Superposition for Finite Domains (Plain Text Version)

Thomas Hillenbrand Christoph Weidenbach

1 Introduction

Standard superposition is not a decision procedure for first-order finite-domain problems. For example, the superposition calculus may not terminate on a given clause set N, even if all function symbols are constants and hence any Herbrand model of N has a finite domain. This is because the well-known reduction rules such as rewriting or subsumption do not exploit the finite domain property and clauses such as

$$x \simeq 1 \vee \ldots \vee x \simeq n$$

forcing a finite domain with at most n digits are the source of a highly prolific infinite search space, if not handled with special care.

This is what this paper is about. By exploiting the finite domain properties via a refined lifting lemma, the superposition calculus for finite domains

(a) restricts the range of inference unifiers to digits or variables,
(b) enables the precise calculation of ordering restrictions,
(c) introduces an effective general semantic redundancy notion, and
(d) incorporates a particular splitting rule for non-Horn clauses

all shown in Sect. 4. The properties (a)–(c) are a consequence of showing that for completeness only ground substitutions to digits need to be considered (Proposition 4.1, Lemma 5.1). Therefore, via the lifting lemma, no complex unifiers need to be considered. The number of ground instances of a clause is finite and hence ordering restrictions can be precisely calculated even for non-ground clauses and the general semantic redundancy notion that any clause semantically entailed by smaller clauses can be deleted, becomes decidable.

But all these refinements do not guarantee termination of the calculus. This can even be shown by a simple ground problem. For example, consider

the two equations $f(a) \simeq a$, $a \simeq f(b)$ ordered by an LPO with precedence $a \succ f \succ b$ [NR01]. These equations generate infinitely many clauses of the form $f^i(b) \simeq a$ by superposition right inferences. By exhaustive rewriting termination can be enforced for this example, but it is not known whether redundancy criteria guarantee termination outside the Horn fragment. Therefore, we introduce a splitting rule that splits non-Horn clauses into clauses with fewer positive literals (d). Eventually, for Horn clauses the calculus is guaranteed to terminate (Theorem 4.21).

The resulting refined superposition calculus for finite domain problems

(e) constitutes a decision procedure for any finite domain problem,
(f) is mostly compatible with the standard superposition redundancy notion, and
(g) can be embedded via a general sort discipline based on monadic predicates in general, potentially infinite domain settings

shown by Theorem 4.21, Sect. 4.3, and Sect. 5, respectively. As a consequence of (e), the refined superposition calculus is a decision procedure for the Bernays-Schoenfinkel class, solving a further classical satisfiability class by the superposition calculus.

Compared to instantiation based methods for finite domain problems [McC03, CS03], superposition for finite domains does not a priori instantiate variables, but exploits the finite domain structure on the clauses with variables level. In particular, this offers advantages if the problem has structure that can be employed by inferences/reduction rules. A first simple example are the two unit clauses $P(x_1, \ldots, x_k, x_1)$, $\neg P(a, x_1, \ldots, x_{k-1}, b)$ where no superposition (resolution) inference is possible but instantiation based methods will generate more than k^n clauses for a finite domain of n digits. In general, the execution or blocking of a superposition inference/reduction with variables simulates up to exponentially many ground steps. Secondly, consider an equation $f(x) \simeq x$ and an atom $P(f(g(x)))$ where a standard rewriting step yields $P(g(x))$. After instantiation with digits this reduction is no longer possible (as $g(\ldots)$ is not a digit). For examples of this form, inferring/reducing at the clauses with variables level has the potential for exponentially shorter proofs and representations of models, compared to instantiation based methods.

Transformation based methods [MB88, BS06] translate a given clause set in a form that afterwards applied standard inference mechanisms, e.g., hyper-resolution, yield the search for a model in a bottom-up way. This work is orthogonal to ours as it transforms the problem where we exploit the finite domain property at the calculus level.

Both the instantiation based as well as the transformation based approach do currently not support the combination with general first-order theories,

shown in Sect. 5. In fact, finite domain sorts are an inherent part of many verification problems arising from software or system analysis. Therefore, the combination has a large application potential.

2 Getting Started

For most logical notions and notations, we refer to [NR01]. In particular we work in a logic with built-in equality. We stipulate a single-sorted signature Σ that contains the constant symbols 1 through n, which we name digits, besides arbitrary other function symbols. Furthermore a set $\mathcal{V}$ provides infinite supply of variables. For a term t we denote by $\mathrm{var}(t)$ the set of variables that occur in t; the set $\mathrm{var}(C)$ is defined correspondingly for every clause C. If σ is a substitution, then $\mathrm{dom}\,\sigma$ is the set of all variables for which $x\sigma \not\equiv x$, $\mathrm{ran}\,\sigma$ is the image of $\mathrm{dom}\,\sigma$ under σ, and $\mathrm{cdom}\,\sigma$ is the set of variables occuring in $\mathrm{ran}\,\sigma$. For simplicity of notation the equality symbol $\simeq$ is supposed to be symmetric. A literal $s \bowtie t$ is either an equation $s \simeq t$ or a disequation $s \not\simeq t$. A clause is a disjunction of literals; a Horn clause is a clause with at most one positive literal.

Semantic entailment is defined in the usual way. We write that a Σ-algebra $\mathcal{A}$ validates a clause C by $\mathcal{A} \models C$. $\mathcal{A}$ contains a valuation for variables and its homorphic extension to functions. By $\mathcal{A}[x/d]$ we denote an interpretation that is identical to $\mathcal{A}$ except that its valuation maps x to d.

The theory $\mathcal{T}$ is given by the formula

$$\forall x.\, x \simeq 1 \vee \ldots \vee x \simeq n$$

We will introduce a superposition-based calculus to tackle the $\mathcal{T}$-satisfiability of clause sets over Σ. Note that this also covers the case that the domain size is exactly n, since one can add clauses $i \not\simeq j$ for any distinct $i, j \in [1; n]$.

The calculus will be described by rule patterns of three different types in a fraction-like notation. Clauses occurring in the numerator are generally called *premises*, and in the denominator *conclusions*. As usually, premises are assumed to share no variables. Finite clause sequences $C_1, \ldots, C_m$ where $m \geq 0$ are abbreviated as $\vec{C}$. If C denotes a clause and M a clause set, then M, C is shorthand notation for $M \cup \{C\}$.

(i) *Inference rules:* $\mathcal{I}\,\dfrac{\vec{C}}{D}$ if *condition*

denotes any transition from a clause set $M, \vec{C}$ to $M, \vec{C}, D$ provided *condition* is fulfilled. Occasionally the rightmost of the premises is named *main premise*, and the remaining ones are the *side premises*.

(ii) *Reduction rules:* $\mathcal{R}\,\dfrac{C}{\vec{C}'}\,\vec{D}$ if *condition*

stands for any transition from a clause set $M, C, \vec{D}$ to a clause set $M, \vec{C}', \vec{D}$ whenever *condition* holds. In essence, the clause C is replaced by the clauses $\vec{C}'$, the sequence of which may be empty.

(iii) *Split rules:* $\mathcal{S}\,\dfrac{C}{D \mid D'}$ if *condition*

describes any transition from a clause set M, C to the pair of clause sets $(M, C, D \mid M, C, D')$ constrained by *condition*. Note that the premise is part of each of the descending clause sets.

In the *condition* part of inference rules, frequently some terms, say s and t, are required to have a most general unifier σ; we stipulate that σ satisfies $\operatorname{dom}\sigma \cup \operatorname{cdom}\sigma \subseteq \operatorname{var}(s,t)$. Furthermore, occurrences of terms or of literals may be restricted to maximal ones. Attention: In the former case this refers to the enclosing literal, in the latter to the enclosing clause. Maximality means that no other occurrence is greater, and is strict if none is greater or equal. Correspondingly we will speak of greatest occurrences, which are greater than or equal to the remaining ones, and of strictly greater ones, that are greater than all the rest. There is no difference between being greatest or maximal in case the underlying ordering is total, as it happens in the case of ground clauses and a reduction ordering total on ground terms.

A *derivation* from a (not necessarily finite) clause set M with respect to a calculus specified that way is a finitely branching tree such that (i) the nodes are sets of clauses, (ii) the root is M, and (iii) if a node N has the immediate descendants $N_1, \ldots, N_k$, respectively, then there is a transition from N to $N_1, \ldots, N_k$ in the calculus. A *complete path* $N_1, N_2, \ldots$ in a derivation tree starts from the root, ends in a leaf in case the path is finite, and has the *limit* $N_\infty = \bigcup_i \bigcap_{j \geq i} N_j$. Given a redundancy notion for inferences and clauses, a derivation is said to be *fair* if for every complete path $N_1, N_2, \ldots$ the following applies to the transitions from N_∞: (i) Every inference is redundant in some N_i, and (ii) for every split, one of its conclusion is in some N_i or redundant with respect to it. A clause set M is *saturated* if it satisfies conditions (i) and (ii) with N_i replaced by M.

3 Ground Horn Superposition

We recapitulate a superposition calculus $\mathcal{G}$ for ground Horn clauses [BG94, NR01]. In every clause with negative literals, at least one of them shall be *selected*. This eager selection leads to a positive unit literal strategy [Der91], where the side premise of superposition inferences is always a positive unit

clause. Furthermore the model construction involves such unit clauses only, which will ease the model extraction in a later stage. From now on, let $\succ$ denote a reduction ordering total on ground terms.

Rules of calculus $\mathcal{G}$:

Ground superposition left

$$\mathcal{I}\,\frac{l \simeq r \quad s[l] \not\simeq t \vee C}{s[r] \not\simeq t \vee C} \qquad \text{if } \begin{array}{l}\cdot\ l \text{ and } s \text{ are strictly greatest}\\ \cdot\ s \not\simeq t \text{ is selected}\end{array}$$

Ground superposition right

$$\mathcal{I}\,\frac{l \simeq r \quad s[l] \simeq t}{s[r] \simeq t} \qquad \text{if } \begin{array}{l}\cdot\ l \text{ and } s \text{ are strictly greatest}\\ \cdot\ l \simeq r \prec s \simeq t\end{array}$$

Ground equality resolution

$$\mathcal{I}\,\frac{C \vee t \not\simeq t}{C} \qquad \text{if } \cdot\ t \not\simeq t \text{ is selected}$$

An inference with maximal premise C and conclusion D is *redundant* with respect to a clause set M if $M^{\prec C} \models D$, where $M^{\prec C}$ contains all elements of M smaller than C. The calculus $\mathcal{G}$ is sound and refutationally complete in the sense that $M \models \bot$ and $\bot \in M$ coincide for every saturated set M. The completeness proof relies on a model functor that associates with M a convergent ground rewrite system R. Let R^* denote the quotient of the free ground term algebra modulo the congruence generated by R; and assume that M is saturated and does not contain the empty clause. Then R^* is a model of M. In detail, for every clause C let $\mathrm{Gen}(C) = \{l \to r\}$ if (i) $C \equiv l \simeq r \in M$, (ii) l is strictly maximal, (iii) l is R_C-irreducible; and let $\mathrm{Gen}(C) = \{\}$ otherwise. Furthermore R_C is $\bigcup_{D \prec C} \mathrm{Gen}(D)$, and finally R is $\bigcup_D \mathrm{Gen}(D)$. Notably R^* is the unique minimal Herbrand model of M [BG91]. For ground terms l and r over Σ we have $M \models l \simeq r$ iff $R^* \models l \simeq r$ iff $l \downarrow_R r$.

