Journal on Satisfiability, Boolean Modeling and Computation 4 (2007) 1-14

Disjoint DNF Tautologies with Conflict Bound Two

Balázs Szörényi szorenyi@inf.u-szeged.hu
Hungarian Academy of Sciences and University of Szeged
Research Group on Artificial Intelligence
Szeged, Hungary

Abstract

Many aspects of the relation of different decision tree and DNF complexity measures of Boolean functions have been more or less substantially explored. This paper adds a new detail to the picture: we prove that DNF tautologies with terms conflicting in one or two variables pairwise possess a tree-like structure. An equivalent reformulation of this result (adopting the terminology of [7, 8, 9]) is the following. Call a clause-set (or CNF) a *hitting clause-set* if any two distinct clauses of it clash in at least one literal, and call a hitting clause-set an *at-most-d hitting clause-set* if any two clauses of it clash in at most d variables. If now an at-most-2 hitting clause-set Φ is unsatisfiable (as a CNF), then, by the above result, there must exist a variable occurring (negated or unnegated) in each clause of Φ.

KEYWORDS: *disjoint DNF tautologies, decision trees, unsatisfiable hitting clause sets, read-once resolution refutation, conflict bound*

Submitted June 2007; revised September 2007; published December 2007

1. Introduction

A decision tree naturally encodes a DNF tautology—each term of which correspond to a unique leaf of the tree—, which holds the following special properties (for the formal definitions see the next section):

(a) the terms are pairwise conflicting: for each pair there exists at least one variable appearing negated in one of them, and unnegated in the other; and

(b) the terms possess a hierarchical structure: there is a variable x that appears in each of them; there is a variable y that appears in every term containing literal x and there is a variable z that appears in every term containing literal $\overline{x}$ (y and z may be identical); and so on.

Such DNFs are called *decision tree generated DNFs*, or DT-DNFs for short[1.]; meanwhile DNFs possessing property (a) but not necessarily property (b) are called *disjoint DNFs*, or DDNFs. The question thus naturally arises, how special do these properties make a decision tree, regarding complexity. (It is important to note that this is a purely complexity question, based on the syntax of the above classes.) This question was investigated by Lovász *et al.* in [10]. More precisely they were interested in the following problem: given a DNF tautology

1. DT-DNFs also have a recursive definition, see Section 2.

F, the task is to construct a decision tree such that for each term of the DNF generated by it there is a term of F that is a subterm of it. They have shown that for some very "small" DNF tautologies this problem can be solved only with "extremely large" decision trees[2].

On the other hand, as it has been proved by Kullmann [8, 9] (and, independently by Sloan *et al.* [13]), when restricting the DNFs to the subclass possessing property (a) (i.e., the class of DDNFs), *and* further bounding the number of conflicts between the terms to one (i.e., for each pair of terms there is *exactly* one variable appearing negated in one of them and unnegated in the other), it turns out that the resulting class consists of DNFs that can all be generated by decision trees:

Theorem 1 ([8, 9] and [13])**.** *If F is a DDNF tautology with terms conflicting in exactly one variable pairwise, then F is a DT-DNF.*

This problem arose in connection with characterizing strongly minimal tautologies[3] with the additional property that the number of terms is one more than the number of variables [8, 9] (Aharoni and Linial [1], Davydov *et al.* [4], Kullmann [7]), and also in connection with maximal DNFs[4] [13]. This DDNF class comes up in other context as well, for example in connection with the complexity of analytic tableaux (Urquhart [15], referring to earlier unpublished work of Cook, and Arai *et al.* [2]).

In this paper we give a strengthening of the above result, showing that the conflict bound can be relaxed to two:

Theorem 2. *If F is a DDNF tautology with terms conflicting in one or two variables pairwise, then F is a DT-DNF.*

Example 3. *The DNF*[5]

$$F_{\text{ex}3} = \overline{x_2}\,\overline{x_4} \vee x_2\overline{x_3}\,\overline{x_4} \vee x_2x_3\overline{x_4} \vee \overline{x_1}x_4 \vee x_1\overline{x_2}\,\overline{x_3}x_4 \vee x_1x_2\overline{x_3}x_4 \vee x_1x_3x_4$$

is a DDNF with conflict bound two, and Figure 2 proves that it is also a DT-DNF—which is also apparent writing $F_{\text{ex}3}$ in the form

$$F_{\text{ex}3} = \overline{x_4}\,\overline{x_2} \vee \overline{x_4}x_2\overline{x_3} \vee \overline{x_4}x_2x_3 \vee x_4\overline{x_1} \vee x_4x_1\overline{x_3}\,\overline{x_2} \vee x_4x_1\overline{x_3}x_2 \vee x_4x_1x_3,$$

or also from Figure 1, visualizing the relations of the truth sets (set of satisfying assignments, denoted by $\mathcal{T}(\cdot)$) of the various terms.

Note however that the result of Theorem 2 does not generalize to conflict bound three, as the following example demonstrates.

2. They measure the complexity by the *depth* of the DNF (resp. decision tree), which is the maximal number of literals appearing in a term of the given DNF (resp. of the DT-DNF generated by the tree). What they show is that for some constant depth DNFs one needs decision trees of depth linear (thus maximal) in the number of variables.
3. A DNF tautology is *strongly minimal* if deleting any term of it, or adding any literal to a term of it results in a non-tautology.
4. A DNF consisting of t terms can have at most $2^t - 1$ prime implicants; a DNF having t terms and $2^t - 1$ prime implicants is called *maximal*.
5. Note that, for simplicity (and following the conventions) we omit the "$\wedge$" signs when giving a term in an explicit form. For example, for the term $x_1 \wedge x_3 \wedge \overline{x_5} \wedge x_7$ we write $x_1x_3\overline{x_5}x_7$.

