ELSEVIER Discrete Applied Mathematics 75 (1997) 217–230

DISCRETE APPLIED MATHEMATICS

Variable and term removal from Boolean formulae

Yves Crama[a], Oya Ekin[b], Peter L. Hammer[b,*]

[a] *Ecole d'Administration des Affaires, Université de Liège, 4000 Liège, Belgium*
[b] *RUTCOR, Rutgers University, P.O. Box 5062, New Brunswick, NJ 08903, USA*

Received 26 June 1995; revised 15 December 1995

Abstract

Given a Boolean formula in disjunctive normal form, the variable deletion control set problem consists in finding a minimum cardinality set of variables whose deletion from the formula results in a DNF satisfying some prescribed property. Similar problems can be defined with respect to the fixation of variables or the deletion of terms in a DNF. In this paper, we investigate the complexity of such problems for a broad class of DNF properties.

1. Introduction

A large number of algorithmic problems on Boolean disjunctive normal forms (DNFs), such as SAT, $MAX\text{-}SAT$, complementation, etc., are known to be computationally difficult, so that polynomial solutions to most of these problems are only available for very special classes of Boolean formulae. On the other hand, a trivial but fruitful observation is that, by fixing a sufficiently large subset of variables to specific values or by removing a sufficiently large subset of terms, any DNF can be reduced to a highly structured one, displaying many special properties. Accordingly, given a DNF Φ and a property π, let us call *control set* any subset of variables or terms whose "removal" from Φ results in a new DNF having property π ("removal" means here either "fixation" or "deletion"; we will be more specific below).

The notion of control set has been exploited by numerous researchers either to decompose an original, hard problem into a collection of simpler ones, or to approximate the original problem by a simpler one (see e.g. 2, 4–7). In most of these investigations a central issue is then to identify control sets of small cardinality. This

* Corresponding author.
Email- ekin@rutcor.rutgers.edu, hammer@rutcor.rutgers.edu
Preprint submitted to Elsevier Science

PII S0166-218X(96)00028-5

concern is often addressed in practice through various heuristics which rely on some intuitive criterion in order to successively select the variables or terms to be included in the control set.

By contrast, our main objective in this paper is to investigate the complexity of computing *minimum cardinality* control sets, under various specifications of this concept and for a wide range of properties π. In particular, we prove several generic results which assert that finding minimum cardinality control sets is NP-hard for all properties π satisfying certain natural assumptions. (Similar results regarding "control sets" of vertices or edges in graphs can be found in [12, 15].)

The remainder of the paper is organized as follows. In Section 2, we recall some basic definitions pertaining to Boolean formulae. Section 3 gives a precise description of the problems to be investigated. Finally, Sections 4 and 5 contain our main results.

2. Definitions and notations

We assume that the reader is familiar with the basic concepts of Boolean algebra, and we only introduce here the notions that we explicitly use in the paper (see e.g.[4] or [13] for more details).

A *disjunctive normal form* (*DNF*) is a Boolean formula of the type

$$\Phi(x_1, \dots, x_n) = \bigvee_{k=1}^{m} \left(\bigwedge_{i \in P_k} x_i \bigwedge_{i \in N_k} \bar{x}_i \right), \tag{1}$$

where $V = \{x_1, \dots, x_n\}$ is a set of n Boolean *variables* (each of which can take value either 0 or 1) and $\bar{x}_i = 1 - x_i$ is the *complement* of x_i, for $i = 1, \dots, n$. The variables x_i and their complements $\bar{x}_i$ are called *literals*. The sets $P_1, \dots, P_m$, $N_1, \dots, N_m$ are subsets of $\{1, \dots, n\}$ and we assume that $P_k \cap N_k = \emptyset$ for $k = 1, \dots, m$.

A conjunction of literals of the form $T_k = \wedge_{i \in P_k} x_i \wedge_{i \in N_k} \bar{x}_i$ is called a *term* of Φ, for $k = 1, \dots, m$. The *degree* of term T_k is given by $|P_k \cup N_k|$, and the *degree* of Φ is the maximum degree of its terms. We denote by Λ the DNF without terms and by Ω the DNF consisting of a single term of degree zero (or, by abuse of notation, any DNF containing at least one zero degree term).

It is customary to view any DNF Φ of the form (1) (or, more generally, any Boolean expression) as defining a *Boolean function*, i.e. a mapping from $B^n = \{0, 1\}^n$ into $\{0, 1\}$: for any assignment of 0–1 values to the variables $(x_1, \dots, x_n)$, the value of $\Phi(x_1, \dots, x_n)$ is simply computed according to the usual rules of Boolean algebra (see e.g. [10]). In particular, $\Lambda(x) = 0$ and $\Omega(x) = 1$ for all $x \in \{0, 1\}^n$.

