A Maximal-Literal Unit Strategy for Horn Clauses*

Nachum Dershowitz
Department of Computer Science
University of Illinois at Urbana-Champaign
1304 West Springfield Avenue
Urbana, IL 61801, U.S.A.
nachum@cs.uiuc.edu

Abstract

A new positive-unit theorem-proving procedure for equational Horn clauses is presented. It uses a term ordering to restrict paramodulation to potentially maximal sides of equations. Completeness is shown using proof orderings.

1. Introduction

A *conditional equation* is a universally-quantified Horn clause in which the only predicate symbol is equality. We write such a clause in the form

$$e_1 \wedge \cdots \wedge e_n \Rightarrow s \simeq t$$

($n \geq 0$), meaning that the equality $s \simeq t$ holds whenever all the equations e_i, called *conditions*, hold. If $n = 0$, then the (positive unit) clause, $s \simeq t$, will be called an *unconditional equation.* Conditional equations are important for specifying abstract data types and expressing logic programs with equations. Our interest here is in procedures for proving validity of equations in all models of a given finite set E of conditional equations. Note that a conditional equation $e_1 \wedge \cdots \wedge e_n \Rightarrow s \simeq t$ is valid for E iff $s \simeq t$ is valid for $E \cup \{e_1, \ldots, e_n\}$. Hence, proving validity of conditional equations reduces to proving validity of unconditional ones.

The completeness of positive-unit resolution for Horn clauses is well-known. An advantage of positive-unit resolution is that the number of conditions never grows; it suffers from the disadvantage of being a bottom-up method. Ordinary Horn clauses

$$p_1 \wedge \cdots \wedge p_n \Rightarrow p_{n+1}$$

where the p_i are not equality literals, can be expressed as conditional equations, by turning each literal p_i into a Boolean equation $p_i \simeq T$, for the truth constant T. Ordered resolution, in which the literals of each clause are arranged in a linear order $>$, and only the largest literal may serve as a resolvent, is also complete for Horn clauses (see Boyer, 1971).

Positive-unit resolution can be expressed by means of the following inference rule:

$$\frac{E \cup \left\{ \begin{array}{rcl} q \wedge s \simeq T & \Rightarrow & u \simeq T, \\ & & l \simeq T \end{array} \right\}}{E \cup \left\{ \begin{array}{rcl} q \wedge s \simeq T & \Rightarrow & u \simeq T, \\ & & l \simeq T, \\ q\sigma & \Rightarrow & u\sigma \simeq T \end{array} \right\}}$$

*This research supported in part by the National Science Foundation under Grant CCR-9007195.

where σ is the most general unifier (*mgu*) of l and s. Here, the positive unit clause $l \simeq T$ is resolved with the negative literal $s \simeq T$ in the clause $q \wedge s \simeq T \Rightarrow u \simeq T$, and produces a new Horn clause $q\sigma \Rightarrow u\sigma \simeq T$. The new clause is a logical consequence of the two given clauses, since $s\sigma = l\tau\sigma$, where τ renames variables in l so that it shares none with s. Any unit clause that is a logical consequence of a set of Horn clauses E is an instance of a unit clause producible by repeated application of this rule of inference.

Horn clauses with both equality and non-equality literals can be expressed as conditional equations with equality literals only. The equality axioms, including functional reflexivity, are also Horn clauses. Positive-unit resolution, or any other complete variation of resolution, could be used to prove theorems in equational Horn theories, but the cost of treating equality axioms like any other clause is prohibitively high. For this reason, special inference mechanisms for equality, notably paramodulation (Robinson and Wos, 1969), have been devised. In the Horn case, a unit strategy can be combined with paramodulation (Henschen and Wos, 1974; Furbach, 1987).

In this paper, we describe a new complete theorem-proving method for equational Horn theories. It utilizes orderings of terms and atoms to restrict inferences, and is a generalization of *ordered completion* (Hsiang and Rusinowitch, 1987; Bachmair, *et al.*, 1989), an "unfailing" extension of the "completion procedure" in Knuth and Bendix (1970) for unconditional equational inference. Completion operates on asymmetrical equations, that is, on *rewrite rules*, and has as its goal the production of confluent (Church-Rosser) systems of rules that can be used to decide validity. For a survey of rewriting, see Dershowitz and Jouannaud (1990).

Brown (1975) and Lankford (1975) first suggested combining completion for oriented equations, with paramodulation for unorientable ones and resolution for non-equality atoms. Paul (1986) studied the application of completion to sets of Horn clauses with equality. Completion was extended to conditional equations by Kaplan (1987) and Ganzinger (1987). Unit strategies, such the one given here, do not seem to be appropriate for completion. Recently, several restrictions of paramodulation based on term orderings have been proposed for the full first-order case, including Zhang and Kapur (1988) and Rusinowitch (1989). Kounalis and Rusinowitch (1987) and Bachmair, *et al.* (1989) improved upon the earlier Horn-clause methods in various ways.

Our method severely restricts resolution with paramodulation by incorporating an ordering on (atoms and) terms. Inferences are limited in the following ways:

- The functional reflexive axioms are not needed and, at the same time, paramodulation into variables is avoided (as for some versions of paramodulation).
- For all (resolution and paramodulation) inferences, at least one of the equations must be unconditional (as in positive unit resolution and positive unit paramodulation).
- Unless an equation is unconditional only its conditional part is used for paramodulation (analogous to positive-unit resolution).
- Only maximal terms (with respect to a given ordering) are used (analogous to ordered resolution).