Notably every inference conclusion makes the corresponding main premise redundant and hence can be turned into a simplification. This way the calculus decides satisfiability of finite ground Horn clause sets, which via splitting extends to the non-Horn case. Therefore it is an attractive basis for techniques to reason modulo $\mathcal{T}$.

4 A Calculus for $\mathcal{T}$-unsatisfiability

4.1 Calculus Rules

We now introduce a calculus $\mathcal{C}$ that shall detect unsatisfiability modulo $\mathcal{T}$. It works on finite or infinite sets of arbitrary clauses, ground or non-ground. For a substitution τ we say that it *numbers* if $\operatorname{cdom}\tau \subseteq [1;n]$, and that it in addition *minimally numbers* with respect to a set of conditions if these are satisfied with τ, but with no other numbering τ' more general than τ. Furthermore τ *ground numbers* a clause C if τ numbers and $C\tau$ is ground. The set of all ground instances of C under such substitutions is denoted by $\Omega(C)$, and its elements are called the Ω-*instances* of C.

The calculus $\mathcal{C}$ is specific in case that more than one literal is maximal in a premise under the unifier: Then we instantiate just as much as is necessary with elements of $[1;n]$ to dissolve this ambiguity. Therefore every ground instance of an inference satisfies the ordering conditions that make it an inference itself. In this sense, lifting is more economical than without instantiation.

As a second specialty, if a most general unifier is involved in an inference rule, then its range consists only of variables and digits. Hence many of the inferences in the standard calculus are not necessary here. For example, with the lexicographic path ordering [KL80] to the precedence $+ \succ s$, from the two clauses $(x+y)+z \simeq x+(y+z)$ and $u+s(v) \simeq s(v+u)$ one would normally obtain every $s^i(x+y)+z \simeq x+(s^i(y)+z)$. But since y needs to be bound to $s(v)$, no inference is drawn here.

Similar to the calculus $\mathcal{G}$, in every Horn clause with negative literals at least one of them shall be selected. In order to minimize the number of splits, we assume that for every non-Horn clause a partitioning into two subclauses is *designated* where each subclause has strictly less positive literals. Furthermore we stipulate that from now on the smallest ground terms are the digits from $[1;n]$, say such that $n \succ \ldots \succ 1$.

Rules of calculus $\mathcal{C}$:

Superposition left

$$\mathcal{I}\,\frac{l \simeq r \quad s[l'] \not\simeq t \vee C}{(s[r] \not\simeq t \vee C)\sigma\tau}$$

if
- $l' \notin \mathcal{V}$ and $\sigma = \operatorname{mgu}(l, l')$
- $\operatorname{ran}\sigma \subseteq \mathcal{V} \cup [1;n]$
- τ minimally numbers such that l and s are strictly greatest under $\sigma\tau$
- $s \not\simeq t$ is selected
- C is Horn

Superposition right

$$\mathcal{I}\,\frac{l \simeq r \quad s[l'] \simeq t}{(s[r] \simeq t)\sigma\tau}$$

if
- $l' \notin \mathscr{V}$ and $\sigma = \mathrm{mgu}(l, l')$
- $\operatorname{ran}\sigma \subseteq \mathscr{V} \cup [1; n]$
- τ minimally numbers such that l and s are strictly greatest under $\sigma\tau$ and $(l \simeq r)\sigma\tau \prec (s \simeq t)\sigma\tau$

Equality resolution

$$\mathcal{I}\,\frac{C \vee t \not\simeq t'}{C\sigma}$$

if
- $\sigma = \mathrm{mgu}(t, t')$
- $\operatorname{ran}\sigma \subseteq \mathscr{V} \cup [1; n]$
- $t \not\simeq t'$ is selected
- C is Horn

Split

$$\mathcal{S}\,\frac{C \vee s \simeq t \vee l \simeq r \vee D}{(C \vee s \simeq t)\tau \mid (l \simeq r \vee D)\tau}$$

if
- the partitioning is designated
- τ minimally numbers such that the conclusions share no variables

Regarding the two superposition rules, if for two given premises the number of substitutions τ that satisfy the side conditions is high, one could alternatively add only a single conclusion $(s[r] \not\simeq t \vee C)\sigma$ respectively $(s[r] \simeq t)\sigma$, which would preserve refutational completeness. The decision procedure that will be developed later relies on the former version, however.

We consider a clause C *redundant* with respect to a set M of clauses if $\Omega(M)^{\prec C\rho} \models C\rho$ holds for every ground numbering ρ; that is, if every Ω-instance of C follows from smaller clauses in $\Omega(M)$ already. By compactness no more than a finite subset of M is necessary. *Simplification,* in its general form, is making a clause redundant by adding (zero or more) entailed smaller clauses. Here it is already enough if these conditions hold on the Ω-instances.

Simplification

$$\mathcal{R}\,\frac{C}{\vec{C}'}\,\vec{D}$$

if
- C is redundant w.r.t. $\vec{C}', \vec{D}$
- $\Omega(C, \vec{D}) \models \Omega(\bigwedge \vec{C}')$
- $\Omega(C) \succ \Omega(\vec{C}')$

An inference with premises $\vec{C}$, most general unifier σ, minimally numbering substitution τ (identity in case of equality resolution), and conclusion D is *redundant* with respect to a set M of clauses if for every ground numbering ρ we have $\Omega(M)^{\prec \max\{\vec{C}\sigma\tau\rho\}} \models D\rho$. In the standard superposition calculus, the notions of redundancy and simplification refer to all ground instances,

not with respect to Ω-instances only. Many standard simplifications are also simplifications in our sense, which will be made precise in Sect. 4.3.

Derivations from an unaugmented clause set M do not necessarily produce the empty clause. For example, if $n = 2$, then there exist exactly four unary functions: a negation-like, two constant ones, and the identity. Each of these satisfies $f^3 = f$. Hence $f^3(c) \not\simeq f(c)$ is $\mathcal{T}$-unsatisfiable although no calculus rule is applicable to this disequation. We will therefore consider derivations from $M \cup \mathcal{T}'$ where $\mathcal{T}'$ consists of the following clauses:

$$f(\vec{x}) \simeq 1 \vee \ldots \vee f(\vec{x}) \simeq n \qquad \text{for any } f \in \Sigma \setminus [1; n]$$

$\mathcal{T}'$ is weaker than $\mathcal{T}$ in the sense that the upper cardinality bound is only applied to function values, but satisfies the same universal formulae.

4.2 Soundness and refutational completeness

Many parts of the correctness proof are rather standard, i. e., as for any superposition calculus with splitting. The first proposition relates $\mathcal{T}$-satisfiability with satisfiability of Ω-instances and justifies the exchange of $\mathcal{T}'$ for $\mathcal{T}$, since all Ω-instances of $\mathcal{T}$ are tautologies.

Proposition 4.1 A clause set M is $\mathcal{T}$-satisfiable iff $\Omega(M \cup \mathcal{T}')$ is satisfiable.

Proof: On the one hand, since $M, \mathcal{T} \models \Omega(M \cup \mathcal{T}')$, every $\mathcal{T}$-model of M is a model of $\Omega(M \cup \mathcal{T}')$ as well. On the other hand, consider any model $\mathcal{A}$ of $\Omega(M \cup \mathcal{T}')$. Its restriction to $\{1^{\mathcal{A}}, \ldots, n^{\mathcal{A}}\}$ is a Σ-algebra because of the range restriction on the functions, and it is a $\mathcal{T}$-model by construction. Finally every clause C is $\mathcal{T}$-equivalent to $\bigwedge \Omega(C)$. □

Next one has to show that within a derivation, satisfiability is inherited from each parent node to one of its immediate descendants.

Proposition 4.2 Let N denote a node in a derivation, with successors $N_1, \ldots, N_k$. If $\Omega(N)$ is satisfiable, so is some $\Omega(N_i)$.

Proof: According to the type of calculus step, we distinguish three cases.

- An inference: Here k equals 1, and N_1 is $N \cup \{C\}$ where C is N-valid. Hence N and N_1 are even equivalent.
- A simplification: Again k is 1, but N has a presentation $N = N' \cup \{C, \vec{D}\}$ such that $N_1 = N' \cup \{\vec{C'}, \vec{D}\}$. The side conditions imply $\Omega(\vec{D}) \models (\bigwedge \Omega(C)) \leftrightarrow (\bigwedge \Omega(\vec{C'}))$, such that the clause sets $\Omega(N)$ and $\Omega(N_1)$ are equivalent.

- A split: In our concrete split rule k equals 2. Let $C' \equiv (C \vee s \simeq t)\tau$ and $D' \equiv (l \simeq r \vee D)\tau$ denote the first and the second conclusion, respectively. Then $C' \vee D'$ is N-valid, and the disjuncts share no variables. If $\mathcal{A}$ is an N-model, then $\mathcal{A}$ satisfies at least one of C' and D', and therefore at least one of $N_1 = N \cup \{C'\}$ and $N_2 = N \cup \{D'\}$.

□

Accordingly, the clause set at the root is $\mathcal{T}$-satisfiable iff the derivation has a path each element of which is satisfiable.

Proposition 4.3 For every clause set M, the following are equivalent:
(i) M is $\mathcal{T}$-satisfiable.
(ii) Every derivation from $M \cup \mathcal{T}'$ contains a complete path $N_1, N_2, \ldots$ such that every $\Omega(N_i)$ is satisfiable.

Proof: If M is $\mathcal{T}$-satisfiable, then by Prop. 4.1 the set $\Omega(N_1) = \Omega(M \cup \mathcal{T}')$ is satisfiable, from which we can recursively construct a complete path as required by Prop. 4.2. The converse implication follows from $N_1 = M \cup \mathcal{T}$ by Prop. 4.1. □

If a clause C occurs at some point in a path, then each of its Ω-instances follows in the limit N_∞ from smaller or equal Ω-instances. Furthermore satisfiability of N_∞ with respect to Ω-instances is the conjunction of this property over all path elements.

Proposition 4.4 Consider a complete path $N_1, N_2, \ldots$ in some derivation.
(i) If $C \in N_i$ is ground numbered by ρ, then $\Omega(N_\infty)^{\preceq C\rho} \models C\rho$.
(ii) Every $\Omega(N_i)$ is satisfiable iff $\Omega(N_\infty)$ is.
(iii) N_∞ is saturated in case the derivation is fair.

Proof:
(i) The proof is by induction on Ω-instances. If $C \in N_\infty$ we are done. Otherwise C is contained in some N_j, but not in N_{j+1}, and $\Omega(\vec{C}', \vec{D})^{\prec C\rho} \models C\rho$ for appropriate $\vec{C}', \vec{D} \in N_{j+1}$. Either both vectors $\vec{C}', \vec{D}$ are empty and $C\rho$ is a tautology, or there is a greatest clause D' in $\Omega(\vec{C}', \vec{D})^{\prec C\rho}$. Inductively all elements of $\Omega(\vec{C}', \vec{D})^{\prec C\rho}$ are valid in $\Omega(N_\infty)^{\preceq D'}$, and so is $C\rho$.
(ii) Assume that every $\Omega(N_i)$ is satisfiable. By compactness $\Omega(N_\infty)$ is satisfiable iff each of its finite subsets is. Given one such subset M, for every Ω-instance $C\rho$ within there is an index j such that C is contained in N_j and all successors thereof. Since M is finite, these indices have a finite maximum k. Now $\Omega(N_k)$ comprises M and is satisfiable by assumption.

