Strong Skolemization

Andreas Nonnengart

MPI–I–96–2–010 December 1996

Author's Address

Max-Planck-Institut für Informatik
and Saarbrücken University
Im Stadtwald, 66123 Saarbrücken, Germany
Email: Andreas.Nonnengart@mpi-sb.mpg.de
WWW: http://www.mpi-sb.mpg.de/~nonnenga/

Acknowledgements

Thanks to Harald Ganzinger, Hans Jürgen Ohlbach and Christoph Weidenbach for their comments on earlier versions of this paper.

Abstract

Skolemization is a means to eliminate existential quantifiers within predicate logic sentences and that by replacing existentially quantified variables with Skolem function applications. The arguments of these Skolem functions are variables which are quantified outside the sub-formula under consideration. In this paper a Skolemization technique is introduced which abstracts from some of the arguments of the Skolem functions. It shows that the Skolemization result obtained this way is usually more general than what can be achieved from standard (classical) Skolemization. This can be of quite some importance since such generalizations often lead to a reduction of both search space and proof length.

Keywords

Skolemization, Skolem functions, Optimized Skolemization, First-order logic, predicate logic theorem proving.

1 Introduction

In automated theorem proving much effort is spent on the development of more and more sophisticated reasoning calculi. Not so much effort is spent on the optimization of certain normal form transformations as they are, for example, necessary for resolution-based theorem proving.

One part of the usual clause normal form transformation for resolution is *Skolemization* (see the historical paper (Skolem 1920)). Its major effect lies in substituting so-called Skolem function applications for existentially quantified variables. Usually, there are two kinds of Skolemization techniques mentioned in the standard literature on automated theorem proving (see e.g. (Loveland 1978, Chang and Lee 1973)) which are called *Inner* and *Outer* Skolemization in this paper. The two differ mainly in the choice of the arguments for the Skolem functions (see below).

Anti-prenex transformation turned out to be a fairly useful preparatory process for Skolemization[1]. Its purpose is to minimize the dependencies of existentially quantified variables from other, universally quantified variables. The value of anti-prenexing can be of quite some significance (see, e.g., (Egly 1994, Loveland 1978)) and that regardless of whether we perform Inner or Outer Skolemization. The Strong Skolemization introduced in this work is shown to be a generalization of both standard Skolemization methods.

The paper is organized as follows: First the usual notion of Skolemization is recapitulated. After that the Strong Skolemization technique is introduced and it is shown that it preserves both satisfiability and unsatisfiability. This is followed by some application examples and the paper concludes by comparing Strong Skolemization with a related technique, the so-called Optimized Skolemization.

2 Preliminaries

For simplicity, we assume formulas to be in negation normal form and that any two different quantifiers inside a formula bind variables with different names. This guarantees that the occurrence of a sub-formula $\exists x\, \Phi$ or $\forall x\, \Phi$

[1] A formula is said to be in anti-prenex form if all quantifiers are moved inwards as far as possible. There is only one exception, namely, existential quantifiers are *not* distributed over disjunctions. It is even the case that formulas of the form $\exists x\, \Phi[x] \vee \exists y\, \Psi[y]$ are transformed into $\exists x\, (\Phi[x] \vee \Psi[x])$ in order to reduce the number of existential quantifiers.

in a given formula Δ is unique.

By $\Delta[\Phi]$ we indicate that Φ is a sub-formula of Δ and $\Delta[\Phi \,/\, \Psi]$ is a representation of Δ with sub-formula Φ replaced by Ψ[2].

We refer to the set of free (i.e. not bound by any quantifier) variables in Φ by $\mathcal{V}(\Phi)$. For any variable x occurring in Δ, the (unique) sub-formula starting with the quantifier which binds x is referred to by Δ^x. Thus, if $\Delta = \Delta[\exists x\, \Phi]$ then $\Delta^x = \exists x\, \Phi$.

For readability we abbreviate sequences of the form $u_1, \ldots, u_k$ of length k by $\overline{u}_k$.

A (first-order) interpretation $\mathcal{M} = (\mathcal{D}, \Im, \phi)$ consists of a non-empty domain of discourse $\mathcal{D}$, a mapping $\Im$ which associates functions and relations to function symbols and predicate symbols respectively, and a variable valuation ϕ which maps variables to elements of $\mathcal{D}$. Given a valuation ϕ we mean by $\phi[x/d]$ the variable valuation which is identical to ϕ except for the variable x which is to be interpreted as the domain value d. Similarly, by $\mathcal{M}[x/d]$ we understand the triple $(\mathcal{D}, \Im, \phi[x/d])$ provided $\mathcal{M} = (\mathcal{D}, \Im, \phi)$.

Finally, whenever we write $\Phi\{x/t\}$ we mean Φ with every occurrence of the variable x replaced by the term t.

Definition 2.1 (inner and outer skolemization)
We obtain the Skolemization *of a sentence Δ by replacing every occurrence of sub-formulas $\exists x\, \Phi$ in Δ with a corresponding $\Phi\{x/f(\overline{y}_n)\}$, where f is an n-place function symbol which is new to the problem under consideration. We speak of* Outer Skolemization *in case the variables $\overline{y}_n$ are all the universally quantified variables such that $\exists x\, \Phi$ is a sub-formula of Δ^{y_i} for each $1 \leq i \leq n$. If $\{y_1, \ldots, y_n\} = \mathcal{V}(\exists x\, \Phi)$, i.e. the variables $\overline{y}_n$ are* all *the free variables in the sub-formula $\exists x\, \Phi$, then we speak of an* Inner Skolemization.

Skolemization is not an equivalence transformation. It preserves satisfiability and unsatisfiability, though. This is stated in the following Lemma (proofs are omitted, for they can be found in most common textbooks and survey articles on first-order predicate logic theorem proving as, e.g., in (Chang and Lee 1973, Loveland 1978, Andrews 1981)).

[2]This definition won't be ambiguous since we are only going to replace sub-formulas of the form $\exists x\, \Phi$ in the sequel which, under the above assumption, are unique inside the formulas we consider.

