Expressiveness and Complexity of Subclasses of Quantified Boolean Formulas

Uwe Bubeck[1], Hans Kleine Büning[2], Anja Remshagen[3], and Xishun Zhao[4]

[1] University of Paderborn, Paderborn (Germany), bubeck@upb.de
[2] University of Paderborn, Paderborn (Germany), kbcsl@upb.de
[3] University of West Georgia, Carrollton (USA), anja@westga.edu
[4] Sun Yat-sen University, Guangzhou (PR China), hsszxs@mail.sysu.edu.cn

Abstract. We give a brief overview of expressiveness and complexity results for a hierarchy of subclasses of quantified Boolean formulas with close connections to Boolean circuits and minimal unsatisfiability.

1 Introduction

Quantified Boolean formulas (*QBF*) generalize propositional formulas by allowing variables to be quantified universally or existentially. We also allow free variables which are not quantified and indicate this with a star (QBF^*). An interesting property of QBF^* formulas is that it is possible to define an equivalence between such formulas and propositional formulas. We say that $\Phi \in QBF^*$ is equivalent to $\psi \in PROP$ ($\Phi \approx \psi$) if and only if the free variables in Φ correspond to the propositional variables in ψ and both formulas have the same truth value for each assignment to the free/propositional variables. This means that quantified variables inside of Φ are not taken into consideration here, so these can be thought of as local or auxiliary variables. An important application of auxiliary variables is to introduce abbreviations for repeating parts in a given formula, such as multiple copies of transition or reachability relations in verification problems [5, 8]. Accordingly, QBF^* encodings are often much more compact than their propositional equivalents. Many problems also have a natural forall-exists semantics which can elegantly be modeled by quantifiers.

Unfortunately, quantified Boolean formulas appear to be much harder to solve than propositional formulas, with *QBF* and QBF^* satisfiability being *PSPACE*-complete. This makes it worthwhile to investigate subclasses with a lower decision complexity. In particular, we are interested to find out how restrictions on the formula structure affect the expressiveness of the corresponding classes of formulas, that means the ability to provide short encodings of propositional formulas. Such results can help to find suitable encodings with a good balance of compactness and complexity, and they may also be beneficial to *QSAT* solvers by allowing them to deal more efficiently with subproblems of restricted structure.

Well-known restrictions in propositional logic are the Horn property of having at most one literal per clause and the class of minimal unsatisfiable formulas, that means unsatisfiable *CNF* formulas which become satisfiable when removing any of the clauses. In the context of quantified Boolean formulas, various interesting subclasses arise depending on which kinds of variables these restrictions are applied to, e.g. to all variables or only to the quantified variables. In the following, we give a brief overview of expressiveness and complexity results for these classes of quantified Horn and minimal unsatisfiable formulas and present some close connections to Boolean circuits. For a more detailed discussion and the proofs which have been omitted here, we refer the reader to the original publications [3, 9, 11], from which the following results have been taken.

2 Horn Formulas and Circuits

When lifting the Horn property to quantified formulas, the easiest case is that all literals have to respect the Horn property, such that the matrix of the formula is a propositional Horn formula (*HORN*). These *QHORN** formulas can easily be solved in at most quadratic time by Q-unit resolution [7]. Another idea is to enforce the Horn property only on the literals over quantified variables. For a clause ϕ_i, we write $\phi_i = \phi_i^b \vee \phi_i^f$ where the *bound part* ϕ_i^b contains all literals over quantified variables in the clause, and the *free part* ϕ_i^f contains all free literals. Then $QHORN^b$ is the class of formulas $\Phi = Q \bigwedge_i (\phi_i^b \vee \phi_i^f)$ for which $\bigwedge_i \phi_i^b \in HORN$. By universal quantifier elimination, it can be shown that these formulas have an NP-complete satisfiability problem [12]. Finally, *QEHORN** is the class of quantified Boolean formulas with Horn clauses after removing all universal literals, so the Horn property affects only existential and free variables. These formulas appear to be very difficult to solve: even for a fixed number of quantifier alternations, the corresponding decision problem is *coNP*-complete [7].

It is known that there are $\exists HORN^b$ formulas for which every equivalent propositional *CNF* formula has exponential length. On the other hand, any propositional non-*CNF* formula can be transformed into a $QHORN^b$ formula of linear length with only existential quantifiers ($\exists HORN^b$), e.g. by the well-known Tseitin transformation $\pi \vee (\phi \wedge \psi) \approx \exists x\ (\pi \vee \neg x) \wedge (\phi \vee x) \wedge (\psi \vee x)$ [14]. Interestingly, a linear transformation from propositional non-*CNF* to $\exists HORN^b$ is still possible if we allow at most two quantified literals per clause ($\exists 2\text{-}HORN^b$).