Unlike Kounalis and Rusinowitch (1987), we use only unit clauses when paramodulating into conditions; unlike Bachmair, *et al.* (1989), all our inference rules use only the maximal side of an equation. Thus, our method is the first to combine a unit strategy with one based on maximal terms. It also allows for (virtually unrestricted) simplification (demodulation) by unconditional equations. Since some of the rules we consider delete or simplify antecedent clauses, the above format for inference rules, with the equations that participated in the inference also appearing as part of the consequent, is advantageous.

Limiting inference partially controls growth; keeping clauses fully simplified stunts growth even further. Such restrictions are of paramount importance in any practical theorem prover, but their (refutational) completeness has been difficult to establish. For our completeness proof, we adapt the proof-ordering method of Bachmair, *et al.* (1986) to conditional proofs. Proof orderings allow us to limit narrowing to negative literals, something that appears impossible with the recent transfinite-tree proof method of Hsiang and Rusinowitch (1986). The crux of our method is the observation that any conditional equational proof not in "normal form" must either have an unconditional "peak", that is, two applications of unconditional equations such that the middle term is the largest of the three involved, or an unconditional "drop", that is, an application of an unconditional equation (or reflexivity of equals) to an instance of a condition. The proof procedure is designed to eliminate peaks and drops, thereby reducing the complexity assigned to the proof. The refutational completeness of this strategy, but not the more general proof normalization result, follows from concurrent work of Bachmair and Ganzinger (1990) on first-order proofs. A unit Horn-clause strategy with simplification is also proved complete in Bachmair (1991).

2. Orderings

Let $\mathcal{T}$ be a set of (first-order) terms, with variables taken from a set $\mathcal{X}$, and $\mathcal{G}$ be its subset of *ground* (variable-free) terms. If t is a term in $\mathcal{T}$, by $t|_\pi$ signifies the subterm of t rooted at position π; then by $t[s]_\pi$, for some term s, we denote the term t with its subterm $t|_\pi$ replaced by s.

Term orderings are of central importance in the proposed method. A total ordering $>$ on ground terms $\mathcal{G}$ is called a *complete simplification ordering* if it has (a) the "replacement property", $s > t$ implies that any term $u[s]_\pi$, with subterm s located at some position π, is greater under $>$ than the term $u[t]_\pi$ with that occurrence of s replaced by t, and (b) the "subterm property", $t \geq t|_\pi$ for all subterms $t|_\pi$ of t. Such a ground-term ordering must be a well-ordering (Dershowitz, 1982). A *completable simplification ordering* on all terms $\mathcal{T}$ is a well-founded *partial* ordering $\succ$ (c) that can be extended to a complete simplification ordering $>$ on ground terms, such that (d) $s \succ t$ implies that $s\sigma > t\sigma$ for all ground substitutions σ. Furthermore, we will assume (e) that the constant T is minimal in $\succ$.

With a total ordering of atoms and with no equations, *per se*, the method of the next section is just selected positive-unit resolution, in which the largest negative literal is chosen. The appropriate inference rule would be expressed as:

$$\frac{E \cup \left\{ \begin{array}{rcl} q \wedge s \simeq T & \Rightarrow & u \simeq T, \\ & & l \simeq T \end{array} \right\}}{E \cup \left\{ \begin{array}{rcl} q \wedge s \simeq T & \Rightarrow & u \simeq T, \\ & & l \simeq T, \\ q\sigma & \Rightarrow & u\sigma \simeq T \end{array} \right\}}$$

where $\sigma = mgu(s, l)$, and would only be applied when $s > q$, by which we mean that s is the largest negative literal in its clause. A total simplification ordering on non-ground literals is not actually possible (which is why the ordering of the parent clause is inherited in ordered resolution), but can be approximated by a partial ordering. If only a partial ordering $\succ$ is given, we resolve negative literals that are potentially maximal. That is, we apply the above rule if $s\sigma \not\prec q\sigma$, or, in other words, if the instance $s\sigma$ of s created by resolution is not necessarily smaller than the other instantiated negative literals.

Suppose E is a set of Horn clauses in conditional equation form. To handle equality literals we need to unify at subterms of conditions, not just at the literal level. Suppose $l \simeq r$ is an equation in E. Note that whenever we refer to equations in a set, we mean that it, or the symmetric equation (with l and r exchanged), or a variant with variables renamed uniformly, actually appears in the set. With that in mind, if l unifies with a non-variable subterm $s|_\pi$ of a maximal term s in a condition $s \simeq t$ of a conditional equation $q \wedge s \simeq t \Rightarrow u \simeq v$, then a new Horn clause is created by applying the most general unifying substitution σ to the conditional equation, and then replacing $l\sigma$ with $r\sigma$, as per the unit clause $l \simeq r$. More precisely, we infer the clause $q\sigma \wedge s\sigma[r\sigma]_\pi \simeq t\sigma \Rightarrow u\sigma \simeq v\sigma$, provided that $s\sigma$ is not smaller under $\succ$ than the other side of the condition $t\sigma$, or of either side of the other conditions $q\sigma$, or of the new term $s\sigma[r\sigma]_\pi$. Thus, the conditions ensure that $s\sigma$ is the (potentially) larger side of the condition that is being paramodulated into and that the replacement yields a (potentially) smaller condition.

