete set U of ACU-unifiers of u and u_1 contains a τ such that θ is an instance of τ. Therefore, $C_0\theta$ 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 occurring in $C_0\tau$ there is a variable x occurring in 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. 5.46, 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. 5.40, 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. □

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

LEMMA 5.49 *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 5.50 *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:*

- *$[\neg]\, e_1$ has the same polarity as $[\neg]\, e$,*
- *there is an atomic term u_1 occurring at the top of e_1,*
- *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.

Integration of the Elimination Algorithm. Using the results above, we can now transform the inference system *CS-Inf*$_{\mathbf{N}^{>0}}$ into a new inference system that is closed under clauses without unshielded variables. The new system *DS-Inf* 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:

- $\dfrac{C_n \quad \ldots \quad C_1}{C_0}$ is a *CS-Inf*$_{\mathbf{N}^{>0}}$ inference.

Instantiating Inference

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

if the following conditions are satisfied:

- $\dfrac{C_n \quad \dots \quad C_1}{C_0}$ is a *CS-Inf*$_{\mathbf{N}^{>0}}$ inference with pivotal literal $[\neg]\, e$ and pivotal term u.
- The multiset difference $\mathrm{elim}(C_0) \setminus C_0$ contains a literal $[\neg]\, e_1$ with the same polarity as $[\neg]\, e$.
- An atomic term u_1 occurs at the top of e_1.
- τ is contained in a minimal complete set of ACU-unifiers of u and u_1.

LEMMA 5.51 *Let N be a set of clauses without unshielded variables. If every DS-Inf inference from clauses in N is redundant with respect to N, then every CS-Inf$_{\mathbf{N}^{>0}}$ inference from clauses in N is redundant with respect to N.*

PROOF. This follows from the definition of *DS-Inf*, Lemmas 5.48 and 5.50, and Cor. 5.41. □

EXAMPLE 5.52 Let us consider once more the cancellation inference ι from Ex. 5.42:

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

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 $\iota\theta$ redundant that satisfy $C\theta \succ_{\mathrm{C}} \mathrm{elim}(C_0)\theta$, that is, in particular, all ground instances with $x\theta \succ z\theta$. The only remaining ground instances are those where $x\theta = z\theta$; these are made redundant by $C_0\tau$.

THEOREM 5.53 *If a set of clauses is saturated with respect to DS-Inf and none of the clauses contains unshielded variables, then it is also saturated with respect to* $CS\text{-}Inf_{\mathbf{N}^{>0}}$.

Of course, the calculus *DS-Inf* is sound only for sets of clauses that contain the axioms of non-trivial divisible torsion-free abelian groups, that is $\mathrm{ACUKT}_{\mathbf{N}^{>0}} \cup \mathrm{DivInvNt}$. The axioms $\mathrm{ACUKT}_{\mathbf{N}^{>0}}$ are already integrated into the cancellative superposition calculus; hence, no inferences with these axioms are required. Does the same hold for 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 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.

Abstraction. 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 $\mathrm{abs}(C)$ by iterated rewriting

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

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.

Let us define a new inference system $DS^{abs}\text{-}Inf$ that contains exactly the inference rules of *DS-Inf* except of the *abstraction* rule. As *abstraction* inferences from fully abstracted clauses are impossible, the following theorem is an obvious consequence of Thm. 5.53.

THEOREM 5.54 *If a set of fully abstracted clauses is saturated with respect to* $DS^{abs}\text{-}Inf$ *and none of the clauses contains unshielded variables, then it is also saturated with respect to* $CS\text{-}Inf_{\mathbf{N}^{>0}}$.

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

LEMMA 5.55 *Let C be a fully abstracted clause. Then* $\mathrm{elim}(C)$ *is fully abstracted.*

LEMMA 5.56 *Let ι be a DS^{abs}-Inf (or more generally, DS-Inf) 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 the clauses produced by DS^{abs}-*Inf* inferences are again fully abstracted and do not contain unshielded variables.

Full abstraction is not unproblematic from an efficiency point of view. It increases the number of variables and the number of incomparable terms in a clause, which both add to the number of inferences in which this clause can participate.[39]

On the other hand, the cancellative superposition calculus requires many *abstraction* inferences anyway, Furthermore, full abstraction has several important advantages for an implementation: First, if all clauses are fully abstracted, then the terms that have to be compared or unified during the saturation have the property that they do not contain the operator $+$. For such terms, ACU-unification and syntactic unification are equivalent. Thus we may reformulate DS^{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. Secondly, full abstraction 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 [98]), 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. Finally we note that, if all clauses are fully abstracted, then the negation function or the symbols $\textit{divided-by}_k$ 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

[39]Note that *equality resolution* inferences with the new variables that are introduced by abstraction are prohibited by the ordering restrictions, though.

no need for further inferences with the theory clauses DivInvNt during the saturation.

5.7 Ordered Abelian Monoids

Torsion-freeness is defined by an infinite set of first-order axioms

$$\psi x \not\approx \psi y \ \vee\ x \approx y$$

(for every $\psi \in \mathbf{N}^{>0}$). While these axioms may be explicitly given in schematized form in the (finite) input of a theorem prover, they may also be given implicitly using the total ordering axioms.

DEFINITION 5.57 *Let $<$ be a binary predicate symbol.*[40] *The clauses*

$$\neg\, x < y \ \vee\ \neg\, y < z \ \vee\ x < z \qquad \text{(Transitivity)}$$

$$\neg\, x < x \qquad \text{(Irreflexivity)}$$

$$x < y \ \vee\ y < x \ \vee\ x = y \qquad \text{(Totality)}$$

$$\neg\, x < y \ \vee\ x + z < y + z \qquad \text{(Compatibility)}$$

are the compatible total ordering axioms, they are denoted by TotOrd.