Theorem 1. ([3]) *Every propositional formula Ψ can be transformed into an equivalent $\exists 2\text{-}HORN^b$ formula Φ_G of linear length.*

The transformation is based on a representation of propositional formulas as series-parallel graphs. We label their edges with propositional formulas and their nodes with new auxiliary variables. For a given input formula, the graph is constructed by successively replacing conjunctions with parallel connections and disjunctions with serial connections, as shown in Figure 1.

Afterwards, an $\exists 2\text{-}HORN^b$ formula is extracted from the graph by creating for each edge $u \to w$ with label α a clause $(\neg u \vee w \vee \alpha)$. For the source x and the sink y, additional unit clauses (x) and $(\neg y)$ are added, and all auxiliary variables

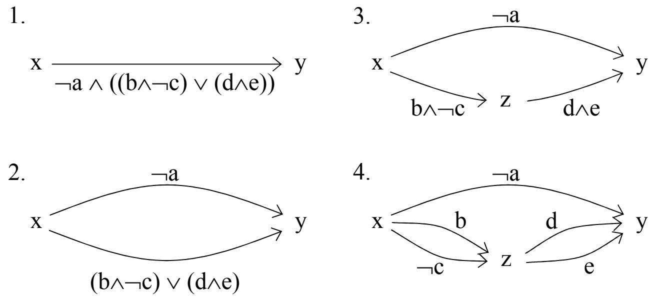

Fig. 1: Construction of a PS-graph for $\psi = \neg a \wedge ((b \wedge \neg c) \vee (d \wedge e))$

are quantified existentially. For the example from Figure 1, we obtain:

$$\begin{aligned}\Phi_G = \exists x \exists y \exists z\ x \wedge \neg y \wedge (\neg x \vee y \vee \neg a) \wedge (\neg x \vee z \vee b) \wedge (\neg x \vee z \vee \neg c) \\ \wedge(\neg z \vee y \vee d) \wedge (\neg z \vee y \vee e)\end{aligned}$$

A remaining open question is whether $\exists HORN^b$ (or even $\exists 2\text{-}HORN^b$) is exponentially more concise than propositional formulas without the *CNF* restriction. One possibility to approach this question might be by considering circuits.

A Boolean circuit c is a directed acyclic graph with one outgoing edge. A node represents either an input or a $\wedge$-gate (AND), $\vee$-gate (OR), or $\neg$-gate (NOT). The input nodes do not have ingoing edges. The nodes representing gates have at most two ingoing edges. A circuit represents a Boolean function, and for any assignment of truth values to the input variables z_i, the circuit returns the corresponding truth value. The size of a circuit is the number of gates. We also define an equivalence between circuits and formulas. A circuit c is equivalent to a formula α if and only if for every truth assignment to the free variables (input variables resp.) both return the same truth value.

Clearly, propositional formulas correspond to circuits in which every gate has exactly one outgoing edge (we say that the maximum fan-out is 1). It can be shown that $\exists HORN^b$ formulas have equivalent circuits of polynomial length when gates may have more than one outgoing edge (i.e. fan-out greater than 1), and vice versa [1, 12]. It is widely assumed that circuits with fan-out greater than 1 are exponentially more powerful than circuits with fan-out 1. In the following section, the transformation from $\exists HORN^b$ to circuits will be extended to more powerful classes of existentially quantified Boolean formulas.

3 Minimal Unsatisfiability

We first recall the definition of deficiency for *CNF* and quantified *CNF* ($QCNF^*$). Let φ be a *CNF* formula over n variables with $n + k$ clauses, then we say k is the *deficiency* of φ. For the deficiency of a formula φ, we write $d(\varphi)$.

Definition 1. *For $\Phi = Q\varphi \in$ QCNF*, the deficiency of Φ, denoted $d(\Phi)$, is the difference between the number of clauses and the number of* existential or free *variables.*
The maximal deficiency of Φ is defined as $d^*(\Phi) := \max\{d(\Phi') \mid \Phi' \subseteq \Phi\}$.

A formula $\phi_1 \wedge \cdots \wedge \phi_q$ in *CNF* is called minimal unsatisfiable if the formula is unsatisfiable and for any clause ϕ_i, the formula $\phi_1 \wedge \cdots \wedge \phi_{i-1} \wedge \phi_{i+1} \wedge \cdots \wedge \phi_q$ is satisfiable. The class of minimal unsatisfiable formulas is denoted by *MU*.

Analogously, a formula $Q(\phi_1 \wedge \cdots \wedge \phi_q)$ in *QCNF** is called minimal false if the formula is false and for any clause ϕ_i, the formula $Q(\phi_1 \wedge \cdots \wedge \phi_{i-1} \wedge \phi_{i+1} \wedge \cdots \wedge \phi_q)$ is true. The class of minimal false formulas is denoted by *MF*.