As to the converse implication, consider an Ω-instance $C\rho$ of a clause $C \in N_i$. Then $\Omega(N_\infty)$ entails $C\rho$ by Prop. 4.4 (i). In other words, any model of $\Omega(N_\infty)$ is a model of $\Omega(N_i)$.

(iii) Firstly we consider an inference with premises $\vec{C}$ from N_∞ and conclusion D with ground numbering substitution ρ. Because of fairness $\Omega(N_i)^{\prec \max\{\vec{C}\rho\}} \models D\rho$ holds for some i, which can be rephrased as $C'_1\rho_1, \ldots, C'_k\rho_k \models D\rho$ for clause instances $C'_j\rho_j$ from $\Omega(N_i)$ below $\max\{\vec{C}\rho\}$. By Prop. 4.4 (i) these clause instances are valid in $\Omega(N_\infty)$ below $\max\{\vec{C}\rho\}$, and so is $D\rho$.
Secondly we study a split from a persistent clause $C \equiv C_1 \vee C_2$ with designated partitioning as indicated and minimally numbering substitution τ. Because of fairness, one split conjunct, say $C_1\tau$, is contained in some N_i or redundant with respect to it. So either $C_1\tau$ is persistent, or $C_1\tau$ is redundant with respect to some N_j where $j \geq i$. In the former case the proof is finished. In the latter we have $\Omega(N_j)^{\prec C_1\rho} \models C_1\rho$ for every ground numbering $\rho = \tau\tau'$, which extends to $\Omega(N_\infty)^{\prec C_1\rho} \models C_1\rho$ with an argument like in the preceding paragraph.

□

For any clause set M, let $\widehat{M}$ denote its Ω-instances which are Horn clauses; for paths let $\widehat{N} = \widehat{N_\infty}$. Then $\Omega(M)$ and $\widehat{M}$ are equivalent for $\mathcal{C}$-saturated clause sets M.

Proposition 4.5 $\Omega(M)$ and $\widehat{M}$ are equivalent for $\mathcal{C}$-saturated clause sets M.

Proof: We show by induction on clause instances that every non-Horn clause $C\rho \in \Omega(M)$ is entailed by $\widehat{M}$. Now, C has a presentation $C \equiv C_1 \vee C_2 \equiv (C'_1 \vee s \simeq t) \vee (l \simeq r \vee C'_2)$ such that the partitioning into C_1 and C_2 is designated. Then ρ numbers the clause C such that the subclauses C_1 and C_2 are variable disjoint. More general such substitutions τ have to satisfy $\tau \subseteq \rho$. There exists a $\subset$-minimal such τ because all $\subset$-chains are finite. Then $C \vdash C_1\tau \mid C_2\tau$ is a valid $\mathcal{C}$-split. Because M is saturated, one split conjunct, say $C_1\tau$, is contained in M or redundant with respect to M. In both cases we have $\Omega(M) \models C_1\rho$, and we obtain inductively $\widehat{M} \models C_1\rho$. Finally $C_1\rho$ entails $C\rho$. □

The crucial lifting result is the following:

Proposition 4.6 If a clause set M is $\mathcal{C}$-saturated, then $\widehat{M}$ is $\mathcal{G}$-saturated.

Proof: We adapt the usual lifting arguments to our calculus, inspecting $\mathcal{G}$-inferences with premises from $\widehat{M}$. If a clause $D \in \widehat{M}$ contains negative

literals, then let the literal selection be inherited from one arbitrary $C \in M$ that instantiates into D.

- Ground superposition right: Given two clauses $l \simeq r$ and $s \simeq t$ from M with ground numbering substitution ρ, consider the $\mathcal{G}$-inference with premises $l\rho \simeq r\rho$ and $s\rho[l\rho]_p \simeq t\rho$, and conclusion $s\rho[r\rho]_p \simeq t\rho$. The position p is within s because the range of ρ consists of digits only. This $\mathcal{G}$-inference corresponds to a variable overlap if $s|_p \equiv x \in \mathscr{V}$, and to a non-variable overlap otherwise.
  In the former case we have $x\rho \equiv l\rho$, such that $l\rho$ is a digit. Because $l\rho \succ r\rho$ and the digits are the smallest ground terms, the term $r\rho$ must be a digit as well. Let ρ' denote the substitution identical to ρ except that $x\rho' \equiv r\rho$. Then $(s \simeq t)\rho'$ is contained in $\Omega(M)$ and makes the inference redundant.
  Now we come to non-variable overlaps. Let $l' \equiv s|_p$, furthermore $\sigma = \mathrm{mgu}(l, l')$ with $\mathrm{dom}\,\sigma \subseteq \mathrm{var}(l, l')$, and $\rho = \sigma\sigma'$. Because ρ is ground numbering, we know that $x\rho$ is a digit for every $x \in \mathrm{dom}\,\sigma$. Given $\rho = \sigma\sigma'$, every $x\sigma$ is either a digit or a variable.
  The substitution σ' numbers the clauses $s\sigma \simeq t\sigma$ and $l\sigma \simeq r\sigma$ such that the literals $l\sigma$ and $s\sigma$ are greatest under σ', respectively, and that $(l \simeq r)\sigma\sigma' \prec (s \simeq t)\sigma\sigma'$. If τ is a more general such substitution, then it satisfies $\mathrm{dom}\,\tau \subseteq \mathrm{dom}\,\sigma'$ and $x\tau \equiv x\sigma'$ for every $x \in \mathrm{dom}\,\tau$, which implies $\tau \subseteq \sigma'$. There exists a $\subset$-minimal such τ because all $\subset$-chains are finite. Summing it up: $l \simeq r,\ s[l'] \simeq t \vdash (s[r] \simeq t)\sigma\tau$ is a $\mathcal{C}$-inference with premises from M, and is redundant with respect to M because M is saturated. If $\sigma' = \tau\tau'$, then the inference instance under τ' is redundant with respect to $\Omega(M)$.
- Ground equality resolution: Consider a Horn clause $C \vee t \not\simeq t' \in M$ with ground numbering substitution ρ such that $C\rho \vee t\rho \not\simeq t'\rho \vdash C\rho$ is a $\mathcal{G}$-inference. We may assume that $t \not\simeq t'$ is selected in $C \vee t \not\simeq t'$. As usually, t and t' have a most general unifier σ, which specializes into ρ say via σ'. We obtain $\mathrm{cdom}\,\sigma \subseteq \mathscr{V} \cup [1; n]$ like for ground superposition right. So $C \vee t \not\simeq t' \vdash C\sigma$ is a $\mathcal{C}$-inference with premises from M; and its redundancy carries over to that of the above instance.
- Ground superposition left: similar to ground superposition right, but taking selectedness into account like for ground equality resolution.

□

Putting everything together, the calculus $\mathcal{C}$ is sound and refutationally complete:

Lemma 4.7 For every clause set M, the following are equivalent:
(i) M is $\mathcal{T}$-satisfiable.

(ii) Every fair derivation from $M \cup \mathcal{T}'$ contains a complete path $N_1, N_2, \ldots$ such that the empty clause is not in N_∞.

Proof: We successively transform the first characterization into the second. By Prop. 4.3 the clause set M is $\mathcal{T}$-satisfiable iff there exists a complete path $N_1, N_2, \ldots$ such that every $\Omega(N_i)$ is satisfiable, or such that $\Omega(N_\infty)$ is, by Prop. 4.4 (ii). Because of Prop. 4.4 (iii) every N_∞ is saturated with respect to $\mathcal{C}$. Hence by Prop. 4.5 the sets $\Omega(N_\infty)$ and $\widehat{N}$ are equivalent, and the latter is saturated with respect to $\mathcal{G}$. Since $\mathcal{G}$ is sound and complete, the satisfiability of $\widehat{N}$ is equivalent to $\bot \notin \widehat{N}$, which is the same as $\bot \notin N_\infty$. □

Notably the minimality of the digits is indispensable for refutational completeness: Assume that $\succ$ is the lexicographic path ordering to the precedence $n \succ \ldots \succ 1 \succ f \succ c$. Then from the unsatisfiable clause set $\{f(x) \simeq 1,\ 1 \simeq c,\ 1 \not\simeq f(c)\}$ nothing but the clause $f(c) \not\simeq c$ is inferable. We have needed this minimality in the proposition on lifting to show that variable overlaps are non-critical; and indeed the variable overlap from $1 \simeq c$ into $f(x) \simeq 1$ would produce $f(c) \simeq 1$ and eventually lead to the empty clause.

4.3 Redundancy in more detail

In the calculus $\mathcal{C}$, redundancy on the general level is defined via redundancy of Ω-instances on the ground level, whereas in standard superposition one goes back to redundancy of all ground instances. Let us compare under which conditions a clause C is redundant with respect to a clause set M. In the calculus $\mathcal{C}$ we require $\Omega(M)^{\prec C\rho} \models C\rho$ for every ground numbering ρ. The condition in standard superposition is $\mathrm{gnd}(M)^{\prec C\sigma} \models C\sigma$ for every ground substitution σ, where $\mathrm{gnd}(M)$ denotes the set of all ground instances of M. So for redundancy in the sense of $\mathcal{C}$ fewer instances need to be shown redundant, but on the other hand there are fewer premises for doing so. For example, $f(g(1)) \simeq 1$ is not redundant with respect to $f(x) \simeq 1$, since it is not entailed from $f(1) \simeq 1, \ldots, f(n) \simeq 1$. Fortunately, in $\mathcal{C}$-derivations the set M with respect to which redundancy is studied always contains the clauses of $\mathcal{T}'$, possibly simplified. Therefore we additionally have $g(1) \simeq 1 \vee \ldots \vee g(1) \simeq n$ at hand, with which $f(g(1)) \simeq 1$ does become redundant.

In this subsection, we develop two results that generalize this obeservation. Firstly, if every digit instance $C\rho$ is entailed from smaller ground instances of M except some problematic ones, then C is redundant in the sense of $\mathcal{C}$. Secondly, if every $C\rho$ follows from arbitrary smaller ground instances, but C is not of a particular form, then C is also redundant. These results permit to adapt concrete simplification techniques like rewriting or

subsumption to our calculus. The subsection ends with a demonstration that $\mathcal{C}$ should not be mixed with the standard notion of redundancy.

4.3.1 Deducing ground instances from digit instances

In the following we will prove that a ground instance $C\sigma$ of a clause C follows from $\Omega(C, \mathcal{T}')$, and give a criterion when this entailment is from smaller instances. We reserve the identifier f for non-digit function symbols, whereas i, j, k denote digits and $\vec{i}$, $\vec{j}$ vectors thereof. For any term t, let $\mathrm{Dig}(t)$ denote the clause $t \simeq 1 \vee \ldots \vee t \simeq n$.