By Thm. A.13, an abelian monoid is cancellative and torsion-free if and only if it can be extended to a totally ordered abelian monoid. We can use this fact to prove certain theorems in a totally ordered abelian monoid without actually using the ordering axioms.

LEMMA 5.58 *Let N be a set of clauses in which the predicate $<$ does not occur. Then $N \cup \mathrm{ACU} \cup \mathrm{TotOrd}$ is satisfiable if and only if $N \cup \mathrm{ACUKT}_{\mathbf{N}^{>0}}$ is satisfiable.*

PROOF. To show the "if" part, let $\mathfrak{M}$ be a term-generated normal model of $N \cup \mathrm{ACUKT}_{\mathbf{N}^{>0}}$. Then $(S^{\mathfrak{M}}_{\mathrm{CAM}}, +^{\mathfrak{M}}, 0^{\mathfrak{M}})$ is a torsion-free cancellative abelian monoid. By Thm. A.13, there exists a binary relation $\sqsubset$ over $S^{\mathfrak{M}}_{\mathrm{CAM}}$, such that $(S^{\mathfrak{M}}_{\mathrm{CAM}}, +^{\mathfrak{M}}, 0^{\mathfrak{M}}, \sqsubset)$ is a totally ordered abelian monoid. Define the function $<^{\mathfrak{M}}$ and the constant $\mathrm{true}^{\mathfrak{M}}_{<}$ in such a way that for all $\mathfrak{m}, \mathfrak{m}' \in S^{\mathfrak{M}}_{\mathrm{CAM}}$ we have $<^{\mathfrak{M}}(\mathfrak{m}, \mathfrak{m}') = \mathrm{true}^{\mathfrak{M}}_{<}$ if and only if $\mathfrak{m} \sqsubset \mathfrak{m}'$. The resulting structure is a term-generated normal model of $N \cup \mathrm{ACU} \cup \mathrm{TotOrd}$.

To show the "only if" part, let $\mathfrak{M}$ be a term-generated normal model $N \cup \mathrm{ACU} \cup \mathrm{TotOrd}$. By Thm. A.13, $(S^{\mathfrak{M}}_{\mathrm{CAM}}, +^{\mathfrak{M}}, 0^{\mathfrak{M}})$ is a torsion-free cancellative abelian monoid, hence $\mathfrak{M}$ is a normal model of $N \cup \mathrm{ACUKT}_{\mathbf{N}^{>0}}$. □

[40] Recall that $t < t'$ is to be taken as an abbreviation for $<(t, t') \approx \mathrm{true}_{<}$.

As a consequence of this lemma, we can use the cancellative superposition calculus 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. To see the effect of this transformation, consider the two clauses $2b \approx 2c$ and $b \not\approx c$. Using the calculus $\mathit{CS}\text{-}\mathit{Inf}_{\mathbf{N}^{>0}}$, the only possible inference from these two clauses is a *negative cancellative superposition* inference. It yields $2c \not\approx 2c$, which allows to derive $\bot$ by *cancellation* and *equality resolution.* On the other hand, to refute $\{\, 2b \approx 2c,\ b \not\approx c \,\} \cup \mathrm{ACU} \cup \mathrm{TotOrd}$ using the standard superposition calculus, the prover SPASS (Weidenbach, Gaede, and Rock [99]) derived 10509 clauses; the proof that was found has length 13 (excluding the premises). If we replace the first clause by $3b \approx 3c$, the improvement is still more drastic: While the proof by means of $\mathit{CS}\text{-}\mathit{Inf}_{\mathbf{N}^{>0}}$ requires again 3 inference steps, SPASS had to derive 54069 clauses to find a proof of length 19.

6 Conclusions

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

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 constraints 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 divisiblity 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 likewise, 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 [11]. This is a matter of further investigation.

At the time of writing this paper, we cannot yet report about practical experiences with our calculus. An implementation within the SPASS system [99] is planned.

Appendix 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 [59] and Gilmer [44]. Theorems A.6 and A.10 are new (to the best of my knowledge).

DEFINITION A.1 *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 $x, y, z \in G$.*

A semigroup $(G, +)$ is said to be abelian (or commutative), if $x + y = y + x$ for all $x, y \in G$.

Let $(G, +)$ be a semigroup. An element $0 \in G$ 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.

DEFINITION A.2 *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.*

DEFINITION A.3 *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.4 *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}, i. e., $x + 0_{\mathrm{R}} = x$ 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 an abelian group G, namely its Grothendieck group. The group $(G,+,0)$ 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 $(x,y) \cong (x',y') \;\Leftrightarrow\; x+y' = x'+y$. 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, if G' itself is already a group, then G and G' are isomorphic.

LEMMA A.5 *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}, i. e., 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 $x+y = x$. This can be easily shown 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.6 *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.*[41]

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.4, G is a group.

(ii) As 0_{L} is a left identity, $0_{\mathrm{L}} = 0_{\mathrm{L}} + 0_{\mathrm{L}}$. Consequently, 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. □

DEFINITION A.7 *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 $x = 0$.*

DEFINITION A.8 *A cancellative abelian semigroup $(G, +)$ is called torsion-free, if its Grothendieck group is torsion-free.*

LEMMA A.9 *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.

[41] Part (iii) of this theorem appeared in a draft of [43] submitted for CADE-13 and in [42], 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 [79].

