Skolemization Modulo Theories

Konstantin Korovin[1] and Margus Veanes[2]

[1] The University of Manchester
korovin@cs.man.ac.uk
[2] Microsoft Research
margus@microsoft.com

Abstract. Combining classical automated theorem proving techniques with theory based reasoning, such as satisfiability modulo theories, is a new approach to first-order reasoning modulo theories. Skolemization is a classical technique used to transform first-order formulas into *equisatisfiable* form. We show how Skolemization can benefit from a new satisfiability modulo theories based simplification technique of formulas called *monadic decomposition*. The technique can be used to transform a theory dependent formula over multiple variables into an equivalent form as a Boolean combination of unary formulas, where a unary formula depends on a single variable. In this way, theory specific variable dependencies can be eliminated and consequently, Skolemization can be refined by minimizing variable scopes in the decomposed formula in order to yield simpler Skolem terms.

1 The role of Skolemization

In classical automated theorem proving, Skolemization [9, 2, 4, 8] is a technique used to transform formulas into *equisatisfiable* form by replacing existentially quantified variables by Skolem terms. In resolution based methods using clausal normal form (CNF) this is a necessary preprocessing step of the input formula. A CNF represents a universally quantified conjunction of clauses, where each clause is a disjunction of literals, a literal being an atom or a negated atom. The arguments of the atoms are terms, some of which may contain Skolem terms as subterms where a Skolem term has the form $f(\bar{x})$ for some Skolem function symbol f and a sequence $\bar{x}$ of variables; f may also be a constant ($\bar{x}$ is empty). The input to Skolemization is a formula ψ in *prenex normal form*: $Q_1x_1Q_2x_2\ldots Q_nx_n\varphi$ where $Q_i \in \{\exists, \forall\}$ and φ is quantifier free and the free variables of φ, $FV(\varphi)$, form a subset of $\{x_i\}_{i=1}^n$; φ is called the *matrix* of ψ. In its most basic form, one Skolemization step *Skolemize1* is a transformation that is applied to the outermost prefix of the given prenex formula, $\mathit{Skolemize1}(\forall\bar{x}\,\exists y\,\chi(\bar{x}, y)) \stackrel{\text{def}}{=} \forall\bar{x}\,\chi(\bar{x}, f_{\chi,\bar{x}}(\bar{x}))$, whose output is another prenex formula with one less existential quantifier and where $f_{\chi,\bar{x}}$ is a new function symbol called a *Skolem function* (or a *Skolem constant* when $n = 0$). *Skolemize1* is applied repeatedly, denoted here by *Skolemize*, until no more existential quantifiers remain. *Skolem Normal Form Theorem* [4, Corollary 3.1.3] implies that *Skolemize*(ψ) is satisfiable if and only if ψ is satisfiable.

In the context of theorem proving it is assumed that the Skolem functions are uninterpreted.

There are several important techniques related to Skolemization. The main objective is to *minimize the arity* of the Skolem functions. *Mini-scoping* [1], also called *antiprenexing* or creating an *antiprenex normal form* [2, 8], is the main Skolemization technique that is used in theorem proving as a method to minimize quantifier scopes by shifting quantifiers from the prenex back into the formula. Mini-scoping can be seen as a separate preprocessing step prior to Skolemization, consisting of the following rewrite steps that correspond to standard equivalence preserving laws of logic. The formula is first transformed into an equivalent negation normal form (NNF), so that all quantifiers occur in a positive context.

$$
\begin{array}{lll}
Qx(\varphi \diamond \psi) & \overset{x\notin FV(\psi), x\in FV(\varphi)}{\Longrightarrow_{\mathit{mini\text{-}scope}}} & Qx\varphi \diamond \psi, \quad \diamond \in \{\vee, \wedge\} \\
\forall x(\varphi \wedge \psi) & \Longrightarrow_{\mathit{mini\text{-}scope}} & \forall x\varphi \wedge \forall x\psi \\
\exists x(\varphi \vee \psi) & \Longrightarrow_{\mathit{mini\text{-}scope}} & \exists x\varphi \vee \exists x\psi
\end{array}
$$