Proposition 4.8 For every clause C and term t, the following entailment holds: $C\{x \mapsto 1\}, \ldots, C\{x \mapsto n\}, \mathrm{Dig}(t) \models C\{x \mapsto t\}$

Proof: Consider a model $\mathcal{A}$ of the premises. Then there exists a digit i fulfilling $\mathcal{A} \models t \simeq i$. This identity inductively lifts to term contexts, and as equivalence to clause contexts. In particular $\mathcal{A} \models C\{x \mapsto i\}$ implies $\mathcal{A} \models C\{x \mapsto t\}$. □

Proposition 4.9 Let C denote a clause with ground substitution $\sigma = \{x_1 \mapsto t_1, \ldots, x_m \mapsto t_m\}$. Then $\Omega(C), \mathrm{Dig}(t_1), \ldots, \mathrm{Dig}(t_m) \models C\sigma$ holds.

Proof: The proof is by induction on m. If σ is the identity we are done. Otherwise we decompose σ according to $\sigma = \{x_1 \mapsto t_1, \ldots, x_m \mapsto t_m\} \cup \{x_{m+1} \mapsto t_{m+1}\} = \sigma_1 \cup \sigma_2$. Since the substitutions are ground we have $\sigma_1 \cup \sigma_2 = \sigma_1 \circ \sigma_2$. Inductively we obtain $\Omega(C\sigma_1), \mathrm{Dig}(t_1), \ldots, \mathrm{Dig}(t_m) \models C\sigma_1$. Proposition 4.8 gives $C\sigma_1, \mathrm{Dig}(t_{m+1}) \models C\sigma_1\sigma_2$. □

Proposition 4.10 Ground terms t obey $\Omega(\mathcal{T}') \models \mathrm{Dig}(t)$.

Proof: We induct on the structure of t. In case $t \equiv i$ the clause $\mathrm{Dig}(t)$ is a tautology. In case $t \equiv f(\vec{t})$ the proposition $\Omega(\mathcal{T}') \models \mathrm{Dig}(t_j)$ is inductively true for every j. Furthermore $\mathcal{T}'$ contains $\mathrm{Dig}(f(\vec{x}))$. Let $\sigma = \{x_1 \mapsto t_1, \ldots, x_m \mapsto t_m\}$, such that $f(\vec{t}) \equiv f(\vec{x})\sigma$. With Prop. 4.9 we obtain $\Omega(\mathrm{Dig}(f(\vec{x}))), \mathrm{Dig}(t_1), \ldots, \mathrm{Dig}(t_m) \models \mathrm{Dig}(f(\vec{x}))\sigma$. □

Proposition 4.11 $\Omega(C, \mathcal{T}') \models C\sigma$ is true for every clause C with ground substitution σ.

Proof: Assume $\sigma = \{x_1 \mapsto t_1, \ldots, x_m \mapsto t_m\}$. Then Prop. 4.10 implies $\Omega(\mathcal{T}') \models \mathrm{Dig}(t_i)$ for every i, such that from Prop. 4.9 finally we obtain $\Omega(C), \mathrm{Dig}(t_1), \ldots, \mathrm{Dig}(t_m) \models C\sigma$. □

We have seen in Prop. 4.10 that every ground term t is subject to $\Omega(\mathcal{T}') \models \mathrm{Dig}(t)$. In the following we will exploit that usually not all of $\Omega(\mathcal{T}')$ is needed for this entailment. There exist subsets $T \subseteq \Omega(\mathcal{T}')$ such that $T \models \mathrm{Dig}(t)$ holds. By compactness there are finite such T even in case the signature is infinite. Let $\Delta(t)$ denote the smallest of these finite T, with respect to the ordering on clause sets. Let furthermore $\delta(t)$ denote the greatest clause in $\Delta(t) \cup \{\bot\}$, and for ground substitutions σ let $\delta(\sigma)$ stand for the greatest clause in $\delta(\mathrm{ran}\,\sigma) \cup \{\bot\}$. Actually one can construct $\Gamma(t)$ recursively, but this is not necessary for our purposes.

Proposition 4.12 Entailment from $\Omega(\mathcal{T}')$ can be restricted by the bounds $\delta(t)$ and $\delta(\sigma)$:

(i) Every ground term t satisfies $\Omega(\mathcal{T}')^{\preceq\delta(t)} \models \mathrm{Dig}(t)$.

(ii) If σ is a ground substitution for C, then $\Omega(C), \Omega(\mathcal{T}')^{\preceq\delta(\sigma)} \models C\sigma$ holds.

Proof:

(i) By definition we have $\Delta(t) \subseteq \Omega(\mathcal{T}')^{\preceq\delta(t)}$ and $\Delta(t) \models \mathrm{Dig}(t)$.

(ii) Let $\sigma = \{x_1 \mapsto t_1, \ldots, x_m \mapsto t_m\}$. Then we obtain $\Omega(\mathcal{T}')^{\preceq\delta(t_i)} \models \mathrm{Dig}(t_i)$ from Prop. 4.12 (i) for every i, and $\Omega(\mathcal{T}')^{\preceq\delta(\sigma)} \models \mathrm{Dig}(t_i)$ by definition of $\delta(\sigma)$. Finally we apply Prop. 4.9 to C and σ.

□

Proposition 4.13 For ground terms t we have $\delta(t) \equiv \bot$ iff t is a digit.

Proof: In case t is a digit, then $\mathrm{Dig}(t)$ is a tautology and $\Delta(t)$ is empty. Otherwise $\mathrm{Dig}(t)$ is not a tautology. □

Proposition 4.14 If t is a ground term and δ a ground substitution, then we can give estimates for $\delta(t)$ and $\delta(\sigma)$ as follows:

(i) $\delta(t) \equiv \mathrm{Dig}(u)$ implies $t \succeq u$.

(ii) $\delta(\sigma) \equiv \mathrm{Dig}(u)$ entails $\max(\mathrm{ran}\,\sigma) \succeq u$.

Proof:

(i) The proof is by induction on the term structure. If t is a digit, then we have $\delta(t) \equiv \bot$ by Prop. 4.13, and there is nothing to show. The case $t \equiv f(\vec{t})$ remains. Let $i_1, \ldots, i_k$ denote exactly the indices for which t_j is not a digit, and let $t' \equiv f(\vec{t})[x_1]_{i_1} \ldots [x_k]_{i_k}$. So t' is obtained from t replacing every non-digit t_j with a fresh variable. Conversely, using $\sigma = \{x_1 \mapsto t_{i_1}, \ldots, x_k \mapsto t_{i_k}\}$ one can instantiate t' back into t again. In case $k = 0$ the argument vector $\vec{t}$ contains only digits. Choosing $T = \{\mathrm{Dig}(t)\}$ implies $T \subseteq \Omega(\mathcal{T}')$ and $T \models \mathrm{Dig}(t)$. Therefore we have

$T \succeq \Delta(t)$ and $\max T \succeq \max \Delta(t) \equiv \delta(t)$, hence $\mathrm{Dig}(t) \succeq \mathrm{Dig}(u)$ and finally $t \succeq u$.
In case $k > 0$ every $\delta(t_{i_j})$ is distinct from $\bot$ by Prop. 4.13, and there exists a ground term v such that $\mathrm{Dig}(v) \equiv \max_j \delta(t_{i_j})$. By induction hypothesis and the subterm property of t we obtain $t \succ v$. Here we choose $T = \Omega(\mathrm{Dig}(t')) \cup \Omega(\mathcal{T}')^{\preceq \mathrm{Dig}(v)}$, which satisfies $T \subseteq \Omega(\mathcal{T}')$. By construction $T \models \mathrm{Dig}(t_{i_j})$ holds for every j. Proposition 4.9 yields $\Omega(\mathrm{Dig}(t')), \mathrm{Dig}(t_{i_1}), \ldots, \mathrm{Dig}(t_{i_k}) \models \mathrm{Dig}(t'\sigma)$. Hence we may conclude that $T \succeq \Delta(t)$ and $\max T \succeq \mathrm{Dig}(u)$. Next we compare T with $\{\mathrm{Dig}(t)\}$. We have $\Omega(\mathrm{Dig}(t')) \prec \{\mathrm{Dig}(t)\}$ by minimality of the digits, and furthermore $\Omega(\mathcal{T}')^{\preceq \mathrm{Dig}(v)} \prec \{\mathrm{Dig}(t)\}$ because of $v \prec t$. Hence we may conclude that $\mathrm{Dig}(t) \succ \max T \succeq \mathrm{Dig}(u)$ holds, such that $t \succ u$ is true.

(ii) Let $\sigma = \{x_1 \mapsto t_1, \ldots, x_m \mapsto t_m\}$. Because of $\delta(\sigma) \not\equiv \bot$ we have $\delta(\sigma) \equiv t_i$ for some i. Using Prop. 4.14 (i) we may conclude that $\max_j t_j \succeq t_i \succeq u$ holds.

□

Given a clause C with ground substitution σ, we call the pair C, σ *problematic* if $x\sigma \equiv f(\vec{\imath})$ for some $x \in \mathrm{var}(C)$ and $C\sigma \preceq \mathrm{Dig}(f(\vec{\imath}))$. Otherwise the pair is called *unproblematic*. Let furthermore denote $\mathring{\mathrm{gnd}}(C)$ the set of all ground instances $C\sigma$ for which C, σ is unproblematic, and let $\mathring{\mathrm{gnd}}$ extend to clause sets in the usual way.

Here are two necessary and quite restrictive conditions for C, σ to be problematic: Firstly some variable $x \in \mathrm{var}(C)$ may occur only in literals of the form $x \simeq i$ and $x \simeq y$. Secondly the greatest literal of $C\sigma$ must have the form $f(\vec{\imath}) \simeq j$.

Proposition 4.15 Let C denote a clause with ground substitution σ such that σ is not numbering, and that C, σ is unproblematic. Then $\Omega(C, \mathcal{T}')^{\prec C\sigma} \models C\sigma$ holds.

Proof: We decompose $\sigma = \sigma_1 \cup \sigma_2$ such that the range of σ_1 contains only digits and the range of σ_2 only non-digits. Since the substitutions are ground we have $\sigma = \sigma_1 \circ \sigma_2$. Proposition 4.12 (ii) implies $\Omega(C\sigma_1), \Omega(\mathcal{T}')^{\preceq \delta(\sigma_2)} \models C\sigma_1\sigma_2$. The substitution σ_2 is not empty because σ is not numbering. Hence we have by minimality of the digits $\Omega(C\sigma_1) \prec \{C\sigma_1\sigma_2\}$. We still have to show $\delta(\sigma_2) \prec C\sigma$. Let t denote the greatest term in $\mathrm{ran}\,\sigma_2$. By Prop. 4.13 the clause $\delta(t)$ equals $\mathrm{Dig}(f(\vec{\imath}))$ for some term $f(\vec{\imath})$. By Prop. 4.14 (ii) we have $t \succeq f(\vec{\imath})$. If $t \succ f(\vec{\imath})$, then the greatest term of $C\sigma$ is above the greatest of $\delta(\sigma_2)$. Otherwise we obtain $C\sigma \succ \mathrm{Dig}(f(\vec{\imath}))$ from the requirement that C, σ is unproblematic. □

4.3.2 Using ground instances in redundancy proofs

We have seen in the preceding proposition that unproblematic ground instances of clauses follow from smaller digit instances of the same clause and of $\mathcal{T}'$. Hence these ground instances can safely be used in redundancy proofs as if they were digit instances. Alternatively we study for which clauses it is safe to use arbitrary ground instances when showing them redundant. A clause C is called *critical* if it has an Ω-instance $C\rho$ with greatest term $f(\vec{\imath})$ such that $C\rho \preceq \mathrm{Dig}(f(\vec{\imath}))$. Otherwise C is called *noncritical.*

Lemma 4.16 Consider a path in a $\mathcal{C}$-derivation from $M \cup \mathcal{T}'$ to N and a clause C. Then C is redundant with respect to N if one of the following conditions holds, where ρ ranges over all ground numbering substitutions:

(i) $\mathring{\mathrm{gnd}}(N)^{\prec C\rho} \models C\rho$ for all ρ,

(ii) $\mathrm{gnd}(N)^{\prec C\rho} \models C\rho$ for all ρ and C is noncritical.