	$(1,1)$	$(2,1)$	$(3,1)$	$(4,1)$	$(5,1)$	$(6,1)$	$(7,1)$	$\cdots$
$(0,0)$	$(1,0)$	$(2,0)$	$(3,0)$	$(4,0)$	$(5,0)$	$(6,0)$	$(7,0)$	$\cdots$

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.10 *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.*

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.6.

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.11 *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.*

DEFINITION A.12 *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 exists no $x \in G$ such that $x < x$, (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 in an analogous way.

The following theorem due to Levi [62] gives the connection between totally ordered and torsion-free cancellative abelian semigroups. A proof can be found in (Gilmer [44]).

THEOREM A.13 *For an abelian semigroup* $(G, +)$ *(or monoid, or group), the following two properties are equivalent:*

- $(G, +)$ *is cancellative and torsion-free.*
- *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.

DEFINITION A.14 *Let* Ψ *be a subset of* $\mathbf{N}^{>0}$. *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* $x = y$.

LEMMA A.15 *Every cancellative abelian semigroup* $(G, +)$ *is* $\{1\}$*-torsion-free.*

LEMMA A.16 *A cancellative abelian semigroup* $(G, +)$ *is torsion-free if and only if it is* $\mathbf{N}^{>0}$*-torsion-free.*

LEMMA A.17 *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, i. e., for all* $\psi, \psi' \in \mathbf{N}^{>0}$

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

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 $\psi\psi' \in \Psi$.

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 $\psi\psi' \notin \Psi$. □

Bibliography

[1] Peter B. Andrews. *An Introduction to Mathematical Logic and Type Theory: To Truth through Proof.* Academic Press, Orlando, FL, USA, 1986.

[2] Yves Auffray and Patrice Enjalbert. Modal theorem proving: An equational viewpoint. *Journal of Logic and Computation*, 2(3):247–295, 1992.

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

[4] Franz Baader and Klaus U. Schulz. Combination of constraint solving techniques: An algebraic point of view. In Jieh Hsiang, ed., *Rewriting Techniques and Applications, 6th International Conference, RTA-95*, Kaiserslautern, Germany, April 5–7, 1995, LNCS 914, pp. 352–366. Springer-Verlag.

[5] Matthias Baaz, Christian G. Fermüller, and Alexander Leitsch. A nonelementary speed-up in proof length by structural clause form transformation. In *Ninth Annual IEEE Symposium on Logic in Computer Science*, Paris, France, July 4–7, 1994, pp. 213–219. IEEE Computer Society Press.

[6] Leo Bachmair. *Canonical Equational Proofs.* Birkhäuser, Boston, MA, USA, 1991.

[7] Leo Bachmair and Harald Ganzinger. Completion of first-order clauses with equality by strict superposition (extended abstract). In Stéphane Kaplan and Mitsuhiro Okada, eds., *Conditional and Typed Rewriting Systems, 2nd International Workshop*, Montreal, Canada, June 11–14, 1990, LNCS 516, pp. 162–180. Springer-Verlag.

[8] Leo Bachmair and Harald Ganzinger. On restrictions of ordered paramodulation with simplification. In Mark E. Stickel, ed., *10th International Conference on Automated Deduction*, Kaiserslautern, Germany, July 24–27, 1990, LNAI 449, pp. 427–441. Springer-Verlag.

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

[10] Leo Bachmair and Harald Ganzinger. Buchberger's algorithm: A constraint-based completion procedure. In Jean-Pierre Jouannaud, ed., *Constraints in Computational Logics*, Munich, Germany, September 7–9, 1994, LNCS 845, pp. 285–301. Springer-Verlag.

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

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

[13] Leo Bachmair and Harald Ganzinger. Rewrite techniques for transitive relations. In *Ninth Annual IEEE Symposium on Logic in Computer Science*, Paris, France, July 4–7, 1994, pp. 384–393. IEEE Computer Society Press.

[14] Leo Bachmair, Harald Ganzinger, Christopher Lynch, and Wayne Snyder. Basic paramodulation. *Information and Computation*, 121(2):172–192, September 1995.

[15] Leo Bachmair, Harald Ganzinger, and Jürgen Stuber. Combining algebra and universal algebra in first-order theorem proving: The case of commutative rings. In Egidio Astesiano, Gianna Reggio, and Andrzej Tarlecki, eds., *Recent Trends in Data Type Specification, 10th Workshop on Specification of Abstract Data Types Joint with the 5th COMPASS Workshop*, S. Margherita, Italy, May 30–June 3, 1994, LNCS 906, pp. 1–29. Springer-Verlag.

[16] Leo Bachmair, Harald Ganzinger, and Uwe Waldmann. Superposition with simplification as a decision procedure for the monadic class with

equality. In Georg Gottlob, Alexander Leitsch, and Daniele Mundici, eds., *Computational Logic and Proof Theory, Third Kurt Gödel Colloquium*, Brno, Czech Republic, August 24–27, 1993, LNCS 713, pp. 83–96. Springer-Verlag.

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

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

[19] Alexandre Boudet, Evelyne Contejean, and Claude Marché. AC-complete unification and its application to theorem proving. In Harald Ganzin ger, ed., *Rewriting Techniques and Applications, 7th International Conference, RTA-96*, New Brunswick, NJ, USA, July 27–30, 1996, LNCS 1103, pp. 18–32. Springer-Verlag.

[20] Thierry Boy de la Tour. An optimality result for clause form translation. *Journal of Symbolic Computation*, 14(4):283–301, October 1992.

[21] R[obert] S. Boyer and J S[trother] Moore. Integrating decision procedures into heuristic theorem provers: A case study of linear arithmetic. In Jean E. Hayes, Donald Michie, and Judith Richards, eds., *Machine Intelligence 11: Logic and the acquisition of knowledge*, ch. 5, pp. 83–124. Oxford University Press, 1988.