Of course, the empty ordering is completable. But the strength of the method, both in minimizing possible inferences and maximizing potential simplifications, is brought to bear by employing more complete orderings. In practice, any efficiently computable ordering should be better than uncontrolled paramodulation. The polynomial and path orderings commonly used in rewrite-based theorem provers (see Dershowitz, 1987) are completable. In particular, the lexicographic variant of the recursive path ordering (Kamin and Lévy, 1980) has decidability properties (Comon, 1990) that make it ideal for this purpose. Choosing an ordering that takes the goal (theorem) into account can impart a top-down flavor to the otherwise bottom-up procedure.

3. Unit Strategy

We formulate our theorem-proving procedure as an inference system operating on a set of conditional equations, and parameterized by a completable ordering $\succ$.

The rules may be classified into three "expansion" rules and four "contraction" rules. The first expansion rule applies to unit clauses:

$$\textbf{Superpose:} \quad \frac{E \cup \left\{ \begin{array}{c} u \simeq v, \\ l \simeq r \end{array} \right\}}{E \cup \left\{ \begin{array}{c} u \simeq v, \\ l \simeq r, \\ u\sigma[r\sigma]_\pi \simeq v\sigma \end{array} \right\}} \quad \text{if} \left\{ \begin{array}{l} u|_\pi \notin \mathcal{X} \\ \sigma = mgu(u|_\pi, l) \\ u\sigma \not\prec v\sigma, u\sigma[r\sigma]_\pi \end{array} \right.$$

Superposition (i.e. oriented paramodulation of positive equational literals) is performed only at non-variable positions ($u|_\pi \notin \mathcal{X}$). Either side of an equation may be used for superposition, but only if, in the context of the paramodulation, it is potentially the largest term involved ($u\sigma \not\prec v\sigma, u\sigma[r\sigma]_\pi$). Note that the two equations, $u \simeq v$ and $l \simeq r$, may actually be the same (except for renaming).

The second rule applies a unit equation to a negative literal:

$$\textbf{Narrow:} \quad \frac{E \cup \left\{ \begin{array}{rcl} q \wedge s \simeq t & \Rightarrow & u \simeq v, \\ & & l \simeq r \end{array} \right\}}{E \cup \left\{ \begin{array}{rcl} q \wedge s \simeq t & \Rightarrow & u \simeq v, \\ & & l \simeq r, \\ q\sigma \wedge s\sigma[r\sigma]_\pi \simeq t\sigma & \Rightarrow & u\sigma \simeq v\sigma \end{array} \right\}} \quad \text{if} \left\{ \begin{array}{l} s|_\pi \notin \mathcal{X} \\ \sigma = mgu(s|_\pi, l) \\ s\sigma \not\prec q\sigma, t\sigma, s\sigma[r\sigma]_\pi \end{array} \right.$$

Whenever this or subsequent rules refer to a conditional equation like $q \wedge s \simeq t \Rightarrow u \simeq v$, the intent is that $s \simeq t$ is any one of the conditions and u is either side of the implied equation.

The last expansion rule in effect resolves a maximal negative literal with reflexivity of equals ($x \simeq x$):

$$\textbf{Reflect}: \quad \frac{E\cup\left\{\begin{array}{rcl} q\wedge s\simeq t & \Rightarrow & u\simeq v\end{array}\right\}}{E\cup\left\{\begin{array}{rcl} q\wedge s\simeq t & \Rightarrow & u\simeq v,\\ q\sigma & \Rightarrow & u\sigma\simeq v\sigma\end{array}\right\}} \quad \text{if } \left\{\begin{array}{l}\sigma = mgu(s,t)\\ s\sigma \not\prec q\sigma\end{array}\right.$$

The remaining, contraction rules all simplify the set of conditional equations. The first deletes trivial conditional equations:

$$\textbf{Delete}: \quad \frac{E\cup\left\{\begin{array}{rcl} q & \Rightarrow & u\simeq u\end{array}\right\}}{E}$$

Here and later, when a rule refers to a clause of the form $q \Rightarrow u \simeq v$, an unconditional equation $(u \simeq v)$ is also intended.

The next rule allows for deletion of conditions that are trivially true:

$$\textbf{Condense}: \quad \frac{E\cup\left\{\begin{array}{rcl} q\wedge s\simeq s & \Rightarrow & u\simeq v\end{array}\right\}}{E\cup\left\{\begin{array}{rcl} q & \Rightarrow & u\simeq v\end{array}\right\}}$$

The last two contraction rules use unit clauses to simplify other clauses. One rule simplifies conditions; the other applies to the equation part. In both cases, the original clause is *replaced* by a version that is equivalent but strictly smaller under $\succ$.

$$\textbf{Simplify}: \quad \frac{E\cup\left\{\begin{array}{rcl} q & \Rightarrow & u\simeq v,\\ & & l\simeq r\end{array}\right\}}{E\cup\left\{\begin{array}{rcl} q[r\sigma]_\pi & \Rightarrow & u\simeq v,\\ & & l\simeq r\end{array}\right\}} \quad \text{if } \left\{\begin{array}{l} q|_\pi = l\sigma\\ q\succ q[r\sigma]_\pi\end{array}\right.$$

$$\textbf{Compose}: \quad \frac{E\cup\left\{\begin{array}{rcl} q & \Rightarrow & u\simeq v,\\ & & l\simeq r\end{array}\right\}}{E\cup\left\{\begin{array}{rcl} q & \Rightarrow & u[r\sigma]_\pi\simeq v,\\ & & l\simeq r\end{array}\right\}} \quad \text{if } \left\{\begin{array}{l} u|_\pi = l\sigma\\ u\succ u[r\sigma]_\pi\\ q\neq T\vee v\succ u\vee u\rhd l\end{array}\right.$$

By $q \neq T$ we mean that the equation has at least one condition; by $u \rhd l$ we mean that u is strictly larger than l in the *encompassment* ordering in which a term is larger than its proper subterms and smaller than its proper instances.