A paradigmatic hard problem for Boolean formulae is the *satisfiability problem* (*SAT*). When working with DNFs, this problem can be formulated as follows: given as input a DNF Φ of the form (1), is there an assignment *satisfying* Φ, i.e. is there a point $x^* \in B^n$ such that $\Phi(x^*) = 0$? (The satisfiability problem is usually posed for conjunctive normal forms rather than for DNFs, but these two versions are trivially equivalent.) SAT is NP-complete for general DNFs, but several classes of DNFs for which SAT is polynomially solvable have been extensively investigated in the literature. Prominent among these are the classes of

- *quadratic* DNFs, i.e. DNFs of degree 2 or less;
- *Horn* DNFs, i.e. DNFs satisfying $|N_k| \leqslant 1$ for $k = 1, \ldots, m$;
- *renamable Horn* DNFs, i.e. DNFs which become Horn after renaming some literals x_i as $\bar{y}_i$ and – simultaneously – $\bar{x}_i$ as y_i (see [11]).

Notice that the class of renamable Horn DNFs subsumes, in particular, the class of *monotone*, or *unate*, DNFs, i.e. the class of those DNFs Φ such that either x_i or $\bar{x}_i$ does not appear in Φ, for $i = 1, \ldots, n$.

A class of DNFs generalizing all previous ones has been introduced in [1] and further investigated in [2, 3]. In order to describe this class, let us consider again the DNF Φ given by (1), and let us associate a real-valued variable $\alpha(u)$ with each literal u. We can now define the following system of linear inequalities:

$$\sum_{i \in P_k} \alpha(x_i) + \sum_{i \in N_k} \alpha(\bar{x}_i) \leqslant 1 \quad (k = 1, \ldots, m), \tag{2}$$

$$\alpha(x_i) + \alpha(\bar{x}_i) = 1 \quad (i = 1, \ldots, n), \tag{3}$$

$$\alpha(x_i) \geqslant 0, \;\; \alpha(\bar{x}_i) \geqslant 0 \quad (i = 1, \ldots, n). \tag{4}$$

The DNF Φ is said to be *q-Horn* if the system (2)–(4) has a solution. In fact, if Φ is q-Horn, then a half-integral solution to the system (2)–(4) always exists (see [1]), so that recognizing q-Horn DNFs amounts to finding a $\{0, \frac{1}{2}, 1\}$ solution of (2)–(4). (An efficient recognition algorithm for q-Horn DNFs is given in [3].) Moreover, it is also shown in [1] that the class of q-Horn DNFs contains all quadratic, Horn, and renamable Horn DNFs, and that SAT is polynomially solvable for q-Horn DNFs.

In order to broaden the scope of our discussion, let us now call *(DNF) property* any subset of DNFs which contains the trivial DNFs Λ and Ω (the proviso regarding Λ and Ω is not essential, but will simplify the ensuing discussion). We say that DNF Φ has property π if $\Phi \in \pi$. In the sequel, we will consider a very broad class of DNF properties, restricted only by certain basic requirements. More precisely, we denote by $\mathscr{C}$ the class of DNF properties π which are:

(a) *nontrivial*: all DNFs of degree 1 have property π, and there exists a DNF which does not have property π;

(b) *hereditary under deletion of terms*: if the DNF $\Phi = \bigvee_{k=1}^{m} T_k$ has property π, then the DNF $\bigvee_{k=1, k \neq j}^{m} T_k$ also has property π, for $j = 1, \ldots, m$.

Observe that, for any property $\pi \in \mathscr{C}$, there are infinitely many DNFs in π as well as infinitely many DNFs not in π. Moreover, the class $\mathscr{C}$ itself is quite large: in particular, all properties discussed above (quadratic, Horn, renamable Horn, q-Horn, monotone) fall in the class $\mathscr{C}$. Some other interesting Boolean properties, however, like thresholdness (see [13]), do not lie in $\mathscr{C}$.

Finally, we will also discuss the following notion: a DNF property π is said to be *term induced* when, for every DNF $\Phi = \bigvee_{k=1}^{m} T_k$, Φ has property π if and only if all its terms $T_1, \dots, T_m$, viewed as DNFs, have property π. Obviously, every term induced property is hereditary under deletion of terms, but the converse relation does not necessarily hold. For example, properties like quadratic or Horn are term induced, while renamable Horn, q-Horn or monotone are not.

3. Removing variables or terms from DNFs

As mentioned in the Introduction, several authors have attempted to attack SAT and other difficult Boolean problems by exploiting the subformulae with "nice" properties that can be produced by "removing" variables or terms from an arbitrary DNF. We have explained above what we mean by "nice" (viz. quadratic, Horn, q-Horn, etc.). Let us now be more specific about the meaning of "removing".

Fixing a variable x_i, $i = 1, \dots, n$, to the value 1 in Φ means removing x_i from each term of Φ where it appears and removing from Φ all terms that contain $\bar{x}_i$. The resulting DNF is called the *restriction* of Φ to $x_i = 1$. If Φ contains a linear term (i.e. a term of degree 1) of the form $T_k = x_i$, then the restriction of Φ to $x_i = 1$ is the DNF $\Omega \equiv 1$. On the other hand, if all terms of Φ contain $\bar{x}_i$, then fixing x_i to 1 produces the DNF $\Lambda \equiv 0$. (Notice that, in terms of functions, all we are doing here is defining the restriction of the function $\Phi(x)$ to $x_i = 1$. But since the relation between Boolean functions and DNFs is not one-to-one, we need the more precise definition just given.)