After mini-scoping, all quantified variables are renamed apart. Finally, *standard Skolemization* [2] is applied to the resulting formula by replacing a subformula $\exists y\chi$ that occurs in the context of universal variables $\bar{x}$, by the formula $\chi\{y \mapsto f(\bar{x})\}$. We refer to the full procedure as mini-scoping. Without theory based reasoning, mini-scoping results in the lowest possible arities of the Skolem functions, and is thus optimal in that sense. In theory based reasoning this is not always true, one can do better, we discuss this below.

2 Working modulo a theory

Theory based automated reasoning is a new area of automated reasoning that combines techniques from propositional satisfiability (SAT) and satisfiability modulo theories (SMT) area into the expressive power of first-order reasoning with quantifiers [5, 6]. Skolemization is one piece of the big picture, it has been considered as a solved problem, much due to the *Skolemization Theorem* [4, Theorem 3.1.2]. Skolemization Theorem implies a much stronger property than equisatisfiability, that allows the use of Skolemization modulo *arbitrary* theories.

Let L be the language of the theory and let Σ be the *Skolem Theory* consisting of axioms $\forall\bar{x}(\exists y\,\chi(\bar{x}, y) \rightarrow \chi(\bar{x}, f_{\chi,\bar{x}}(\bar{x})))$ for all L-formulas $\chi(\bar{x}, y)$ and new function symbols $f_{\chi,\bar{x}}$. In other words, the Skolem theory axiomatizes the intended interpretations of the Skolem functions. Skolemization Theorem says that any L-structure A can be expanded to be a model A^{Σ} of Σ. Therefore, if we work with uninterpreted function symbols, i.e., without assuming Σ, and A is a model of the original formula, then some expansion of A models the Skolemized one: just pick the intended interpretations from A^{Σ} for the uninterpreted Skolem functions. In the other direction, the Skolemized formula always entails the original formula. In particular if the Skolemized formula is satisfiable then so is the original one.

Often the starting point in theory based reasoning is a formula which presumes Skolemization. For example, assume the theory of integer linear arithmetic and consider the following (true) sentence:

$$\forall x \exists y (0 \leq x \leq 1 \implies (0 \leq y \wedge x + y \leq 1)) \tag{1}$$

It is already in prenex form and mini-scoping produces the equisatisfiable formula where f is a Skolem function:

$$\forall x (0 \leq x \leq 1 \implies (0 \leq f(x) \wedge x + f(x) \leq 1)) \tag{2}$$

We will see below how introduction of f can be avoided completely in this case.

3 Using monadic decomposition

We consider theories that satisfy the following conditions. More general theories fall outside the scope of this paper. Let A be a recursively enumerable (re) L-structure with an (re universe) so that all elements can be named by L-terms. As the theory we take the theory of A.

Moreover, let Ψ be an re set of formulas that is closed under Boolean operations, and if a is an element, x a variable, and $\psi \in \Psi$ then $\psi\{x \mapsto a\} \in \Psi$. Furthermore, satisfiability of formulas in Ψ is assumed *decidable*: it is decidable, for $\psi(\bar{x}) \in \Psi$, if $A \models \exists \bar{x} \psi(\bar{x})$. It follows from A being re that concrete witnesses can also be generated for satisfiable formulas. Examples of A are: standard integers or standard rational numbers (or A may be multi-sorted), and an example of Ψ is quantifier free L-formulas where all variables have a fixed sort. These conditions are very natural from the standpoint of modern SMT solvers, because Ψ embodies the basic properties supported by any state-of-the-art SMT solver [3].