Ordered completion, which deals just with unconditional equations, uses the rules, **superpose, delete,** and **compose.**

As a simple example of our unit strategy, consider the following three clauses:

$$\begin{array}{rcl} & & 0 < c(0) \simeq T\\ & & c(y) < c(z) \simeq y < z\\ x<y\simeq T\wedge y<z\simeq T & \Rightarrow & x<z\simeq T\end{array}$$

Using a (left-to-right) lexicographic path ordering, they generate an infinite number of clauses to which contraction rules cannot be applied:

$$\begin{array}{rcll} & & 0 < c^i(0) \simeq T & (i\geq 1) \qquad (1)\\ & & c(y) < c(z) \simeq y < z & \qquad\qquad\; (2)\\ x<c^j(y)\simeq T\wedge y<z\simeq T & \Rightarrow & x<c^j(z)\simeq T & (j\geq 0) \qquad (3)\\ 0<z\simeq T & \Rightarrow & 0<c^k(z)\simeq T & (k\geq 1) \qquad (4)\\ x<c^j(0)\simeq T & \Rightarrow & x<c^{i+j}(0)\simeq T & (j\geq 0) \qquad (5)\\ x<y\simeq T\wedge c^k(y)<z\simeq T & \Rightarrow & c^k(x)<z\simeq T & (k\geq 1) \qquad (6)\end{array}$$

There are no possible reflections, since no condition has unifiable sides. No superposition inferences between clauses of type (1) and (2), or narrowing inferences between (2) and (4), are allowed, because $c(y)$ does not unify with 0.

Unit clauses of type (1) do not unify with the second condition of (6) for the same reason. Narrowing at the first condition of (6) is also not possible. That would produce an instance of (6) in which the chosen condition is not maximal, since, in

$$0 < c^i(0) \simeq T \wedge c^{k+i}(0) < z \simeq T \quad \Rightarrow \quad c^k(0) < z \simeq T$$

the term $c^{k+i}(0)$ is larger than 0.

Unifying $0 < c^i(0)$ from (1) with the condition of (4) and narrowing yields:

$$T \simeq T \quad \Rightarrow \quad 0 < c^{i+k}(0)$$

which condenses to a new equation of type (1):

$$0 < c^{i+k}(0)$$

Similarly, narrowing (5) with (1) generates an equation of type (1).

Unifying $0 < c^i(0)$ from (1) with the first condition $x < c^j(y)$ of (3) can only succeed if $i \geq j$, giving

$$0 < c^i(0) \simeq T \wedge c^{i-j}(0) < z \simeq T \quad \Rightarrow \quad 0 < c^i(z) \simeq T$$

as the relevant instance of (3). For $i > j$, the second condition is larger than the first and the inference is not performed. For $i = j$, we get the new clause

$$T \simeq T \wedge 0 < z \simeq T \quad \Rightarrow \quad 0 < c^i(z) \simeq T$$

which condenses to a clause of type (4).

Unifying $0 < c^i(0)$ with the second condition of (3) gives (after condensation)

$$x < c^i(0) \simeq T \quad \Rightarrow \quad x < c^{i+j}(0) \simeq T$$

which is of type (5).

Unifying (2) with the first condition of (6) gives

$$c(x) < c(y) \simeq T \wedge c^{k+1}(y) < z \simeq T \quad \Rightarrow \quad c^{k+1}(x) < z \simeq T$$

which simplifies to another type (6) clause:

$$x < y \simeq T \wedge c^{k+1}(y) < z \simeq T \quad \Rightarrow \quad c^{k+1}(x) < z \simeq T$$

Unifying (2) with the second condition of (6) gives (after simplification):

$$x < y \simeq T \wedge c^{k-1}(y) < z \simeq T \quad \Rightarrow \quad c^k(x) < c(z) \simeq T$$

which after composition turns out to be an already existing clause:

$$x < y \simeq T \wedge c^{k-1}(y) < z \simeq T \quad \Rightarrow \quad c^{k-1}(x) < z \simeq T$$

Similarly, narrowing (3) with (2) generates bigger clauses of type (3), while narrowing (5) with (2) gives smaller clauses of type (5).

Note that clauses like $c(0) < c(c(c(0))) \simeq T$ are not generated; all the same, the complete set of unit clauses, (1) and (2), will reduce any equation $c^j(0) < c^{i+j}(0) \not\simeq T$ to the contradiction $T \not\simeq T$.

4. Completeness

Let $>$ be any complete simplification ordering extending the given partial ordering $\succ$. We define a symmetric binary relation $\leftrightarrow$, for a particular set of conditional equations E, as the smallest relation satisfying $t[l\sigma]_\pi \leftrightarrow t[r\sigma]_\pi$ for all $u_1 \simeq v_1 \wedge \cdots \wedge u_n \simeq v_n \Rightarrow l \simeq r$ in E such that $u_i\sigma \leftrightarrow^* v_i\sigma$ for each i, where $\leftrightarrow^*$ is the reflexive-transitive closure of $\leftrightarrow$. This relation corresponds to "substitution of equals" according to the axioms in E. We also define a *rewrite* relation on ground terms $\mathcal{G}$ as the intersection of $>$ and $\leftrightarrow$. That is, $t[l\sigma]_\pi \rightarrow t[r\sigma]_\pi$ if $u_1 \simeq v_1 \wedge \cdots \wedge u_n \simeq v_n \Rightarrow l \simeq r$ is in E, $t[l\sigma]_\pi > t[r\sigma]_\pi$, and $u_i\sigma \leftrightarrow^* v_i\sigma$ for each i. We use $\leftarrow$ for the inverse of $\rightarrow$, and $\rightarrow^*$ and $\leftarrow^*$ for the reflexive-transitive closures of $\rightarrow$ and $\leftarrow$, respectively.

