Skolem Functions and Equality in Automated Deduction *

William McCune

Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, Illinois 60439-4801
mccune@mcs.anl.gov

Abstract

We present a strategy for restricting the application of the inference rule paramodulation. The strategy applies to problems in first-order logic with equality and is designed to prevent paramodulation into subterms of Skolem expressions. A weak completeness result is presented (the functional reflexive axioms are assumed). Experimental results on problems in set theory, combinatory logic, Tarski geometry, and algebra show that the strategy can be useful when searching for refutations and when applying Knuth-Bendix completion. The emphasis of the paper is on the effectiveness of the strategy rather than on its completeness.

1 Introduction

Most inference rules and strategies for guiding or restricting searches in automated deduction are syntactic in spirit. Examples are inference rules that generate positive clauses only, unrestricted back-chaining searches, restriction strategies that consider orderings on terms, and guidance strategies that focus on clauses with few symbols. Notable exceptions, which use interpretations or the intended meaning of the symbols, are semantic inference rules, inference rules with built-in knowledge of the intended domain, the set of support strategy for starting the search with specific clauses, and ad hoc weighting methods for controlling searches.

We present a paramodulation strategy which has a semantic motivation in that a distinction is made between Skolem functions and ordinary functions. The strategy is to prevent paramodulation (equality substitution) into proper subterms of Skolem terms. Human reasoners tend to treat the objects corresponding to Skolem terms as atomic, and it appears that programs can benefit by making a related distinction.

One can easily show the completeness of the strategy for otherwise-unrestricted paramodulation in the presence of the functional reflexive axioms. The proof rests on the (not well-known) fact [BKS85, Ben89] that equality axioms are not required for Skolem functions. The focus of this work is to try to determine the practical effectiveness of the strategy.

The restriction strategy was added to the deduction system OTTER [McC90], and experiments were conducted on problems in set theory, combinatory logic, Tarski geometry, and Knuth-Bendix completion. The completeness result does not yet apply to Knuth-Bendix completion problems, but our experiments indicate that the strategy can be valuable for those problems anyway.

Preliminary work on this topic appeared in [McC89].

2 Preliminaries

We assume a resolution/paramodulation refutation system for first-order logic with equality. If the formula representing the problem in question involves existentially quantified variables or is not in conjunctive normal form, it is preprocessed. A standard way to preprocess consists of three steps: conversion to negation normal form, Skolemization, then conversion to conjunctive normal form. The Skolemization procedure is the interesting step. Existentially quantified variables are replaced with new *Skolem functions* and *Skolem constants*. Arguments of the Skolem functions are the universally quantified variables in whose scope the existential quantifier occurs. A term is a *Skolem*

*This work was supported by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

expression if its leading function symbol is a Skolem function symbol, and $Sk(F)$ is any Skolemization of a formula F.

The result of preprocessing is a conjunction of clauses, whose variables are implicitly universally quantified. (Variables in clauses start with 'u'–'z'.) The key property of Skolemization is that unsatisfiability is preserved: $Sk(F)$ unsatisfiable (E-unsatisfiable) if and only if F is unsatisfiable (E-unsatisfiable).

If the problem in question involves equality, one can either apply resolution inference rules with the equality axioms, or apply specialized inference rules that operate on equalities. The *equality axioms* for a set of function and relation symbols are reflexivity, symmetry, transitivity, and a substitution axiom for each argument position of each function symbol and relation symbol.

$$x = x$$
$$x \neq y \ \lor \ y = x$$
$$x \neq y \ \lor \ y \neq z \ \lor \ x = z$$
$$x \neq y \ \lor \ f(\ldots, x, \ldots) = f(\ldots, y, \ldots)$$
$$x \neq y \ \lor \ \neg P(\ldots, x, \ldots) \ \lor \ P(\ldots, y, \ldots)$$

$EqAx(F)$ is the set of equality axioms for the function and relation symbols of a formula F. Although equality substitution axioms for Skolem functions have traditionally been included, it is known [BKS85, Ben89] that they are not necessary.

The most widely used inference rule for equality is *paramodulation* [WOLB84], which generalizes equality substitution to include unification and extra literals in the spirit of resolution. Let N and L be disjunctions of literals, and let M be a literal containing a term t_2. Paramodulation applies *from* clause $t_1 = r \lor L$ *into* clause $M[t_2] \lor N$ if t_1 and t_2 have most a general unifier σ. The paramodulant is $(M[r] \lor N \lor L)\sigma$.

