On variables with few occurrences in conjunctive normal forms

Oliver Kullmann
Computer Science Department
Swansea University
Swansea, SA2 8PP, UK
O.Kullmann@Swansea.ac.uk
`http://cs.swan.ac.uk/~csoliver`

Xishun Zhao*
Institute of Logic and Cognition
Sun Yat-sen University
Guangzhou, 510275, P.R.C.

Abstract

We consider the question of the existence of variables with few occurrences in boolean conjunctive normal forms (clause-sets). Let $\mu\mathrm{vd}(F)$ for a clause-set F denote the minimal variable-degree, the minimum of the number of occurrences of variables. Our main result is an upper bound $\mu\mathrm{vd}(F) \leq \mathrm{nM}(\sigma(F)) \leq \sigma(F) + 1 + \log_2(\sigma(F))$ for *lean clause-sets* F in dependency on the *surplus* $\sigma(F)$. Lean clause-sets, defined as having no non-trivial autarkies, generalise minimally unsatisfiable clause-sets. For the surplus we have $\sigma(F) \leq \delta(F) = c(F) - n(F)$, using the deficiency $\delta(F)$ of clause-sets, the difference between the number of clauses and the number of variables. $\mathrm{nM}(k)$ is the k-th "non-Mersenne" number, skipping in the sequence of natural numbers all numbers of the form $2^n - 1$. As an application of the upper bound we obtain that clause-sets F violating $\mu\mathrm{vd}(F) \leq \mathrm{nM}(\sigma(F))$ must have a non-trivial autarky (so clauses can be removed satisfiability-equivalently by an assignment satisfying some clauses and not touching the other clauses). It is open whether such an autarky can be found in polynomial time.

1 Introduction

We study the existence of "simple" variables in boolean conjunctive normal forms, considered as clause-sets. "Simple" here means a variable occurring not very often. A major use of the existence of such variables is in inductive proofs of properties of minimally unsatisfiable clause-sets, using splitting on a variable to reduce n, the number of variables, to $n - 1$: here it is vital that we have control over the changes imposed by the substitution, and so we want to split on a variable occurring as few times as possible. The background for these considerations is the enterprise of classifying minimal unsatisfiable clause-sets F in dependency on the deficiency $\delta(F) := c(F) - n(F)$, the difference between the number $c(F) := |F|$ of clauses of F and the number $n(F) := |\mathrm{var}(F)|$ of variables of F. The most basic fact is $\delta(F) \geq 1$, as first shown in [1]. For $\delta(F) = 1$ the structure is completely known ([1, 2, 6], for $\delta(F) = 2$ the structure after reduction of singular variables (occurring in one sign only once) is known ([4]), while for $\delta(F) \in \{3, 4\}$ only basic cases have been classified ([15]).

The starting point of our investigation is Lemma C.2 in [6], where it is shown that a minimally unsatisfiable clause-set F must have a variable v with at most $\delta(F)$ positive and at most $\delta(F)$ negative occurrences; we write this as $\mathrm{ld}_F(v) \leq \delta(F)$ and $\mathrm{ld}_F(\overline{v}) \leq \delta(F)$, using the notion of *literal degrees* (the number of occurrences of the literal). Thus we have $\mathrm{vd}_F(v) \leq 2\delta(F)$, using the *variable degree* $\mathrm{vd}_F(v) := \mathrm{ld}_F(v) + \mathrm{ld}_F(\overline{v})$. Using the *minimum variable degree* (min-var-degree) $\mu\mathrm{vd}(F) := \min_{v \in \mathrm{var}(F)} \mathrm{vd}_F(v)$ of F, this becomes $\mu\mathrm{vd}(F) \leq 2\delta(F)$. In this article we show a sharper bound on $\mu\mathrm{vd}(F)$ for a larger class of clause-sets F. More precisely, we show that the worst-cases $\mathrm{ld}_F(v), \mathrm{ld}_F(\overline{v}) \leq \delta(F)$ can not occur at the same time (for a suitable variable), but actually $\mathrm{ld}_F(v) + \mathrm{ld}_F(\overline{v}) - \delta(F)$ only grows logarithmically in $\delta(F)$, and this for a larger class of formulas.