A *proof* of an equation $s \simeq t$ between *ground* terms (any variables in s and t may be treated as Skolem constants) is a "derivation"

$$\begin{array}{ccccccc} s = t_1 & \overset{\pi_1}{\underset{e_1\sigma_1}{\longleftrightarrow}} & t_2 & \overset{\pi_2}{\underset{e_2\sigma_2}{\longleftrightarrow}} & \cdots & \overset{\pi_m}{\underset{e_m\sigma_m}{\longleftrightarrow}} & t_{m+1} = t \\ & | & & | & & | & \\ & P_1 & & P_2 & & P_m & \end{array}$$

of $m+1$ terms ($m \geq 0$), each step $t_k \leftrightarrow t_{k+1}$ of which is either *trivial* ($t_{k+1} = t_k$), or else is justified by a conditional equation e_k in E, a position π_k in t_k, a substitution σ_k for variables in the equation, and subproofs P_k for the conditions $p_k\sigma_k$ of the applied instance $e_k\sigma_k$. Steps employing an unconditional equation do not have subproofs as part of their justification. By the completeness of positive-unit resolution for Horn clauses, any equation $s \simeq t$ that is valid for a set E of conditional equations is amenable to such an equational proof.

We use the notation $E \vdash E'$ to denote one inference step, applying any of the seven rules to a set E of conditional equations to obtain a new set E'. The inference rules are evidently sound, in that the class of provable theorems is unchanged by an inference step.

By a *peak*, we mean a proof segment of the form $s \leftarrow u \rightarrow t$; by a *valley*, we mean a proof segment of the form $u \rightarrow^* w \leftarrow^* t$; by a *drop*, we mean a step $s \rightarrow t$ with valley subproofs; a *plateau* is a trivial subproof of form $s \leftrightarrow s$. The *depth* of a proof is the maximum number of nestings of subproofs; it is one more than the maximum depth of its subproofs. A *normal-form* proof is a valley proof of depth 0. That is the same as saying that a normal-form proof has no peaks, no drops, and no plateaus. Normal-form proofs may be thought of as "direct" proofs; in a refutational framework the existence of such a proof for $s \simeq t$ means that demodulation of s and t using positive unit equations suffices to derive a contradiction between the Skolemized negation $s' \not\simeq t'$ of the given theorem and $x \simeq x$.

The above inference rules are designed to allow any equational proof to be tranformed into normal form. A strategy based on these rules is complete if we can show that, with enough inferences, any theorem has a normal-form proof. We call an inference "fair" if all persistent superpositions, narrowings, and reflections have been considered:

Definition. An inference sequence $E_0 \vdash E_1 \vdash \cdots$ is *fair* if

$$\mathbf{exp}(E_\infty) \quad \subseteq \quad \bigcup_{i \geq 0} E_i,$$

where E_∞ is the set $\liminf_j E_j = \cup_{i \geq 0} \cap_{j \geq i} E_j$ of *persisting* conditional equations and $\mathbf{exp}(E_\infty)$ is the set of conditional equations that may be inferred from persisting equations by one application of an expansion rule (**superpose, narrow,** or **reflect**).

Our goal is to demonstrate that for any proof $s \leftrightarrow^* t$ of $s \simeq t$ in E_0, there eventually exists an unconditional valley proof $s \rightarrow^* w \leftarrow^* t$. Were it not for contraction rules, it would be relatively easy to show that **narrow** and **reflect** eventually provide an unconditional proof of $s \simeq t$, and that **superpose** eventually turns that into a valley.

Theorem (Normalization). *If an inference sequence $E_0 \vdash E_1 \vdash \cdots$ is fair, then for any proof of $s \simeq t$ in E_0, there is a normal-form proof of $s \simeq t$ in E_∞.*

This is shown by transfinite induction on proofs. Proofs are measured in the following way: For each step

$$t_k = w[u\sigma]_\pi \underset{\substack{e\sigma \\ | \\ P}}{\overset{\pi}{\longleftrightarrow}} w[v\sigma]_\pi = t_{k+1}$$

in a proof, where e is a conditional equation $q_1 \wedge \cdots \wedge q_n \Rightarrow u \simeq v$, we consider the quintuple

$$\langle n,\ q_1\sigma \wedge \cdots \wedge q_n\sigma,\ t_k,\ u,\ t_{k+1}\rangle,$$

where (with loss of generality) we are assuming $t_k \geq t_{k+1}$ in the complete ordering $>$. Quintuples are compared left-to-right lexicographically, with the first component (the number of conditions n) compared in the natural ordering of natural numbers, the second (the instantiated conditions $q_1\sigma \wedge \cdots \wedge q_n\sigma$), third ($t_k$), and fifth ($t_{k+1}$) components under the complete ordering $>$, and the fourth component (u) in the encompassment ordering $\trianglerighteq$. An unconditional step has 0 as its first component and T as its second. Finally, proofs are compared by comparing multisets consisting of quintuples for all steps in their top-level proofs *or* subproofs, using the extension to finite multisets of the above ordering on quintuples. We use $\gg$ to denote this proof ordering. It can be shown by standard arguments (Dershowitz and Manna, 1979) that $\gg$ is well-founded.

