Available online at www.sciencedirect.com
ScienceDirect

Discrete Applied Mathematics 155 (2007) 1408 – 1419

DISCRETE APPLIED MATHEMATICS

www.elsevier.com/locate/dam

Satisfiability of mixed Horn formulas☆

Stefan Porschen, Ewald Speckenmeyer

Institut für Informatik, Universität zu Köln, D-50969 Köln, Germany

Received 8 February 2005; received in revised form 9 February 2007; accepted 18 February 2007
Available online 4 March 2007

Abstract

In this paper the class of *mixed Horn formulas* is introduced that contain a Horn part and a 2-CNF (conjunctive normal form) (also called quadratic) part. We show that SAT remains NP-complete for such instances and also that any CNF formula can be encoded in terms of a mixed Horn formula in polynomial time. Further, we provide an exact deterministic algorithm showing that SAT for mixed Horn formulas containing n variables is solvable in time $O(2^{0.5284n})$. A strong argument showing that it is hard to improve a time bound of $O(2^{n/2})$ for mixed Horn formulas is provided. We also obtain a fixed-parameter tractability classification for SAT restricted to mixed Horn formulas C of at most k variables in its positive 2-CNF part providing the bound $O(\|C\|2^{0.5284k})$. We further show that the NP-hard optimization problem minimum weight SAT for mixed Horn formulas can be solved in time $O(2^{0.5284n})$ if non-negative weights are assigned to the variables. Motivating examples for mixed Horn formulas are level graph formulas [B. Randerath, E. Speckenmeyer, E. Boros, P. Hammer, A. Kogan, K. Makino, B. Simeone, O. Cepek, A satisfiability formulation of problems on level graphs, ENDM 9 (2001)] and graph colorability formulas.

Keywords: (Hidden) Horn formula; q-Horn formula; Quadratic formula; (Weighted) Satisfiability; Minimal vertex cover; Fixed-parameter tractability

1. Introduction

In recent time the interest in designing exact algorithms providing better upper time bounds than the trivial ones for NP-complete problems and their NP-hard optimization counterparts has increased. Of particular interest in this context is the investigation of exact algorithms for testing the satisfiability (SAT) of propositional formulas in conjunctive normal form (CNF). This interest stems from the fact that SAT is well known to be a fundamental NP-complete problem appearing naturally or via reduction as the abstract core of many application-relevant problems. Not only the whole class CNF is of interest in this context. In several applications subclasses of CNF are of importance for which SAT unfortunately remains NP-complete. Nevertheless, it is often possible by exploiting the specific structure of such formulas to design fast exact algorithms for their solution. Such subclasses, for instance, can be obtained by composing or *mixing* formulas of two different parts each of which separately is SAT-testable in polynomial time (see also [12]).

In this paper we introduce and study so-called *mixed Horn formulas* which roughly speaking are formulas composed of a quadratic part and a Horn part. More precisely, for a positive monotone 2-CNF formula P (containing only 2-clauses) and a Horn formula H, we call the formula $M = H \wedge P$ a *mixed Horn formula* (*MHF*). It is well known

☆ A preliminary version of this paper appeared in [16].

E-mail addresses: porschen@informatik.uni-koeln.de (S. Porschen), esp@informatik.uni-koeln.de (E. Speckenmeyer).

0166-218X/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2007.02.010

that 2-SAT and Horn-SAT are solvable in linear time [1,14], but SAT for MHFs (MHF-SAT) remains NP-complete. A closely related class generalizing the classes of Horn and quadratic formulas is the class of so-called q-Horn formulas introduced by Boros et al. [2], for which SAT can be solved in linear time also [2]. A q-Horn formula (in partition form, see below), can be considered as a specific mixed Horn formula. The class of q-Horn formulas (in partition form), probably is the largest subclass of MHF that is SAT-solvable in polynomial time.

The main purpose of this paper is to prove a non-trivial worst case upper time bound for solving MHF-SAT, namely $O(2^{0.5284n})$ where n is the number of variables in the instance. Moreover, we obtain a fixed-parameter tractability classification (cf. e.g. [5]) of SAT restricted to MHFs $M = P \wedge H$ where P has a fixed number k of different variables, provided by the polynomial bound $O(\|M\| 2^{0.5284k})$, where $\|M\|$ is the length of M.