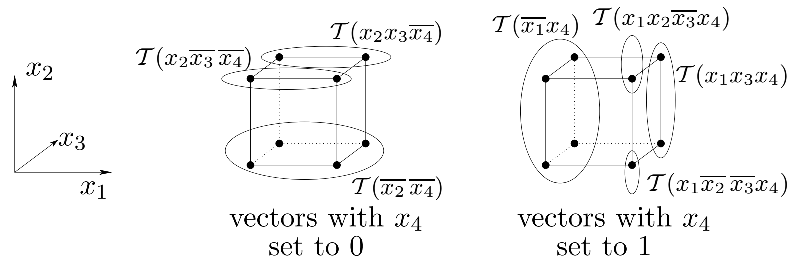

Figure 1. The assignments to variables x_1, x_2, x_3 and x_4 represented as the vertices of the 4-dimensional hypercube and grouped according to which term of $F_{\text{ex}3}$ they satisfy.

Example 4. *DDNF* $F_{\text{ex}4} = x_1x_3 \vee \overline{x_1}x_2 \vee \overline{x_2}\,\overline{x_3} \vee \overline{x_1}\,\overline{x_2}x_3 \vee x_1x_2\overline{x_3}$ *is a tautology and has terms conflicting in at most three variables pairwise, but is not a DT-DNF. (Simply note that there is no variable that appears in every term.)*

In [13] it was also asked, what is the value of α_n, defined by

$$\alpha_n = \min_{F:F \text{ is a DDNF tautology over } n \text{ variables}} \max_{1\le j\le n} v_{F,j}, \tag{1}$$

where

$$v_{F,j} = \sum \left\{2^{-|T|} : T \text{ is a term of } F, x_j \text{ or } \overline{x_j} \text{ appears in } T\right\}.$$

(A DDNF tautology F can be thought of as partitioning the n-dimensional hypercube $\{0,1\}^n$ into smaller dimensional subcubes—again, see Figure 1—, and $v_{F,j}$ as the total volume of the cubes in this partition that are contained in either one of the half cubes obtained by splitting the hypercube along the variable x_j.) In [13] it was proved that

$$\frac{\log n - \log\log n}{n} \le \alpha_n \le O\left(n^{-1/5}\right).$$

(Their upper bound follows from a construction of Savický and Sgall [12].) Considering DDNFs with smaller conflict bound, Sloan et al. also introduced the notation α_n^d, which is the same as α_n except that F is restricted to DDNFs with conflict bound d. Theorem 1 then implies $\alpha_n^1 = 1$, which is also stregthened by Theorem 2 to $\alpha_n^2 = 1$ (for arbitrary positive integer n).

1.1 Related work

1.1.1 CNFs and hitting clause-sets

A clause-set (or CNF) is a *hitting clause-set* if any two distinct clauses of it clash in at least one variable (i.e., the variable occurs unnegated in one of them and negated in the other). Let $\mathcal{HIT}$ denote the set of hitting clause-sets, and $\mathcal{HIT}_{\le d}$ the hitting clause-sets in which no two clauses clash in more than d variable. Note that for each clause in $\mathcal{HIT}_{\le 1}$ it holds that any two distinct clauses of it clash in *exactly* one variable; accordingly we also use $\mathcal{HIT}_1$ to denote this class (which is the notation used for this class in [7, 8, 9]).

Call a clause-set *saturated minimally unsatisfiable* if removing any of its clauses or adding any literal to one of its clauses results in a satisfiable clause-set. Denote the set of unsatisfiable clause-sets by $\mathcal{USAT}$, and the set of saturated minimally unsatisfiable clause-sets by $\mathcal{SMUSAT}$.

Obviously, the notion of hitting clause-set is the dual of the DDNF, $\mathcal{HIT}_{\leq d}$ is the dual of DDNFs with conflict bound d, tautology is the dual of unsatisfiable clause-sets, and saturated minimally unsatisfiable clause formula is the notion of strongly minimal DNF tautologies (introduced previously). Finally, the dual of DT-DNF is based on the notion of *read-once resolution refutation*, an incomplete restriction of resolution in which each clause can be used at most once [3, 5]. Consequently a read-once resolution refutation has a tree structure, and it is also easy to see that hitting clause-sets having read-once resolution refutations are exactly those CNFs which can be obtained by negating some DT-DNF.

Reformulating Theorem 1, it is shown that any clause-set in $\mathcal{USAT} \cap \mathcal{HIT}_1$ (which obviously equals to the set $\mathcal{USAT} \cap \mathcal{HIT}$) contains some variable occurring (either negated or unnegated) in each clause of it. Reformulating the result of this paper (Theorem 2): this also holds for clause-sets in $\mathcal{HIT}_{\leq 2} \cap \mathcal{USAT}$; that is if $\Phi \in \mathcal{HIT}_{\leq 2} \cap \mathcal{USAT}$, then there is a variable occurring (negated or unnegated) in each clause of Φ; what is more, Φ even has a read-once resolution refutation.

1.1.2 DNF and DT complexity for Boolean functions

Finally we note that a related problem is to represent a Boolean function f as a DNF or as a decision tree—that is, when one needs to construct a DNF tautology (resp. decision tree) with each term (resp. with each term of the corresponding DT-DNF) covering only assignments that satisfy f, or only assignments that falsify f—, and one is interested in comparing the complexity of the two class in this setting. See for example [6, 11, 14].

2. Preliminaries

We use standard notations from propositional logic such as variable, literal, assignment, term (or conjunction), subterm, DNF, equivalence of formulas, etc. Throughout let n denote the number of variables in our universe.

In the paper both the syntactical and semantical view is used, switching frequently between the two. For this, we first discuss the two separately, and then discuss some connections of the two used heavily later on.