Now, given a property π, a DNF Φ and a subset of variables $S \subseteq V$, we say that S is a *variable fixation control set* (for short, a *VF set*) of Φ for property π if *all* restrictions of Φ obtained by fixing the variables in S to arbitrary 0–1 values have property π.

Example 1. For illustration, let us consider the DNF

$$\Phi = x_1x_6 \vee x_1x_2x_3 \vee \bar{x}_3x_4x_5x_6. \tag{5}$$

If the property π under consideration is that of being quadratic, then the set $\{x_3, x_4\}$ is a VF set for Φ, since all possible assignments of values to the variables x_3 and x_4 produce quadratic DNFs. It is also easy to check that there is no VF set of cardinality 1 in this example.

When a VF set S is at hand (and if S is not too large), then SAT can be handled by complete enumeration of S: fix the variables of S in all ($2^{|S|}$) possible ways, and solve the SAT subproblems associated with the corresponding restrictions. As a matter of fact, the concept of variable-fixation control set has been used extensively since the inception of Boolean theory. For instance, elimination methods for the solution of satisfiability problems, such as the Davis–Putnam method, rely more or less explicitly on the enumeration of VF sets. These methods usually proceed by fixing variables until the resulting formula becomes either monotone, or quadratic, or Horn (see e.g. [5]). More generally, in [4], VF sets for monotonicity are put to systematic use in the solution of several subproblems arising in logic minimization. VF sets resulting in renamable Horn or q-Horn formulae have also been recently considered in [2, 6, 14].

Several variants of VF sets will also be discussed in the sequel. A *weak VF control set* for property π is a set of variables such that *at least one* assignment of 0–1 values to these variables yields a restriction of Φ having property π. Obviously, weak VF sets generalize VF sets.

Example 1 (cont.). For the DNF (5), the set $\{x_3\}$ is the smallest weak VF set (by setting x_3 to 1, Φ becomes $x_1x_6 \vee x_1x_2$, which is quadratic). Notice, however, that $\{x_3\}$ is not a VF set.

Deleting a variable x_i from a DNF Φ consists in removing all occurrences of x_i and of $\bar{x}_i$ from Φ. If either x_i or $\bar{x}_i$ is a linear term of Φ, then this term vanishes (i.e. becomes 0) when x_i is deleted. We denote by $\Phi\backslash S$ the DNF obtained by deleting a set S of variables from Φ. The set S is a *variable deletion control set* (*VD set*) of Φ for property π if $\Phi\backslash S$ has property π.

Example 1 (cont.). For the DNF (5), deleting the set of variables $\{x_3, x_4\}$ produces the quadratic DNF $x_1x_6 \vee x_1x_2 \vee x_5x_6$. It is easy to notice that this VD set for the "quadraticity" property is of minimum cardinality.

Variable deletion has been recently investigated by Chandru and Hooker [6] and Truemper [14]. Although this operation is not really of an "algebraic" nature, it plays an interesting role in the study of control sets, mostly because of its close relationship with variable fixation. To clarify this relationship, let us first record an elementary observation (see [6] for a special case of this result).

Proposition 1. *Let property π be hereditary under deletion of terms and let Φ be an arbitrary DNF. Then, every variable deletion control set of Φ for property π is a variable fixation control set of Φ for π.*

Proof. It suffices to establish the theorem for a variable deletion control set of size 1, say $S = \{x\}$, $x \in V$. Write Φ in the form

$$\Phi = \Phi_0\bar{x} \vee \Phi_1 x \vee \Phi_2, \tag{6}$$

where Φ_0, Φ_1 and Φ_2 are DNFs not involving x or $\bar{x}$. If x and $\bar{x}$ are not linear terms of Φ, then the DNF obtained after deletion of x is

$$\Phi_0 \vee \Phi_1 \vee \Phi_2. \tag{7}$$

The restriction of Φ to $x = 1$ is

$$\Phi_1 \vee \Phi_2 \tag{8}$$

and the restriction of Φ to $x = 0$ is

$$\Phi_0 \vee \Phi_2. \tag{9}$$

By assumption, (7) has property π. Hence, by heredity, so do (8) and (9).

If x is a linear term of Φ, then Φ_1 does not appear in (7) anymore, and the previous argument does not apply. But in this case, the restriction (8) is simply Ω, which satisfies π (by definition of DNF properties), and (9) can be obtained from (7) by term deletion (or is Ω, if $\bar{x}$ is also a linear term of Φ).

A symmetrical argument applies when $\bar{x}$ is a linear term of Φ. □

The following example shows that, for properties which are hereditary under deletion of terms, the concepts of VD set and VF set are usually distinct.