Intuitively, the first component is designed to decrease with each application of **reflect** or **condense**, the second with applications of **narrow** or **simplify**, the third with **superpose**, the fourth and fifth cater to **compose**. The multiset structure of the proof ordering ensures that decreasing the complexity of subproofs decreases the complexity of the whole proof (and also takes care of **delete**). We need to show that inferences never increase the complexity of proofs and, furthermore, that there are always inferences that can decrease the complexity of non-normal proofs. Then, by induction with respect to $\gg$, the eventual existence of a normal-form proof follows.

Lemma 1. *If $E \vdash E'$, then for any proof P in E of an equation $s \simeq t$, there exists a proof P' in E' of $s \simeq t$, such that $P \gg P'$ or $P = P'$.*

This is established by consideration of the effects of each contracting inference rule that deletes or replaces equations, since for expansion rules, $E \subseteq E'$, and we can take $P' = P$.

Consider any ground proof and look at a step $t_k = w[u\sigma]_\pi \leftrightarrow^{\pi}_{e\sigma} w[v\sigma]_\pi = t_{k+1}$.

> **Delete** can prevent a proof from employing a clause $p \Rightarrow u \simeq u$. There is, however, an alternative proof that splices out the step $t_k \leftrightarrow t_{k+1} = t_k$, leaving just t_k, and omits all its subproofs. This strictly decreases the complexity of the whole proof, by eliminating one or more quintuples from the multiset measure.
>
> **Condense** erases a condition $s \simeq s$ from a clause. If e is such a clause, there is a new proof using the condensed clause instead, which omits any (unnecessary) proofs of that condition. The quintuple associated with the step experiences a decrease in its first component.
>
> **Simplify** changes the second component of the step's cost from $q[l\sigma]$ to the smaller $q[r\sigma]$. Though the cost of the subproofs is increased on account of an additional unconditional step $q[l\sigma] \rightarrow q[r\sigma]$ needed to establish $q[r\sigma]$ (given proofs of $q[l\sigma]$), that unconditional step (with 0 in its first component) is dominated by the simplified conditional one (with $m > 0$ in its first component).
>
> **Compose** replaces a step $t_k = w[l\tau] \leftrightarrow_{p\sigma\Rightarrow u\sigma\simeq v\sigma} t_{k+1}$ with a two-step proof $t_k = w[l\tau] \rightarrow w[r\tau] \leftrightarrow t_{k+1}$. The cost of the replaced step is $\langle n, p\sigma, t_k, u, t_{k+1}\rangle$ if $t_k > t_{k+1}$; otherwise, it is $\langle n, p\sigma, t_{k+1}, v, t_k\rangle$. If $t_k > t_{k+1}$, then the cost of $w[r\tau] \leftrightarrow t_{k+1}$ is smaller in the third component; if $t_k \leq t_{k+1}$, it is smaller in the last. (The first two components are in any case unchanged.) The cost $\langle 0, T, t_k, l, w[r\tau]\rangle$ of the unconditional step $t_k \rightarrow w[r\tau]$ is

smaller than that of the replaced step in the first, third, or fourth component, depending on which of the enabling conditions of the inference rule is satisfied: if the replaced step was conditional ($q \neq T$), then it is smaller in the first; if $q = T$, but $v > u$, then (by the replacement and substitution properties) $t_{k+1} > t_k$, and the reduction is in the third; if $q = T$ and $v \leq u$, but $u \blacktriangleright l$, then the first three components are the same, but the fourth is smaller in the encompassment ordering.

Lemma 2. *If P is a non-normal-form proof in E, then there exists a proof P' in $E \cup \exp(E)$ such that $P \gg P'$.*

The argument depends on a distinction between "non-critical" subproofs, for which there is a proof P' in E itself, and "critical" subproofs, for which equations in $\exp(E)$ are needed. A peak $t_{k-1} \leftarrow^{\pi}_{p\sigma \Rightarrow l\sigma \simeq r\sigma} t_k \rightarrow^{\rho}_{q\tau \Rightarrow u\tau \simeq v\tau} t_{k+1}$, where $t_k = w[l\sigma]_\pi[u\tau]_\rho$, is *critical* if the position π is at or below the position ρ in w at which $u \simeq v$ is applied, but not at or below a position corresponding to any variable in u, or (symmetrically) if ρ falls within the non-variable part of the occurrence of l in w. Similarly, a drop $t_k \rightarrow^{\pi}_{q\sigma \Rightarrow e\sigma} t_{k+1}$ is *critical* if the first or last step of one of the subproofs for $q\sigma$ takes place within the non-variable part of the condition q.

Since any proof must have at least one subproof of depth 0, any non-normal proof must have a plateau, an *unconditional* peak, or a drop of depth 1 with (unconditional) valley subproofs. Thus, we need not worry about peaks involving a conditional rule, nor drops in which the proof of some condition is not unconditional. All plateaus of depth 0 can be spliced out. Critical unconditional peaks, critical drops with non-empty unconditional valley subproofs, and drops with empty proofs of conditions can each be replaced by a smaller proof, using the conditional equation generated by a **superpose**, **narrow**, or **reflect** inference, respectively. Superposition replaces two steps with one that is smaller in the third component (the first two are unchanged); narrowing results in a step that is smaller in the second component (and removes a step from the subproof); reflection causes a decrease in the first component. Narrowing can be restricted to the maximal side of the maximal condition, since a drop with non-empty subproofs must have a step emanating from the larger side of its largest condition.