The larger class of clause-sets considered is the class $\mathcal{LEAN}$ of *lean clause-sets*, which are clause-sets having no non-trivial autarky. For an overview on the theory of minimally unsatisfiable clause-sets and on the theory of autarkies see [5]. The deficiency $\delta(F) \in \mathbb{Z}$ of clause-sets is replaced by the *surplus* $\sigma(F) \in \mathbb{Z}$, which is the minimal deficiency over all clause-sets $F[V]$ for non-empty variable sets $V \subseteq \mathrm{var}(F)$, where $F[V]$ is obtained from F by removing clauses which have no variables in V, and restricting the remaining clauses to V; see [11] for more information on the surplus of (generalised) clause-sets. We need to count multiple occurrences of clauses here (which might arise during the process of removing literals with variables not in V), and thus actually multi-clause-sets F are used here. Note that by considering $V = \mathrm{var}(F)$ we have $\sigma(F) \leq \delta(F)$, and by considering $V = \{v\}$ for $v \in \mathrm{var}(F)$ we get $\sigma(F) \leq \mu\mathrm{vd}(F) - 1$. Now the main result of this article (Theorem 4.1) is

$$\mu\mathrm{vd}(F) \leq \mathrm{nM}(\sigma(F))$$

for lean F, where $\mathrm{nM} : \mathbb{N} \to \mathbb{N}$ (see Definition 3.1) is a super-linear function with $\mathrm{nM}(k) \leq k+1+\log_2(k)$. As an application we obtain (Corollary 4.2), that if a (multi-)clause-set F has no variable occurring with degree at most $\delta(F) + 1 + \log_2(\delta(F))$, then F has a non-trivial autarky. It is an open problem whether such an autarky can be found in polynomial time (for arbitrary clause-sets F); we conjecture (Conjecture 4.3) that this is possible.

Related work   This article appears to be the first systematic study of the problem of minimum variable occurrences in minimally unsatisfiable clause-sets and generalisations, in dependency on the deficiency, asking for the existence of a variable occurring "infrequently" in general, or for extremal examples where all variables occur not infrequently. The problem of maximum variable occurrences (asking for the existence of a variable occurring frequently in general, or for extremal examples where all variables occur not frequently) in uniform (minimally) unsatisfiable clause-sets, in dependency on the (constant) clause-length, has been studied in the literature, starting with [14]; for a recent article see [3].

Overview   In Section 2 basic notions and concepts regarding clause-sets, autarkies and minimal unsatisfiability are reviewed. Section 3 introduces the numbers $\mathrm{nM}(k)$ and proves exact formulas and sharp lower and upper bounds. Section 4 contains the main results. First in Subsection 4.1 the bound is shown for minimally unsatisfiable clause-sets (Theorem 4.5). In Subsection 4.2 the bound then is lifted to lean clause-sets, proving Theorem 4.1. The immediate corollary of Theorem 4.1 is, that if the asserted upper bound on the minimal variable degree is not fulfilled, then a non-trivial autarky must exist (Corollary 4.2). In Subsection 4.3 the problem of finding such autarky is discussed, with Conjecture 4.3 making precise our believe that one can find such autarkies efficiently. In Section 5 we discuss the sharpness of the bound, and the possibilities to generalise it further. Finally, in Section 6

open problems are stated, culminating in the central Conjecture 6.1 about the classification of unsatisfiable hitting clause-sets (or "disjoint tautologies" in the terminology of DNFs).

2 Preliminaries

We follow the general notations and definitions as outlined in [5], where also further background on autarkies and minimal unsatisfiability can be found. We use $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