Example 3. Consider the DNF

$$\Psi = x_1x_2x_3x_4 \vee x_1\bar{x}_2x_3x_4 \vee \bar{x}_1\bar{x}_2x_3x_4 \vee x_1x_2\bar{x}_3\bar{x}_4 \vee \bar{x}_1x_2\bar{x}_3\bar{x}_4 \vee x_1\bar{x}_2\bar{x}_3\bar{x}_4.$$

For the property "renamable Horn", the set $\{x_4\}$ is a minimum size variable fixation control set of Ψ. On the other hand, Ψ has no variable deletion control set of size 1.

The relation between VF and VD control sets becomes even simpler for term induced properties (see Example 1 above for an illustration):

Proposition 2. *For term induced DNF properties, the concepts of variable deletion control set and of variable fixation control set are equivalent.*

Proof. Proceeding as in Proposition 1, we now assume that $S = \{x\}$ is a VF set of Φ for π, and we only need to prove that S necessarily is a VD set of Φ. Since π is term induced, every term of (8) and of (9) has property π. Therefore, (7) has property π. □

Propositions 1 and 2 imply that, for most algorithmic purposes, VD sets are often just as useful as VF sets (although, as illustrated by Example 2, a minimal size VD set can turn out to be larger than a minimal size VF set). On the other hand, VD sets have a subtle advantage over VF sets. To see this, let us define three decision problems associated with an arbitrary property π:

P_0: Given a DNF Φ, verify whether Φ has property π.

P_1: Given a DNF Φ and a subset S of variables, verify whether S is a variable fixation control set of Φ for property π.

P_2: Given a DNF Φ and a subset S of variables, verify whether S is a variable deletion control set of Φ for property π.

Even if problem P_0 is easy, problem P_1 may very well be difficult. By contrast, we get for P_2:

Proposition 3. *For every property π, any instance of problem P_2 can be reduced in polynomial time to an instance of problem P_0.*

Proof. Trivial. □

Proposition 3 is noteworthy to the extent that, for most properties considered in practice, the recognition problem P_0 (and hence, P_2) is polynomially solvable. Together, Propositions 1–3 motivate our interest in variable deletion control sets.

Finally, we introduce the concept of *term deletion control sets* (*TD sets*), i.e. sets of terms which, when deleted from the DNF under consideration, produce a DNF having the specified property π. (Observe that this concept has already been implicitly used in our definition of class $\mathscr{C}$.)

Example 1 (cont.). For DNF (5), the unique minimum cardinality term deletion control set for the "quadraticity property" is $\{x_1x_2x_3,\ \bar{x}_3x_4x_5x_6\}$.

To every DNF Φ, a TD set associates a DNF Ψ such that $\Psi(x) \leqslant \Phi(x)$ for all $x \in \{0, 1\}^n$. Such a lower bounding DNF Ψ can be helpful in establishing inconsistency of the (SAT) equation $\Phi = 0$ (namely, if $\Psi = 0$ is inconsistent, then so is $\Phi = 0$). This observation has been used by Gallo and Urbani [7] in a SAT algorithm, in which they systematically generate TD sets for the "Horn" property.

Term deletion control sets also arise in the following framework. Let σ be the DNF property "Φ is satisfiable" or, more formally, $\sigma = \{\Phi \mid \Phi(x) \not\equiv 1\} \cup \{\Omega\}$. (Note that we have only added Ω to this set in order to abide by our definition of DNF properties.) Then, finding a minimal TD set for σ is equivalent to the well-known Maximum Satisfiability MAX-SAT problem (see e.g. [8]).

Our aim in this note is to investigate the complexity of computing minimal cardinality control sets, for each of the four types of control sets mentioned above and for various DNF properties. Variable fixation and variable deletion problems will be addressed in Section 4, while term deletion problems will be treated in Section 5.

4. Variable deletion and variable fixation control set problems

The proofs in this section closely follow the proofs presented by Lewis and Yannakakis in [12, 15] for vertex deletion problems in graphs. For our purpose, however, we find it useful to extend some of the graph-theoretic notions used in [12, 15] to a Boolean framework.

Consider again an arbitrary DNF of the form

$$\Phi(x_1, \ldots, x_n) = \bigvee_{k=1}^{m} \left(\bigwedge_{i \in P_k} x_i \bigwedge_{i \in N_k} \bar{x}_i \right). \tag{10}$$

With this DNF, we associate a graph $G = G(\Phi)$. The vertex set of G is $X = \{1, \ldots, n\}$ (though, for convenience, we often identify X with $V = \{x_1, \ldots, x_n\}$). The edges of

G are induced by the terms of Φ of degree two or more: more exactly, the edges of G are all pairs of vertices of the form $\{i,j\}$, $i \neq j$, such that $\{i,j\} \subseteq P_k \cup N_k$ for some $k \in \{1, \dots, m\}$.

Now, a *component* of Φ is simply a (connected) component of G. If S is a component of Φ, then, by a slight abuse of terminology, we also call "component of Φ" the DNF obtained by deleting $V \backslash S$ from Φ.