We also analyze the connection of MHF-SAT to unrestricted SAT. Specifically we show that each CNF formula C with n different variables can be transformed in polynomial time into a MHF $M = P \wedge H$, such that P has $k \leqslant 2n$ different variables. Then C is satisfiable if and only if M is satisfiable, and the question, whether $M \in$ SAT, can be answered in time $O(\|C\| 2^{k/2})$. Hence, if there is an $\alpha < \frac{1}{2}$ such that every MHF $M = P \wedge H$ can be solved in time $O(\|C\| 2^{\alpha k})$, then there is $\beta \leqslant 2\alpha < 1$ such that SAT for an arbitrary CNF-formula C can be decided in time $O(\|C\| 2^{\beta n})$. The MHF-formulation of a CNF-formula C yields a partition of all variables in C into the *essential* variables (variables occurring in P) and the remaining ones.

The introduction and investigation of MHFs is by no means artificial. Well known problems for level graphs, like level-planarity test or the NP-hard crossing-minimization problem, can be formulated conveniently in terms of MHFs (for more details see [17]). This was our motivation for considering MHFs. Also graph colorability naturally leads to MHFs. To see this, consider a simple graph $G = (V, E)$ and a set of r colors $[r] := \{1, \ldots, r\}$. The decision whether G is r-colorable, i.e. whether at most r colors can be assigned to all vertices in V such that no two adjacent vertices are colored equally, can be encoded into MHF-SAT as follows: for every vertex $x \in V$ introduce r variables x_i, $i \in [r]$, and one clause $x_1 \vee x_2 \vee \cdots \vee x_r$. For every edge $x - y \in E$, we have to ensure that x and y are colored differently. So we introduce for each color $i \in [r]$ the clause $\neg(x_i \wedge y_i) \equiv (\overline{x_i} \vee \overline{y_i})$ yielding r 2-clauses for each edge. In summary, we obtain a CNF formula $C(G)$ consisting of $|V| + r|E|$ clauses and containing $r|V|$ different variables. Finally complementing all variables in $C(G)$ turns all its r-clauses into negative monotone clauses and its 2-clauses into positive monotone clauses, hence yields a MHF $\tilde{C}(G)$. It is easy to verify that G is r-colorable if and only if the MHF $\tilde{C}(G)$ is satisfiable via the interpretation that setting variable x_i to FALSE means that the corresponding vertex x is colored by i. Notice that introducing only one r-clause for every vertex ensuring at least (instead of exactly) one color for every vertex suffices for deciding r-colorability.

Another source of the interest in Horn clauses contained in CNF formulas stems from recent observations of hidden threshold phenomena [20] according to a fixed fraction of Horn clauses in CNF formulas.

The present paper is structured as follows. Section 2 introduces basic notions and notations used throughout the paper. In Section 3, several versions of MHFs are introduced. Each of these classes is NP-complete w.r.t. SAT as follows by the above described reduction from the NP-complete graph coloring problem [6]. We provide another polynomial time transformation of CNF-SAT to MHF-SAT on which some investigations in this paper rely. In Section 4, a vertex cover based algorithm for determining SAT of a MHF M is presented having running time $O(2^{0.5284n})$, with n being the number of variables in M. The approach also yields a classification of MHFs allowing for a fixed-parameter tractability result. Section 5 provides a strong argument stating that it is hard to improve an $O(2^{n/2})$ time bound for solving MHF-SAT. Section 6, describes a further vertex cover based technique for speeding up the MHF-SAT algorithm. Some experimental results illustrating the usefulness of this approach are presented. Section 7, finally, provides an algorithm for the minimum weight MHF-SAT problem, where weights are assigned to the variables.

2. Basic notions and notation

Let CNF denote the set of formulas (free of duplicate clauses) in CNF over a set $V = \{x_1, \ldots, x_n\}$ of propositional variables $x_i \in \{0, 1\}$. Each variable x induces a positive literal (variable x) or a negative literal (negated variable: $\overline{x}$). Each formula $C \in$ CNF is considered as a clause set $C = \{c_1, \ldots, c_{|C|}\}$. Each clause $c \in C$ is a disjunction of different literals, and is also represented as a set $c = \{l_1, \ldots, l_{|c|}\}$. The length of a formula C is denoted by $\|C\|$ whereas $|C|$ denotes the number of its clauses. A clause containing positive (negative) literals only is called *positive* (*negative*) *monotone*. We denote by $V(C)$ the set of variables occurring in formula C. The satisfiability problem (SAT) asks in its *decision* version, whether a given CNF instance C is *satisfiable*, i.e. whether C has a *model*, which is a truth assignment