[22] D. Brand. Proving theorems with the modification method. *SIAM Journal on Computing*, 4(4):412–430, December 1975.

[23] B[runo] Buchberger. A critical-pair/completion algorithm for finitely generated ideals in rings. In Egon Börger, Gisbert Hasenjaeger, and Dieter Rödding, eds., *Logic and Machines: Decision Problems and Complexity, Proceedings of the Symposium "Rekursive Kombinatorik"*, Münster/Westfalen, Germany, May 23–28, 1983, LNCS 171, pp. 137–161. Springer-Verlag.

[24] Bruno Buchberger. Basic features and development of the critical-pair/completion procedure. In Jean-Pierre Jouannaud, ed., *Rewriting Techniques and Applications, [1st International Conference]*, Dijon, France, May 20–22, 1985, LNCS 202, pp. 1–45. Springer-Verlag.

[25] Bruno Buchberger. Symbolic computation: Computer algebra and logic. In Franz Baader and Klaus U. Schulz, eds., *Frontiers of Combining Systems, First International Workshop*, Munich, Germany, March 26–29, 1996, Applied Logic Series, Vol. 3, pp. 193–219. Kluwer Academic Publishers.

[26] Reinhard Bündgen. Simulating Buchberger's algorithm by Knuth-Bendix completion. In Ronald V. Book, ed., *Rewriting Techniques and Applications, 4th International Conference*, Como, Italy, April 10–12, 1991, LNCS 488, pp. 386–397. Springer-Verlag.

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

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

[29] E[dward] Cardoza, R[ichard] Lipton, and A[lbert] R. Meyer. Exponential space complete problems for petri nets and commutative semigroups: Preliminary report. In *Eighth Annual ACM Symposium on Theory of Computing*, Hershey, PA, USA, May 3–5, 1976, pp. 50–54.

[30] Martin Davis. The prehistory and early history of automated deduction. In Siekmann and Wrightson [88], pp. 1–28.

[31] Martin Davis and Hilary Putnam. A computing procedure for quantification theory. *Journal of the ACM*, 7:201–215, 1960. Reprinted in Siekmann and Wrightson [88], pp. 125–139.

[32] A[natoli] Degtyarev. The monotonic paramodulation strategy [in russian]. In *Proceedings of the 5th All-Union Conference on Mathematical Logic*, Novosibirsk, USSR, 1979. Cited in Bachmair et al. [14].

[33] Nachum Dershowitz and Jean-Pierre Jouannaud. Rewrite systems. In Jan van Leeuwen, ed., *Handbook of Theoretical Computer Science*, vol. B: Formal Models and Semantics, ch. 6, pp. 243–320. Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1990.

[34] Vincent J. Digricoli and Malcolm C. Harrison. Equality-based binary resolution. *Journal of the ACM*, 33(2):253–289, April 1986.

[35] Eric Domenjoud. A technical note on AC-unification. The number of minimal unifiers of the equation $\alpha x_1 + \cdots + \alpha x_p \doteq_{\mathrm{AC}} \beta y_1 + \cdots + \beta y_q$. *Journal of Automated Reasoning*, 8(1):39–44, February 1992.

[36] Uwe Egly. On different structure-preserving translations to normal form. *Journal of Symbolic Computation*, 22(2):121–142, August 1996.

[37] Uwe Egly and Thomas Rath. On the practical value of different definitional translations to normal form. In Michael A. McRobbie and John K. Slaney, eds., *Automated Deduction – CADE-13, 13th International Conference on Automated Deduction*, New Brunswick, NJ, USA, July 30–August 3, 1996, LNAI 1104, pp. 403–417. Springer-Verlag.

[38] M. Fay. First-order unification in an equational theory. In *Fourth Workshop on Automated Deduction*, Austin, TX, USA, February 1979, pp. 161–167. Academic Press.

[39] Melvin Fitting. *First-Order Logic and Automated Theorem Proving*. Springer-Verlag, New York, NY, USA, 1990.

[40] Dov Gabbay and Hans Jürgen Ohlbach. Quantifier elimination in second-order predicate logic. In Bernhard Nebel, Charles Rich, and William Swartout, eds., *Principles of Knowledge Representation and Reasoning, Proceedings of the Third International Conference, KR'92*, Cambridge, MA, USA, October 25–29, 1992, pp. 425–435. Morgan Kaufmann Publishers.

[41] Harald Ganzinger. Order-sorted completion: the many-sorted way. *Theoretical Computer Science*, 89:3–32, 1991.

[42] Harald Ganzinger and Uwe Waldmann. Theorem proving in cancellative abelian monoids. Technical Report MPI-I-96-2-001, Max-Planck-Institut für Informatik, Saarbrücken, Germany, January 1996.

[43] Harald Ganzinger and Uwe Waldmann. Theorem proving in cancellative abelian monoids (extended abstract). In Michael A. McRobbie and John K. Slaney, eds., *Automated Deduction – CADE-13, 13th International Conference on Automated Deduction*, New Brunswick, NJ, USA, July 30–August 3, 1996, LNAI 1104, pp. 388–402. Springer-Verlag.

[44] Robert Gilmer. *Commutative Semigroup Rings*. Chicago Lectures in Mathematics. The University of Chicago Press, Chicago, IL, USA, 1984.

[45] Jieh Hsiang. Refutational theorem proving using term-rewriting systems. *Artificial Intelligence*, 25(3):255–300, 1985.