For every variable $x \in V$, a *component of Φ relative to x* is a set of variables of the form $S \cup \{x\}$, where S is any component of $\Phi \backslash \{x\}$. Variable x is called a *cut variable* of Φ if the number of components of $\Phi \backslash \{x\}$ is strictly larger than the number of components of Φ. (If $\{x\}$ is, in itself, a component of Φ, then we extend the previous definitions by saying that $\{x\}$ is a component of Φ relative to x and that x is a cut variable of Φ.)

We now proceed to introduce a total preorder among DNFs. Let $\Phi^1, \Phi^2, \dots, \Phi^t$ denote the components of Φ (viewed as DNFs). For $i = 1, \dots, t$ and any variable c_i of Φ^i, let $\gamma_i = (n_{i1}, n_{i2}, \dots, n_{ij_i})$, where n_{ik} is the number of variables in the kth component of Φ^i relative to c_i, and $n_{i1} \geqslant n_{i2} \geqslant \cdots \geqslant n_{ij_i}$. Notice that the vector γ_i may depend on the choice of the variable c_i. We shall henceforth assume that c_i is chosen in such a way that γ_i is lexicographically smallest among all possible choices. Observe that, if Φ^i has no cut variable, then γ_i is simply the number of variables of Φ^i and we may pick c_i arbitrarily. Otherwise, our rule ensures that c_i will be a cut variable of Φ^i.

Let now $\beta(\Phi) = (\gamma_1, \gamma_2, \dots, \gamma_t)$, where $\gamma_1 \geqslant_L \gamma_2 \geqslant_L \cdots \geqslant_L \gamma_t$ and $\geqslant_L$ denotes the lexicographic order. Then, the lexicographic order among the sequences $\beta(\Phi)$ induces a total preorder R among DNFs: for any two DNFs Φ and Ψ, we say that Φ is smaller than Ψ in R if $\beta(\Phi) \leqslant_L \beta(\Psi)$ (this generalizes a definition introduced for graphs in [12, 15]).

We now have all the basics needed for the proof of our first theorem.

Theorem 4. *Finding a minimum cardinality variable deletion control set is* NP-*hard for every DNF property in class* $\mathscr{C}$.

Proof. As mentioned before, our proof is inspired from [12, 15]. Therefore, we only give here the main elements of our proof, and we refer the reader to [12, 15] for the missing details.

Let π be any property in $\mathscr{C}$. We are going to provide a polynomial transformation from the vertex cover problem to the VD set problem for π. The vertex cover problem is the following: given a graph G, find a minimum vertex cover of G, i.e. find a smallest subset of vertices that meets all edges of G.

In order to describe the transformation, let Φ_π be a smallest DNF in the preorder R with the property that the disjunction of a sufficiently large number of independent copies of Φ_π violates π (by "independent copies", we mean here "copies on disjoint sets of variables"). The existence of Φ_π follows immediately from the nontriviality of π. We denote by K the smallest number of independent copies of Φ_π whose disjunction violates π.

As above, let $\Phi^1, \Phi^2, \dots, \Phi^t$ denote the (DNF) components of Φ_π, sorted by lexicographically nonincreasing γ-vectors. By nontriviality of π, Φ^1 involves at least two variables. Let c be the variable of Φ^1 that gave rise to γ_1, let Φ^{11} be the largest component of Φ^1 relative to c, and let d be any variable of Φ^{11} other than c. In the sequel, we assume that the variables of Φ_π are ordered so that $\Phi_\pi = \Phi_\pi(x_1, \dots, x_r, c, d)$.

Consider now an instance $G = (U, \mathscr{A})$ of the vertex cover problem, with vertex set $U = \{u_1, \dots, u_n\}$ and edge set $\mathscr{A}$. It will be useful to view every edge of G as an *ordered pair* (u, v), $u, v \in U$. We also assume (without loss of generality) that G is connected.

First, we construct a DNF Ψ_G as follows. We interpret every vertex $u \in U$ as a Boolean variable (bearing the same name) and, for every edge $(u, v) \in \mathscr{A}$, we create r new Boolean variables $x_1^{uv}, \dots, x_r^{uv}$. We then define

$$\Psi_{uv} = \Phi_\pi(x_1^{uv}, \dots, x_r^{uv}, u, v), \tag{11}$$

and

$$\Psi_G = \bigvee_{(u,v) \in \mathscr{A}} \Psi_{uv} = \bigvee_{(u,v) \in \mathscr{A}} \Phi_\pi(x_1^{uv}, \dots, x_r^{uv}, u, v). \tag{12}$$

(Observe that each variable x_i^{uv} appears in only one of the DNFs Ψ_{uv}, whereas the variables u and v may be common to several of these DNFs.)

Next, we consider nK independent copies of G, say $G_1, G_2, \dots, G_{nK}$, and the corresponding copies of Ψ_G, say $\Psi_{G_1}, \Psi_{G_2}, \dots, \Psi_{nK}$. We define

$$\Psi = \bigvee_{i=1}^{nK} \Psi_{G_i}. \tag{13}$$

(Thus, the DNF Ψ consists of $nK|\mathscr{A}|$ copies of Φ_π.)