The *functional reflexive axioms* for a set of function symbols consists of an equality $f(x_1, \ldots, x_n) = f(x_1, \ldots, x_n)$ for each n-ary function symbol f.

Let S be a set of clauses involving equality. The following are basic results in logic and automated theorem proving.

- S is E-unsatisfiable if and only if $S \& EqAx(S)$ is unsatisfiable.
- Paramodulation is a complete inference rule for equality. In particular, if S is an E-unsatisfiable set of clauses, there is a paramodulation/resolution/factoring refutation of S & $(x = x)$. (Some useful restrictions of paramodulation, such as the set of support strategy, require the presence of the functional reflexive axioms for completeness. Even when required for completeness, they are rarely used in practice.)

3 The Restriction Strategy

Although the following result is basic in logic, it appears in just one [Lov78] (as far as we can tell) of the standard texts on resolution-based automated theorem proving, and it is not widely used by the automated theorem proving community. A first-order formula F is E-unsatisfiable if and only if $F \& EqAx(F)$ is unsatisfiable. (Note that F is not necessarily Skolemized.)

An immediate consequence of that result is that the equality substitution axioms can be fixed before Skolemization occurs; in particular, equality substitution axioms are not required for the Skolem functions. In fact, the following five statements are equivalent.

(1) F is E-unsatisfiable.

(2) $F \& EqAx(F)$ is unsatisfiable.

(3) $Sk(F)$ is E-unsatisfiable.

(4) $Sk(F \& EqAx(F))$ is unsatisfiable.

(5) $Sk(F) \& EqAx(Sk(F))$ is unsatisfiable.

Research in automated theorem proving with equality has focused on the equivalence (3) iff (5), because clause sets rather than the first-order formulas are usually taken as given. For example, the widely used and cited problem set [MOW76] presents clauses, including equality axioms for Skolem functions.

The fact that equality axioms are not needed for Skolem functions can be turned into a strategy for restricting paramodulation, because paramodulation into a term corresponds to a sequence of resolution steps with equality axioms. For example, let f be a 3-place function symbol, and consider a term $f(t_1, t_2, t_3)$. Paramodulation into t_2 or one of its subterms corresponds to a sequence of resolution steps with equality axioms. One of the equality axioms is $x \neq y \ \lor \ f(x_1, x, x_3) = f(x_1, y, x_3)$, because paramodulation is into the second argument of f. If f is a Skolem function, that equality axiom need not be present, indicating that the paramodulation inference need not be made.

The Strategy. *Never paramodulate into a proper subterm of a Skolem expression.*

An Outline of the Completeness Proof. If one assumes unrestricted paramodulation as a starting point, it is not difficult to show the completeness of the restriction. Let S be an E-unsatisfiable set of clauses; let F be the set of function and relation symbols, *excluding Skolem function symbols*, of S; let $EqAx$ be the equality axioms for F; and let FRA be the functional reflexive axioms for F. Let R be a hyperresolution/factoring refutation of $S\&EqAx$. Such a refutation exists because $S\&EqAx$ is unsatisfiable and hyperresolution/factoring is complete. R can be transformed—as in [CL73, pp. 171-172]—into a hyperresolution/factoring/paramodulation refutation of $S\&\ (x = x)\ \&FRA$. Each hyperresolution inference with an equality axiom can be directly transformed into a paramodulation inference that satisfies the restriction.

Motivation. The restriction strategy has intuitive appeal as well. The arguments of Skolem functions should serve as place holders for the objects on which the "existing" object depends—equality substitution should not be applied to them. For an example in set theory, if $A \not\subseteq B$ is assumed, one can conclude that there is an object c in A that is not in B. In clauses, c is a Skolem function applied to A and B. If A and B are complicated expressions, one might wish to reason about them by applying equality substitutions, but not to the occurrences in the Skolem expression representing c—those should be fixed when c is "chosen". If A and B are not ground when c is chosen, further inferences should be free to instantiate the Skolem expression. With this view, the arguments of Skolem functions serve as constraints on unification and constraints on inference.

4 Applications and Experiments

OTTER [McC90] is a resolution/paramodulation theorem prover for first-order logic with equality. The paramodulation restriction strategy was installed in OTTER, and a number of experiments were conducted to try to determine the value of the strategy.