2.1 Syntax

For some term T let $\mathrm{Vars}(T)$ denote the set of variables appearing in T. (For example $\mathrm{Vars}(x_1 x_3 \overline{x_5} x_7) = \{x_1, x_3, x_5, x_7\}$.) Terms T and T' *conflict* in variable x if x appears unnegated in one of them, and negated in the other; $T \otimes T'$ denotes the set of variables T and T' conflict in (e.g., if $T = x_1 x_3 \overline{x_5} x_7$ and $T' = x_1 x_2 \overline{x_3} x_5 x_9$ then $T \otimes T' = \{x_3, x_5\}$).

A term is sometimes considered as a set of literals, and a DNF as a set of terms. Accordingly, for some DNF F and variable x, $T \in F$ is used to denote that T is a term of the DNF F, and $x \in T$ (resp. $\overline{x} \in T$) that x (resp. $\overline{x}$) is a literal in term T. (For example, considering terms $T_1 = x_1 x_2$, $T_2 = \overline{x_2} x_3$ and $T_3 = x_1 x_3$, then $x_2 \in T_1$, but $x_2 \notin T_2$, and, of course $x_2 \notin T_3$.) A *disjoint DNF form formula*, or DDNF for short, is a DNF with pairwise conflicting terms. A DDNF formula F has *conflict bound* d if for arbitrary terms $T, T' \in F$ it holds that $|T \otimes T'| \leq d$ (i.e., any two term of F conflict in at most d variables).

A *decision tree* (or DT for short) is a rooted binary tree such that for each inner node the edge leading to its right (resp. left) child is labeled "x" (resp. "$\overline{x}$") for some variable x. (For an example see Figure 2 or Figure 7.) In a decision tree a path from the root to a leaf naturally determines a term obtained by simply conjuncting the literals appearing in the labels of the edges along the path. Thus, given a decision tree, the terms corresponding to its leaves put up a DDNF tautology. Such DDNF tautologies are called *decision tree generated DNF*s, or DT-DNFs for short. Alternatively, one can define DT-DNFs as the smallest subset $\mathcal{DT}\text{-}\mathcal{DNF}$ of the set of DNFs satisfying:

- If x is a variable, then the DNF $x \vee \overline{x}$ is an element of $\mathcal{DT}\text{-}\mathcal{DNF}$.
- If x is a variable and both $T_1 \vee \cdots \vee T_k$ and $T'_1 \vee \cdots \vee T'_\ell$ are elements of $\mathcal{DT}\text{-}\mathcal{DNF}$, then the DNF $(x \wedge T_1) \vee \cdots \vee (x \wedge T_k) \vee (\overline{x} \wedge T'_1) \vee \cdots \vee (\overline{x} \wedge T'_\ell)$ is also an element of $\mathcal{DT}\text{-}\mathcal{DNF}$.

2.2 Semantics

Given an assignment α, its *weight* is defined to be the number of variables it assigns 1 to.

Given some truth assignment α of the variables $x_1, \ldots, x_n$, and some index $1 \leq i \leq n$, $\alpha^{(i)}$ denotes the truth assignment obtained from α by flipping the value it assigns to x_i.

2.3 Connecting syntax and semantics

We say that a term T is *consistent* with partial assignment a, if for every variable $x \in \mathrm{Vars}(T)$ it holds that, if x appears negated (resp. unnegated) in T, then either a assigns no value to x, or a assigns 0 (resp. 1) to it. An assignment α is said to *satisfy* term T if α (as a partial assignment) is consistent with T. Accordingly, the *truth set* of a term T, denoted by $\mathcal{T}(T)$ is the set of assignments satisfying T.

For a DDNF tautology F and an assignment α there is a *unique* term of F satisfied by α; denote it by $T^{\alpha,F}$. When it causes no ambiguity, F is omitted and T^α is used instead. Now, if for some term T and index i it holds that $\alpha^{(i)} \in \mathcal{T}(T)$ but $\alpha \notin \mathcal{T}(T)$ (i.e., if $\alpha^{(i)}$ satisfies T but α does not), then

- if α assigns 0 to x_i then $x_i \in T$ and $\overline{x_i} \in T^\alpha$,
- otherwise $\overline{x_i} \in T$ and $x_i \in T^\alpha$.

3. Proof of Theorem 2

Theorem 2 is proved by induction on the number of terms in F. In case F contains one or two terms, the statement is obvious. Now we show that F is a DT-DNF, assuming:

Induction hypothesis: DDNF F with conflict bound two contains $t \geq 3$ terms, and the statement holds for any DDNF tautology with conflict bound two having less than t terms. (2)

Let T be an arbitrary term of F. Assume without loss of generality that $T = x_1 \cdots x_k$. Of course, if F is a DT-DNF, then for some $1 \le i \le k$ F has a subformula equivalent to $x_1 \cdots x_{i-1} x_{i+1} \cdots x_k$; namely the one induced by the parent node of the leaf corresponding to T. (For example if $F = F_{\text{ex}3}$ from Example 3 and $T = x_1 x_3 x_4$, then $i = 3$, and the subformula $x_1 \overline{x_2}\,\overline{x_3} x_4 \vee x_1 x_2 \overline{x_3} x_4 \vee x_1 x_3 x_4$ of F is equivalent to $T \setminus \{x_i\} = x_1 x_4$.) The next Claim considers the reverse of this implication. (Also, for an example demonstrating the claim see Example 6.)