Let now S^* be a smallest VD set of Ψ and let C^* be a smallest vertex cover of G. We are going to establish the theorem by proving that, for any $l \in \{1, \dots, n\}$,

$$|C^*| \leqslant l \quad \text{if and only if} \quad |S^*| \leqslant nKl. \tag{14}$$

(*Only if*) Perform the following operations on Ψ_G:

(i) delete C^* (viewed as a set of variables), and

(ii) delete all but one copy of Φ^k, for $k = 2, \dots, t$.

Denote the resulting DNF by Ψ_G^*.

After operation (i), since each edge of G has at least one endpoint in C^*, each copy of Φ^1 in (12) has lost either the variable u (corresponding to c), or the variable v (corresponding to d), or both. On the other hand, due to operation (ii), Ψ_G^* and Φ_π both contain exactly one copy of components $\Phi^2, \dots, \Phi^t$.

These observations imply that $\beta(\Psi_G^*) <_L \beta(\Phi_\pi)$. Thus, by choice of Φ_π, any disjunction of independent copies of Ψ_G^* must satisfy property π.

Now, consider the set S obtained by taking nK copies of C^* (one for each copy of Ψ_G in (13)). We claim that S is a variable deletion set of Ψ. Indeed, after deleting S from Ψ, we obtain a DNF which can alternatively be produced by deleting terms from $nK|\mathscr{A}|$ independent copies of Ψ_G^* (we need $nK|\mathscr{A}|$ copies here, in order to restore the copies of $\Phi^2, \ldots, \Phi^t$ removed in operation (ii)). By our previous conclusions, and because property π is hereditary under deletion of terms, this implies that S is a VD set of Ψ for π.

Since $|S| = nK|C^*|$, we have established the "only if" part of (14).

(*If*) Let us now suppose that the optimal vertex cover C^* satisfies

$$l + 1 \leqslant |C^*| \leqslant n - 1, \tag{15}$$

for some $l \in \{1, \ldots, n-2\}$.

For $i = 1, \ldots, nK$, denote by $\Psi^*_{G_i}$ the DNF obtained by deleting S^* from Ψ_{G_i}. Then, deleting S^* from Ψ yields the DNF

$$\Psi^* = \bigvee_{i=1}^{nK} \Psi^*_{G_i} \tag{16}$$

(see (13)). Since Ψ^* has property π, whereas K independent copies of Φ_π violate π, there must be at least $(n-1)K + 1$ indices $i \in \{1, \ldots, nK\}$ such that Φ_π cannot be derived from $\Psi^*_{G_i}$ by deletion of terms. Let I be the set of all such indices:

$$|I| \geqslant (n-1)K + 1. \tag{17}$$

Consider any index $i \in I$. For notational simplicity, assume that Ψ_{G_i} is given by (11)–(12). By choice of i, S^* must contain at least one variable from each copy of Φ_π on the right-hand side of (12). In other words, for each $(u, v) \in \mathscr{A}$, S^* contains either u or v or one of the variables $x_1^{uv}, \ldots, x^{uv}$. This immediately implies that the following set S_i^* is a vertex cover of G_i:

$$S_i^* = \bigcup_{\substack{(u,v) \in \mathscr{A}: \\ S^* \cap (u,v) \neq \emptyset}} (S^* \cap \{u, v\}) \cup \bigcup_{\substack{(u,v) \in \mathscr{A}: \\ S^* \cap (u,v) = \emptyset}} \{u\}$$

(if either u or v is in S^*, then we take it; otherwise, we take arbitrarily u).

From this definition, it follows that S_i^* does not contain more variables of Ψ_{G_i} than S^*. Moreover, the sets S_i^*, $i \in I$, are pairwise disjoint. Hence,

$$|S^*| \geqslant \sum_{i \in I} |S_i^*|.$$

Now, since each S_i^* is a vertex cover, we get in view of (15) and (17):

$$\begin{aligned} |S^*| &\geqslant [(n-1)K + 1]\,|C^*| \geqslant [(n-1)K + 1](l+1) \\ &= nKl + l + 1 + K(n - 1 - l) > nKl. \end{aligned}$$

This establishes the "if" part of (14) and the theorem. $\square$

Observe that Theorem 4 generalizes results in [6] and [14], where it is shown that the VD set problem for the "renamable Horn" property is NP-hard.

Theorem 5. *Finding a minimum cardinality variable fixation control set is* NP-*hard for every DNF property in class* $\mathscr{C}$.

Proof. A proof almost identical to that of Theorem 4 works here as well.

(*Only if*) Using Proposition 1, the VD set S constructed above is also a VF set.

(*If*) Let now S^* be a smallest VF set of Ψ (see 13) and Ψ^* be the DNF obtained from Ψ by fixing the variables in S^* to arbitrary 0–1 values. Following with the exact same arguments as in the necessity proof of Theorem 4 will lead to the same contradiction. □

Theorem 6. *Finding a minimum cardinality weak variable fixation control set is* NP-*hard for every DNF property in class* $\mathscr{C}$.