2.1 Clause-sets

Complementation of literals x is denoted by $\overline{x}$, while for a set L of literals we define $\overline{L} := \{\overline{x} : x \in L\}$. A **clause** C is a finite and clash-free set of literals (i.e., $C \cap \overline{C} = \emptyset$), while a **clause-set** is a finite set of clauses. We use $\mathrm{var}(F) := \bigcup_{C \in F} \mathrm{var}(C)$ for the set of variables of F, where $\mathrm{var}(C) := \{\mathrm{var}(x) : x \in C\}$ is the set of variables of clause C, while $\mathrm{var}(x)$ is the underlying variable for a literal x. For a clause-set F we denote by $n(F) := |\mathrm{var}(F)| \in \mathbb{N}_0$ the number of variables and by $c(F) := |F| \in \mathbb{N}_0$ the number of clauses. The **deficiency** of a clause-set is denoted by $\delta(F) := c(F) - n(F) \in \mathbb{Z}$. We call a clause C **full** for a clause-set F if $\mathrm{var}(C) = \mathrm{var}(F)$, while a clause-set F is called full if every clause is full. For a finite set V of variables let $A(V)$ be the set of all $2^{|V|}$ full clauses over V. Thus full clause-sets are exactly the sub-clause-sets of some $A(V)$. A **partial assignment** is a map $\varphi : V \to \{0, 1\}$ for some (possibly empty) set V of variables. The application of a partial assignment φ to a clause-set F is denoted by $\varphi * F$, which yields the clause-set obtained from F by removing all satisfied clauses (which have at least one literal set to 1), and removing all falsified literals from the remaining clauses. A clause-set F is satisfiable iff there is a partial assignment φ with $\varphi * F = \top := \emptyset$, otherwise F is unsatisfiable. All $A(V)$ are unsatisfiable.

These notions are generalised to **multi-clause-sets**, which are pairs (F, m), where F is a clause-set and $m : F \to \mathbb{N}$ determines the multiplicity of the clauses. Now $c((F, m)) := \sum_{C \in F} m(C)$, while the application of partial assignments φ to a multi-clause-set F yields a *multi*-clause-set $\varphi * F$, where the multiplicity of a clause C in $\varphi * F$ is the sum of all multiplicities of clauses in F which are shortened to C by φ. For example if φ is a total assignment for F (assigns all variables of F) which does not satisfying F (i.e., $\varphi * F \neq \top$), then $\varphi * F$ is $(\{\bot\}, (f)_{C \in \{\bot\}})$, where $\bot := \emptyset$ is the empty clause, while $f \in \mathbb{N}$ is the number of clauses (with their multiplicities) of F falsified by φ.

For the number of occurrences of a literal x in a (multi-)clause-set (F, m) we write $\mathrm{ld}_F(x) := \sum_{C \in F, x \in C} m(C)$, called the **literal-degree**, while the **variable-degree** of a variable v is defined as $\mathrm{vd}_F(v) := \mathrm{ld}_F(v) + \mathrm{ld}_F(\overline{v})$. A **singular variable** in a (multi-)clause-set F is a variable occurring in one sign only once (i.e., $1 \in \{\mathrm{ld}_F(v), \mathrm{ld}_F(\overline{v})\}$). A (multi-)clause-set is called **non-singular** if it does not have singular variables.

For a set V of variables and a multi-clause-set F by $F[V]$ the **restriction** of F to V is denoted, which is obtained by removing clauses from F which have no variables in common with V, and removing from the remaining clauses all literals where the underlying variable is not in V (note that this can increase multiplicities of clauses).

2.2 Autarkies

An **autarky** for a clause-set F is a partial assignment φ which satisfies every clause $C \in F$ it touches, i.e., with $\text{var}(\varphi) \cap \text{var}(C) \neq \emptyset$. The empty partial assignment is always an autarky for every F, the **trivial autarky**. If φ is an autarky for F, then $\varphi * F \subseteq F$ holds, and thus $\varphi * F$ is satisfiability-equivalent to F. A clause-set F is **lean** if there is no non-trivial autarky for F. A weakening is the notion of a **matching-lean** clause-set F, which has no non-trivial **matching autarky**, which are special autarkies given by a matching condition (for every clause touched, a unique variable underlying a satisfied literal must be selectable). The process of applying autarkies as long as possible to a clause-set is confluent, yielding the **lean kernel** of a clause-set. Computation of the lean kernel is NP-hard, but the **matching-lean kernel**, obtained by applying matching autarkies as long as possible, which is also a confluent process, is computable in polynomial time. Note that a clause-set F is lean resp. matching lean iff the lean resp. matching-lean kernel is F itself. For every matching-lean multi-clause-set $F \neq \top$ we have $\delta(F) \geq 1$, while in general a multi-clause-set $F \neq \top$ is matching lean iff $\sigma(F) \geq 1$, where the **surplus** $\sigma(F) \in \mathbb{Z}$ is defined as the minimum of $\delta(F[V])$ for all $\emptyset \neq V \subseteq \text{var}(F)$. Note that while w.r.t. general autarkies there is no difference between a multi-clause-set and the underlying clause-set, for matching autarkies there is a difference, due to the matching condition. For more information on autarkies see [5, 11].