Lemma 2.2
Let Δ be some arbitrary sentence with sub-formula $\exists x\, \Phi$ and let $f, \overline{y}_n$ as in Definition 2.1. Then

$$\Delta\,[\exists x\,\Phi] \ \textit{ is satisfiable}$$
$$\textit{iff}$$
$$\Delta\,[\exists x\,\Phi\,/\,\Phi\{x/f(\overline{y}_n)\}] \ \textit{ is satisfiable}$$

Outer Skolemization generates new Skolem functions which get all the universally quantified variables as arguments the formula under consideration depends on. There might be unnecessarily many such variable dependencies, however, if the formula to be skolemized is not in anti-prenex form. As an example consider the formula $\forall x, y\,[\exists z\, P(x,z) \vee P(x,y)]$ for which Standard Skolemization leads to $\forall x, y\ [P(x, f(x,y)) \vee P(x,y)]$. If preceded with an anti-prenex transformation, however, we would end up with the much simpler $\forall x\ [P(x, f(x)) \vee \forall y\ P(x,y)]$ instead.

During Inner Skolemization, Skolem functions are introduced which take only "necessary" variables as arguments. However, the Skolemization result depends on the order of the Skolemization of the sub-formulas. As an example consider the formula $\forall x\ [\exists y\,(P(x,y) \wedge \exists z\, Q(x,y,z))]$. If we first skolemize on $\exists z\, Q(x,y,z)$ then we finally end up with $\forall x\ [P(x,f(x)) \wedge Q(x,f(x),g(x,f(x)))]$ whereas starting the Skolemization with the outer existentially quantified sub-formula yields $\forall x\ [P(x,f(x)) \wedge Q(x,f(x),g(x))]$. Evidently, the latter is to be preferred. Indeed, Inner Skolemization should always be performed from left to right in order to result in such a preferred form.

There is, however, a possibility to define Skolemization without such a Skolemization strategy and which always ends up with the very same result. The idea is to figure out the overall dependencies of existentially quantified variables and to use these dependencies instead of local occurrences of certain variables.

Definition 2.3
Let x be a variable in the sentence Δ. Then we define

$$Dep_\Delta(x) = \begin{cases} \{x\} & \textit{if } x \textit{ is universally quantified in } \Delta \\ \displaystyle\bigcup_{y \in \mathcal{V}(\Delta^x)} Dep_\Delta(y) & \textit{otherwise} \end{cases}$$

As an example consider the formula $\Delta = \forall x, u\, \exists y\, (P(x, y) \wedge \exists z\, Q(u, z))$. Then $Dep_\Delta(x) = \{x\}$, $Dep_\Delta(u) = \{u\}$, $Dep_\Delta(y) = \{x, u\}$, and $Dep_\Delta(z) = \{u\}$.

Definition 2.4 (standard skolemization)
We obtain the Standard Skolemization *of a sentence Δ by replacing all sub-formulas $\exists x\, \Phi$ within Δ with $\Phi\{x/f(\overline{y}_n)\}$, where f is a new function symbol and*

$$\{y_1, \ldots, y_n\} = \bigcup_{z \in \mathcal{V}(\exists x\, \Phi)} Dep_\Delta(z)$$

Lemma 2.5
Standard Skolemization preserves both satisfiability and unsatisfiability.
Proof: The above variable restriction is identical to the one for Inner Skolemization if Δ^x lies not within the scope of another existential quantifier. Moreover, the result achieved by Standard Skolemization is unique – it does not depend on the order the respective existentially quantified sub-formulas are visited. Hence, what has been claimed follows immediately from Lemma 2.2 (and Skolemization is performed from left to right). □

Note that this Skolemization is still not *optimal* in the sense that the least possible variable dependencies are found. For instance, consider the formula $\forall x, y\, \exists z\, [P(z) \wedge (Q(x, z) \vee R(y, z))]$ which, after Skolemization, becomes $\forall x, y\, [P(f(x, y)) \wedge (Q(x, f(x, y)) \vee R(y, f(x, y)))]$ and that regardless of which of the above Skolemization possibilities is performed. If we first distributed the $\wedge$ over the $\vee$, however, we would be able to perform an anti-prenex step and finally we were faced with the formula $\forall x, y\, [\exists z\, (P(z) \wedge Q(x, z)) \vee \exists z\, (P(z) \wedge R(y, z))]$. In this case the Skolemization would result in $\forall x, y\, [(P(f(x)) \wedge Q(x, f(x))) \vee (P(g(y)) \wedge R(y, g(y)))]$. Thus we have fewer dependencies for the price of having to deal with more Skolem functions. Whether or not this is still of some advantage varies from case to case; we shall not discuss this issue in this paper.

3 The Strong Skolemization Technique

Strong Skolemization applies to sub-formulas of the form $\exists \overline{x}_l\,(\Phi \wedge \Psi)$[3]; in all other cases it behaves just like Standard Skolemization. Recall that Standard Skolemization of a formula Δ replaces each of its sub-formulas of the form $\exists x\,(\Phi \wedge \Psi)$ by $\Phi\{x/f(\overline{u}_k)\} \wedge \Psi\{x/f(\overline{u}_k)\}$, where f is a new function symbol and

$$\{u_1, \ldots, u_k\} = \bigcup_{z \in \mathcal{V}(\exists x\,(\Phi \wedge \Psi))} Dep_\Delta(z)$$

Now, we can split the variables $\overline{u}_k$ into two disjoint subsets $\overline{y}_n$ and $\overline{z}_m$ such that

$$\{y_1, \ldots, y_n\} = \bigcup_{y \in \mathcal{V}(\Phi) \setminus \{x\}} Dep_\Delta(z)$$

and

$$\{z_1, \ldots, z_m\} = \left(\bigcup_{z \in \mathcal{V}(\Phi) \setminus \mathcal{V}(\Psi)} Dep_\Delta(z) \right) \setminus \{y_1, \ldots, y_n\}$$

With this we can reformulate the Standard Skolemization step as follows: We replace $\exists x\,(\Phi \wedge \Psi)$ with $\Phi\{x/f(\overline{y}_n, \overline{z}_m)\} \wedge \Psi\{x/f(\overline{y}_n, \overline{z}_m)\}$. As it will turn out, it is allowed to abstract from $\overline{z}_m$ in one of the two conjuncts resulting in fresh universally quantified variables $\overline{w}_m$. This is what Strong Skolemization is about.