Definition 2. *Let k be fixed. Then we define*
MU(k) $:= \{\varphi : \varphi \in MU \text{ and } d(\varphi) = k\}$, *and*
MF(k) $:= \{\Phi : \Phi \in MF \text{ and } d(\Phi) = k\}$.

The class *MU* is D^P-complete [13], whereas *MF* is *PSPACE*-complete [10]. Any minimal unsatisfiable formula has deficiency greater than 0 [2]. Moreover, it has been shown that $MU(k)$ is solvable in polynomial time [6].

Theorem 2. ([11]) *MF*(1) *can be solved in polynomial time.*

Theorem 3. ([11]) *Let k be fixed.*

1. *The satisfiability problem for QCNF* with maximal deficiency k is in NP.*
2. *The minimal falsity problem MF(k) for QCNF* is in D^P.*
3. *The satisfiability problem for formulas in QEHORN* and QE2-CNF* with maximal deficiency k is solvable in polynomial time,*
4. *The minimal falsity problem for formulas in QEHORN* and QE2-CNF* with deficiency k can be decided in polynomial time.*

We say that a *CNF* formula ϕ is in $MU^*(k)$ if every minimal unsatisfiable subset of clauses of ϕ has at most deficiency k. Note that a satisfiable subset may have an arbitrary deficiency and that every satisfiable *CNF* formula is in $MU^*(k)$ for every integer k.

An interesting property of *QCNF** formulas is that the minimal false subsets of the bound parts of the clauses determine which combinations of the free parts must be true.

Lemma 1. ([4]) *Let $\Phi = Q \bigwedge_{1 \le i \le q} (\phi_i^b \vee \phi_i^f)$ be a QCNF* formula with non-empty bound parts ϕ_i^b and free parts ϕ_i^f. Let*

$$S(\Phi) := \left\{\Phi' \mid \Phi' = Q\phi_{i_1}^b \wedge ... \wedge \phi_{i_r}^b \text{ is minimal false}, 1 \le i_1, ..., i_r \le q\right\}$$

be the set of minimal false subformulas of the quantified bound parts of Φ. Then we have the following equivalence:

$$\Phi \approx \bigwedge_{(Q\phi_{i_1}^b \wedge ... \wedge \phi_{i_r}^b) \in S(\Phi)} (\phi_{i_1}^f \vee \ldots \vee \phi_{i_r}^f)$$

This motivates us to consider formulas in which only the bound parts of clauses are in $MU^*(k)$. To keep things simple, we only allow existential quantifiers.

$$\exists MU^*(k)^b := \left\{ \Phi \mid \Phi = \exists x_1 ... \exists x_n \phi_1 \wedge ... \wedge \phi_q \text{ and } \bigwedge_i \phi_i^b \in MU^*(k) \right\}$$

For technical reasons, we allow the free parts to be arbitrary formulas over free literals, not just disjunctions of literals, so the ϕ_i are not clauses in a strict sense. It can easily be seen by induction on the number of variables that every minimal unsatisfiable Horn formula has deficiency 1, so we obtain the following hierarchy:

$$PROP \subset \exists HORN^b \subset \exists MU^*(1)^b \subset \exists MU^*(2)^b \subset ... \subset \exists CNF^b$$

To see that the inclusion $\exists HORN^b \subset \exists MU^*(1)^b$ is proper, consider the following:

$$\Phi = \exists x_1 \exists x_2 \exists x_3 \ (x_1 \vee x_2 \vee z_1) \wedge (x_1 \vee \neg x_2 \vee z_2) \wedge (\neg x_1 \vee x_3 \vee z_3) \wedge (\neg x_1 \vee \neg x_3 \vee z_4)$$

Φ is in $\exists MU^*(1)^b$, since the bound parts $(x_1 \vee x_2) \wedge (x_1 \vee \neg x_2) \wedge (\neg x_1 \vee x_3) \wedge (\neg x_1 \vee \neg x_3)$ are minimal unsatisfiable and have deficiency 1. But $(x_1 \vee x_2) \wedge (x_1 \vee \neg x_2) \wedge (\neg x_1 \vee x_3) \wedge (\neg x_1 \vee \neg x_3)$ is not a Horn formula, and there is no suitable renaming into a Horn formula.

In [9], a transformation from $\exists MU^*(k)^b$ formulas into equivalent Boolean circuits has been presented. The idea is to apply restricted hyperresolution.