Four application areas were chosen for study: several versions of a problem in set theory, two problems in combinatory logic, a problem in Tarski geometry, and a benchmark algebra problem in complete sets of reductions. The set theory and geometry problems are non-Horn and require a mixture of resolution and paramodulation; in addition, the set theory problems use defined concepts. The combinatory logic and algebra problems require no resolution because they are presented as equality units; the algebra problem requires demodulation (term rewriting) and is not a search for a refutation.

4.1 Set Theory

The problem, to show that the composition of two functions is also a function, is the naive version of one of the lemmas in [BLM+86]. It is an easy problem, but OTTER has difficulty with defined concepts, non-Horn clauses, and mixtures of equality and other relations, all of which occur in this kind of set theory problem.

Definitions of relation, single-valued set, function, and composition.

$$\forall R(relation(R) \leftrightarrow \forall u(u \in R \rightarrow \exists x \exists y(u = \langle x,y\rangle)))$$

$$\forall S\left(singval(S) \leftrightarrow \forall x\forall y\forall z\left(\begin{bmatrix} \langle x,y\rangle \in S\& \\ \langle x,z\rangle \in S \end{bmatrix} \rightarrow y = z\right)\right)$$

$$\forall F(function(F) \leftrightarrow relation(F)\ \&\ singval(F))$$

$$\forall u\forall F\forall G\left(u \in F\circ G \leftrightarrow \exists x\exists y\exists z\begin{bmatrix} u = \langle x,z\rangle\ \& \\ \langle x,y\rangle \in G\ \& \\ \langle y,z\rangle \in F \end{bmatrix}\right)$$

Property of ordered pair.

$$\forall x\forall y\forall z\forall w(\langle x,y\rangle = \langle z,w\rangle \rightarrow x = z\ \&\ y = w)$$

Theorem.

$$\forall F\forall G(function(F)\&function(G) \rightarrow function(F\circ G))$$

Four versions of the theorem were considered. Versions ST-1 and ST-2 use the clauses shown below. Version ST-3 does not use the defined relations *relation*, *singval*, or *function* (the theorem is stated in terms of the *composition*, *ordered pair*, and equality). Version ST-4 is entirely in terms of *ordered pair* and equality. Table 1 contains a summary of the results.

Clause form of the denial of the theorem. (Function symbols starting with the letter f are Skolem function symbols.)

1. $x = x$
2. $\neg relation(z) \lor u \notin z \lor u = \langle f7(z,u), f8(z,u)\rangle$
3. $relation(z) \lor f9(z) \in z$
4. $relation(z) \lor f9(z) \neq \langle x,y\rangle$
5. $\neg singval(x) \lor \langle u,v\rangle \notin x \lor \langle u,w\rangle \notin x \lor v = w$
6. $singval(x) \lor \langle f1(x), f2(x)\rangle \in x$
7. $singval(x) \lor \langle f1(x), f3(x)\rangle \in x$
8. $singval(x) \lor f2(x) \neq f3(x)$
9. $\neg function(x) \lor relation(x)$
10. $\neg function(x) \lor singval(x)$
11. $function(x) \lor \neg relation(x) \lor \neg singval(x)$
12. $z \notin y \circ x \lor z = \langle f4(z,x,y), f6(z,x,y)\rangle$
13. $z \notin y \circ x \lor \langle f4(z,x,y), f5(z,x,y)\rangle \in x$
14. $z \notin y \circ x \lor \langle f5(z,x,y), f6(z,x,y)\rangle \in y$
15. $z \in y \circ x \lor z \neq \langle u,w\rangle \lor \langle u,v\rangle \notin x \lor \langle v,w\rangle \notin y$
16. $\langle x,y\rangle \neq \langle u,v\rangle \lor x = u$
17. $\langle x,y\rangle \neq \langle u,v\rangle \lor y = v$
18. $function(F)$
19. $function(G)$
20. $\neg function(F \circ G)$

Table 1: Set Theory

Version	Time	Time with restriction
ST-1	319 seconds	32 seconds
ST-2	93 seconds	45 seconds
ST-3	53 seconds	11 seconds
ST-4	52 seconds	35 seconds

4.2 Combinatory Logic and Fragments

Combinatory Logic (CL) is closely related to the untyped λ-calculus. There are two constants S and K, and one binary operator *apply*. Terms are normally written without the operator, and when parentheses are omitted, association to the left is assumed. For example, the term $apply(apply(a,b), apply(c,d))$ is normally written $ab(cd)$. Many interesting first-order equational theorems in CL can be found in [Smu85].