Definition 3.1 (strong skolemization)
Let Δ be a first-order sentence in anti-prenex negation normal form. Strong Skolemization replaces existentially quantified sub-formulas exactly as Standard Skolemization does except for sub-formulas of the form $\exists \overline{x}_k\,(\Phi \wedge \Psi)$. In this case let

$$\{y_1, \ldots, y_n\} = \bigcup_{y \in \mathcal{V}(\Phi) \setminus \{\overline{x}_k\}} Dep_\Delta(y)$$

$$\{z_1, \ldots, z_m\} = \left(\bigcup_{z \in \mathcal{V}(\Psi) \setminus \mathcal{V}(\Phi)} Dep_\Delta(z) \right) \setminus \{y_1, \ldots, y_n\}$$

Then $\exists \overline{x}_k\,(\Phi \wedge \Psi)$ is to be replaced by $\forall \overline{w}_m\,\Phi\{x_i/f_i(\overline{y}_n, \overline{w}_m)\} \wedge \Psi\{x_i/f_i(\overline{y}_n, \overline{z}_m)\}$ for all $1 \leq i \leq k$. Such transformation steps are called Strong Skolemization steps on the sub-formula Φ.

[3]Without loss of generality we assume that both Φ and Ψ contain the variables $\overline{x}_l$, for otherwise we could apply an anti-prenex step.

The following theorem ensures that Strong Skolemization behaves as desired.

Theorem 3.2
Let $\overline{y}_n$ and $\overline{z}_m$ be as in Definition 3.1. Then

$$\Delta\,[\exists \overline{x}_k\,(\Phi \wedge \Psi)] \text{ is satisfiable}$$
$$\text{iff}$$
$$\Delta\,[\exists \overline{x}_k\,(\Phi \wedge \Psi)\,/\,(\forall \overline{w}_m\;\Phi\{x_i/f_i(\overline{y}_n, \overline{w}_m)\} \wedge \Psi\{x_i/f_i(\overline{y}_n, \overline{z}_m)\})] \text{ is satisfiable}$$

Proof: See page 8. □

Given a sub-formula $\exists x\,(\Phi \wedge \Psi)$, it indeed matters whether we perform a Strong Skolemization step on Φ or on Ψ. For instance, if

$$\bigcup_{z \in \mathcal{V}(\Psi)\setminus\{x\}} Dep_\Delta(z) \subseteq \bigcup_{z \in \mathcal{V}(\Phi)\setminus\{x\}} Dep_\Delta(z)$$

it turns out that by doing such a step on Φ the number m becomes 0. In such a case Strong Skolemization behaves as Standard Skolemization and we have no gain at all over the classical methods. Quite evidently, the greater this m, the more general is the Strong Skolemization outcome. Hence, if the one choice results in a value of $m = 0$ and the other in a value of $m > 0$ then the latter is to be preferred, for it is the one that covers the possibility to gain something more general than what Standard Skolemization could produce.
But even in case either choice leads to a value of $m > 0$ it often matters which of the two takes part. Some such examples can be found in the next section.
But first we have to show that Theorem 3.2 holds. To this end we make use of the following auxiliary lemma.

Lemma 3.3
Let Δ be some arbitrary sentence with sub-formula $\exists \overline{x}_k\,(\Phi \wedge \Psi)$. Then[4]

$$\Delta\,[\exists \overline{x}_k\,(\Phi \wedge \Psi)] \text{ is satisfiable}$$
$$\text{implies}$$
$$\Delta\,[\exists \overline{x}_k\,(\Phi \wedge \Psi)\,/\,(\forall w\,\Phi\{x_i/h_i(\overline{y}_n, w)\} \wedge \exists z\,\Psi\{x_i/h_i(\overline{y}_n, z)\})] \text{ is satisfiable}$$

[4] In fact, the Lemma holds for both directions. We only need this one in the sequel, however.

where h_i are new function symbols, w and z are fresh variables and

$$\{y_1, \ldots, y_n\} = \bigcup_{u \in \mathcal{V}(\Phi) \setminus \{\overline{x}_k\}} Dep_\Delta(u)$$

Proof: It suffices to consider the case where Δ^{x_1} lies not within the scope of another existential quantifier in Δ. The other cases then follow by some applications of Lemma 2.2 (or Lemma 2.5) (Skolemization and de-Skolemization). Thus we can assume that $\{\overline{y}_n, \overline{x}_k\}$ are the free variables of Φ. Now, for any interpretation $\mathcal{M} = (\mathcal{D}, \Im, \phi)$ and elements $\overline{\alpha}_n$ in $\mathcal{D}$ we define

$$S^\Phi(\alpha_1, \ldots, \alpha_n) = \{(\beta_1, \ldots, \beta_k) \in \mathcal{D}^k \mid \mathcal{M}[y_i/\alpha_i][x_j/\beta_j] \models \Phi\}$$

S^Φ is a representation of the set of tuples that make Φ true under the interpretation $\mathcal{M}$ (given suitable values for the universally quantified free variables in Φ). Therefore, in case of $\mathcal{M} \models \exists \overline{x}_k\, (\Phi \wedge \Psi)$ we know that

- $S^\Phi(\phi(y_1), \ldots, \phi(y_n)) \neq \emptyset$ and
- there is a $(\beta_1, \ldots, \beta_k) \in S^\Phi(\phi(y_1), \ldots, \phi(y_n))$ with $\mathcal{M}[x_i/\beta_i] \models \Psi$