[46] Jieh Hsiang and Michaël Rusinowitch. Proving refutational completeness of theorem-proving strategies: The transfinite semantic tree method. *Journal of the ACM*, 38(3):559–587, July 1991.

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

[48] Jean-Marie Hullot. Canonical forms and unification. In Wolfgang Bibel and Robert Kowalski, eds., *5th Conference on Automated Deduction*, Les Arcs, France, July 8–11, 1980, LNCS 87, pp. 318–334. Springer-Verlag.

[49] Costas S. Iliopoulos. Worst-case complexity bounds on algorithms for computing the canonical structure of finite Abelian groups and the Hermite and Smith normal forms of an integer matrix. *SIAM Journal on Computing*, 18(4):658–669, August 1989.

[50] Joxan Jaffar and Jean-Louis Lassez. Constraint logic programming. In *Fourteenth Annual ACM Symposium on Principles of Programming Languages*, Munich, Germany, January 21–23, 1987, pp. 111–119.

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

[52] Abdelilah Kandri-Rody, Deepak Kapur, and Paliath Narendran. An ideal-theoretic approach to word problems and unification problems over finitely presented commutative algebras. In Jean-Pierre Jouannaud, ed., *Rewriting Techniques and Applications, [1st International Conference]*, Dijon, France, May 20–22, 1985, LNCS 202, pp. 345–364. Springer-Verlag.

[53] Deepak Kapur and Paliath Narendran. Complexity of unification problems with associative-commutative operators. *Journal of Automated Reasoning*, 9(2):261–288, October 1992.

[54] Deepak Kapur and Paliath Narendran. Double-exponential complexity of computing a complete set of AC-unifiers. In *Seventh Annual IEEE Symposium on Logic in Computer Science*, Santa Cruz, CA, USA, June 22–25, 1992, pp. 11–21. IEEE Computer Society Press.

[55] Claude Kirchner, Hélène Kirchner, and Michaël Rusinowitch. Deduction with symbolic constraints. *Revue d'Intelligence Artificielle*, 4(3):9–52, 1990.

[56] Donald E. Knuth. *The Art of Computer Programming*, vol. 2: Seminumerical Algorithms. Addison-Wesley, Reading, MA, USA, 1969.

[57] Donald E. Knuth and Peter B. Bendix. Simple word problems in universal algebras. In J. Leech, ed., *Computational Problems in Abstract Algebra*, pp. 263–297. Pergamon Press, Oxford, United Kingdom, 1970. Reprinted in Siekmann and Wrightson [89], pp. 342–376.

[58] R[obert] Kowalski and P. J. Hayes. Semantic trees in automatic theorem proving. In Bernard Meltzer and Donald Michie, eds., *Machine Intelligence 4*, pp. 87–101. Edinburgh University Press, Edinburgh, United Kingdom, 1969. Reprinted in Siekmann and Wrightson [89], pp. 217–232.

[59] Serge Lang. *Algebra*. Addison-Wesley, Redwood City, CA, USA, second edition, 1984.

[60] D[allas S.] Lankford, G. Butler, and A. [Michael] Ballantyne. A progress report on new decision algorithms for finitely presented abelian groups. In Robert E. Shostak, ed., *7th International Conference on Automated Deduction*, Napa, CA, USA, May 14–16, 1984, LNCS 170, pp. 128–141. Springer-Verlag.

[61] Philippe Le Chenadec. Canonical forms in finitely presented algebras. In Robert E. Shostak, ed., *7th International Conference on Automated Deduction*, Napa, CA, USA, May 14–16, 1984, LNCS 170, pp. 142–165. Springer-Verlag.

[62] F. Levi. Arithmetische Gesetze im Gebiete discreter Gruppen. *Rend. Circ. Mat. Palermo*, 35:225–236, 1913. Cited in Gilmer [44].

[63] Claude Marché. Normalized rewriting: an alternative to rewriting modulo a set of equations. *Journal of Symbolic Computation*, 21(3):253–288, March 1996.

[64] Ernst W. Mayr and Albert R. Meyer. The complexity of the word problems for commutative semigroups and polynomial ideals. *Advances in Mathematics*, 46(3):305–329, December 1982.

[65] Greg Nelson and Derek C. Oppen. Simplification by cooperating decision procedures. *ACM Transactions on Programming Languages and Systems*, 1(2):245–257, October 1979.

[66] Robert Nieuwenhuis. First-order completion techniques. Technical report, UPC-LSI, 1991. Cited in Nieuwenhuis and Rubio [67].

[67] Robert Nieuwenhuis and Albert Rubio. Basic superposition is complete. In Bernd Krieg-Brückner, ed., *ESOP'92, 4th European Symposium on Programming*, Rennes, France, February 26–28, 1992, LNCS 582, pp. 371–389. Springer-Verlag.

[68] Robert Nieuwenhuis and Albert Rubio. Theorem proving with ordering constrained clauses. In Deepak Kapur, ed., *11th International Conference on Automated Deduction*, Saratoga Springs, NY, USA, June 15–18, 1992, LNAI 607, pp. 477–491. Springer-Verlag.

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

[70] Robert Nieuwenhuis and Albert Rubio. Theorem proving with ordering and equality constrained clauses. *Journal of Symbolic Computation*, 19(4):321–351, April 1995.

[71] Andreas Nonnengart. Resolution-based calculi for modal and temporal logics. In Michael A. McRobbie and John K. Slaney, eds., *Automated Deduction – CADE-13, 13th International Conference on Automated Deduction*, New Brunswick, NJ, USA, July 30–August 3, 1996, LNAI 1104, pp. 598–612. Springer-Verlag.