Claim 5. *Assume (2), and let $T = x_1 \cdots x_k$ be a term of F. Suppose that for some $i \in \{1, \ldots, k\}$ it holds that every term in F that conflicts with T only in x_i contains $x_1 \cdots x_{i-1} x_{i+1} \cdots x_k$ as a subterm. Then F is a DT-DNF.*

Proof. Consider the following sets

$$\begin{aligned}
S_1 =& \left\{ \alpha \in \{0,1\}^n : \alpha^{(i)} \in \mathcal{T}(T) \right\}, \\
S_2 =& \mathcal{T}(x_1 \cdots x_{i-1} \overline{x_i} x_{i+1} \cdots x_k), \\
S_3 =& \cup_{T' \in F : x_1 \cdots x_{i-1} \overline{x_i} x_{i+1} \cdots x_k \text{ is a subterm of } T'} \mathcal{T}(T'), \\
S_4 =& \cup_{T' \in F : T \otimes T' = \{x_i\}} \mathcal{T}(T').
\end{aligned}$$

Then $S_1 = S_2$ and $S_2 \supseteq S_3$ always hold, and $S_3 \supseteq S_4$ follows from the condition of the Claim. However, $S_4 \supseteq S_1$ is also true because

- since F is a tautology, each element β of S_1 appears in some $\mathcal{T}(T')$ for some $T' \in F$—recall that this T' is the term we denote as T^β—, and
- since F is a DDNF, each of these T^β terms must conflict with T in some variable. But this variable must be x_i, and only x_i, as the first k bit of each $\beta \in S_1$ is 1, *except for the i-th bit.*

Thus all of the above sets are identical. Then defining

$$F_1 := \{T' \in F : x_1 \cdots x_{i-1} \overline{x_i} x_{i+1} \cdots x_k \text{ is a subterm of } T'\}$$

and

$$F_2 := \big(F \setminus (F_1 \cup \{T\})\big) \cup \{x_1 \cdots x_{i-1} x_{i+1} \cdots x_k\}$$

it holds that both $F_1' := \{T' \setminus \{x_1, \cdots, x_{i-1}, \overline{x_i}, x_{i+1}, \cdots, x_k\} : T' \in F_1\}$ and F_2 are DDNF tautologies (because of the $S_2 = S_3$ equality). Furthermore both have less terms then F, thus by the induction hypothesis both are DT-DNFs. This immediately implies the Claim: pick a DT τ_1 for F_1' and a DT τ_2 for F_2, expand τ_1 to a decision tree for $x_i \vee \{\overline{x_i} \wedge T' : T' \in F_1'\}$ in the natural way, and paste it into τ_2 in the place of the leaf corresponding to $x_1 \cdots x_{i-1} x_{i+1} \cdots x_k$. □

Example 6. *Demonstrating Claim 5, let $F = F_{\text{ex}3}$ from Example 3 and let $T = x_1 x_3 x_4$. Then $i = 3$, $F_1 = x_1 \overline{x_2}\,\overline{x_3} x_4 \vee x_1 x_2 \overline{x_3} x_4$, $F_1' = \overline{x_2} \vee x_2$ and $F_2 = (\overline{x_2}\,\overline{x_4} \vee x_2 \overline{x_3}\,\overline{x_4} \vee x_2 x_3 \overline{x_4} \vee \overline{x_1} x_4) \vee x_1 x_4$. See also Figure 2 for the decision tree τ_1 (resp. τ_2) for F_1' (resp. F_2).*

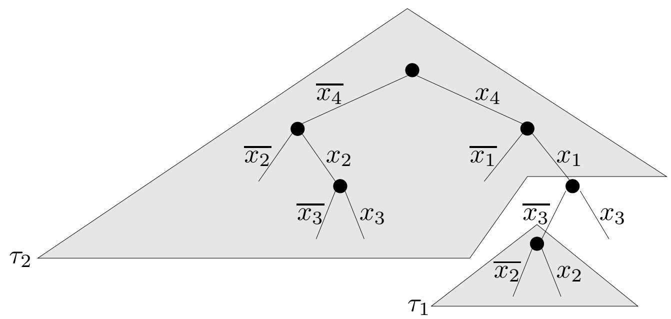

Figure 2. Marking τ_1 and τ_2 on the decision tree generating F_{ex3} from Example 3.

Defining the following directed graph $G(V,E) = G_{F,T}(V_{F,T}, E_{F,T})$:

$$\begin{aligned} V &= \{T' \in F : |T \otimes T'| = 1 \text{ and } \mathrm{Vars}(T') \not\supseteq \mathrm{Vars}(T)\}, \\ E &= \{(T', T'') \in V^2 : \overline{x_i} \in T' \text{ and } x_i \notin \mathrm{Vars}(T'') \text{ for some } 1 \le i \le k\}, \end{aligned} \tag{3}$$

based on Claim 5 one can give the following sufficient condition for F being a DT-DNF (which, as one can easily show, is also a necessary condition):

Claim 7. *Assume (2), let $T = x_1 \cdots x_k$ be a term of F, and let $G = G_{F,T}$ be the graph defined as in (3). If G contains no cycle, then F is a DT-DNF.*

Proof. We show that if F is not a DT-DNF, then G contains a cycle. Suppose thus that F is not a DT-DNF. By Claim 5 this can only be if for $i = 1, \ldots, k$ there is a term $T_i \in F$ containing $\overline{x_i}$, containing no other variable from T negated, and having at least one of the variables in T missing. Consequently $T_1, \ldots, T_k \in V$, and in the subgraph induced by them, each vertex has indegree at least one. The subgraph has thus no sink, implying that it contains a cycle. (For example if $F = F_{\text{ex4}}$ from Example 4 and $T = x_1x_3$, then V consists of the terms $T_1 = \overline{x_1}x_2$ and $T_2 = \overline{x_2}\,\overline{x_3}$, and there is an edge in E both from T_1 to T_2 and from T_2 to T_1—and thus G contains a cycle[6]: T_1, T_2, T_1.) □

In the rest of the paper we show that G indeed contains no cycle. Assume for the contradiction that this is not the case, and let $T_1, \ldots, T_\ell, T_1$ be a cycle of minimal length (then of course $\ell \le k$), and assume without loss of generality that $\overline{x_i} \in T_i$, $i = 1, \ldots, \ell$. (Note that no other variable of T appears unnegated in T_i, as $T_i \in V$.) Then for any distinct indices $i, j \in \{1, \ldots, \ell\}$,

- if T_j follows T_i in the cycle[7], then $x_i \notin T_j$ (by the construction of E),
- if not, then $x_i \in T_j$, as otherwise $(T_i, T_j) \in E$, which would shortcut the cycle, and contradict that it is of minimal length.