Now let $\widehat{h}\colon \mathcal{D}^{n+k} \to \mathcal{D}^k$ be such that

$$\widehat{h}(\overline{\alpha}_n, \overline{\gamma}_k) = \begin{cases} (\overline{\gamma}_k) & \text{if} \quad (\overline{\gamma}_k) \in S^\Phi(\overline{\alpha}_n) \\ c_1 & \text{if } (\overline{\gamma}_k) \notin S^\Phi(\overline{\alpha}_n) \text{ and } S^\Phi(\overline{\alpha}_n) \neq \emptyset \\ c_2 & \text{otherwise} \end{cases}$$

where c_1 is an arbitrary element of $S^\Phi(\overline{\alpha}_n)$ and c_2 is an arbitrary element of $\mathcal{D}^k$ respectively. Thus, if $S^\Phi(\overline{\alpha}_n) \neq \emptyset$ for some interpretation $\mathcal{M}$ and appropriate values for the respective α_i, then $\widehat{h}(\overline{\alpha}_n, \overline{\gamma}_k)$ serves as a 'selector function' on $S^\Phi(\overline{\alpha}_n)$. I. e., $\widehat{h}(\overline{\alpha}_n, \overline{\gamma}_k)$ always provides us with an element of $S^\Phi(\overline{\alpha}_n)$ and, just as important, all elements of $S^\Phi(\overline{\alpha}_n)$ can be reached by $\widehat{h}$. Now let π_i denote the usual projection functions on tuples of elements from $\mathcal{D}$ such that $\pi_i(\ldots, \beta_i, \ldots) = \beta_i$. Then

$$\begin{gathered} S^\Phi(\phi(y_1), \ldots, \phi(y_n)) \neq \emptyset \\ \text{implies} \\ \mathcal{M}[x_i/\pi_i(\widehat{h}(\phi(y_1), \ldots, \phi(y_n), \overline{\alpha}_k))] \models \Phi \text{ for all } (\overline{\alpha}_k) \in \mathcal{D}^k \end{gathered}$$

and also

$$\begin{gathered} \mathcal{M}[x_i/\alpha_i] \models \Psi \text{ for some } (\overline{\alpha}_k) \in S^\Phi(\phi(y_1), \ldots, \phi(y_n)) \\ \text{implies} \\ \mathcal{M}[x_i/\pi_i(\widehat{h}(\phi(y_1), \ldots, \phi(y_n), \overline{\alpha}_k))] \models \Psi \text{ for some } (\overline{\alpha}_k) \in \mathcal{D}^k \end{gathered}$$

Thus, if $\mathcal{M} \models \exists \overline{x}_k \, (\Phi \wedge \Psi)$ then

$$\mathcal{M}' \models \forall \overline{w}_k \, \Phi\{x_i/h_i(\overline{y}_n, \overline{w}_k)\} \wedge \exists \overline{z}_k \, \Psi\{x_i/h_i(\overline{y}_n, \overline{z}_k)\}$$

where $\mathcal{M}'$ is like $\mathcal{M}$ except for the interpretation of the function symbols h_i which is $\mathcal{M}'(h_i) = \pi_i \circ \widehat{h}$, i. e. the ith element of the result of $\widehat{h}$. $\square$

Proof of Theorem 3.2:

$$\begin{array}{ll}
\multicolumn{2}{l}{\Delta\,[\exists \overline{x}_k \, (\Phi \wedge \Psi)] \ \text{ is satisfiable}} \\
\text{implies} & \Delta\,[\exists \overline{x}_k \, (\Phi \wedge \Psi) \,/\, \forall \overline{w}_k \, \Phi\{x_i/h_i(\overline{y}_n, \overline{w}_k)\} \wedge \exists \overline{u}_k \, \Psi\{x_i/h_i(\overline{y}_n, \overline{u}_k)\}] \\
& \qquad\qquad \text{is satisfiable (by Lemma 3.3)} \\
\text{implies} & \Delta\,[\exists \overline{x}_k \, (\Phi \wedge \Psi) \,/\, \forall \overline{w}_k \, \Phi\{x_i/h_i(\overline{y}_n, \overline{w}_k)\} \\
& \qquad\qquad\qquad\qquad \wedge \ \Psi\{x_i/h_i(\overline{y}_n, g_i(\overline{y}_n, \overline{z}_m))\}] \\
& \qquad\qquad \text{is satisfiable (by Lemma 2.5)} \\
\text{implies} & \Delta\,[\exists \overline{x}_k \, (\Phi \wedge \Psi) \,/\, \forall \overline{w}_m \, \Phi\{x_i/h_i(\overline{y}_n, g_i(\overline{y}_n, \overline{w}_m))\} \\
& \qquad\qquad\qquad\qquad \wedge \ \Psi\{x_i/h_i(\overline{y}_n, g_i(\overline{y}_n, \overline{z}_m))\}] \\
& \qquad\qquad \text{is satisfiable (by instantiation)} \\
\text{implies} & \Delta\,[\exists \overline{x}_k \, (\Phi \wedge \Psi) \,/\, \forall \overline{w}_m \ \Phi\{x_i/f_i(\overline{y}_n, \overline{w}_m)\} \wedge \Psi\{x_i/f_i(\overline{y}_n, \overline{z}_m)\}] \\
& \qquad\qquad \text{is satisfiable (let } f_i(\overline{\alpha}_n, \overline{\gamma}_m) = h_i(\overline{\alpha}_n, g_i(\overline{\alpha}_n, \overline{\gamma}_m))) \\
\text{implies} & \Delta\,[\exists \overline{x}_k \, (\Phi \wedge \Psi) \,/\, \Phi\{x_i/f_i(\overline{y}_n, \overline{z}_m)\} \wedge \Psi\{x_i/f_i(\overline{y}_n, \overline{z}_m)\}] \\
& \qquad\qquad \text{is satisfiable (by instantiation)} \\
\text{implies} & \Delta\,[\exists \overline{x}_k \, (\Phi \wedge \Psi)] \ \text{ is satisfiable (by Lemma 2.5)}
\end{array}$$

and this implication-cycle completes the proof. $\square$

4 Examples

As mentioned earlier, there are cases where Strong Skolemization cannot be distinguished from Standard Skolemization. For instance, if a Strong Skolemization step is to be performed on a sub-formula $\exists x \, \Phi$, where Φ is *not* a conjunction, Strong Skolemization in fact degenerates to Standard Skolemization. Also, even if Φ is a conjunction, the Strong Skolemization step might be a mere variant of a corresponding Standard Skolemization step.