[72] H[ans] J[ürgen] Ohlbach. Semantics-based translation methods for modal logics. *Journal of Logic and Computation*, 1(5):691–746, 1991.

[73] H[ans] J[ürgen] Ohlbach, R[enate A.] Schmidt, and U[llrich] Hustadt. Translating graded modalities into predicate logics. In H. Wansing, ed., *Proof Theory of Modal Logic*, Applied Logic Series, Vol. 2, pp. 253–291. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[74] Hans Jürgen Ohlbach and Christoph Weidenbach. A note on assumptions about Skolem functions. *Journal of Automated Reasoning*, 15(2):267–275, 1995.

[75] Etienne Paul. A general refutational completeness result for an inference procedure based on associative-commutative unification. *Journal of Symbolic Computation*, 14(6):577–618, December 1992.

[76] Gerald E. Peterson. A technique for establishing completeness results in theorem proving with equality. *SIAM Journal on Computing*, 12(1):82–100, February 1983.

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

[78] Gordon D. Plotkin. Building-in equational theories. In Bernard Meltzer and Donald Michie, eds., *Machine Intelligence 7*, ch. 4, pp. 73–90. American Elsevier, New York, NY, USA, 1972.

[79] Péter Prőhle. Which of the cancellative semigroups are groups? *Semigroup Forum*. Accepted for publication in 1996.

[80] Christophe Ringeissen. Cooperation of decision procedures for the satisfiability problem. In Franz Baader and Klaus U. Schulz, eds., *Frontiers of Combining Systems, First International Workshop*, Munich, Germany, March 26–29, 1996, Applied Logic Series, Vol. 3, pp. 121–139. Kluwer Academic Publishers.

[81] G[eorge] Robinson and L[arry] Wos. Paramodulation and theorem-proving in first-order theories with equality. In Bernard Meltzer and Donald Michie, eds., *Machine Intelligence 4*, ch. 8, pp. 135–150. Edinburgh University Press, Edinburgh, United Kingdom, 1969. Reprinted in Siekmann and Wrightson [89], pp. 298–313.

[82] J[ohn] A[lan] Robinson. Automatic deduction with hyper-resolution. *International Journal of Computer Mathematics*, 1:227–234, 1965. Reprinted in Siekmann and Wrightson [88], pp. 416–423.

[83] J[ohn] A[lan] Robinson. A machine-oriented logic based on the resolution principle. *Journal of the ACM*, 12(1):23–41, January 1965. Reprinted in Siekmann and Wrightson [88], pp. 397–415.

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

[85] Michaël Rusinowitch. Theorem-proving with resolution and superposition. *Journal of Symbolic Computation*, 11(1&2):21–49, January/February 1991.

[86] Michaël Rusinowitch and Laurent Vigneron. Automated deduction with associative-commutative operators. *Applicable Algebra in Engineering, Communication and Computing*, 6(1):23–56, January 1995.

[87] Robert E. Shostak. A practical decision procedure for arithmetic with function symbols. *Journal of the ACM*, 26(2):351–360, April 1979.

[88] Jörg Siekmann and Graham Wrightson, eds. *Automation of Reasoning*, vol. 1: Classical Papers on Computational Logic 1957–1966. Springer-Verlag, Berlin, Germany, 1983.

[89] Jörg Siekmann and Graham Wrightson, eds. *Automation of Reasoning*, vol. 2: Classical Papers on Computational Logic 1967–1970. Springer-Verlag, Berlin, Germany, 1983.

[90] Thoralf Skolem. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen. *Videnskopsselskapits skifter, I. Matematik-naturvidenskabelig klasse*, 4, 1920. Cited in Davis [30].

[91] Thoralf Skolem. Über die mathematische Logik. *Norsk matematisk tideskrift*, 10:125–142, 1928. Cited in Davis [30].

[92] James R. Slagle. Automatic theorem proving with renamable and semantic resolution. *Journal of the ACM*, 14(4):687–697, October 1967. Reprinted in Siekmann and Wrightson [89], pp. 55–65.

[93] James R. Slagle. Automatic theorem proving with built-in theories including equality, partial ordering, and sets. *Journal of the ACM*, 19(1):120–135, January 1972.

[94] James R. Slagle. Automated theorem-proving for theories with simplifiers, commutativity, and associativity. *Journal of the ACM*, 21(4):622–642, October 1974.

[95] Mark E. Stickel. Automated deduction by theory resolution. *Journal of Automated Reasoning*, 1(4):333–355, 1985.

[96] Jürgen Stuber. Superposition theorem proving for abelian groups represented as integer modules. In Harald Ganzinger, ed., *Rewriting Techniques and Applications, 7th International Conference, RTA-96*, New

Brunswick, NJ, USA, July 27–30, 1996, LNCS 1103, pp. 33–47. Springer-Verlag.

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

[98] Uwe Waldmann. Extending reduction orderings to ACU-compatible reduction orderings. `http://www.mpi-sb.mpg.de/~uwe/paper/ACUExt.ps.gz`. Submitted, 1997.

[99] Christoph Weidenbach, Bernd Gaede, and Georg Rock. SPASS & FLOTTER, version 0.42. In Michael A. McRobbie and John K. Slaney, eds., *Automated Deduction – CADE-13, 13th International Conference on Automated Deduction*, New Brunswick, NJ, USA, July 30–August 3, 1996, LNAI 1104, pp. 141–145. Springer-Verlag.

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

[101] L[arry] Wos and L[awrence] Henschen. Automated theorem proving 1965–1970. In Siekmann and Wrightson [89], pp. 1–24.