These observations are summarized in Figure 3.

6. Which is in accordance with the fact that F_{ex4} is not a DT-DNF.
7. That is, $j = i + 1$ if $i < \ell$, and $j = 1$ if $i = \ell$.

	x_1	x_2	x_3	x_4	$\cdots$	$x_{\ell-2}$	$x_{\ell-1}$	x_ℓ
T	$+$	$+$	$+$	$+$	$\cdots$	$+$	$+$	$+$
T_1	$-$	$+$	$+$	$+$	$\cdots$	$+$	$+$	$\cdot$
T_2	$\cdot$	$-$	$+$	$+$	$\cdots$	$+$	$+$	$+$
T_3	$+$	$\cdot$	$-$	$+$	$\cdots$	$+$	$+$	$+$
T_4	$+$	$+$	$\cdot$	$-$	$\cdots$	$+$	$+$	$+$
$\vdots$					$\ddots$			
T_ℓ	$+$	$+$	$+$	$+$	$\cdots$	$+$	$\cdot$	$-$

Figure 3. The cycle $T_1, \ldots, T_\ell, T_1$. In the row of a term: "$+$" means that the given variable appears unnegated in it, "$-$" means that it appears negated in it, and " $\cdot$ " means that it does not appear in it. Consecutive elements of the cycle might conflict in other variables too, but non-consecutive elements have no more conflict.

Let us now investigate how these terms "behave" on the rest of the variables. The above observation obviously implies that if terms T_i and T_j are not consecutive elements of the cycle, then they do not conflict in variables $x_{\ell+1}, \ldots, x_n$, as otherwise they would conflict in at least three variables: x_i, x_j and $x_{\ell'}$ for some $\ell \leq \ell' \leq n$. The question is, whether two consecutive elements of the cycle can (or have to) have some further conflicts. An equivalent (semantic) formulation of this question is whether there exists a (partial) assignment to variables $x_{\ell+1}, \ldots, x_n$ consistent with the two terms. (Again, for an example demonstrating the claim see Example 9.)

Lemma 8. *Assume (2), let $T = x_1 \cdots x_k$ be a term of F with $k < n$, let $G = G_{F,T}$ defined as in (3), and let $T_1, \ldots, T_\ell$ be a cycle of minimal length in G as in Figure 3. Then there is no partial assignment for variables $x_{\ell+1}, \ldots, x_n$ that is consistent with T and all of $T_1, \ldots, T_\ell$.*

Proof. Suppose that T is of length less then n and a is a partial assignment for variables $x_{\ell+1}, \ldots, x_n$ consistent with $T, T_1, T_2, \ldots, T_\ell$. Let F' be the DDNF consisting of the terms of F that are consistent with a (thus T and $T_1, \ldots, T_\ell$ are all in F'), and from this construct F'' by removing all occurrences of variables $x_{\ell+1}, \ldots, x_n$. Then F'' is a DDNF tautology[8]. By the induction hypotheses F'' is a DT-DNF[9], consequently for some $i \in \{1, \ldots, \ell\}$ variable x_i occurs (either negated or unnegated) in every term of F'', and thus also in every term of F'—specifically in each of $T_1, \ldots, T_\ell$. But the term following T_i in the cycle contains neither x_i nor $\overline{x_i}$—a contradiction. (The condition $k < n$ is necessary since the partial assignment with empty domain is consistent with all terms.) □

Example 9. *Let $F = F_{\text{ex}3}$ from Example 3, and let $T = x_1x_3x_4$. Then V contains terms $T_1 = \overline{x_1}x_4$ and $T_2 = x_2x_3\overline{x_4}$, and E contains the edge (T_1, T_2). As F is a DT-DNF, by*

8. F'' is obviously a DDNF, since F' is a DDNF and the omitted variables occur only with one orientation in F' (i.e., are unate). To see that F'' is also a tautology consider an arbitrary assignment β. Let β' be the assignment that agrees with β on components corresponding to $x_1, \ldots, x_\ell$, and with a on the rest. Since F is a tautology, it has a term $T^{\beta'}$ satisfied by β'. By construction, $T^{\beta'}$ is consistent with a, thus removing from it all the occurences of the variables $x_{\ell+1}, \ldots, x_n$, the resulting term will be a term of F''—which is still satisfied by β', and thus also by β.
9. Here it is is used that $k < n$ and is assumed implicitly that every variable occurs in some of the terms of F.

Lemma 8 (or, more precisely, by the proof of the lemma), some variable of T (i.e., one of x_1, x_3 and x_4) must occur in T_1 and T_2—and indeed: x_4 occurs unnegated in T_1 and negated in T_2.

The next lemma rules out another case: when there is exactly one pair of consecutive elements of the cycle that conflict in two variables.