Proof. Trivially follows from the proof of Theorem 5, since weak VF sets generalize VF sets. □

5. Term deletion control set problems

We start our discussion of TD set problems with an easy observation:

Theorem 7. *For every term induced DNF property* π, *the unique minimum cardinality term deletion control set for* π *is the set of terms not having property* π.

Proof. This is obvious. □

The above result implies that, for every term induced property, the TD set problem is no more difficult than the problem of verifying whether a single term has the property or not (problem P_0 in Section 3). By contrast, for DNF properties which are not term induced, finding control sets of terms may become hard. We will illustrate this for three specific properties.

First we prove a simple lemma.

Lemma 8. *A quadratic DNF* Φ *without linear terms is renamable Horn if and only if the equation* $\Phi = 0$ *is consistent.*

Proof. A quadratic DNF Φ without linear terms is Horn if and only if the point $x = (0, 0, \ldots, 0)$ is a solution of $\Phi(x) = 0$. The result directly follows from this observation. □

Theorem 9. *Finding a minimum cardinality term deletion control set for the property "renamable Horn" is* NP-*hard.*

Proof. Due to the preceding lemma, this result is a corollary of the NP-completeness of *MAX 2-SAT* (i.e., given as input a quadratic DNF Φ and a positive integer k, does Φ have a TD set of size k for the property "satisfiable"?). □

Theorem 10. *Finding a minimum cardinality term deletion control set for the property "monotone" is* NP-*hard.*

Proof. We provide a polynomial transformation from the vertex cover problem to our problem.

Consider an instance $G = (U, \mathscr{A})$ of the vertex cover problem, with vertex set $U = \{u_1, \dots, u_n\}$ and edge set $\mathscr{A}$. We now construct a DNF Φ, which we will consider as input to the TD control set problem for monotonicity. Every edge $(u_i, u_j) \in \mathscr{A} (i < j)$ corresponds to a variable x_{ij}, and every vertex $u_i \in U$ gives rise to a term T_i of Φ. More precisely, for every edge $(u_i, u_j) \in \mathscr{A} (i < j)$, we include literal x_{ij} in term T_i and literal $\bar{x}_{ij}$ in term T_j. Thus,

$$\Phi = \bigvee_{i=1}^{n} \left(\bigwedge_{(u_i, u_j) \in \mathscr{A} \mid i < j} x_{ij} \bigwedge_{(u_k, u_i) \in \mathscr{A} \mid k < i} \bar{x}_{ki} \right).$$

As a result, each variable appears twice in Φ, once complemented and once uncomplemented.

If we let S^* be a smallest TD set of Φ for the monotonicity property and C^* be a smallest vertex cover of G, there immediately follows from the construction that $|S^*| = |C^*|$. □

Theorem 11. *Finding a minimum cardinality term deletion control set for the property "q-Horn" is NP-hard.*

Proof. We shall proceed by reducing 3-*SAT* to our problem. Indeed, let Φ be an input DNF for 3-*SAT* and assume without loss of generality that each term T_i has exactly 3 literals. Let $T_i = a_i b_i c_i$ where each a_i, b_i and c_i represents either a variable or its complement ($i = 1, \dots, m$). Consider the following DNF Φ' as input to the TD control set problem for the property q-Horn (where e_i, $i = 1, \dots, m$, are new variables):

$$\begin{aligned}\Phi' = \bigvee_{i=1}^{m} (&a_ib_i \vee a_ic_i \vee b_ic_i \vee a_ib_ic_i \vee \bar{a}_i\bar{b}_i\bar{c}_i \vee \bar{a}_i\bar{b}_i\bar{c}_ie_i \vee a_ib_ie_i \vee a_ib_i\bar{e}_i \vee a_ic_ie_i \vee a_ic_i\bar{e}_i \\ &\vee b_ic_ie_i \vee b_ic_i\bar{e}_i \vee a_i\bar{b}_ie_i \vee a_i\bar{b}_i\bar{e}_i \vee a_i\bar{c}_ie_i \vee a_i\bar{c}_i\bar{e}_i \vee b_i\bar{c}_ie_i \vee b_i\bar{c}_i\bar{e}_i \vee \bar{a}_ib_ie_i \\ &\vee \bar{a}_ib_i\bar{e}_i \vee \bar{a}_ic_ie_i \vee \bar{a}_ic_i\bar{e}_i \vee \bar{b}_ic_ie_i \vee \bar{b}_ic_i\bar{e}_i \vee \bar{a}_i\bar{b}_ie_i \vee \bar{a}_i\bar{b}_i\bar{e}_i \vee \bar{a}_i\bar{c}_ie_i \vee \bar{a}_i\bar{c}_i\bar{e}_i \\ &\vee \bar{b}_i\bar{c}_ie_i \vee \bar{b}_i\bar{c}_i\bar{e}_i).\end{aligned}$$

Table 1 lists all possible half-integral values of $\alpha(a_i)$, $\alpha(b_i)$, $\alpha(c_i)$ and $\alpha(e_i)$ (up to obvious symmetries), as well as the number of inequalities of type (2) can be satisfied in the system of inequalities (2)–(4).