Proof:

(i) Given an arbitrary ground numbering substitution ρ, there exist clauses $D_1, \ldots, D_m \in N$ and ground substitutions $\sigma_1, \ldots, \sigma_m$ such that every D_i, σ_i is unproblematic and $D_i\sigma_i \prec C\rho$, and that $D_1\sigma_1, \ldots, D_m\sigma_m \models C\rho$. We have shown $\Omega(N)^{\prec C\rho} \models C\rho$ if $\Omega(N)^{\prec C\rho} \models D_i\sigma_i$ holds for every i. If $D_i\sigma_i$ is a digit instance of D_i, then we have $D_i\sigma_i \in \Omega(N)^{\prec C\rho}$. Otherwise Prop. 4.15 ensures $\Omega(D_i, \mathcal{T}')^{\prec D_i\sigma_i} \models D_i\sigma_i$ because D_i, σ is unproblematic. If some element of $\mathcal{T}'$ is no longer present in N, then by a monotonicity argument similar to Prop. 4.4 (i) each of its digit instances follows from smaller clauses in $\Omega(N)$. Hence we get $\Omega(N)^{\prec D_i\sigma_i} \models D_i\sigma_i$ and $\Omega(N)^{\prec C\rho} \models D_i\sigma_i$.

(ii) Similar to the proof of Lem. 4.16 (i), for every ground numbering substitution ρ there exist clauses $D_1, \ldots, D_m \in N$ and ground substitutions $\sigma_1, \ldots, \sigma_m$ such that always $D_i\sigma_i \prec C\rho$, and that $D_1\sigma_1, \ldots, D_m\sigma_m \models C\rho$. If $C\rho$ is a tautology we are done. Otherwise we decompose every $\sigma_k = \sigma'_k \cup \sigma''_k$ such that the range of σ'_k contains only digits and the range of σ''_k only non-digits. Proposition 4.12 (ii) guarantees that $\Omega(D_k\sigma'_k), \Omega(\mathcal{T}')^{\preceq \delta(\sigma''_k)} \models D_k\sigma_k$. By minimality of the digits we obtain $\Omega(D_k\sigma'_k) \preceq \{D_k\sigma_k\} \prec \{C\rho\}$.
Next we show that $\delta(\sigma''_k) \prec C\rho$. The clause C is not empty since otherwise $\models \bot$; so $C\rho$ has a greatest term s. Let t denote the greatest term of $D_k\sigma_k$, then we have $s \succeq t$. If $\delta(\sigma''_k) \equiv \bot$ then $\bot \prec C\rho$. Otherwise $\delta(\sigma''_k)$ has the shape $\mathrm{Dig}(f(\vec{\imath}))$. Because of Prop. 4.14 (ii) we have $\max(\operatorname{ran}\sigma''_k) \succeq f(\vec{\imath})$, and because of $t \succeq \max(\operatorname{ran}\sigma''_k)$ we have $s \succeq f(\vec{\imath})$ as well. Now $s \succ f(\vec{\imath})$ directly entails $C\rho \succ \delta(\sigma''_k) \equiv \mathrm{Dig}(f(\vec{\imath}))$. Other-

wise s equals $f(\vec{\imath})$, and $C\rho \succ \mathrm{Dig}(f(\vec{\imath}))$ holds because C is noncritical by assumption.
Summing it up, we obtain $\Omega(D_k\sigma'_k, \mathcal{T}')^{\prec C\rho} \models D_k\sigma_k$ and therefore as well $\Omega(D_k, \mathcal{T}')^{\prec C\rho} \models D_k\sigma_k$. If some element of $\mathcal{T}'$ is no longer present in N, then by a monotonicity argument similar to Prop. 4.4 (i) each of its digit instances follows from smaller clauses in $\Omega(N)$. Hence we get $\Omega(N)^{\prec C\rho} \models D_k\sigma_k$.

□

4.3.3 Incompatibility of $\mathcal{C}$ with standard redundancy

The difference of our redundancy notion to the one of standard superposition may show up in practice: Assume $n = 2$ and some input M which via $\mathcal{C}$ eventually leads to the clause set $N = \{x \simeq 1,\ f(1) \simeq 2,\ f(2) \simeq 2,\ f(1) \not\simeq 1\}$. Now the clause $x \simeq 1$ has the ground instances $2 \simeq 1$ and $f(1) \simeq 1$ which make the second and the third clause redundant in the standard sense. Since $f(1) \simeq 1$ is not an Ω-instance of $x \simeq 1$, these clauses are not redundant in the sense of $\mathcal{C}$. Going further, the example shows that combining $\mathcal{C}$ with standard redundancy is problematic: If $f(1) \simeq 2$ and $f(2) \simeq 2$ were deleted from N, then the rest $\{x \simeq 1,\ f(1) \not\simeq 1\}$ would be $\mathcal{C}$-saturated, despite the apparent unsatisfiability. Summing it up, refutational completeness would be lost. However, because of Lem. 4.16 only in rare cases standard redundancy is stronger than redundancy in the sense of $\mathcal{C}$.

4.4 Model extraction

Here we study the case that a fair derivation from $M \cup \mathcal{T}'$ contains a complete path $N_1, N_2, \ldots$ without the empty clause. Let R denote the rewrite system that the superposition model functor of Sect. 3 produces from $\widehat{N}$. Then R^* is a witness that M is $\mathcal{T}$-satisfiable:

Proposition 4.17 R^* is subject to the following properties:
(i) $C \in N_i$ implies $R^* \models \bigwedge \Omega(C)$.
(ii) For every ground term t there exists some digit j such that $R^* \models t \simeq j$.
(iii) R^* is a $\mathcal{T}$-model of M.

Proof:
(i) From Prop. 4.4 (i) we obtain $\Omega(N_\infty) \models \bigwedge \Omega(C)$. Because of Prop. 4.4 (iii), the limit N_∞ is $\mathcal{C}$-saturated. So by Prop. 4.5 the clause sets $\Omega(N_\infty)$ and $\widehat{N}$ are equivalent. Because of Prop. 4.6, the set $\widehat{N}$ is $\mathcal{G}$-saturated. Finally R^* is a model of $\widehat{N}$.

(ii) If t is a digit itself, then we are done. Otherwise $t \equiv f(\vec{t})$; and inductively $R^* \models \bigwedge_k t_k \simeq i_k$ for some digit vector $\vec{\imath}$. Because of $\mathcal{T}' \subseteq N_1$ there is a clause $\bigvee_j f(\vec{x}) \simeq j$ in N_1, which has an Ω-instance under the substitution $\rho = \{\vec{x} \mapsto \vec{\imath}\}$. This Ω-instance $f(\vec{\imath}) \simeq 1 \vee \ldots \vee f(\vec{\imath}) \simeq n$ holds in R^* by Prop. 4.17 (i); and necessarily R^* satisfies one disjunct.

(iii) Firstly we show that R^* satisfies $\forall x. \bigvee_i x \simeq i$. Consider an R^*-assignment μ such that $\mu(x) = [t]_R$. Then t is a ground term by construction of R^*, such that $[t]_R = [j]_R$ by Prop. 4.17 (ii). Secondly, given $C \in M \subseteq N_1$, we get $R^* \models \bigwedge \Omega(C)$ from Prop. 4.17 (i). Now, C and $\bigwedge \Omega(C)$ are $\mathcal{T}$-equivalent, and R^* is a $\mathcal{T}$-model.

□

Next we will demonstrate that ordered rewriting is sufficient to extract the $\mathcal{T}$-model R^*. By Prop. 4.17 (ii) the carrier of R^* is given by $\{[1]_R, \ldots, [n]_R\}$, where some of the classes may coincide. Since the digits from $[1; n]$ are the smallest ground terms, every non-digit ground term $f(\vec{t})$ is R-reducible.

In order to rephrase this reducibility in terms of N_∞, let $E_\infty \subseteq N_\infty$ denote the set of all persistent unit equations, and E_k the corresponding subset of every N_k. For an arbitrary set E of such equations, the ordered rewrite relation $\to_E$ is the smallest relation on terms such that $u[s\sigma] \to_E u[t\sigma]$ whenever $s \simeq t \in E$, $s\sigma \succ t\sigma$ and $(\mathrm{var}(t) \setminus \mathrm{var}(s))\sigma \equiv \{1\}$. The third condition ensures that given say $f(x) \simeq f(y)$, the term $f(n)$ can only be rewritten to $f(1)$, thus eliminating the need to search decreasing y-instances. This restricted version of ordered rewriting with respect to E_∞ reduces every ground term to its R-normal form:

Proposition 4.18 On ground terms, E_∞-normal forms are unique and coincide with R-normal forms.

Proof: Let $E = E_\infty$. To start with, we prove $\to_R \subseteq \to_E$: Assume t is R-reducible say with $l \to r$ generated from $u \simeq v \in E$ instantiated via some ground numbering ρ. If every variable of v occurs in u, then we have directly $t \equiv t[l] \equiv t[u\sigma] \to_E t[v\sigma] \equiv t[r]$. Otherwise we still have to show that $x\rho \equiv 1$ for any x exclusive to v. Imagine $x\rho \succ 1$, and let ρ' coincide with ρ except that $x\rho' \equiv 1$. Since $l \to r$ has been generated, we know that $u\rho$ is $R_{(u \simeq v)\rho}$-irreducible. Because of $v\rho \succ v\rho'$, the term $u\rho$ is $R_{(u \simeq v)\rho'}$-irreducible as well. But then $u\rho \to v\rho'$ would have been generated and not $u\rho \to v\rho$.

Consider now a ground rewrite step with $u \simeq v \in E$ instantiated via σ. Let ρ denote the substitution that maps every x to $x\sigma{\downarrow_R}$. Then $R^* \models u\rho \simeq v\rho$ holds because R^* is a model of $\Omega(E) \subseteq \widehat{N}$. Because of $u\rho{\downarrow_R} u\sigma$ we obtain $R^* \models u\sigma \simeq v\sigma$. So we have $\to_E \subseteq \downarrow_R$ on ground terms.