In this section we examine some interesting, yet typical, examples which show the effect of Strong Skolemization.

Consider the formula

$$\forall x\,[C(x) \vee \forall y\,\exists z\,(A(y,z) \wedge B(x,z))] \tag{1}$$

Standard Skolemization and clause normal form transformation leads to[5]

$$\begin{array}{l} C(x) \vee A(y, f(x,y)) \\ C(x) \vee B(x, f(x,y)) \end{array}$$

After Strong Skolemization (on $A(y,z)$), however, we obtain the clauses

$$\begin{array}{l} C(x) \vee A(y, f(w,y)) \\ C(x) \vee B(x, f(x,y)) \end{array}$$

Observe the difference between the two Skolemization outcomes. The latter (after Strong Skolemization) is more general than the former, for it has a new variable w where the former has a variable x which occurs somewhere else in the clause. Thus the latter should be preferred. This is particularly valuable in case a $\neg C(t)$ follows from the problem under consideration, because then we are able to derive $A(y, f(w,y))$ which even subsumes the first clause from above, something which would not be possible after Standard Skolemization.

Note that the result of Strong Skolemization (on $A(y,z)$) is identical to what we would achieve if we (standard) skolemized the formula

$$\forall x, y\,\exists z\,[(\forall u\,C(u) \vee A(y,z)) \wedge (C(x) \vee B(x,z))] \tag{2}$$

Indeed, the two formulas (1) and (2) are logically equivalent as the reader might easily check.

But what would happen if we (strongly) skolemized on $B(x,z)$ instead of $A(y,z)$? In this case we would end up with the same result as Standard Skolemization and hence there would be no gain in using Strong Skolemization. This shows how the choice on the conjuncts can influence the final result and that it is by no means irrelevant on which of the conjunctive elements the Strong Skolemization is performed.

Formula (1) is an example in which we have no problems in deciding on which conjunct to perform the Strong Skolemization step. However, if we slightly change (1) to

$$\forall x, y\,[C(x,y) \vee \exists z\,(A(y,z) \wedge B(x,z))] \tag{3}$$

[5] Note that Δ is already in anti-prenex negation normal form.

then it is not any more clear whether we should choose the literal $A(y,z)$ or the literal $B(x,z)$ for the next Strong Skolemization step. In the one case we produce the clauses

$$\begin{array}{l} C(x,y) \vee A(y,f(w,y)) \\ C(x,y) \vee B(x,f(x,y)) \end{array}$$

in the other case we end up with the clauses

$$\begin{array}{l} C(x,y) \vee A(y,g(x,y)) \\ C(x,y) \vee B(x,g(x,w)) \end{array}$$

Both are more general than the Standard Skolemization result, nonetheless the two cannot be compared with each other. For example, the former is more valuable in case we are able to prove $\forall x\, \exists y\, C(y,x)$, the latter is of particular interest if $\forall x\, \exists y\, C(x,y)$ can be shown.

The above example is not artificial. As an interesting instance consider the so-called "density property" of some binary relation R, i.e.,

$$\forall x,y\, [R(x,y) \supset \exists z\, (R(x,z) \wedge R(z,y))] \tag{4}$$

Here, as in the previous example, Strong Skolemization ends up with either

$$\begin{array}{l} \neg R(x,y) \vee R(x,f(x,w)) \\ \neg R(x,y) \vee R(f(x,y),y) \end{array} \quad \text{or} \quad \begin{array}{l} \neg R(x,y) \vee R(x,g(x,y)) \\ \neg R(x,y) \vee R(g(w,y),y) \end{array}$$

The former is particularly interesting if R can be proven to be forward serial, whereas the latter is more important in case backward seriality of R holds. Note that the results achieved from Strong Skolemization are identical to the application of Standard Skolemization on

$$\forall x,y\, \exists z\, [\forall w\, (R(x,w) \supset R(x,z)) \wedge (R(x,y) \supset R(z,y))] \tag{5}$$

and

$$\forall x,y\, \exists z\, [(R(x,y) \supset R(x,z)) \wedge \forall w\, (R(w,y) \supset R(z,y))] \tag{6}$$

respectively. I leave it to the reader to verify that (4), (5), and (6) are pairwise equivalent.

The above examples show that the Strong Skolemization outcome can depend on the choice of the conjunct. Quite often, however, this choice turns out to be irrelevant because the possible cases are somewhat symmetric. One such example can be found in the following formula

$$\forall x,y,z\, [(R(x,y) \wedge R(x,z)) \supset \exists u\, (R(y,u) \wedge R(z,u))] \tag{7}$$

which expresses something like a "confluence" property of the relation R. Standard Skolemization (plus clausification) results in

$$\begin{array}{l}\neg R(x,y) \vee \neg R(x,z) \vee R(y,f(y,z)) \\ \neg R(x,y) \vee \neg R(x,z) \vee R(z,f(y,z))\end{array}$$

whereas Strong Skolemization (on $R(y,u)$) yields

$$\begin{array}{l}\neg R(x,y) \vee \neg R(x,z) \vee R(y,f(y,w)) \\ \neg R(x,y) \vee \neg R(x,z) \vee R(z,f(y,z))\end{array}$$

Here, a condensing step is possible on the literals $\neg R(x,y)$ and $\neg R(x,z)$ with substitution $\sigma = \{z/y\}$ and so we end up with

$$\begin{array}{l}\neg R(x,y) \vee R(y,f(y,w)) \\ \neg R(x,y) \vee \neg R(x,z) \vee R(z,f(y,z))\end{array}$$

Similarly, Strong Skolemization (on $R(z,u)$) leads to (again after condensing)

$$\begin{array}{l}\neg R(x,y) \vee \neg R(x,z) \vee R(y,g(y,z)) \\ \neg R(x,z) \vee R(z,g(w,z))\end{array}$$

In contrast to the earlier examples the two outcomes are mere variants of each other. Nevertheless, both are more general than the Standard Skolemization result.