Definition 3. *Let k be an integer. Then k-bound hyperresolution is the hyperresolution operation with the additional restriction that the resolvent and the side-clauses consist of at most k literals, in symbols $\alpha_1, \ldots, \alpha_n \vdash_{k\text{-}b\text{-}Hyper\text{-}Res} \sigma$.*

For fixed k, k-bound hyperresolution is not refutation-complete for *CNF*. But for Horn formulas, for example, even 1-bound hyperresolution is refutation-complete.

Lemma 2. ([9]) *For $k \geq 1$, let ϕ be a formula in $MU(k)$ over n variables. Then there is a $(\log_2(n) + k)$-bound hyperresolution refutation of at most n^{k+1} resolution steps.*

Theorem 4. ([9]) *There are constants a and c, such that for every $\exists MU^*(k)^b$ formula with a free part of length l and a bound part of length m, there is an equivalent circuit of size less than $l + a \cdot m^{c(log_2(m)+k)^2}$.*

4 Conclusion

By placing restrictions on the occurrences of bound variables in $QCNF^*$ formulas, we can obtain a hierarchy of subclasses of quantified Boolean formulas with an interesting connection to Boolean circuits. There also appears to be a close relationship between formula expressiveness and the structure of minimal false subformulas of the bound parts.

An open problem is how the above results on existentially quantified $\exists MU^*(k)^b$ formulas could be generalized to cases in which we also allow universal quantifiers, perhaps restricted to a prefix with a fixed number of quantifier alternations. Furthermore, the question arises whether the upper bound for the circuit size from Theorem 4 can be improved to $O(l) + O(m^{p(k)})$ for some polynomial p.

References

[1] S. Aanderaa and E. Börger. *The Horn Complexity of Boolean Functions and Cook's Problem.* Proc. 5th Scandinavian Logic Symposium 1979, Aalborg University Press, pp. 231–256, 1979.

[2] R. Aharoni and N. Linial. *Minimal NON-Two-Colorable Hypergraphs and Minimal Unsatisfiable Formulas.* Journal of Cominatorial Theory, 43(2): 196–204, 1986.

[3] U. Bubeck and H. Kleine Büning. *A New 3-CNF Transformation by Parallel-Serial Graphs.* Journal Information Processing Letters, 109(7):376–379, 2009.

[4] U. Bubeck and H. Kleine Büning. *Rewriting (Dependency-)Quantified 2-CNF with Arbitrary Free Literals into Existential 2-HORN.* Proc. 13th Intl. Conf. on Theory and Applications of Satisfiability Testing (SAT 2010), to appear, 2010.

[5] N. Dershowitz, Z. Hanna, and J. Katz. *Bounded Model Checking with QBF.* Proc. 8th Intl. Conf. on Theory and Applications of Satisfiability Testing (SAT'05). Springer LNCS 3569, pp. 408–414, 2005.

[6] H. Fleischner, O. Kullmann, and S. Szeider. *Polynomial-time Recognition of Minimal Unsatisfiable Formulas with Fixed Clause-Variable Deficiency.* Theoretical Computer Science, 289(1):503–516, 2002.

[7] A. Flögel, M. Karpinski, and H. Kleine Büning. *Resolution for Quantified Boolean Formulas.* Information and Computation, 117(1):12–18, 1995.

[8] T. Jussila and A. Biere. *Compressing BMC Encodings with QBF.* Electronic Notes in Theoretical Computer Science, 174(3):45–56, 2007.

[9] H. Kleine Büning and A. Remshagen. *An Upper Bound for the Circuit Complexity of Existentially Quantified Boolean Formulas.* Theoretical Computer Science, to appear, doi:10.1016/j.tcs.2010.04.017, 2010.

[10] H. Kleine Büning and X. Zhao. *Minimal False Quantified Boolean Formulas.* Proc. 9th Intl. Conf. on Theory and Applications of Satisfiability Testing (SAT 2006), Springer LNCS 4121, pp. 339–352, 2006.

[11] H. Kleine Büning and X. Zhao. *Computational Complexity of Quantified Boolean Formulas with Fixed Maximal Deficiency.* Theoretical Computer Science 407(1-3): 448–457, 2008.

[12] H. Kleine Büning, X. Zhao, and U. Bubeck. *Resolution and Expressiveness of Subclasses of Quantified Boolean Formulas and Circuits.* Proc. 12th Intl. Conf. on Theory and Applications of Satisfiability Testing (SAT 2009), Springer LNCS 5584, pp. 391–397, 2009.

[13] C. Papadimitriou and D. Wolfe. *The Complexity of Facets Resolved.* Journal of Computer Systems Science, 37(1):2–13, 1988.

[14] G. Tseitin. *On the Complexity of Derivation in Propositional Calculus.* In A. Silenko (ed.): Studies in Constructive Mathematics and Mathematical Logic, Part II, pp. 115–125, 1970.