This extends to ${}_E\!\leftarrow \circ \rightarrow_E \subseteq \downarrow_R \circ \downarrow_R \subseteq \leftrightarrow_R^* \subseteq \downarrow_R \subseteq \downarrow_E$, by the Church-Rosser property of R. Since the relation $\rightarrow_E$ is terminating by construction, it is also ground confluent, such that ground normal forms are unique.

Finally, E-irreducibility implies R-irreducibility because of $\rightarrow_R \subseteq \rightarrow_E$; and the converse holds because of $\rightarrow_E \subseteq \downarrow_R$. □

4.5 Termination

The calculus $\mathcal{C}$ is refutationally complete. If M is $\mathcal{T}$-unsatisfiable, then in every path of any fair derivation, eventually the empty clause will show up, even for infinite M. Since the derivation tree is finitely branching, any such derivation is finite. If M is $\mathcal{T}$-satisfiable, however, then derivations without suitable simplification steps may become infinite. We will design a simplification rule that can enforce termination. In order to make this effective, naturally the input clause set M must be finite, and so is the signature Σ; and the ordering $\succ$ must be decidable.

We have seen in the preceding section that, unless the empty clause has been derived, every function application to digits, i.e., every $f(\vec{\imath})$, is reducible with respect to the limit E_∞. The key observation now is that E_∞ can sufficiently be approximated finitely: Only finitely many of the persistent equations can actually reduce the terms $f(\vec{\imath})$; and these are all present from some E_κ on. Given a non-ground function occurrence, each of its finitely many Ω-instances can then be simplified into a digit, which simplifies the non-ground expression provided some ordering restriction is met. In the end, non-digit function symbols only occur on top-level.

Formally, given a clause set N with unit equations $E \subseteq N$, we say that N *reduces to digits* if $f(\vec{\imath}) \rightarrow_E^* j$ for every digit vector $\vec{\imath}$. Inductively every ground term can then be rewritten to a digit as well. Furthermore, a term is called $[1;n]$-*shallow* if it is either a variable, or a digit, or a non-digit function symbol applied to a sequence of digits or variables. That is, no nested function applications are allowed. We speak of a $[1;n]$-*shallow clause* if non-digit function symbols occur only at the top-level of positive literals.

For any set E of unit equations, we extend the ordered rewrite relation $\rightarrow_E$ from terms to clauses in the obvious way. As such, it is a simplification in the sense of the calculus $\mathcal{C}$ only if the clause to be simplified is above the simplifying equation instances. For example, $f(3) \simeq 1 \rightarrow_{\{f(3)\simeq 2\}} 2 \simeq 1$ is a rewrite step, but not a simplification, because the clause to be rewritten is smaller than the one used for rewriting. In order to capture this, let $\rightarrow_E^\succ$ denote the smallest relation on clauses such that $C[s\sigma] \rightarrow_E^\succ C[t\sigma]$ when-

ever $s \simeq t \in E$, $s\sigma \to_E t\sigma$ and $C\rho \succ (s \simeq t)\sigma\rho$ for all ground numbering ρ. The latter condition weakens the one of the standard calculus, which is $C \succ (s \simeq t)\sigma$. A further condition is necessary to meet the requirements of Lem. 4.16: Rewriting $C \to_E^* D$ is called Ω-*admissible* for any noncritical C. If C is critical, however, then it contains literals of the shape $f(\vec{s}) \simeq t$ where t and every s_i is a digit or a variable, such that with a suitable ground numbering substitution ρ the term $f(\vec{s})\rho$ is the greatest of $C\rho$. Then $C \to_E^* D$ is Ω-*admissible* only if rewrite steps on the left-hand side of such literals $f(\vec{s}) \simeq t$ with equations $x \simeq i \in E$ or $x \simeq y \in E$ only take place below f. Summing it up, the following is an instance of $\mathcal{C}$-simplification:

Ordered unit rewriting

$$\mathcal{R}\,\frac{C}{D}\,E \qquad \text{if} \begin{array}{l} \cdot\ E \text{ is a set of unit equations} \\ \cdot\ C \to_E^{\succ} \circ \to_E^* D \\ \cdot\ C \to_E^* D \text{ is } \Omega\text{-admissible} \end{array}$$

Alas, even when every term $f(\vec{\imath})$, and every ground term $f(\vec{t})$, is reducible, non-ground terms still may resist simplification. Hence we need to combine instantiation and rewriting. If C is a clause and Γ a set of numbering substitutions with $\operatorname{dom}\tau \subseteq \operatorname{var}(C)$ for every $\tau \in \Gamma$, then we say that Γ *covers* C if every ρ that ground numbers C can be obtained as specialization of some $\tau \in \Gamma$.

Instance rewriting

$$\mathcal{R}\,\frac{C}{\{D_\rho \colon \rho \in \Gamma\}}\,E \qquad \text{if} \begin{array}{l} \cdot\ E \text{ is a set of unit equations} \\ \cdot\ \Gamma \text{ covers } C \\ \cdot\ \text{for every } \rho \in \Gamma\colon\ C\rho \to_E^{\succ} \circ \to_E^* D_\rho \\ \quad \text{and } C\rho \to_E^* D\rho \text{ is } \Omega\text{-admissible} \end{array}$$

Proposition 4.19 Consider a complete path $N_1, N_2, \ldots$ in a fair derivation from $M \cup \mathcal{T}'$, where M is finite.

(i) For some index κ, all $N_{\kappa+i}$ contain $\bot$; or they all reduce to digits.

(ii) If $C \in N_{\kappa+i}$ is not $[1;n]$-shallow, then C can effectively be simplified into a finite set of $[1;n]$-shallow clauses.

Proof:

(i) If M is $\mathcal{T}$-unsatisfiable, then $\bot$ is continuously present from some N_κ on, by Lem. 4.7. Otherwise the E_∞-normal form of every term $f(\vec{\imath})$ is a digit, because of Prop. 4.17 (ii) and Prop. 4.18. Since the signature is finite and $\to_{E_\infty}$ is terminating, only a finite portion E of E_∞ is needed for these reductions. For every element $s \simeq t$ of E, there is an index

$k_{s\simeq t}$ such that $s \simeq t \in E_i$ for every $i \geq k_{s\simeq t}$. By finiteness of E, the maximum κ of all $k_{s\simeq t}$ is finite.

(ii) Let $E = E_{\kappa+i}$. If N_κ contains $\bot$, which is $[1;n]$-shallow itself, then any other clause is redundant. Otherwise every ground term can be rewritten, with respect to E, into a digit. Accordingly, for every ground clause D there exists a ground clause $\Delta(D)$ containing only digits such that $D \to_E^* \Delta(D)$. Let now Γ denote the finite set of all substitutions ρ that ground number C and satisfy $\operatorname{dom}\rho \subseteq \operatorname{var}(C)$. The clause C is not $[1;n]$-shallow and therefore of the shape (α) $C \equiv f(\vec{s}) \not\simeq t \vee C'$ or (β) $C \equiv l[f(\vec{s})]_{j.p} \bowtie r \vee C'$. For an arbitrary $\rho \in \Gamma$, let $f(\vec{s})\rho \to_E s_\rho$, say with $u \simeq v \in E$ under the substitution σ such that $u\sigma \succ v\sigma$. In case (α) we have $C\rho \succeq (f(\vec{s}) \not\simeq t)\rho \succ (f(\vec{s}) \simeq f(\vec{s}))\rho \succeq (u \simeq u)\sigma \succ (u \simeq v)\sigma$, and in case ($\beta$) similarly $C\rho \succeq (l[f(\vec{s})]_{j.p} \bowtie r)\rho \succ (f(\vec{s}) \simeq f(\vec{s}))\rho$. In both cases, we obtain a rewrite chain $C\rho[f(\vec{s})] \to_E^\succ C\rho[s_\rho] \to_E^* \Delta(C\rho[s_\rho]) \equiv D_\rho$. Putting everything together, we obtain an instance rewriting step $\mathcal{R}\ \dfrac{C}{\{D_\rho \colon \rho \in \Gamma\}}\ E$ as required.

□

One may want to test explicitly whether a given N_k reduces to digits already (and if so, perhaps test immediately whether E_k describes a $\mathcal{T}$-model of M). Notably the property is not always inherited from N_k to N_{k+1}. Consider for example the following simplification steps in the sense of the calculus $\mathcal{C}$: $\mathcal{R}\ \dfrac{f(3) \simeq f(1)}{1 \simeq f(1)}\ 1 \not\simeq 1 \vee f(3) \simeq 1$, or $\mathcal{R}\ \dfrac{f(3) \simeq 1 \quad f(1) \simeq 3}{f(2) \simeq 1 \quad f(1) \simeq 2}$. The term $f(3)$ is E_k-reducible, but not necessarily E_{k+1}-reducible. As the second example shows, this may even occur if unit equations are simplified with respect to E_k only. In case this is not desired, one has to restrict the simplification of unit equations. For example, ordered unit rewriting, instance rewriting, subsumption and tautology elimination are compatible.

Now we come to the second ingredient of our argumentation towards termination: Inference and split conclusions have no more variables than one of the premises.

Proposition 4.20 The number of variables in clauses develops as follows:

(i) Let $\sigma = \operatorname{mgu}(u,v)$ with $\operatorname{ran}\sigma \subseteq \mathscr{V} \cup [1;n]$ and $\operatorname{dom}\sigma \cup \operatorname{cdom}\sigma \subseteq \operatorname{var}(u,v)$. Then σ exists such that $\operatorname{var}(v\sigma) \subseteq \operatorname{var}(v)$.

(ii) If $C \vdash D$ is a unary inference or a split, then $\operatorname{var}(D) \subseteq \operatorname{var}(C)$ holds.

(iii) If $l \simeq r$, $C \vdash D$ is a binary inference, then $|\operatorname{var}(D)| \leq |\operatorname{var}(C)|$ is true.

Proof:

(i) Let $\mathcal{P}(m)$ hold iff there exists an mgu σ of u and v with $\operatorname{ran}\sigma \subseteq \mathscr{V} \cup [1;n]$, $\operatorname{dom}\sigma \cup \operatorname{cdom}\sigma \subseteq \operatorname{var}(u,v)$, and $|\operatorname{var}(v\sigma) \setminus \operatorname{var}(v)| = m$. By

assumption $\mathcal{P}$ holds for some $m \geq 0$. We will now show that $\mathcal{P}(j+1)$ implies $\mathcal{P}(j)$.

Assume σ is a witness for $\mathcal{P}(j+1)$. Because of $j+1 > 0$ there exists a variable y in $\mathrm{var}(v\sigma) \setminus \mathrm{var}(v)$. By the shape of σ, this variable is the σ-image of another variable $x \in var(v)$. Consider now the substitutions $\tau = \{x \mapsto y, y \mapsto x\}$ and $\sigma' = \sigma \circ \tau$. The latter is a unifier of u and v. Because of $\sigma'\tau = \sigma\tau^2 = \sigma$, it is even a most general one. The image of a variable z under σ' is x if $z\sigma \equiv y$, and $z\sigma$ otherwise; in particular $x\sigma' \equiv x$ and $y\sigma' \equiv x$. That is, going from σ to σ', the variable x moves from the dom-part to the cdom-part, and y in the opposite direction, which are all effects in terms of dom and cdom. The identity $\mathrm{var}(v\sigma') = (\mathrm{var}(v\sigma) \cup \{x\}) \setminus \{y\}$ concludes the proof of $\mathcal{P}(j)$.