We need some additional notions before defining monadic decomposition formally. A *unary* formula is a formula with at most one free variable. An *explicitly monadic* formula is a Boolean combination of unary formulas. A *monadic* formula is a formula for which there exists an *equivalent* explicitly monadic formula. Now, *monadic decomposition* (for Ψ) is the following problem: given a monadic formula $\psi \in \Psi$, construct and explicitly monadic formula that is equivalent to ψ. It is shown in [10] that this problem is solvable, the given algorithm ***mondec*** relies solely on the assumptions of Ψ as stated above. *Deciding* if a formula is monadic is shown decidable in two cases but is an open problem in general.

Now, monadic decomposition can be applied as a preprocessing step to mini-scoping. This can happen in several different ways. First, several variables can be grouped together and viewed as a single variable by using tuples; the structure A as well as Ψ can, without loss of gererality, be extended with tuples. Second, the decomposition can be applied selectively to some subformulas only. Finally, if the formula is not known to be monadic then ***mondec*** might not terminate and thus heuristics need to be developed to decide when to abandon the decomposition attempt.

As an example assume that the theory is linear arithmetic and pick the matrix of the prenex formula (1). This formula is monadic (which can also be decided [10] with a Presburger formula but there is currently no particular implementation for this decision procedure). If we apply ***mondec*** to the matrix of (1) we get the following concrete output as the result of running the python script from [10]:

```
And(Or(Not(And(x >= 0, x <= 1)), x <= 1),
    Or(Not(And(x >= 0, x <= 1)), x <= -1,
       And(y >= 0, y <= 1, Or(Not(And(x >= 0, x <= 1)), x <= 0)),
       And(y >= 0, y <= 0, Or(Not(And(x >= 0, x <= 1)), x <= 1))))
```

This formula is explicitly monadic and equivalent to the matrix of (1). Thus, we can replace the matrix of (1) by this formula. Now mini-scoping will produce a formula where y is not in the scope of x any more. So the final Skolemized formula will use *Skolem constants* for y.

In general, if all maximal quantifier free subformulas are monadic (as is the case with formula (1)) and mini-scoping is slightly modified and applied so that quantifiers are pushed all the way to the unary sub-formulas, then the quantifiers are effectively eliminated and the final formula will be a Boolean combination of $\forall x\varphi(x)$ or $\exists x\varphi(x)$ where $\varphi(x)$ is a formula in Ψ and thus decidable. So the final formula is essentially propositional (modulo the theory of A). Overall, this implies that the full first-order fragment over monadic formulas is decidable, as an extension of of the Löwenheim class [7].

References

1. P. B. Andrews. *An Introduction to Mathematical Logic and Type Theory: To Truth through Proof.* Academic Press, 1986.
2. M. Baaz, U. Egly, and A. Leitsch. Normal form transformations. In A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*, volume I, chapter 5, pages 273–333. North Holland, 2001.
3. L. De Moura and N. Bjørner. Satisfiability modulo theories: introduction and applications. *Commun. ACM*, 54(9):69–77, 2011.
4. W. Hodges. *Model theory.* Cambridge Univ. Press, 1995.
5. K. Korovin. Instantiation-based automated reasoning: From theory to practice. In *Automated Deduction - CADE-22, 22nd International Conference on Automated Deduction*, volume 5663 of *LNAI*, pages 2–7. Springer, 2009.
6. K. Korovin. Inst-gen – a modular approach to instantiation-based automated reasoning. In A. Voronkov and C. Weidenbach, editors, *Programming Logics, Essays in Memory of Harald Ganzinger*, volume 7797 of *LNCS*, pages 239–270. Springer, 2013.
7. L. Löwenheim. Über Möglichkeiten im Relativkalkül. *Math. Annalen*, 76:447–470, 1915.
8. A. Nonnengart and C. Weidenbach. Computing small clause normal forms. In A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*, volume I, chapter 6, pages 335–367. North Holland, 2001.
9. T. Skolem. *Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme ber dichte Mengen.* Skrifter utgitt av Videnskapsselskapet i Kristiania. 1920.
10. M. Veanes, N. Bjørner, L. Nachmanson, and S. Bereg. Monadic decomposition. In *CAV'14*, volume 8559 of *LNCS*. Springer, 2014.