The above examples all are of the form $\exists x\,(\Phi \wedge \Psi)$. However, the definition of Strong Skolemization also covers the slightly more general cases in which there may be more than just one existentially quantified variable. This more general case *cannot* be simulated by a sequence of Strong Skolemization steps of the above kind as the following example shows. Let

$$\Delta = \forall x\,[C(x) \vee \forall y\,\exists u,v\,(A(y,u,v) \wedge B(x,u,v))] \tag{8}$$

If we first (strongly) skolemized on the sub-formula Δ^u we would obtain

$$\forall x\,[C(x) \vee \forall y\,\exists v\,(A(y,f(x,y),v) \wedge B(x,f(x,y),v))]$$

and so would finally end up with

$$\forall x\,[C(x) \vee \forall y\,(A(y,f(x,y),g(x,y)) \wedge B(x,f(x,y),g(x,y)))]$$

Hence Strong Skolemization would behave as Standard Skolemization in this case. However, according to Definition 3.1, we can treat the two existentially quantified variables u and v simultaneously and so actually achieve the result

$$\forall x\,[C(x) \vee \forall y\,(A(y, f(w, y), g(w, y)) \wedge B(x, f(x, y), g(x, y)))]$$

which subsumes the Standard Skolemization outcome and (after clause normal form transformation) allows for clause splitting.

Summarizing: The effect of Strong Skolemization compared to Standard Skolemization is not that the structure after clausification is changed; rather it lies in the abstraction from some of the arguments of the Skolem functions (resulting in new universally quantified variables).

5 Related Work

There is a somewhat related technique to Strong Skolemization, the so-called Optimized Skolemization, that has been developed by Christoph Weidenbach and Hans Jürgen Ohlbach in (Ohlbach and Weidenbach 1995). Their approach can briefly be described as follows:
Let Δ be a sentence with sub-formula $\exists x\,(\Phi \wedge \Psi)$. Moreover, suppose that we can prove that $\Delta \models \forall y_1, \ldots, y_n \exists x\,\Phi$ where $\{y_1, \ldots, y_n\} = \mathcal{V}(\exists x\,(\Phi \wedge \Psi))$. Then

$$\begin{gathered}\Delta[\exists x\,(\Phi \wedge \Psi)] \text{ is satisfiable} \\ \text{iff} \\ \Delta[\exists x\,(\Phi \wedge \Psi)\,/\,\Psi\{x/f(\overline{y}_n)\}] \wedge \forall \overline{y}_n\,\Phi\{x/f(\overline{y}_n)\} \text{ is satisfiable}\end{gathered}$$

where f is a new (Skolem) function symbol.
As an example consider again the density property for R, i.e., $\forall u, v\,[R(u, v) \supset \exists w\,(R(u, w) \wedge R(w, v))]$. In case we know (or can easily prove) that the binary relation R is serial, i.e., $\forall x \exists y\,R(x, y)$, the Optimized Skolemization of the density property results in the clauses

$$\begin{aligned}&R(x, f(x, y))\\ &\neg R(u, v) \vee R(f(u, v), v)\end{aligned}$$

The very same two clauses could be obtained by Standard Skolemization from

$$\forall x, y\,\exists z\,[R(x, z) \wedge (R(x, y) \supset R(z, y))] \tag{9}$$

We leave it to the reader to verify that indeed formula (9) is logically equivalent to the conjunction of the formulas for seriality and density of R.

Now let us compare this with the method proposed in this paper. Recall that Strong Skolemization on the literal $R(u, w)$ yields the clauses

$$\begin{array}{l}\neg R(u,v) \vee R(u, f(u,w)) \\ \neg R(u,v) \vee R(f(u,v), v)\end{array}$$

Moreover, suppose that there is a further clause, say $R(x, g(x))$, which states the seriality of R. Then a resolution step with the first clause from above yields the unit clause $R(u, f(u, w))$ which even subsumes its two-literal parent clause. Thus, we finally ended up with the same outcome as in the case of Optimized Skolemization, i.e., Optimized Skolemization has in a sense been simulated by Strong Skolemization and resolution.

The above shows some similarity between Strong Skolemization and Optimized Skolemization. However, the similarities are not always as strong as the above example might suggest.

Consider again formula (7) and assume that we are able to prove that $\forall y \exists u\, R(y, u)$, i.e., R is serial. Optimized Skolemization then results in the clauses

$$\begin{array}{l}R(y, f(y,z)) \\ \neg R(x,y) \vee \neg R(x,z) \vee R(z, f(y,z))\end{array}$$

whereas Strong Skolemization on $R(y, u)$ leads to

$$\begin{array}{l}\neg R(x,y) \vee R(y, f(y,w)) \\ \neg R(x,y) \vee \neg R(x,z) \vee R(z, f(y,z))\end{array}$$

Note that – in contrast to the "density example" – the seriality of R does not help here to simulate Optimized Skolemization, although, somewhat surprisingly, backward seriality does; and after all, seriality follows from backward seriality in the example under consideration (although not the other way round). A major reason for this is, that Optimized Skolemization takes the whole problem into account and not only the sub-problem in which the subformula to be skolemized occurs. In other words, Optimized Skolemization works globally on the problem under consideration whereas Strong Skolemization applies only locally to the various sub-problems. This leads us to the conclusion that, if Optimized Skolemization applies at all, then it produces a more general result than Strong Skolemization and thus in particular a more general result than Standard Skolemization. This fact might indicate that Optimized Skolemization should always be performed, however, there still is the restriction that it requires the provability of a certain intermediate conjecture (seriality of R in the examples above). Hence, if automated, Optimized Skolemization itself involves the application of theorem provers, so that clause normal form generation can become a significantly complicated procedure. In fact, such proofs might be arbitrary complicated and since such desired requirements cannot be decided (they are only semi-decidable, just as first-order theorem proving is in general) some restrictions are to be imposed on the intermediate prover, be it by time limits or by a bounded number of inference steps (see (Weidenbach, Gaede and Rock 1996) for a description of FLOTTER, an implementation of Optimized Skolemization). In case the desired property cannot be proved – regardless of whether it simply does not hold or the given restrictions prevent a proof to be found – Optimized Skolemization performs a Standard Skolemization step and thus shows no advantage at all. These are situations where Strong Skolemization would nicely come into play, for then its superiority over Standard Skolemization could show its significance.