We claim that $16m$ of the inequalities of type (2) that they satisfy simultaneously if and only if the equation $\Phi = 0$ is consistent. Indeed, if there exists a solution of $\Phi = 0$,

Table 1

$\alpha(a_i)$	$\alpha(b_i)$	$\alpha(c_i)$	$\alpha(e_i)$	Number of inequalities satisfied
0	0	0	0	16
			$\frac{1}{2}$	10
			1	16
0	0	$\frac{1}{2}$	0	12
			$\frac{1}{2}$	14
			1	12
0	0	1	0	16
			$\frac{1}{2}$	10
			1	16
0	$\frac{1}{2}$	$\frac{1}{2}$	0	12
			$\frac{1}{2}$	12
			1	12
0	$\frac{1}{2}$	1	0	10
			$\frac{1}{2}$	12
			1	10
0	1	1	0	16
			$\frac{1}{2}$	9
			1	15
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	15
			$\frac{1}{2}$	3
			1	15
$\frac{1}{2}$	$\frac{1}{2}$	1	0	11
			$\frac{1}{2}$	10
			1	10
$\frac{1}{2}$	1	1	0	10
			$\frac{1}{2}$	12
			1	9
1	1	1	0	14
			$\frac{1}{2}$	8
			1	14

then at least one of the literals a_i, b_i or c_i must be set to zero in this solution for each $i = 1, \ldots, m$. If we set $\alpha(a_i) = a_i$, $\alpha(b_i) = b_i$, $\alpha(c_i) = c_i$ and $\alpha(e_i) = 0$, then it can be checked from Table 1 that we obtain an assignment satisfying $16m$ of the inequalities (2). Conversely, no matter what values are assigned to $\alpha(a_i)$, $\alpha(b_i)$, $\alpha(c_i)$ and $\alpha(e_i)$, at most 16 inequalities can be satisfied simultaneously for $i = 1, \ldots, m$. Moreover if all $\alpha(a_i) = \alpha(b_i) = \alpha(c_i) = 1$ or if at least one of these is assigned a value $\frac{1}{2}$, then at most 15 inequalities can be satisfied for $i = 1, \ldots, m$.

Therefore, the identification of a minimum cardinality term deletion control set of Φ' for the q-Horn property provides a solution of the equation $\Phi = 0$, or proves that the equation is inconsistent. □

Acknowledgements

The authors gratefully acknowledge the partial support by the Office of Naval Research (grants N00014-92-J1375 and N00014-92-J4083), by the Air Force Office of Scientific Research (grant F49620-93-1-0041), by NATO (grant CRG 931531) and by NSF (grant INT 9321811).

References

[1] E. Boros, Y. Crama and P.L. Hammer, Polynomial-time inference of all valid implications for Horn and related formulae, Ann. Math. Artificial Intelligence 1 (1990) 21–32.
[2] E. Boros, Y. Crama, P.L. Hammer and M. Saks, A complexity index for satisfiability problems, SIAM J. Comput. 23 (1994) 45–49.
[3] E. Boros, P.L. Hammer and X. Sun, Recognition of q-Horn formulae in linear time, Discrete Appl. Math. 55 (1994) 1–13.
[4] R.K. Brayton, G.D. Hachtel, C.T. McMullen and A.L. Sangiovanni-Vincentelli, Logic Minimization Algorithms for VLSI Synthesis (Kluwer, Dordrecht, 1984).
[5] M. Buro and H. Kleine Büning, Report on a SAT competition, Technical Report 'Informatik Nr. 110', Universität Paderborn, 1992.
[6] V. Chandru and J.N. Hooker, Detecting embedded Horn structure in propositional logic, Inform. Process. Lett. 42 (1992) 109–111.
[7] G. Gallo and G. Urbani, Algorithms for testing the satisfiability of propositional formulae, J. Logic Programming 7 (1989) 45–61.
[8] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness (Freeman, San Francisco, CA, 1979).
[9] M.R. Garey, D.S. Johnson and L. Stockmeyer, Some simplified NP-complete graph problems, Theoretical Comput. Sci. 1 (1976) 237–267.
[10] P.L. Hammer and S. Rudeanu, Boolean Methods in Operations Research (Springer, Berlin, 1968).
[11] H.R. Lewis, Renaming a set of clauses as a Horn set. J. ACM 25 (1978) 134–135.
[12] J.M. Lewis and M. Yannakakis, The node-deletion problem for hereditary properties is NP-complete, J. Comput. System Sci. 20 (1980) 219–230.
[13] S. Muroga, Threshold Logic and its Applications (Wiley-Interscience, New York, 1971).
[14] K. Truemper, Monotone decomposition of matrices, Technical Report UTDCS-1-94, 1994.
[15] M. Yannakakis, Node- and edge-deletion NP-complete problems. Proceedings of Tenth Annual ACM Symposium on Theory of Computing (1978) 253–264.