Non-critical unconditional peaks $t_{k-1} \leftarrow t_k \rightarrow t_{k+1}$ have alternative, smaller proofs $t_{k-1} \rightarrow^* t_k \leftarrow^* t_{k+1}$ in E by the version of the Critical Pair Lemma of (Knuth and Bendix, 1970) in (Lankford, 1975). Consider a non-critical drop $w[u\sigma] \leftrightarrow_{q\sigma \Rightarrow u\sigma \simeq v\sigma} w[v\sigma]$, with unconditional subproof $p\sigma \rightarrow p' \rightarrow^* p'' \leftarrow^* p'''$, where $p\sigma$ is no smaller than any other term in the subproof $q\sigma$. Suppose p has a variable x at position π and the first step applies within the variable part $p|_\pi$. That is, $p\sigma = p\sigma[x\sigma]_\pi \rightarrow p\sigma[r]_\pi = p'$. Let τ be the same substitution as σ except that $\tau : x \mapsto r$. There is a smaller proof (smaller, vis-a-vis $\gg$) in E: $w[u\sigma] \leftarrow^* w[u\tau] \leftrightarrow_{q\sigma \Rightarrow u\sigma \simeq v\sigma} w[v\tau] \rightarrow^* w[v\sigma]$. The new conditional step $w[u\tau] \leftrightarrow w[v\tau]$ is cheaper (in the second component) than the original, since $q\tau$ must be strictly smaller than $q\sigma$. The steps $w[u\sigma] \leftarrow^* w[u\tau]$ and $w[v\tau] \rightarrow^* w[v\sigma]$ are also cheaper than $w[u\sigma] \leftrightarrow w[v\sigma]$, since they are unconditional (hence smaller in the first component). Also any rewrites $x\sigma \rightarrow r$ that need to be added to turn a proof of $q\sigma$ into a proof of $q\tau$ are unconditional.

The Normalization Theorem follows. If $s \simeq t$ is provable in E_0, then (by Lemma 1) it has a proof P in the limit E_∞. If P is non-normal, then (by Lemma 2) it admits a smaller proof P' using (in addition to E_∞) a finite number of equations in $\exp(E_\infty)$. By fairness, each of those equations appeared at least once along the way. Subsequent inferences (by Lemma 1) can only decrease the complexity of the proof of such an equation once it appears in a set E_j (and has a one-step proof). Thus, each equation needed in P' has a proof of no greater complexity in E_∞ itself, and hence (by the multiset nature of the proof measure), there is a proof of $s \simeq t$ in E_∞ that is strictly smaller than P. Since the ordering on proofs is well-founded, by induction there must be a normal proof in E_∞.

5. Extensions

In the above method, the same ordering is used for simplification as for choosing the maximal literal. In fact, a different selection strategy can be used for choosing the literal to narrow, as in (Ganzinger, 1987; Sivakumar, 1989), but then the term ordering must be used to choose the larger side of the equality.

We used only unconditional equations for simplification and composition. Conditional equations can also be used—but only in those cases where the proof ordering shows a decrease anyway. An alternative is to design an inference system that distinguishes between different kinds of non-unit clauses. An instance $p\sigma \Rightarrow u\sigma \simeq v\sigma$ of a conditional equation is "decreasing" (in the terminology of Dershowitz and Okada, 1990) if $u\sigma \succ v\sigma, p\sigma$ in the completable ordering. This is the same condition as imposed on conditional rewrite rules by the completion-like procedures of Kaplan (1987) and Ganzinger (1987). In these methods, superposition is used when the left-hand side is larger than the conditions; narrowing, when a condition dominates the left-hand side. As theorem provers, however, they are refutationally *in*complete, since they make no provision for "unorientable" equations $s \simeq t$ such that $s \not\succ t$ and $t \not\succ s$. For a complete method, the inference rules given here must be modified to use the largest positive *or* negative clause in each expansion, and to treat decreasing equations like unit equations. In particular, superposition is needed between decreasing conditional rules. We must redefine a normal-form proof of $s \simeq t$ to be a valley proof in which each subproof is also in normal form and each term in a subproof is smaller than the larger of s and t; see (Dershowitz and Okada, 1988). (The normal forms of the previous section are a special case.) Any non-normal-form proof has a peak made from decreasing instances with normal-form subproofs, or else has a non-decreasing step with a drop.

References

[1] L. Bachmair, *Canonical Equational Proofs.* Boston: Birkhäuser, 1991. To appear.

[2] L. Bachmair, N. Dershowitz, and J. Hsiang, "Orderings for equational proofs," in *Proceedings of the IEEE Symposium on Logic in Computer Science*, (Cambridge, MA), pp. 346–357, June 1986.

[3] L. Bachmair, N. Dershowitz, and D. A. Plaisted, "Completion without failure," in *Resolution of Equations in Algebraic Structures 2: Rewriting Techniques* (H. Aït-Kaci and M. Nivat, eds.), ch. 1, pp. 1–30, New York: Academic Press, 1989.

[4] L. Bachmair and H. Ganzinger,."On restrictions of ordered paramodulation with simplification," in *Proceedings of the Second International Workshop on Conditional and Typed Rewriting Systems* (M. Okada, ed.), (Montreal, Canada), June 1990. *Lecture Notes in Computer Science*, Springer, Berlin.