Problem W, *versions* W-1 *and* W-2. In combinatory logic, find a combinator W with the property $Wxy = xyy$. Version W-1 uses a finely tuned search strategy for problems of this type, and W-2 uses a more naive search strategy.

$\forall x \forall y \forall z (Sxyz = xz(yz))$
$\forall x \forall y (Kxy = x)$

Theorem. $\exists W \forall x \forall y (Wxy = xyy)$

Denial of the theorem in clauses, with explicit function symbol for apply (f and g are Skolem functions, and Ans(z) is an answer literal).

1. $a(a(a(S,x),y),z) = a(a(x,z),a(y,z))$
2. $a(a(K,x),y) = x$
3. $a(a(z,f(z)),g(z)) \neq a(a(f(z),g(z)),g(z)) \lor Ans(z)$

If axioms different from those for S and K are used, the system is in general weaker than CL and is called a fragment of CL.

Problem FP, *versions* FP-1 *and* FP-2. Find a fixed point combinator in the fragment $\{Q, M\}$, with $Qxyz = y(xz)$ and $Mx = xx$. Version FP-1 uses a finely tuned search strategy for problems of this type, and FP-2 uses a more naive search strategy.

$\forall x \forall y \forall z (Qxyz = y(xz))$
$\forall x (Mx = xx)$

Theorem. $\exists \Theta \forall x (\Theta x = x(\Theta x))$

Denial of the theorem in clauses, with explicit function symbol for apply. (f is a Skolem function).

1. $a(a(a(Q,x),y),z) = a(y,a(x,z))$
2. $a(M,x) = a(x,x)$
3. $a(z,f(z)) \neq a(f(z),a(z,f(z))) \lor Ans(z)$

Table 2: Combinatory Logic

Version	Time	Time with restriction
W-1	9 seconds	9 seconds
W-2	(no proof)	12 seconds
FP-1	1 second	1 second
FP-2	50 seconds	3 seconds

4.3 Complete Sets of Reductions

By setting the appropriate options, OTTER can be made to search for a complete set of reductions with the Knuth-Bendix procedure. The completeness argument as presented does not directly apply in this context, because the goal is a canonical term-rewriting system rather than a refutation. However, the restriction strategy can be directly applied, because the computation of critical pairs is itself a restricted form of paramodulaion.

A benchmark completion problem [Chr89]. Given an associative system with 24 left identities and 24 right inverses, find a complete set of reductions. The 24 inverses can be clearly interpreted as Skolem functions. Paramodulation into proper subterms of inverse expressions was prevented, but simplification of those subterms was allowed.

$$f(f(x,y),z) = f(x,f(y,z))$$

$$\begin{array}{ll} f(e_1,x) = x & f(x,g_1(x)) = e_1 \\ f(e_2,x) = x & f(x,g_2(x)) = e_2 \\ \vdots & \vdots \\ f(e_{24},x) = x & f(x,g_{24}(x)) = e_{24} \end{array}$$

The time required to find a complete set of reductions was 474 seconds without the restriction and 56 seconds with it. The same set was found.

4.4 Tarski Geometry

Tarski developed and studied several first-order axiomatizations of elementary geometry—we used the version reproduced in [Wos88]. The domain is points in the plane, and the primitives are a 3-place relation "between" and a 4-place relation "equidistance of 2 pairs of points". We experimented with test problem 10 in [Wos88], the five point theorem. Even though equality relation and several Skolem functions are present, the restriction strategy had little or no effect in any of the comparisons we made. Part of the reason is that there are few occurrences of equality in the axioms and equality plays a small role in the proof of the theorem.

4.5 Summary of Experimental Results

It is difficult to evaluate the effectiveness and generality of new ideas in automated theorem proving. The fact that a new strategy performs well on a particular problem is little indication of its performance on semantically or syntactically similar problems. The Skolem function restriction strategy had a positive effect in three of the four areas that were considered. In *none* of the experiments did the strategy have a negative effect. In particular, we have not found any cases in which the strategy blocks a refutation or interferes in any other way with the search for a refutation (even though the functional reflexive axioms are never included and other restrictions are sometimes applied).