Another slight difference between Strong and Optimized Skolemization can be seen in the kind of sub-formulas they are applied to. Recall that Strong Skolemization applies to sub-formulas of the form $\exists \overline{x}_k \, (\Phi \wedge \Psi)$, where k may be any natural number, whereas Optimized Skolemization requires sub-formulas of the form $\exists x \, (\Phi \wedge \Psi)$, i.e., it considers only single existentially quantified variables. As an example consider again formula (8)

$$\forall x \, [C(x) \vee \forall y \, \exists u, v \, (A(y, u, v) \wedge B(x, u, v))]$$

In a first step Optimized Skolemization would eliminate the variable u, yielding

$$\forall x \, [C(x) \vee \forall y \, \exists v \, (A(y, f(x, y), v) \wedge B(x, f(x, y), v))]$$

After that there is an attempt to prove $\forall x, y \, \exists v \, A(y, f(x, y), v)$, and, if this does not succeed, it is tried to show that $\forall x, y \, \exists v \, B(x, f(x, y), v)$. Of course, it might very well be that neither attempt succeeds, and so we would in fact finally end up with the standard result. But, actually, it would have sufficed to test the much simpler $\forall y \exists u, v \, A(y, u, v)$ (or $\forall x \exists u, v \, B(x, u, v)$ respectively). Unfortunately, this possibility is not given in the definition of Optimized Skolemization described in (Ohlbach and Weidenbach 1995). Nonetheless it is absolutely correct and we fix this in the following theorem.

Theorem 5.1 (improved optimized skolemization)
Let $\Delta[\exists\overline{x}_k\,(\Phi \wedge \Psi)]$ *be a sentence in anti-prenex negation normal form and, moreover, assume that* $\Delta \models \forall\overline{y}_n\,\exists\overline{x}_k\,\Phi$ *where* $\{y_1, \ldots, y_n\} = \mathcal{V}(\exists\overline{x}_k\,(\Phi \wedge \Psi))$. *Then*

$$\Delta \text{ is satisfiable}$$
$$\text{iff}$$
$$\forall\overline{y}_n\,\Phi\{x_i/f_i(\overline{y}_n)\} \;\wedge \Delta[\exists\overline{x}_k\,(\Phi \wedge \Psi)/\Psi\{x_i/f_i(\overline{y}_n)\}] \text{ is satisfiable}$$

where $f_1, \ldots, f_k$ *are new (Skolem) function symbols.*

Proof: The proof presented in (Ohlbach and Weidenbach 1995) for the single variable case can easily be extended to the k variable case. □

$\mathcal{ALC}$ Experiments

Typical situations where Optimized Skolemization turns out to be of almost no use at all are ones in which the intermediate proofs that have to be performed during the overall Skolemization process are numerous and/or difficult. In such cases the whole reasoning process might get stuck already within the clause normal form generation.

In order to compare Strong Skolemization with Standard Skolemization (and that without ignoring the possibilities of Optimized Skolemization) we decided to run several rather huge examples where FLOTTER and thus Optimized Skolemization, in its current implementation[6] has almost no chance to terminate within reasonable time. The problems we considered come from the area of knowledge representation and are described in the terminological language $\mathcal{ALC}$ (see (Schmidt-Schauß and Smolka 1991)) which is essentially a multi-modal logic K and thus can be translated into a decidable fragment of first-order predicate logic. I do not want to go into detail what this language is concerned. What is interesting about it, however, is that, after performing a so-called relational translation into first-order predicate logic, we are faced with formulas that contain sub-formulas of the kind $\exists x\,[R(y, x) \wedge \Phi[x]]$ where R is a binary predicate symbol and $\Phi[x]$ is meant to represent a complicated formula with x as its only free variable. When faced with such sub-formulas, Optimized Skolemization first tries to prove that $\forall y \exists x\, R(y, x)$ follows from the problem under consideration, and, if this does not succeed, attempts to show $\exists x\,\Phi[x]$. However, it is usually not the case in $\mathcal{ALC}$ problems that

[6] We ran FLOTTER version 0.42

$\forall y \exists x\, R(y,x)$ holds and therefore Optimized Skolemization will usually have to deal with the very complicated problem of showing that $\exists x\, \Phi[x]$. Evidently, the bigger the sub-formula Φ the more expensive is such an attempt. In fact, typical $\mathcal{ALC}$ files contain several hundreds of such sub-formulas and so Optimized Skolemization, at least in its current implementation, quite easily gets lost.

Strong Skolemization, however, works nicely on these kinds of problems. Why this is so, is justified by the following observation.
Let Δ be a sentence with sub-formula $\exists x\,(\Phi \wedge \Psi)$ such that x is the only free variable in Φ and let $\{y_1, \ldots, y_n\} = \bigcup_{z \in \mathcal{V}(\Psi) \setminus \{x\}} Dep_\Delta(z)$. Then Strong Skolemization transforms $\Delta[\exists x\,(\Phi \wedge \Psi)]$ into

$$\Delta[\exists x\,(\Phi \wedge \Psi)\,/\,\forall \overline{w}_n\, \Phi\{x/f(\overline{w}_n)\} \wedge \Psi\{x/f(\overline{y}_n)\}]$$

Note that the sub-formula $\forall \overline{w}_n\, \Phi\{x/f(\overline{w}_n)\}$ obtained this way has no free variables at all, something which is particularly important in case the theorem prover at hand has a built-in splitting rule.