(ii) In case of an equality resolution step $C \vee t \simeq t' \vdash C\sigma$ we have $\mathrm{cdom}\,\sigma \subseteq \mathrm{var}(t, t')$. Given a split $C \vee s \simeq t \vee l \simeq r \vee D \vdash (C \vee s \simeq t)\tau \mid (l \simeq r \vee D)\tau$, the substitution τ is numbering, such that $\mathrm{cdom}\,\tau \subseteq [1; n]$.

(iii) We will prove that $\mathrm{var}(D) \subseteq \mathrm{var}(C)$ holds in case the most general unifier is chosen according to Prop. 4.20 (i). All mgu's are equal up to variable renaming; and the number of variables in a clause is invariant under such renamings. This yields the estimate stated above.

We jointly treat superposition left and right inferences via the pattern $l \simeq r,\ C[l'] \vdash C[r]\sigma\tau \equiv D$. Because of $l'\sigma\tau \equiv l\sigma\tau \succ r\sigma\tau$ we know that $\mathrm{var}(l'\sigma\tau) \supseteq \mathrm{var}(r\sigma\tau)$ is true, and hence $\mathrm{var}(D) \subseteq \mathrm{var}(C\sigma\tau) \subseteq \mathrm{var}(C\sigma)$. Applying Prop. 4.20 (i), without loss of generality σ can be chosen such that $\mathrm{var}(l'\sigma) \subseteq \mathrm{var}(l')$. Let σ' denote the restriction of σ to $\mathrm{var}(l')$. By this definition we have $\mathrm{cdom}\,\sigma' \subseteq \mathrm{var}(l'\sigma) \subseteq \mathrm{var}(l') \subseteq \mathrm{var}(C)$. Since the premises are variable disjoint, we obtain $\mathrm{var}(C\sigma) = \mathrm{var}(C\sigma') \subseteq \mathrm{var}(C) \cup \mathrm{cdom}\,\sigma' = \mathrm{var}(C)$, which completes the proof of $\mathrm{var}(D) \subseteq \mathrm{var}(C)$.

□

Simplifying a $[1; n]$-shallow clause with respect to other such clauses can arbitrarily increase the number of variables and need not preserve $[1; n]$-shallowness, as witnessed by

$$\mathcal{R}\,\frac{f(x) \simeq 2}{g(1) \not\simeq g(1) \vee y_1 \not\simeq y_1 \vee \ldots \vee y_m \not\simeq y_m \vee f(x) \simeq 1}\,2 \simeq 1$$

if $f(1) \succ g(1)$. Clearly this counteracts our efforts towards termination; so a strategy is needed that guides the execution of calculus steps. We say that a $\mathcal{C}$-derivation is a $\mathcal{C}_\kappa$-*derivation* from a clause set M if (i) it is fair, (ii) the root node is $M \cup \mathcal{T}'$, and in every path eventually (iii) simplifications do not increase the number of variables, (iv) $[1; n]$-shallowness is preserved under

simplifications, (v) inferences and splits are not repeated, (vi) every fresh inference conclusion which is not $[1;n]$-shallow, is immediately simplified into a set of $[1;n]$-shallow clauses, (vii) no duplicate literals occur in $[1;n]$-shallow clauses, and (viii) $[1;n]$-shallow clauses equal up to variable renaming are identified. Indeed such derivations exist for every finite M: The crucial item (vi) can be satisfied because of Prop. 4.19. Condition (v) does not conflict with (i) because repeated inferences and splits are redundant.

Theorem 4.21 $\mathcal{C}_\kappa$-derivations decide $\mathcal{T}$-satisfiability of finite clause sets.

Proof: Consider a κ-derivation from a finite clause set M. Then M by Lem. 4.7 is $\mathcal{T}$-satisfiable if and only if the derivation contains a complete path without the empty clause in the limit. The derivation tree is finitely branching. It remains to show that every path $N_1, N_2, \ldots$ is finite. Let $\|N_i\| = \max\{|\mathrm{var}(C)| : C \in N_i\}$.

There exists an index κ such that from N_κ on, the conditions (iii) through (vi) of the definition of $\mathcal{C}_\kappa$-derivation are satisfied. We form a subsequence of $N_1, N_2, \ldots$ that starts from $N'_1 = N_\kappa$. If in N_i a new clause C is inferred and, according to condition (vi), immediately simplified into $[1;n]$-shallow clauses $\vec{D}$ until N_{i+k}, then for $(N'_j)_j$ all sequence elements but N_{i+k} are dropped, and the latter shows up only if $\vec{D}$ is not empty. Assume now $(N_i)_i$ is infinite. Inferences with empty $\vec{D}$ are not repeated because of condition (v), as well as splits; so there must be infinitely many simplifications or inferences with non-empty $\vec{D}$. Since simplifications are decreasing with respect to Ω-instances, the latter occur infinitely many times; so $(N'_j)_j$ is then infinite as well.

Inductively $\|N'_j\| \leq \|N'_1\|$ holds for all j: If a clause $C \in N'_j$ is simplified to some non-empty $\vec{D}$, then we know that $|\mathrm{var}(D_i)| \leq |\mathrm{var}(C)|$ by condition (iii). In case of a split or an inference, we additionally apply Prop. 4.20 (ii) and Prop. 4.20 (iii).

Assume now $(N'_j)_j$ were infinite; then we can argue like above for $(N_i)_i$ and obtain that infinitely many inferences are drawn. The inference conclusions are simplified according to condition (vi), such that they become $[1;n]$-shallow and have no more than $\|N'_1\|$ variables. Because of conditions (vii) and (viii), only finitely many such clauses exist. Moreover the number of clauses that are produced from simplification and splitting alone is finite. Therefore, eventually an inference has to be repeated, but this contradicts condition (v). Hence $(N'_j)_j$ is finite, and so is $(N_i)_i$. □

4.6 Extensions

Let us have a short look at a many-sorted setting where $\mathcal{T}$ consists of size restrictions for every sort, each built over an individual set of digits. One has to employ the usual typing constraints for equations, terms and substitutions. Then the calculus $\mathcal{C}$, and the results obtained for it so far, straightforwardly extends to this situation.

Up to now, our calculus did not deal with predicates. Of course one could extend $\mathcal{C}$ with an ordered resolution rule, and consider predicate atoms in the superposition and split rules. Alternatively, we can introduce a two-element sort Bool, say over the digits I and II, and provide a clause $\mathrm{I} \not\simeq \mathrm{II}$. As usually we can now encode predicate atoms $P(\vec{t})$ of any other sort as equations $P(\vec{t}) \simeq \mathrm{I}$. Notably $\mathcal{T}'$ need not contain an axiom $P(\vec{x}) \simeq \mathrm{I} \vee P(\vec{x}) \simeq \mathrm{II}$: Given an algebra $\mathcal{A}$ such that at some point $P^{\mathcal{A}}$ does not map into $\{\mathrm{I}^{\mathcal{A}}, \mathrm{II}^{\mathcal{A}}\}$, let the algebra $\mathcal{B}$ coincide with $\mathcal{A}$ except that $P^{\mathcal{B}}$ maps all such points onto $\mathrm{II}^{\mathcal{B}}$. Then $\mathcal{A}$ and $\mathcal{B}$ satisfy the same encoded atoms $P(\vec{t}) \simeq \mathrm{I}$.

As an application, consider the validity problem for a formula $\phi \equiv \forall x_1 \ldots \forall x_n \, \exists y_1 \ldots \exists y_m \phi'$ where ϕ' is quantifier-free and contains no function symbols. This problem was proven decidable by Bernays and Schönfinkel [BS28]. Now, ϕ is valid iff $\psi \equiv \forall y_1 \ldots \forall y_m \neg \phi' \{x_1 \mapsto 1, \ldots, x_n \mapsto n\}$ is unsatisfiable iff ψ is $\mathcal{T}$-unsatisfiable. Since no function symbols are present, the set $\mathcal{T}'$ is empty. That is, derivations start from unaugmented clause sets.

Corollary 4.22 $\mathcal{C}_\kappa$-derivations decide the Bernays-Schönfinkel class.

5 Combinations with First-Order Theories

So far, we only considered the case where actually the overall Herbrand domain of a formula is finite. The interesting question is whether the techniques developed in the previous sections can be generalized to a setting where the overall Herbrand domain may be infinite, but finite subsorts exists and for these subsorts we can exploit the advanced technology. The answer we give in the section is "yes" the combination is possible in exactly this way, i.e., for the finite sorts we only need to consider finite instances and there are no inferences with the finite sort axiom needed.

The idea is to code the finite subsets via sorts. Our modeling of sorts are general monadic predicates, so in expressiveness it goes far beyond the standard many or order-sorted setting [Wei01]. The theory $\mathcal{T}$ under consideration for a sort S of cardinality n is

$$S(1) \wedge \ldots \wedge S(n)$$
$$\forall x.\, \neg S(x) \vee x \simeq 1 \vee \ldots \vee x \simeq n$$

where an atom $S(t)$ is as usual an abbreviation for $S(t) \simeq \mathrm{I}$ (see also Sect. 4.6). We write $C \vee_S S(t)$ to express that S does not occur in a positive literal in C. For a clause containing a negative literal $\neg S(x)$ we say that x *is of sort* S. As in the restricted case of Sect. 4, instances of negative literals $\neg S(x)$ only need to be considered with respect to $[1; n]$. However, if the clause set N under consideration in addition to $S(1) \wedge \ldots \wedge S(n)$ contains clauses with positive literals $C \vee S(t)$ we need to consider inferences with the clause $C \vee t \simeq 1 \vee \ldots \vee t \simeq n$. This is expressed by the following proposition.

Lemma 5.1 Let N be a clause set containing the monadic predicate S. Let $M = \{C \vee S(t_1) \vee \ldots \vee S(t_m) \mid C \vee_S S(t_1) \vee \ldots \vee S(t_m) \in N\}$, $M' = \{C \vee t_1 \simeq 1 \vee \ldots \vee t_1 \simeq n \vee \ldots \vee t_m \simeq 1 \vee \ldots \vee t_m \simeq n \mid C \vee_S S(t_1) \vee \ldots \vee S(t_m) \in M\}$ and $\mathcal{T}' = \{S(1), \ldots, S(n)\}$. Furthermore, let Ω_S be the restriction of Ω to variables x of sort S and $N' = ((N \setminus M) \cup M')$. Then the clause set $N \cup \mathcal{T}$ is satisfiable iff $\Omega_S(N') \cup \mathcal{T}'$ is satisfiable.