Lemma 10. *Assume (2), let $T = x_1 \cdots x_k$ be a term of F with $k < n$, let $G = G_{F,T}$ defined as in (3), and let ℓ be the length of the smallest cycle in G. Unless $\ell = 2$, there is no cycle in G of length ℓ with the property that one pair of consecutive elements of the cycle conflict in two variables, and all other consecutive pairs conflict in one.*

Proof. Assume for the contradiction that $T_1, \ldots, T_\ell, T_1$ is such a cycle in G with $\ell > 2$, and suppose that T_1 and T_ℓ are the only consecutive elements conflicting in two variables, namely in x_1 and in some $z \in \{x_{\ell+1}, \ldots, x_n\}$[10.]. Assume without loss of generality that $T_1, \ldots, T_\ell$ behave as in Figure 3, and that $z \in T_1$ and $\overline{z} \in T_\ell$. (Note that neither T nor $T_2, \ldots, T_{\ell-1}$ contains z or $\overline{z}$: if T contained z (resp. $\overline{z}$) it would conflict with T_ℓ (resp. T_1) in two variables; if any of $T_2, \ldots, T_{\ell-2}$ (resp. $T_3, \ldots, T_{\ell-1}$) contained z, it would conflict with T_ℓ (resp. T_1) in three variables; finally if T_2 (resp. $T_{\ell-1}$) contained $\overline{z}$ (resp. z), then it would conflict with T_1 (resp. T_ℓ) in two variables, contradicting the assumption of the lemma.) Then there is some partial assignment to the variables $\{x_{\ell+1}, \ldots, x_n\} \setminus \{z\}$ consistent with $T_1, \ldots, T_\ell$ and T. Denote one such by a.

	x_1	x_2	x_3	$\cdots$	$x_{\ell-2}$	$x_{\ell-1}$	x_ℓ	z
T	$+$	$+$	$+$	$\cdots$	$+$	$+$	$+$	$\cdot$
T_1	$-$	$+$	$+$	$\cdots$	$+$	$+$	$\cdot$	$+$
T_2	$\cdot$	$-$	$+$	$\cdots$	$+$	$+$	$+$	$\cdot$
$\vdots$				$\ddots$				
T_ℓ	$+$	$+$	$+$	$\cdots$	$+$	$\cdot$	$-$	$-$
α	$-$	$+$	$+$	$\cdots$	$+$	$+$	$+$	$-$
β	$+$	$+$	$+$	$\cdots$	$+$	$+$	$-$	$+$

Figure 4. The cycle $T_1, \ldots, T_\ell, T_1$. In the row of a term: "$+$" means that the given variable appears unnegated in it, "$-$" means that it appears negated in it, and " $\cdot$ " means that it does not appear in it. In the row of an assignment: "$+$" means that it assigns 1 to the given variable, "$-$" means that it assigns 0. Terms $T, T_1, \ldots, T_\ell$ do not conflict in other variables.

Let α be the assignment consistent with a assigning 0 to x_1 and z, and assigning 1 to $x_2, \ldots, x_\ell$ (see Figure 4). Then one can make the following observations:

- $\overline{x_1} \in T^\alpha$, since $\alpha \notin \mathcal{T}(T)$ and $\alpha^{(x_1)} \in \mathcal{T}(T)$,
- $\overline{z} \in T^\alpha$, since $\alpha \notin \mathcal{T}(T_1)$ and $\alpha^{(z)} \in \mathcal{T}(T_1)$
- $x_\ell \notin T^\alpha$, as otherwise T^α and T^β conflicts in three variables—where β is the assignment that is consistent with a and assigns 0 to x_ℓ and 1 to the rest of the variables—, because

10. If $\ell = 2$, then T_1 and T_ℓ does not conflict in x_1—which is the reason for handling this case separately.

  - $\overline{x_\ell} \in T^\beta$, as $\beta \notin \mathcal{T}(T)$ and $\beta^{(x_\ell)} \in \mathcal{T}(T)$,
  - $x_1 \in T^\beta$, as $\beta \notin \mathcal{T}(T_1)$ and $\beta^{(x_1)} \in \mathcal{T}(T_1)$,
  - $z \in T^\beta$, as $\beta \notin \mathcal{T}(T_\ell)$ and $\beta^{(z)} \in \mathcal{T}(T_\ell)$.

  Consequently (as T^α conflicts with T in exactly one variable and does not contain x_ℓ) $T^\alpha \in V$, furthermore $(T_\ell, T^\alpha), (T^\alpha, T_2) \in E$.

- $x_i \in T^\alpha$ for $i = 2, \ldots \ell - 1$, as otherwise $(T_i, T^\alpha) \in E$, which would mean that $T_2, \ldots, T_i, T^\alpha, T_2$ is a cycle in G shorter then ℓ—a contradiction.

But then $T^\alpha, T_2, \ldots, T_\ell, T^\alpha$ is a cycle of length ℓ (thus also of minimal length) such that all consecutive elements conflict in exactly one variable, contradicting Lemma 8. □

Based on the two previous Lemmas we can prove the following:

Lemma 11. *Assume (2), let $T = x_1 \cdots x_k$ be a term of F with $k < n$, and let $G = G_{F,T}$ defined as in (3). Then the smallest cycle in G has length at most two.*

Proof. Assume for the contradiction that $T_1, \ldots, T_\ell, T_1$ is a cycle in G of minimal length with $\ell > 2$. Assume furthermore without loss of generality that $T_1, \ldots, T_\ell$ is as in Figure 3. Then, by the above lemmas, there is some $1 \le i \le \ell - 1$ such that T_i and T_{i+1} conflict in two variables: in x_{i+1} and in some $z \in \{x_{k+1}, \ldots, x_n\}$. ($T$ contains neither z nor $\overline{z}$ as otherwise it would conflict with T_{i+1} or T_i in two variables.) Suppose i is the smallest such index. Then there is some partial assignment of the variables $\{x_1, \ldots, x_n\} \setminus \{x_i, x_{i+1}, z\}$ consistent with T, T_i and T_{i+1}. Denote one such by a, and assume without loss of generality that T_i contains z, and T_{i+1} contains $\overline{z}$. (See Figure 5.)

	x_i	x_{i+1}	z
T	$+$	$+$	$\cdot$
T_i	$-$	$+$	$+$
T_{i+1}	$\cdot$	$-$	$-$
α	$-$	$+$	$-$
β	$+$	$-$	$+$

Figure 5. Terms T_i, T_{i+1}, T and assignments α and β.