In order to compare Standard and Strong Skolemization we ran about 100 $\mathcal{ALC}$ examples and it showed that in more than 50% there was a significantly better behavior of Strong Skolemization[7]. In about one fourth of the examples, Strong Skolemization made it at all possible for the theorem prover to stop its reasoning process (be it with a proof, or with a saturated set of clauses) so that we can conclude that Strong Skolemization shows a significant improvement over Standard Skolemization.

Summarizing: Optimized Skolemization has the advantage that it produces better results than any other Skolemization technique, provided it can at all be applied. Its disadvantages are, first, that it requires a theorem prover on its own to perform the Skolemization, and second, that it performs Standard Skolemization steps whenever the theorem prover is not able to solve the intermediate goals. Strong Skolemization comes into play whenever either the user (or implementer) of automated theorem provers does not want to spend that much effort for the clause normal form generation, or it is likely that Optimized Skolemization gets stuck because the sub-formulas to be considered are too complicated (or there are too many of them), or, otherwise, if none of the intermediate goals is provable from the clause set.

[7] Sure, there are specially designed inference mechanisms and translation techniques for such kind of logics and it would be nonsense to compare such a general framework as the Strong Skolemization with these special methods. But this is not our purpose anyway.

Hence the two techniques Optimized Skolemization and Strong Skolemization can nicely be combined. I.e., whenever applicable we should perform an Optimized Skolemization step. In all other cases a Strong Skolemization step is to be preferred over Standard Skolemization.

References

Andrews, P. B.: 1981, Theorem Proving via General Mating, *Journal of the Association for Computing Machinery* **28**(2), 193–214.

Chang, C.-L. and Lee, R. C.: 1973, *Symbolic Logic and Mechanical Theorem Proving*, Computer Science and Applied Mathematics Series, Academic Press, New York.

Egly, U.: 1994, On the Value of Antiprenexing, *Proceedings of the LPAR'94*, Springer, LNAI 822.

Loveland, D.: 1978, *Automated Theorem Proving: A Logical Basis*, Fundamental Studies in Computer Science, North-Holland.

Ohlbach, H. J. and Weidenbach, C.: 1995, A Note on Assumptions about Skolem Functions, *Journal of Automated Reasoning* **15**(2), 267–275.

Schmidt-Schauß, M. and Smolka, G.: 1991, Attributive Concept Description with Complements, *Artificial Intelligence* **48**, 1–26.

Skolem, T.: 1920, Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze, nebst einem Theoreme über dichte Mengen, *Skrifter utgit av Videnskappsellkapet i Kristiania* **4**, 4–36. see also: (Skolem 1976).

Skolem, T.: 1976, Logico-combinatorial Investigations in the Satisfiability and Provability of Mathematical Propositions, *in* J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Harvard University Press.

Weidenbach, C., Gaede, B. and Rock, G.: 1996, SPASS & FLOTTER Version 0.42, *Proceedings of the 13th International Conference on Automated Deduction, CADE 13*, Springer, LNAI 1104, pp. 141–145.

I N F O R M A T I K

Below you find a list of the most recent technical reports of the research group *Logic of Programming* at the Max-Planck-Institut für Informatik. They are available by anonymous ftp from our ftp server `ftp.mpi-sb.mpg.de` under the directory `pub/papers/reports`. Most of the reports are also accessible via WWW using the URL `http://www.mpi-sb.mpg.de`. If you have any questions concerning ftp or WWW access, please contact `reports@mpi-sb.mpg.de`. Paper copies (which are not necessarily free of charge) can be ordered either by regular mail or by e-mail at the address below.

Max-Planck-Institut für Informatik
Library
attn. Birgit Hofmann
Im Stadtwald
D-66123 Saarbrücken
GERMANY
e-mail: `library@mpi-sb.mpg.de`

MPI-I-96-2-009	D. Basin, N. Klarlund	Beyond the Finite in Automatic Hardware Verification
MPI-I-96-2-007	A. Herzig	SCAN and Systems of Conditional Logic
MPI-I-96-2-006	D. Basin, S. Matthews, L. Vigano	Natural Deduction for Non–Classical Logics
MPI-I-96-2-005	A. Nonnengart	Auxiliary Modal Operators and the Characterization of Modal Frames
MPI-I-96-2-004	G. Struth	Non-Symmetric Rewriting
MPI-I-96-2-003	H. Baumeister	Using Algebraic Specification Languages for Model-Oriented Specifications
MPI-I-96-2-002	D. Basin, S. Matthews, L. Vigano	Labelled Propositional Modal Logics: Theory and Practice
MPI-I-96-2-001	H. Ganzinger, U. Waldmann	Theorem Proving in Cancellative Abelian Monoids
MPI-I-95-2-011	P. Barth, A. Bockmayr	Modelling Mixed-Integer Optimisation Problems in Constraint Logic Programming
MPI-I-95-2-010	D. A. Plaisted	Special Cases and Substitutes for Rigid E-Unification
MPI-I-95-2-009	L. Bachmair, H. Ganzinger	Ordered Chaining Calculi for First-Order Theories of Binary Relations
MPI-I-95-2-008	H. J. Ohlbach, R. A.Schmidt, U. Hustadt	Translating Graded Modalities into Predicate Logic
MPI-I-95-2-007	A. Nonnengart, A. Szalas	A Fixpoint Approach to Second-Order Quantifier Elimination with Applications to Correspondence Theory
MPI-I-95-2-006	D. Basin, H. Ganzinger	Automated Complexity Analysis Based on Ordered Resolution
MPI-I-95-2-005	F. Baader, H. J. Ohlbach	A Multi-Dimensional Terminological Knowledge Representation Language
MPI-I-95-2-003	P. Barth	A Davis-Putnam Based Enumeration Algorithm for Linear Pseudo-Boolean Optimization
MPI-I-95-2-002	H. J. Ohlbach, R. A. Schmidt	Functional Translation and Second-Order Frame Properties of Modal Logics