[102] Hantao Zhang. A case study of completion modulo distributivity and Abelian groups. In Claude Kirchner, ed., *Rewriting Techniques and Applications, 5th International Conference, RTA-93*, Montreal, Canada, June 16–18, 1993, LNCS 690, pp. 32–46. Springer-Verlag.

[103] Hantao Zhang. A new method for the boolean ring based theorem proving. *Journal of Symbolic Computation*, 17(2):189–211, 1994.

List of Symbols

$\perp$, 6
$\neg$, 6
$[\neg]$, 6
$+$, 31
$-$, 99, 110
α^*, 6
$C \cdot \rho$, 21
$\cup$, 8

$=_E$, 13
$=_{\text{ACU}}$, 31
$\sigma = \sigma'$ over $\mathcal{V}$, 8
$\doteq$, 9, 34
$\approx$, 6, 33
$\approx^{\mathfrak{M}}$, 6
$\not\approx$, 6

$<$, 135
$\succ$, 9, 12, 33
$\succ_{\text{C}}$, 34
$\succ_{\text{L}}$, 34
$\succeq$, 12
$\Pi^{\prec s}$, 12
$\Pi^{\preceq s}$, 12
$\Pi^{\succ s}$, 12
$\Pi^{\succeq s}$, 12

$\rightarrow$, 5, 11
$\rightarrow^*$, 11
$\rightarrow^+$, 11
$\rightarrow_{\text{elim}}$, 126

$\rightarrow_R$, 12, 68
$\rightarrow_{\delta,R}$, 68
$\rightarrow_{\gamma,R}$, 68
$\rightarrow_{\kappa,R}$, 68
$\rightarrow_x$, 125
$\leftarrow$, 11
$\leftrightarrow$, 11
$\leftrightarrow_E$, 13
$\downarrow$, 32
$\mapsto$, 8
$\rightsquigarrow_N$, 100

$\models$, 13
$\models_{\approx}$, 11
$\models_{\Psi}$, 34
$\vdash$, 15

$[\_]_{\text{ACU}}$, 31
$[\![\_]\!]$, 8

0, 31
0_{L}, 140
0_{R}, 139

α (assignment), 6
γ, 68
δ, 68
θ (substitution)
ι (inference), 13
κ, 68
λ (natural number), 106

μ (natural number)
ν (natural number)
Π (ordered set), 12
ρ (substitution)
σ (substitution)
Σ (set of function symbols), 5
τ (substitution)
ϕ, 23
χ (natural number), 37
ψ (element of Ψ), 31
Ψ, 31

A, 31
abs, 133
AC, 31
ACK, 31
ACU, 31
ACUK, 31
ACUKT_Ψ, 31

b (constant symbol)

c (constant symbol)
C (formula or clause)
C, 31
cl_Ψ, 35
concl, 13
CS-Inf$_\Psi$, 45
CS-Red$_\Psi$, 62
CS-Red$_\Psi^{\mathrm{C}}$, 62
CS-Red$_\Psi^{\mathrm{I}}$, 62

d (constant symbol)
D (formula or clause)
divided-by$_n$, 110
DivInvNt, 124
Dom, 8
DS-Inf, 131
DS^{abs}*-Inf*, 133

e (equation), 6
E (set of equations)
E_C, 69
E_C^Ψ, 69
$E_{\mathfrak{M}}$, 9
elig, 36
elim, 126
EQ, 35

f (function symbol)
$f^{\mathfrak{M}}$, 6
false, 9

g (function symbol)
G (semigroup or group), 139

h (function symbol)
H (subset of a semigroup), 140
H, 140

i (index or natural number)
I (set of indices)
Inf, 13

j (index or natural number)
J (set of indices)

k (index or natural number)
K (set of indices)
K, 31

l (index or natural number)
L (set of indices)
lhs, 6

m (natural number)
$\mathfrak{m}$ (element of an interpretation)
M (set of clauses)
$\mathfrak{M}$ (interpretation or model), 6
$\mathfrak{M}(\_)$, 6
ml, 34
ms, 34
mt, 31
mt, 33
$\mathrm{mt}_\#$, 33

n (natural number)
N (set of formulae or clauses)
$\overline{N}$, 51
N_∞, 15
$\mathbf{N}$, 5
$\mathbf{N}^{>0}$, 5
neg, 6

o (position), 11

p (function symbol), 6
P (predicate symbol), 6
pos, 11

q (term)
Q (syntactic object), 5

r (term)
R (set of rewrite rules), 12
R_C, 69
R_∞, 70
R_C^Ψ, 69
R_∞^Ψ, 70
Red (redundancy criterion), 14
Red^{C}, 14
Red^{I}, 14
rhs, 6

s (term)
S (sort), 5
S_{CAM}, 31
$S^{\mathfrak{M}}$, 6
$\mathcal{S}$ (set of sorts), 5

t (term)
$t(o)$, 11
$t|_o$, 11
$t[t']$, 11
$t[t']_o$, 11
T (constraint), 8
$T_{\mathrm{O}}^{\mathrm{Lit}}$, 45
T_Ψ, 31
tr, 69
tr_Ψ, 69
true, 8
true_p, 6

u (atomic term)
U (set of substitutions), 32
U, 31

v (atomic term)
$\mathcal{V}$ (set of variables), 5
var, 5
vm_R, 60

w (term)
$\mathit{WCS\text{-}Inf}_{\{1\}}$, 111
$\mathit{WCS\text{-}Inf}_{\mathbf{N}^{>0}}$, 119

x (variable)
$\hat{x}$, 44
$\check{x}$, 44

y (variable)

z (variable)