Let α and β be the assignments that are consistent with a, with α assigning 0 to x_i and z and 1 to x_{i+1}, and β assigning 1 to x_i and z and 0 to x_{i+1}. Then

- $\overline{x_i} \in T^\alpha$, since $\alpha \notin \mathcal{T}(T)$ but $\alpha^{(x_i)} \in \mathcal{T}(T)$,
- $x_{i+1} \in T^\alpha$, since $\alpha \notin \mathcal{T}(T_{i+1})$ but $\alpha^{(x_{i+1})} \in \mathcal{T}(T_{i+1})$,
- $\overline{z} \in T^\alpha$, since $\alpha \notin \mathcal{T}(T_i)$ but $\alpha^{(z)} \in \mathcal{T}(T_i)$,
- $\overline{x_{i+1}} \in T^\beta$, since $\beta \notin \mathcal{T}(T)$ but $\beta^{(x_{i+1})} \in \mathcal{T}(T)$, and
- $z \in T^\beta$, since $\beta \notin \mathcal{T}(T_{i+1})$ but $\beta^{(z)} \in \mathcal{T}(T_{i+1})$.

Thus T^β does not contain x_i, as otherwise T^α and T^β would conflict in three variables. But then $T^\beta \in V$, furthermore $(T_i, T^\beta), (T^\beta, T_{i+2}) \in E$, so $T_1, \dots, T_i, T^\beta, T_{i+2}, \dots, T_\ell, T_1$ is also a cycle in G of minimal length, but with T_i and T^β conflicting only in one variable. That is, in this new cycle one gets further (starting from T_1) than in the original cycle without using an edge that's two endpoints conflict in two variables.

Iterating the above process if necessary, proceeding from the smaller indices to the larger ones, one obtains a cycle $T'_1, \dots, T'_\ell, T'_1$ with consecutive elements conflicting in only one variable (apart maybe from T_ℓ and T_1), contradicting Lemma 10. □

Now all that is left to prove is that G contains no cycle of length 2.

Lemma 12. *Assume (2), let $T = x_1 \cdots x_k$ be a term of F with $k < n$, and let $G = G_{F,T}$ defined as in (3). Then G contains no cycle.*

Proof. By Lemma 11, as noted, it suffices to show that G contains no cycle of length 2. Assume for the contradiction that T_1, T_2, T_1 is a cycle in G and assume furthermore without loss of generality that $\overline{x_1} \in T_1$, $x_2 \notin T_1$, $x_1 \notin T_2$ and $\overline{x_2} \in T_2$. There are two cases: when T_1 and T_2 conflict in only one variable and when they conflict in two.

Let us consider the first case. Then T_1 and T_2 conflict in some $z \in \{x_{k+1}, \dots, x_n\}$ (just like before, T cannot contain variable z, as otherwise it would conflict with T_1 or T_2 in at least two variables), and let us assume without loss of generality that $z \in T_1$ and $\overline{z} \in T_2$. Then there is some partial assignment to variables $\{x_3, \dots, x_n\} \setminus \{z\}$ consistent with T_1 and T_2. Denote one such by a. Let furthermore α and β be the assignments consistent with a, with α assigning 0 to x_1 and z and 1 to x_2, and β assigning 1 to x_1 and z and 0 to x_2 (see Figure 6(a)). Using a similar argument as before one can see that $\overline{x_1}, x_2, \overline{z} \in T^{(\alpha)}$ and $x_1, \overline{x_2}, z \in T^{(\beta)}$, thus the two terms conflict in three variables, contradiction.

	x_1	x_2	z
T	+	+	·
T_1	−	·	+
T_2	·	−	−
α	−	+	−
β	+	−	+

(a)

	x_1	x_2	z	v
T	+	+	·	·
T_1	−	·	+	+
T_2	·	−	−	−
α	−	+	−	+
β	+	−	+	−

(b)

Figure 6. Terms T_i, T_{i+1}, T and assignments α and β.

The second case is when T_1 and T_2 conflict in some $z, v \in \{x_{k+1}, \dots, x_n\}$ (as in the previous case T contains neither z nor v). Let us assume without loss of generality that $z, v \in T_1$ and $\overline{z}, \overline{v} \in T_2$. Similarly as above, let α and β be two assignments that are consistent with T, T_1 and T_2 on those variables in which these terms don't conflict, and otherwise behave as in Figure 6(b). Again, one can show that $\overline{z}, \overline{x_1} \in T^\alpha$ and $z, \overline{x_2} \in T^\beta$. Furthermore $x_2 \in T^\alpha$ (resp. $x_1 \in T^\beta$), as otherwise $T^\alpha \in V$ (resp. $T^\beta \in V$) and with T_2 (resp. with T_1) they would form a cycle of length two conflicting with each other in only one variable, which was ruled out in the previous case. Consequently T^α and T^β conflicts in three variables, contradiction. □

The proof of the Theorem now follows from Claim 7 and Lemma 12, noting that if F is a DDNF with conflict bound two that only has terms of length n, then $n \leq 2$, in which case the statement obviously holds.

4. Concluding remarks

Theorem 2 considers a very limited class of DDNFs—for which a somewhat surprising property is proved. Nevertheless this does not bring us any closer to determining α_n^d in the general case (recall (1), the definition of α_n^d), or to deriving a sharp bound for α_n; this problem thus remains open.