[5] R. S. Boyer, *Locking: A restriction of resolution.* PhD thesis, University of Texas at Austin, Austin, TX, 1971.

[6] T. C. Brown, Jr., *A structured design-method for specialized proof procedures.* PhD thesis, California Institute of Technology, Pasadena, CA, 1975.

[7] H. Comon, "Solving inequations in term algebras (Preliminary version)," in *Proceedings of the Fifth Annual IEEE Symposium on Logic in Computer Science* (Philadelphia, PA), pp. 62—69, June 1990.

[8] N. Dershowitz, "Orderings for term-rewriting systems," *Theoretical Computer Science*, vol. 17, pp. 279–301, March 1982.

[9] N. Dershowitz, "Termination of rewriting," *J. of Symbolic Computation*, vol. 3, pp. 69–115, February/April 1987. Corrigendum: *4*, 3 (December 1987), 409–410.

[10] N. Dershowitz and J.-P. Jouannaud, "Rewrite systems," in *Handbook of Theoretical Computer Science B: Formal Methods and Semantics*, (J. van Leeuwen, ed.), ch. 6, Amsterdam: North-Holland, 1990.

[11] N. Dershowitz and Z. Manna, "Proving termination with multiset orderings," *Communications of the ACM*, vol. 22, pp. 465–476, August 1979.

[12] N. Dershowitz and M. Okada, "Proof-theoretic techniques and the theory of rewriting," in *Proceedings of the Third IEEE Symposium on Logic in Computer Science* (Edinburgh, Scotland), pp. 104–111, July 1988.

[13] N. Dershowitz and M. Okada, "A rationale for conditional equational programming," *Theoretical Computer Science*, vol. 75, pp. 111—138, 1990.

[14] U. Furbach, "Oldy but goody: Paramodulation revisited," in *Proceedings of the GI Workshop on Artificial Intelligence* (Morik, ed.), pp. 195–200, 1987. Vol. 152 of *Informatik Fachberichte.*

[15] H. Ganzinger, "A completion procedure for conditional equations," in *Proceedings of the First International Workshop on Conditional Term Rewriting Systems* (S. Kaplan and J.-P. Jouannaud, eds.), (Orsay, France), pp. 62–83, July 1987. Vol. 308 of *Lecture Notes in Computer Science*, Springer, Berlin (1988).

[16] L. Henschen and L. Wos, "Unit refutations and Horn sets," *J. of the Association for Computing Machinery*, vol. 21, pp. 590–605, 1974.

[17] J. Hsiang and M. Rusinowitch, "A new method for establishing refutational completeness in theorem proving," in *Proceedings of the Eighth International Conference on Automated Deduction* (J. H. Siekmann, ed.), (Oxford, England), pp. 141–152, July 1986. Vol. 230 of *Lecture Notes in Computer Science*, Springer, Berlin.

[18] J. Hsiang and M. Rusinowitch, "On word problems in equational theories," in *Proceedings of the Fourteenth EATCS International Conference on Automata, Languages and Programming* (T. Ottmann, ed.), (Karlsruhe, West Germany), pp. 54–71, July 1987. Vol. 267 of *Lecture Notes in Computer Science*, Springer, Berlin.

[19] S. Kamin and J.-J. Lévy, "Two generalizations of the recursive path ordering," Unpublished note, Department of Computer Science, University of Illinois, Urbana, IL, February 1980.

[20] S. Kaplan, "Simplifying conditional term rewriting systems: Unification, termination and confluence," *J. of Symbolic Computation*, vol. 4, pp. 295–334, December 1987.

[21] D. E. Knuth and P. B. Bendix, "Simple word problems in universal algebras," in *Computational Problems in Abstract Algebra* (J. Leech, ed.), pp. 263–297, Oxford, U. K.: Pergamon Press, 1970. Reprinted in *Automation of Reasoning 2*, Springer, Berlin, pp. 342–376 (1983).

[22] E. Kounalis and M. Rusinowitch, "On word problems in Horn theories," in *Proceedings of the First International Workshop on Conditional Term Rewriting Systems* (S. Kaplan and J.-P. Jouannaud, eds.), (Orsay, France), pp. 144–160, July 1987. Vol. 308 of *Lecture Notes in Computer Science*, Springer, Berlin (1988).

[23] D. S. Lankford, "Canonical inference," Memo ATP-32, Automatic Theorem Proving Project, University of Texas, Austin, TX, December 1975.

[24] E. Paul, "On solving the equality problem in theories defined by Horn clauses," *Theoretical Computer Science*, vol. 44, no. 2, pp. 127–153, 1986.

[25] G. Robinson and L. Wos, "Paramodulation and theorem-proving in first order theories with equality," in *Machine Intelligence 4* (B. Meltzer and D. Michie, eds.), pp. 135–150, Edinburgh, Scotland: Edinburgh University Press, 1969.

[26] M. Rusinowitch, *Démonstration Automatique: Techniques de réécriture.* Paris, France: InterEditions, 1989.

[27] G. Sivakumar, *Proofs and Computations in Conditional Equational Theories.* PhD thesis, Department of Computer Science, University of Illinois, Urbana, IL, 1989.

[28] H. Zhang and D. Kapur, "First-order theorem proving using conditional equations," in *Proceedings of the Ninth International Conference on Automated Deduction* (E. Lusk and R. Overbeek, eds.), (Argonne, Illinois), pp. 1–20, May 1988. Vol. 310 of *Lecture Notes in Computer Science*, Springer, Berlin.