Index

abelian, 139
abelian group, 23, 27
abelian monoid, 27
abelian semigroup, 27
abstraction, 41, 48, 50, 112, 118, 133
AC-paramodulation, 27
AC-superposition, 27
AC-unification, 27
ACU-compatible, 32, 134
ACU-unification, 27, 32, 134
ACUKT_Ψ-closure, 35
assignment, 6
associative, 23, 27, 31
atomic term, 31

basic superposition, 21
below, 32
black box, 25

cancellation, 31, 37, 46
cancellative, 139
cancellative abelian monoid, 28
cancellative equality factoring, 37, 39, 49, 119, 121
cancellative rewrite rule, 67
cancellative superposition, 36
cancellative superposition calculus, 45
CancelVar, 125
case splitting, 102
chaining, 26
clause, 6
clause normal form transformation, 24
clause ordering, 34
Coalesce, 126
commutative, 23, 27, 31, 139
compact, 7
compatible, 135
complete set of unifiers, 32, 35
completion, 19
conclusion, 13
confluent, 12, 72
congruent, 10
constant symbol, 5
constrained clause, 8
constraint, 8, 57
constraint language, 34
current interpretation, 51

decision procedure, 102
δ-reducible, 68
δ-rewriting, 68
demodulation, 95
derivation, 15
derivation relation, 15
divisible, 110
divisible torsion-free abelian group, 124
domain (of a substitution), 8

domain (of an interpretation), 6

eligible, 36, 107
eliminating inference, 131
ElimNeg, 125
ElimPos, 126
empty clause, 6
entailment, 11, 35
equal, 9
equality axioms, 10
equality constraint, 22
equality factoring, 20
equality interpretation, 10
equality model, 10
equality resolution, 37, 40, 46
equation, 6
essential, 35
Euclid's algorithm, 106
EXPSPACE, 28
extended clause, 27

fair, 16
false, 7
free function symbol, 31
full abstraction, 133
function symbol, 5

γ-reducible, 68
γ-rewriting, 68
Gaussian elimination, 107
gcd superposition, 105
Grothendieck group, 140
ground, 5
ground instance, 8, 9
ground instance (of an inference), 57, 60
group, 139

Herbrand interpretation, 9
hierarchic technique, 25
hyperresolution, 17

idempotent, 8
identity, 27, 31
identity element, 139
immediately below, 32
inference, 13
inference rule, 13
inference system, 13
instance, 8
instantiating inference, 132
integrating technique, 25
interpretation, 6
inverse, 110
irreducible, 12, 58
irreflexive, 135

κ-reducible, 68
κ-rewriting, 68
Knuth-Bendix completion, 19

left identity element, 139
left-cancellative, 139
lifting, 39, 43, 56, 64
limit, 15
literal, 6
literal ordering, 34
lower bounded, 63

maximal atomic subterm, 33
maximal literal, 34
merging paramodulation, 20
model, 7, 9
model construction, 69
model generalizable, 58, 63
monoid, 139
most general unifier, 8
multiplication, 31
multiset property, 33, 134

negative cancellative superposition, 37, 39, 47, 111, 118
negative literal, 6
noetherian, 12

non-essential, 35, 50
non-group, 108
non-trivial, 109
normal form, 12
normal interpretation, 9
normalized completion, 28, 38

occurrence, 11
ordered abelian monoid, 29, 135
ordered resolution, 18
ordering, 12, 26
ordering constraint, 22

paramodulation, 18
partial solution, 102
persisting, 15
pivotal literal, 128
pivotal substitution, 128
pivotal term, 128
position, 11
positive cancellative superposition, 37, 40, 48
positive literal, 6
positive maximal, 59
predicate, 6
premise, 13
primary rule, 55, 69
productive, 51
proper sum, 32
Ψ-torsion-free, 31, 143
Ψ-truth set, 69

quantifier elimination, 125
quasisimplification, 99

reducible, 68
reduction ordering, 12
reductive, 69
redundant, 14, 20, 45, 52, 56, 60, 95
reflexive, 10
refutationally complete, 13, 16, 51, 89
resolution, 16
rewrite relation, 11, 67
rewrite rule, 12, 54, 67
rewrite system, 12
right identity element, 139
right-cancellative, 139

satisfiable (constraint), 9
satisfiable (set of clauses), 7
satisfy, 7, 9
saturated, 15
secondary rule, 55, 69
selection function, 36, 61
semigroup, 139
shielded, 36, 40, 107
signature, 5
simplification, 95
skeleton, 21
skolemization, 24
solution, 9
sort, 5
sound, 13, 50
stable, 11
standard equality factoring, 20, 40, 49
standard superposition, 19, 40, 47
stratified, 59, 88
substitution, 7
subsumption, 102
sufficiently complete, 25
superposition, 19
superposition calculus, 20
symmetric, 10
symmetrization, 28
syntactic object, 5

term, 5
term-generated, 6
terminating, 12
theorem proving calculus, 15
theory reasoning, 23

top occurrence, 32
torsion element, 141
torsion-free, 29, 38, 108, 135, 141
total, 135
total ordering axioms, 135
totally ordered abelian group, 142
totally ordered abelian monoid, 142
totally ordered abelian semigroup, 142
transformation technique, 24
transitive, 10, 135
true, 6
truth set, 69

unifier, 8
unsatisfiable (set of clauses), 7
unshielded, 36, 43, 107

variable, 5
variable elimination, 107, 124
variable minimal, 58, 59

weak, 111, 118
weak abstraction, 112, 114, 118, 121
weak cancellative equality factoring, 112, 115
weak negative cancellative superposition, 111, 113, 118–120
word problem, 28, 104