For another problem consider the direction took by Kullmann [7, 8, 9]: examining the role of deficiency in various combinatorial aspects of different classes of clause-sets, where the *deficiency* of a clause-set is the difference of the number of its clauses and the number of its variables. Let us denote by $\mathcal{SMUSAT}(k)$ the set of saturated minimally unsatisfiable clause-sets with deficiency at most k. (Restricting our focus to $\mathcal{SMUSAT}$, it is enough to consider only positive deficiency: Aharoni and Linial [1] has shown that clause-sets in $\mathcal{SMUSAT}$ have deficiency at least 1.) Using these notations a further characterization of the class $\mathcal{HIT}_{\leq 1} \cap \mathcal{USAT}$, given in [9], can be formalized as follows:

$$\mathcal{HIT}_{\leq 1} \cap \mathcal{USAT} = \mathcal{SMUSAT}(1).$$

Although there doesn't seem to be a similarly nice characterization for the set $\mathcal{HIT}_{\leq 2} \cap \mathcal{USAT}$ (for instance the DT-DNF in Example 13 with conflict bound two consists of twelve terms and has five variables, meanwhile replacing the last two (resp. four) terms with the single term vw (resp. v) would result in a DT-DNF with conflict bound two consisting of eleven (resp. nine) terms and having 5 variables), it would be interesting to find a bound, say some $f : \mathbb{N} \to \mathbb{N}$, such that if a clause-set in $\mathcal{HIT}_{\leq 2} \cap \mathcal{USAT}$ has m variables then it has at most $f(m)$ clauses.

Example 13. *The DNF*

$$\begin{aligned} F_{\text{ex}13} =& \overline{v}\,\overline{x}\,\overline{y}\,\overline{z} \vee \overline{v}\,\overline{x}\,\overline{y}z \vee \overline{v}\,\overline{x}y\overline{z} \vee \overline{v}\,\overline{x}yz \vee \overline{v}x\overline{w}\,\overline{y} \vee \overline{v}x\overline{w}y \vee \overline{v}xw\overline{z} \vee \overline{v}xwz \vee \\ & v\overline{w}\,\overline{y} \vee v\overline{w}y \vee vw\overline{z} \vee vwz \end{aligned}$$

is a DDNF with conflict bound two over five variables (namely v, w, x, y, z), having twelve clauses. See Figure 7 for the decision tree generating $F_{\text{ex}13}$.

Finding answers to these problems requires further investigations.

5. Acknowledgement

I would like to thank my advisor, György Turán, for his help during the preparation of this paper. I also owe thanks to Oliver Kullmann for his comments and suggestions from which the presentation of the paper had benefited a lot. Finally I would like to thank an anonymous referee for pointing out the duality between DT-DNFs and hitting clause-sets having read-once resolution refutations.

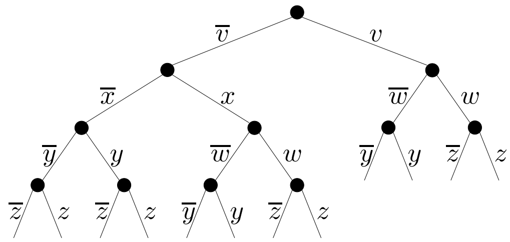

Figure 7. The decision tree generating the DNF $F_{\mathrm{ex}13}$ from Example 13.

References

[1] R. Aharoni and N. Linial. Minimal non-two-colorable hypergraphs and minimal unsatisfiable formulas. *J. Comb. Th. A*, **43**:196–204, 1986.

[2] N. H. Arai, T. Pitassi, and A. Urquhart. The complexity of analytic tableaux. In *Proc. 33rd Annu. ACM Sympos. Theory Comput. (STOC)*, pages 356–363, 2001.

[3] H. K. Büning and X. Zhao. The complexity of read-once resolution. *Annals of Mathematics and Artificial Intelligence*, **36**(4):419–435, 2002.

[4] G. Davydov, I. Davydova, and H. K. Büning. An efficient algorithm for the minimal unsatisfiability problem for a subclass of CNF. *Ann. Math. and Artif. Intell.*, **23**:229–245, 1998.

[5] K. Iwama and E. Miyano. Intractability of read-once resolution. In *Proceedings of the 10th Structure in Complexity Theory Conference (SCT)*, pages 29–36, Los Alamitos, CA, USA, 1995. IEEE Computer Society.

[6] S. Jukna, A. Razborov, P. Savický, and I. Wegener. On P versus NP∩co-NP for decision trees and read-once branching programs. *Comput. Complex.*, **8**(4):357–370, 1999.

[7] O. Kullmann. An application of matroid theory to the SAT problem. In *15th IEEE Conf. Computational Complexity*, pages 116–124, 2000.

[8] O. Kullmann. The combinatorics of conflicts between clauses. In *Theory and Appl. of Satisfiability Testing*, volume **2919** of *LNCS*, pages 426–440. Springer, 2003.

[9] O. Kullmann. On the conflict matrix of clause-sets. Technical Report CSR 7-2003, Univ. Wales at Swansea, 2003.

[10] L. Lovász, M. Naor, I. Newman, and A. Wigderson. Search problems in the decision tree model. *SIAM J. Discret. Math.*, **8**(1):119–132, 1995.

[11] P. Savický. On determinism versus unambiquous nondeterminism for decision trees. *Electronic Colloquium on Computational Complexity (ECCC)*, 009, 2002.

[12] P. Savický and J. Sgall. DNF tautologies with a limited number of occurrences of every variable. *Theoret. Comput. Sci.*, **238**(1–2):495–498, 2000.

[13] R. H. Sloan, B. Szörényi, and G. Turán. On k-term DNF with the maximal number of prime implicants. Accepted for publication at *SIAM J. Discrete Math.* Preliminary version available as Electronic Colloquium on Computational Complexity (ECCC) Technical Report TR05-023.

[14] G. Tardos. Query complexity, or why is it difficult to separate $NP^A \cap coNP^A$ from P^A by random oracles A? *Combinatorica*, **9**(4):385 – 392, 1989.

[15] A. Urquhart. The complexity of propositional proofs. *Bull. Symb. Logic*, **1**:425–467, 1995.