The consequences and completeness of the restriction strategy have not yet been analyzed for Knuth-Bendix completion. However, the algebra completion experiment shows that the restriction can be useful in practice even if it is not complete. When the restricted procedure terminates, we do not in general know whether the resulting set is canonical. However, the unrestricted procedure can then be applied to the resulting set to check whether it is canonical; the cost of such a check is very small.

5 Remarks

Researchers in automated deduction usually start with or are presented with clauses rather than the fully quantified formulas from which they came (for example [MOW76]). In the algebra completion problem it is clear that the 24 inverses can be interpreted as Skolem functions, but it is not always obvious whether a function symbol is a Skolem function symbol. A partial solution is presented in [McC88], which contains a procedure that attempts to "un-Skolemize" a set of clauses—that is, eliminate function symbols by introducing existentially quantified variables, while maintaining unsatisfiability. Such function symbols can be interpreted as Skolem functions, and the strategy can be applied to them.

We conclude with several points on enhancements and related work.

- We believe that the strategy can be shown to be complete without the presence of the functional reflexive axioms. One attack, suggested by Dan Benanav [Ben90], is to define paramodulation so that all occurrences of the instantiated "into" term are replaced and use a version of the lifting lemma that does not require the functional reflexive axioms.
- A refinement of the strategy was suggested, also by Benanav [Ben90], in which the relevant functions are the ones with just variables as arguments in the input clauses, rather than Skolem functions. This is analogous to the unnecessary equality axioms studied in [Ben89].
- Should demodulation (term rewriting, simplification) be prohibited inside of Skolem expressions when searching for a refutation?
- Is the strategy compatible with the Knuth-Bendix completion procedure? In particular, can critical pair computation and/or simplification be prevented inside of Skolem expression? If not, is there a related restriction that is compatible?
- Can the restriction shed any light on the problem of searching for models/counterexamples of clauses containing Skolem functions?

References

[Ben89] Dan Benanav. *Recognizing Unnecessary Inference*. PhD thesis, Rensselaer Polytechnic Institute, 1989.

[Ben90] Dan Benanav, January 1990. communication.

[BKS85] W. Bledsoe, K. Kunen, and R. Shostak. Completeness results for inequality provers. *Artificial Intelligence*, 27:255–288, 1985.

[BLM+86] Robert Boyer, Ewing Lusk, William McCune, Ross Overbeek, Mark Stickel, and Larry Wos. Set theory in first-order logic: Clauses for Gödel's axioms. *Journal of Automated Reasoning*, 2(3):287–327, 1986.

[Chr89] Jim Christian. Fast Knuth-Bendix completion: A summary. In N. Dershowwitz, editor, *Proceedings of the 3rd International Conference on Rewriting Techniques and Applications, Springer-Verlag Lecture Notes in Computer Science, Vol. 355*, pages 551–555, New York, 1989. Springer-Verlag.

[CL73] C.-L. Chang and R. C.-T. Lee. *Symbolic Logic and Mechanical Theorem Proving.* Academic Press, New York, 1973.

[Lov78] Donald Loveland. *Automated Theorem Proving: A Logical Basis.* North-Holland, Amsterdam, 1978.

[McC88] William McCune. Un-skolemizing clause sets. *Information Processing Letters*, 29:257–263, November 1988. An early version appears as Argonne National Laboratory Tech. Memo ANL/MCS-TM-93, 1987.

[McC89] William McCune. Skolem functions and equality. Association for Automated Reasoning Newsletter no. 12, Argonne National Laboratory, May 1989.

[McC90] William McCune. OTTER 2.0 users guide. Tech. Report ANL-90/9, Argonne National Laboratory, Argonne, Ill., March 1990.

[MOW76] John McCharen, Ross Overbeek, and Larry Wos. Problems and experiments for and with automated theorem-proving programs. *IEEE Transactions on Computers*, C-25(8):773–782, August 1976.

[Smu85] Raymond Smullyan. *To Mock a Mockingbird.* Knopf, New York, 1985.

[WOLB84] Larry Wos, Ross Overbeek, Ewing Lusk, and James Boyle. *Automated Reasoning: Introduction and Applications.* Prentice-Hall, Englewood Cliffs, N.J., 1984.

[Wos88] Larry Wos. *Automated Reasoning: 33 Basic Research Problems.* Prentice-Hall, Englewood Cliffs, N.J., 1988.