Proof: "$\Rightarrow$" Let $\mathcal{A}$ be a model for $N \cup \mathcal{T}$, i.e., $\mathcal{A} \models N \cup \mathcal{T}$ and so $\mathcal{A} \models \mathcal{T}'$. For each clause $C \in N$ we have $\mathcal{A} \models C$. We need to show $\mathcal{A} \models C'$ for all $C' \in \Omega_S(N')$. By construction $S^{\mathcal{A}} = \{1^{\mathcal{A}}, \ldots, n^{\mathcal{A}}\}$. We distinguish the following cases: (i) $C \in (N \setminus M)$ and C does not contain a literal $\neg S(x)$. Then, as $C = C'$, $C \in \Omega_S(N')$ and we are done. (ii) let $C = \neg S(x_1) \vee \ldots \vee \neg S(x_n) \vee D$ and $C \in (N \setminus M)$ where D does not contain a literal $\neg S(x)$. Assume $\mathcal{A} \not\models C'$ for some $C' \in \Omega_S(C)$ where $C' = C\sigma$ and $x_i\sigma \in \{1, \ldots, n\}$. Then $\mathcal{A}[x_1/(x_1\sigma)^{\mathcal{A}}, \ldots, x_n/(x_n\sigma)^{\mathcal{A}}] \not\models C$ which is a contradiction. (iii) $C \in M$, $C = \neg S(x_1) \vee \ldots \vee \neg S(x_n) \vee S(t_1), \ldots, S(t_m) \vee D$ and D does not contain a positive occurrence of S nor a literal $\neg S(x)$. Obviously $\mathcal{A}[x/(x\sigma)^{\mathcal{A}}] \models \neg S(x)$ iff $\mathcal{A} \models \neg S(x)\sigma$. For $C' = (\neg S(x_1) \vee \ldots \vee \neg S(x_n) \vee t_1 \simeq 1 \vee \ldots \vee t_1 \simeq n \vee \ldots \vee t_m \simeq 1 \vee \ldots \vee t_m \simeq n \vee D)\sigma$ with $C' \in \Omega_S(C)$ we have $\mathcal{A}[x_1/(x_1\sigma)^{\mathcal{A}}, \ldots, x_n/(x_n\sigma)^{\mathcal{A}}] \models S(t_i)$ iff $\mathcal{A} \models (t_i \simeq 1 \vee \ldots \vee t_i \simeq n)\sigma$ because $S^{\mathcal{A}} = \{1^{\mathcal{A}}, \ldots, n^{\mathcal{A}}\}$ for all i. Hence $\mathcal{A} \models C'$.

"$\Leftarrow$" Let $\mathcal{A} \models \Omega_S(N')$ and let $\mathcal{A}'$ be identical to $\mathcal{A}$, except that $S^{\mathcal{A}'} = \{1^{\mathcal{A}}, \ldots, n^{\mathcal{A}}\}$. As $\mathcal{T}'$ contains the only positive occurrences of S in $\Omega_S(N') \cup \mathcal{T}'$ and by construction $\mathcal{A}' \models \mathcal{T}'$, we also have $\mathcal{A}' \models \Omega_S(N')$. We need to show that $\mathcal{A}' \models N$ and $\mathcal{A}' \models \mathcal{T}$. Again it holds that $\mathcal{A}'[x_1/(x_1\sigma)^{\mathcal{A}'}, \ldots, x_n/(x_n\sigma)^{\mathcal{A}'}] \models S(t_i)$ iff $\mathcal{A}' \models (t_i \simeq 1 \vee \ldots \vee t_i \simeq n)\sigma$ and $\mathcal{A}'[x/(x\sigma)^{\mathcal{A}'}] \models \neg S(x)$ iff $\mathcal{A}' \models \neg S(x)\sigma$ proving $\mathcal{A}' \models N$. By construction $S^{\mathcal{A}'} = \{1^{\mathcal{A}'}, \ldots, n^{\mathcal{A}'}\}$ and hence $\mathcal{A}' \models \mathcal{T}$. $\square$

Note that the four clauses $S(1)$, $S(2)$, $\neg S(x) \vee x \simeq 1 \vee x \simeq 2$, $f^3(x) \not\simeq f(x)$ are satisfiable as not the input, nor the output of f is of sort S. Adding the clause $\neg S(x) \vee S(f(x))$ declares f to be a function from and into S and hence causes unsatisfiability of the five clauses. For the latter clause, the transformation of Lem. 5.1 applies.

Now by the lifting theorem for standard superposition, we know by Lem. 5.1 that $N \cup \mathcal{T}$ has a superposition refutation iff $N' \cup \mathcal{T}'$ has one. The open question is how we can exploit the fact that we considered solely numbering substitutions for variables of sort S. Note that although S has a finite domain, the overall domain of N may be infinite. Therefore, we cannot take the approach of Sect. 4 where we used the numbering substitution available for all variables to require that inferences are only performed on strictly greatest terms and literals. Furthermore, the abstract superposition redundancy notion is no longer effective and satisfiability is of course not decidable anymore. Therefore, the idea is to restrict the range of substitutions for variables of sort S to $\mathcal{V} \cup [1; n]$, and to require that (strict) maximality is preserved under any numbering substitution for the finite sort S. If the digits are minimal in the ordering $\succ$, then this yields a sound and complete refinement of the standard superposition calculus for finite sorts.

We exemplify the refinement for the superposition right inference rule. The other inference rules are defined accordingly.

Superposition right

$$\mathcal{I}\ \frac{C \vee l \simeq r \quad s[l'] \simeq t \vee D}{(C \vee s[r] \simeq t \vee D)\sigma} \quad \text{if} \begin{array}{l} \cdot\ l' \notin \mathcal{V} \text{ and } \sigma = \mathrm{mgu}(l, l') \\ \cdot\ \mathrm{ran}\,\sigma|_S \subseteq \mathcal{V} \cup [1; n] \\ \cdot\ \text{there exists a minimally numbering } \tau \\ \quad \text{of sort } S \text{ such that } l,\ l \simeq r,\ s \text{ and} \\ \quad s \simeq t \text{ are strictly maximal under } \sigma\tau \\ \cdot\ (C \vee l \simeq r)\sigma\tau \not\succeq (s \simeq t \vee D)\sigma\tau \end{array}$$

Theorem 5.2 Consider the notions of Lem. 5.1 and let the digits $1, \ldots, n$ be minimal in the ordering $\succ$. Then the clause set $N \cup \mathcal{T}$ is unsatisfiable iff there is a derivation of the empty clause from $N' \cup \mathcal{T}'$ by the superposition calculus defined above.

Proof: By Lem. 5.1 $N \cup \mathcal{T}$ is unsatisfiable iff $\Omega_S(N') \cup \mathcal{T}'$ is unsatisfiable. The set $\Omega_S(N') \cup \mathcal{T}'$ is unsatisfiable iff there is a derivation of the empty clause by the standard superposition calculus. As the digits are minimal in the ordering, they might only be replaced by each other. For any clause $\neg S(x) \vee C \in N'$, all instances of $\neg S(x)$ in the proof are generated by substitutions

from x into $[1; n]$. Hence, all steps can be lifted to steps of the above refined superposition calculus on N'. □

Here is an example for the refined maximality condition. Let $\succ$ denote the lexicographic path ordering to the precedence $f \succ g \succ n \succ \ldots \succ 1$. Then in the clause $\neg S(x) \vee g(x, y) \simeq y \vee f(y) \simeq y$, the literal $g(x, y) \simeq y$ is not maximal, because $g(x, y)\tau \prec f(y)\tau$ for any ground numbering substitution τ.

We can even stay with the clause set N. The additional clause $\neg S(x) \vee x \simeq 1 \vee \ldots \vee x \simeq n$ does not harm, i.e., the proof of Lem. 5.1 goes also through with this extra clause. Once the clause is added by selecting $\neg S(x)$ in this clause we produce together with a clause $C \vee S(t)$ the clause $C \vee t \simeq 1 \vee \ldots \vee t \simeq n$, i.e., it eventually generates the clauses from N'. All inferences between $\neg S(x) \vee x \simeq 1 \vee \ldots \vee x \simeq n$ and the facts $S(1), \ldots, S(n)$ are redundant. So we could also remove the clauses M' and stay with M, meaning that superposition with the above restrictions is also complete for $N \cup \mathcal{T}$.

6 Conclusion and Future Work

We have presented a light-weight adaptation of superposition calculi to the first-order theory of finite domains. The achievement is a superposition calculus for finite domains that

(a) restricts the range of inference unifiers to digits or variables,
(b) enables the precise calculation of ordering restrictions,
(c) introduces an effective general semantic redundancy criterion,
(d) incorporates a particular splitting rule for non-Horn clauses,
(e) constitutes a decision procedure for any finite domain problem,
(f) is mostly compatible with the all standard superposition redundancy criteria,

and can in particular

(g) be embedded via a general sort discipline based on monadic predicates in any general first-order setting.

We have already done some promising experiments on the basis of the superposition calculus for finite domains [HTW06] and a full-fledged integration into SPASS is on the way. To this end, ordering computation, inference computation, redundancy notions will be refined for the case of variables with a finite domain.

References

[BG91] L. Bachmair and H. Ganzinger. Perfect model semantics for logic programs with equality. In *Proceedings of the 8th International Conference on Logic Programming*, pages 645–659. MIT Press, 1991.

[BG94] L. Bachmair and H. Ganzinger. Rewrite-based equational theorem proving with selection and simplification. *Journal of Logic and Computation*, 4(3):217–247, 1994.

[BS28] P. Bernays and M. Schönfinkel. Zum Entscheidungsproblem der mathematischen Logik. *Mathematische Annalen*, 99:342–372, 1928.

[BS06] P. Baumgartner and R. Schmidt. Blocking and other enhancements for bottom-up model generation methods. In U. Furbach and N. Shankar, editors, *Proceedings of the 3rd International Joint Conference on Automated Reasoning*, volume 4130 of *LNAI*, pages 125–139. Springer-Verlag, 2006.

[CS03] K. Claessen and N. Sörensson. New techniques that improve MACE-style finite model finding. In P. Baumgartner and Chr. Fermueller, editors, *Proceedings of the Workshop on Model Computation*, 2003.

[Der91] N. Dershowitz. A maximal-literal unit strategy for Horn clauses. In S. Kaplan and M. Okada, editors, *Proceedings of the 2nd International Workshop on Conditional and Typed Rewriting Systems*, volume 516 of *LNCS*, pages 14–25. Springer-Verlag, 1991.

[HTW06] Th. Hillenbrand, D. Topic, and Chr. Weidenbach. Sudokus as logical puzzles. In W. Ahrendt, P. Baumgartner, and H. de Nivelle, editors, *Proceedings of the Third Workshop on Disproving*, pages 2–12, 2006.

[KL80] S. Kamin and J.-J. Levy. Attempts for generalizing the recursive path orderings. Available electronically from `http://perso.ens-lyon.fr/pierre.lescanne/not_accessible.html`. University of Illinois, Department of Computer Science. Unpublished note, 1980.

[MB88] R. Manthey and F. Bry. SATCHMO: a theorem prover implemented in Prolog. In E. Lusk and R. Overbeek, editors, *Proceedings of the 9th International Conference on Automated Deduction*, volume 310 of *LNCS*, pages 415–434. Springer-Verlag, 1988.

[McC03] W. McCune. MACE4 reference manual and guide. Technical Report ANL/MCS-TM-264, Argonne National Laboratory, 2003.

[NR01] R. Nieuwenhuis and A. Rubio. Paramodulation-based theorem proving. In A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*, volume I, chapter 7, pages 371–443. Elsevier, 2001.

[Wei01] Chr. Weidenbach. Combining superposition, sorts and splitting. In A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*, volume II, chapter 27, pages 1965–2012. Elsevier, 2001.