UNIVERSITÉ D'ARTOIS

August, 7th 2014.
To whom it may concern:

Ph.D. Dissertation Review

Special Classes of Boolean Functions
with Respect to the Complexity of their Minimization

by Štefan Gurský
Charles University in Prague

Summary

This thesis is about Boolean functions, represented as propositional formulae in conjunctive normal form (CNF). It focuses on two main concepts: *minimization* and *propagation completeness*.

Let L be any representation language for Boolean functions. Minimizing a given L-representation Σ aims at generating an L-representation of minimal size, equivalent to Σ. Several notions of size can be considered depending on L, for instance when L is a subset of CNF, the number of clauses (the clause size measurement), or the number of literal occurrences (the literal size measurement). Because Boolean functions are ubiquitous in Computer Science (CS) and the CNF representation format is "natural", the minimization problem for CNF languages is a fundamental issue and has a number of significant applications. What makes it also valuable for CS researchers is that it is computationally difficult: the decision problem associated with the minimization problem (does there exist an L-representation of Σ of size at most k?) is at the second level of the polynomial hierarchy in the general case. Thus, this problem has been investigated by several researchers from various communities within CS since the 80's. Its complexity has been identified in the general case and in some specific cases (for instance, it is NP-complete when L is the language of Horn CNF formulae).

In addition, propagation completeness is a valuable property of CNF formulae Σ: when it holds, it is possible to derive using unit propagation all the literals which are entailed by Σ conjoined with any given consistent term. Since unit propagation can be achieved efficiently (i.e., in deterministic polynomial time), the clausal entailment problem for propagation complete CNF formulae is in P(TIME) while it is coNP-complete in the general case. However, though it is satisfied by several important fragments of CNF (including the Horn one), propagation completeness is not offered in the general case of CNF formulae. Adding some implied clauses (so called "empowering implicates") to a CNF formula during an off-line compilation phase in order to make it propagation complete is then a way to offer some guaranteed on-line response times, provided that the number of clauses to be added is not huge. Such a knowledge compilation perspective has been considered by several researchers from the Artificial Intelligence (AI) community since the 80's as well.

The dissertation addresses several important issues and presents a number of significant results, pertaining to the minimization and the propagation completeness of CNF formulae; mainly: a study of the minimization of Boolean functions represented as CNF formulae, the identification of the computational complexity of the minimization problem for several subsets of CNF for which the satisfiability problem is tractable (i.e., solvable in deterministic polynomial time), and the identification of the complexity of several computational issues related to the existence or the derivation of empowering implicates for CNF formulae. Especially, it is shown that the minimization problem for the language of matched CNF formulae is at the second level of the polynomial hierarchy; this exemplifies the fact that focusing on a CNF language for which the satisfiability issue is trivial is not sufficient to remove a source of complexity for the minimization problem. As to propagation completeness, the key results are that the recognition of propagation complete CNF formulae is a coNP-complete problem and that the number of empowering implicates to be added to a CNF formula Σ in order to make it propagation complete is provably exponential in the worst case. In the specific case when clausal entailment is made tractable via unit propagation, this strengthens a “standard”, yet non-constructive uncompilability result stating that the number of clauses to be added to Σ is not polynomial in the size of Σ and requiring the additional proviso that the polynomial hierarchy does not collapse at the second level. It also refines a more recent result stating that any propagation complete CNF encoding Σ of the ALLDIFFERENT constraint is of size at least superpolynomial in the size of Σ.

Comments

The dissertation is divided into four main chapters, completed by an introduction, a conclusion, and a bibliography; the four main chapters are:

- *Boolean functions*
- *Complexity of SAT and minimization for certain CNF classes*
- *Matched CNFs*
- *Propagation completeness*

The dissertation is well organized and structured; especially, the most sophisticated results are decomposed into a number of intermediate propositions or lemmata, which make the results easier to understand.

In what follows, I elaborate on the contents and significant contributions of the four main chapters of the dissertation.

<u>Boolean functions.</u>
In this chapter of the dissertation, Štefan Gurský presents some general background on Boolean functions, and their representations in terms of formulae; the focus is laid on the dual propositional languages, CNF and DNF, and a couple of dual concepts related to them, implicate and implicant, resolution rule and consensus rule. Štefan Gurský then states in formal terms some key decision problems, including the satisfiability one and the minimization one (in terms of number of clauses/terms and in terms of the number of occurrences of literals).

<u>Complexity of SAT and minimization for certain CNF classes.</u>
In this chapter of the dissertation, Štefan Gurský first recalls the definitions of some subsets of CNF for which the satisfiability problem can be solved in deterministic polynomial time: monotone CNF, Horn CNF, renamable Horn CNF, quadratic CNF (also known as the language of Krom formulae), qHorn CNF, component-wise quadratic Horn CNF. For all those languages but the latter, the minimization problem is known as NP-complete (for the two size measurements considered here); it is also known that the problem can be solved in polynomial time for component-wise quadratic Horn CNF. Finally, Štefan Gurský presents a first contribution of the thesis, under the form of three conditions on a CNF fragment L which prove sufficient to ensure that the minimization problem for L is in NP when L satisfies them; those properties are tractable satisfiability, tractable conditioning and stability w.r.t. prime representations. The result is quite easy but significant since many subsets of CNF satisfies those three conditions.

<u>Matched CNFs.</u>
In this chapter of the dissertation, Štefan Gurský focuses on the language of matched CNFs (already introduced in the literature), namely those CNF formulae where each clause can be associated in an injective way with a specific literal in it. The language of matched CNFs trivially satisfies the tractable satisfiability condition, but none of the two other conditions. The complexity of the minimization problem for the language of matched CNFs when the literal size measurement is considered can be easily established from previous results from the literature. As a further contribution, Štefan Gurský proves that the problem is also Σ_2^p-complete when the clause size measurement is considered. While it is not the case that every prime irredundant CNF representation of a given matched CNF formula Σ is matched, Štefan Gurský also shows that every minimal CNF representation (w.r.t. the clause size measurement) of a Boolean function which has a matched CNF representation is also matched.

<u>Propagation completeness.</u>
In this chapter of the dissertation, Štefan Gurský addresses a number of open questions related to propagation

complete CNF formulae. He first shows that deciding whether a given clause is an empowering implicate for a given CNF formula is coNP-complete. He then shows that any CNF formula Σ which is not propagation complete has an empowering implicate which can be derived using a resolution proof of length linear in the number of literals occurring in Σ. Furthermore, the linear bound is tight (it can be reached for some CNF formulae). Afterwards, Štefan Gurský proves that the problem of deciding whether a given CNF formula has an empowering implicate is NP-complete; the hardness result comes from a reduction from the 3-Dimensional Matching problem, and is quite evolved. Finally, Štefan Gurský shows that the number of empowering implicates to be added to a CNF formula Σ in order to make it propagation complete is provably exponential in the worst case, which is one of the main results of the dissertation.

Suggested corrections

Globally speaking, the results are suitably contrasted to related work but some additional references could nevertheless be added at some places, especially when considering notions which are not that “usual” (e.g., merge resolution, the qHorn CNF fragment).

While the investigation of the thesis problems is thorough, there is room for improvement from the point of view of presentation. Indeed, this dissertation is very synthetic and, though it is not difficult to read for a specialist of the topic, it is surely not that easy to read for a “naive” reader. Accordingly, more examples would be welcome in order to enhance the level of illustration of the document (there are very few examples in the dissertation). More details would be expected at some places (for instance, only a couple of lines are devoted to the idea of knowledge compilation despite it is a key motivation for some investigations in the thesis). A few pictures would also be appreciated as well for improving the readibility of the manuscript (especially for illustrating the concept of matched CNF which is graph-based, or the matchings at Example 5.4.2, and also for making the results based on resolution derivations easier to grasp, e.g., in the proof of Proposition 5.3.3 – such derivations could be depicted graphically). In the same vein, the definition environment should be modified so that the definitions are numbered, and an index could be added to the document.

There are also a few typos in some mathematical statements (e.g. the recognition of renamable Horn formulae or the set of literals associated with a given set of variables), but none of them is harmful. These flaws can be easily fixed and do not question the overall quality of the thesis.

Conclusion

To sum up, Štefan Gurský's dissertation *"Special Classes of Boolean Functions with Respect to the Complexity of their Minimization"* is a very good piece of work. The problems studied by Štefan Gurský are important for a number of disciplines within CS. Some of the results he obtained are undoubtedly significant and required some complex proofs. Parts of the work have already been published in two international journals, including the top-ranked one in AI, *Artificial Intelligence.* Subsequently, I recommend the thesis to be defended.

Would you have any further questions regarding this review, please feel free to contact me by e-mail.

Prof. Pierre Marquis, Ph.D.
Professor of Computer Science,
Université d'Artois, France
e-mail: marquis@cril.univ-artois.fr