2.3 Minimally unsatisfiable clause-sets

The set of minimally unsatisfiable clause-sets is $\mathcal{MU}$, the set of all clause-sets which are unsatisfiable, while removal of any clause makes them satisfiable. Furthermore the set of saturated minimally unsatisfiable clause-sets is $\mathcal{SMU} \subset \mathcal{MU}$, which is the set of minimally unsatisfiable clause-sets such that addition of any literal to any clause renders them satisfiable. We recall the fact that every minimally unsatisfiable clause-set $F \in \mathcal{MU}$ can be **saturated**, i.e., by adding literal occurrences to F we obtain $F' \in \mathcal{SMU}$ with $\text{var}(F') = \text{var}(F)$ such that there is a bijection $\alpha : F \to F'$ with $C \subseteq \alpha(C)$ for all $C \in F$. Some basic properties of $\mathcal{MU}$ and $\mathcal{SMU}$ w.r.t. the application of partial assignments are given in the following lemma.

Lemma 2.1 *For all clause-sets F we have:*

1. *$F \in \mathcal{SMU}$ iff for all $v \in \text{var}(F)$ and $\varepsilon \in \{0, 1\}$ we have $\langle v \to \varepsilon \rangle * F \in \mathcal{MU}$.*

2. *If for some variable v holds $\langle v \to 0 \rangle * F \in \mathcal{SMU}$ and $\langle v \to 1 \rangle * F \in \mathcal{SMU}$, then $F \in \mathcal{SMU}$.*

3. *If for some variable v holds $\langle v \to 0 \rangle * F \in \mathcal{MU}$ and $\langle v \to 1 \rangle * F \in \mathcal{MU}$, then $F \in \mathcal{MU}$.*

For more information on minimal unsatisfiability see [5, 12].

3 Non-Mersenne numbers

Splitting on variables with minimum occurrence in minimally unsatisfiable clause-sets leads by Theorem 4.5 to the following recursion. The understanding of this recursion is the topic of this section. On a first reading, only Definition 3.1 and the main results, Lemma 3.8 and Corollary 3.9, need to be considered.

Definition 3.1 *For $k \in \mathbb{N}$ let $\mathrm{nM}(k) := 2$ if $k = 1$, while else*

$$\mathrm{nM}(k) := \max_{i \in \{2,\ldots,k\}} \min(2 \cdot i, \mathrm{nM}(k - i + 1) + i).$$

Remarks:

1. This is sequence `http://oeis.org/A062289` in the "On-Line Encyclopedia of Integer Sequences". It can be defined as the enumeration of those natural numbers containing the string "10" (at consecutive positions). The sequence leaves out exactly the number of the form $2^n - 1$ for $n \in \mathbb{N}$, and thus the name. The sequence consists of arithmetic progressions of slope 1 and length $2^m - 1$, $m = 1, 2, \ldots$, each such progression separated by an additional step of $+1$. The recursion in Definition 3.1 is new, and so we can not use these characterisations, but must directly prove the basic properties.

2. The value of $\mathrm{nM}(k)$ for $k = (1), (2, 3, 4), (5, \ldots, 11), (12, \ldots, 26)$ is $(2), (4, 5, 6),$ $(8, \ldots, 14), (16, \ldots, 30)$.

3. For $k \geq 2$ we have $\mathrm{nM}(k) \geq 4$. This holds since $\mathrm{nM}(2) = 4$, while the induction step for $k \geq 3$ is $\mathrm{nM}(k) = \max_{i \in \{2,\ldots,k\}} \min(2i, \mathrm{nM}(k - i + 1) + i) \geq \min(4, \min(4 + 2, 1 + 3)) = 4$.

4. By induction and by definition we have $k + 1 \leq \mathrm{nM}(k) \leq 2 \cdot k$ for $k \in \mathbb{N}$.

For a sequence $a : \mathbb{N} \to \mathbb{R}$ and $k \in \mathbb{N}$ let $\boldsymbol{\Delta a(k)} := a(k + 1) - a(k)$ be the step in the value of the sequence from k to $k + 1$. The next number in the sequence of non-Mersenne numbers is obtained by adding 1 or 2 to the previous number:

Lemma 3.2 *For $k \in \mathbb{N}$ holds $\Delta \mathrm{nM}(k) \in \{1, 2\}$.*

Proof For $k = 1$ we get $\Delta \mathrm{nM}(1) = 2$. Now consider $k \geq 2$. We have $\mathrm{nM}(k + 1) = \max(\min(4, \mathrm{nM}(k) + 2), \max_{i \in \{3,\ldots,k+1\}} \min(2i, \mathrm{nM}(k - i + 2) + i)) = \max_{i \in \{3,\ldots,k+1\}} \min(2i, \mathrm{nM}(k - i + 2) + i) = \max_{i \in \{2,\ldots,k\}} \min(2(i + 1), \mathrm{nM}(k - (i + 1) + 2) + (i + 1)) = \max_{i \in \{2,\ldots,k\}} \min(2i + 2, \mathrm{nM}(k - i + 1) + i + 1) = 1 + \max_{i \in \{2,\ldots,k\}} \min(2i + 1, \mathrm{nM}(k - i + 1) + i)$.

Thus on the one hand we have $\mathrm{nM}(k + 1) \geq 1 + \max_{i \in \{2,\ldots,k\}} \min(2i, \mathrm{nM}(k - i + 1) + i) = 1 + \mathrm{nM}(k)$, and on the other hand $\mathrm{nM}(k + 1) \leq 1 + \max_{i \in \{2,\ldots,k\}} \min(2i + 1, \mathrm{nM}(k - i + 1) + i + 1) = 2 + \mathrm{nM}(k)$. ∎

Corollary 3.3 $\mathrm{nM} : \mathbb{N} \to \mathbb{N}$ *is strictly increasing.*

Corollary 3.4 *We have $\mathrm{nM}(a + b) \geq \mathrm{nM}(a) + b$ for $a \in \mathbb{N}$ and $b \in \mathbb{N}_0$, and thus $\mathrm{nM}(a - b) \leq \mathrm{nM}(a) - b$ for $b \leq a$.*

Instead of considering the maximum over $k - 1$ cases $i \in \{2, \ldots, k\}$ to compute $\mathrm{nM}(k)$, we can now simplify the recursion to only one case $i(k) \in \{2, \ldots, k\}$, and for that case also consideration of the minimum is dispensable:

Lemma 3.5 *For $k \in \mathbb{N}$, $k \geq 2$, let $i(k) \in \mathbb{N}$ be the smallest $i \in \{2, \ldots, k\}$ with $i \geq \mathrm{nM}(k - i + 1)$ (note that $k \geq \mathrm{nM}(k - k + 1) = 2$, and thus $i(k)$ is well-defined). For example we have $i(2) = 2$, $i(3) = 3$, $i(4) = 4$ and $i(5) = 4$. Then we have:*

1. $i(k) - \mathrm{nM}(k - i(k) + 1) \leq 2$.

2. $\mathrm{nM}(k) = \mathrm{nM}(k - i(k) + 1) + i(k)$.

3. $\Delta i(k) \in \{0, 1\}$.

Proof We have $i(k) = 2$ iff $k = 2$, while for $k = 2$ the assertions hold trivially; so assume $k \geq 3$ and $i(k) \geq 3$. Part 1 follows by Lemma 3.2 from the facts that the sequence $i \in \{2, \ldots, k\} \mapsto i$ moves up in steps of $+1$, while the sequence $i \in \{2, \ldots, k\} \mapsto \mathrm{nM}(k - i + 1)$ moves down in steps of -1 or -2. It remains to show Part 2. By Lemma 3.2 the sequence $i \in \{2, \ldots, k\} \mapsto \mathrm{nM}(k - i + 1) + i$ is monotonically decreasing, and thus by definition we obtain $\mathrm{nM}(k) = \max(2 \cdot (i(k) - 1), \mathrm{nM}(k - i(k) + 1) + i(k))$. That the maximum here is actually always attained in the second component follows by Part 1. Finally Part 3 follows again from Lemma 3.2. ∎

After these preparations we are able to characterise the "jump positions", the set $J \subset \mathbb{N}$ of $k \in \mathbb{N}$ with $\Delta \mathrm{nM}(k) = 2$. Thus $\Delta \mathrm{nM}(k) = 1$ iff $k \notin J$, and $J = \{1, 4, 11, 26, \ldots\}$. Note $\mathrm{nM}(k) = 1 + k + |\{k' \in J : k' < k\}|$.

Lemma 3.6 *Let $i'(k) := k - i(k) + 1$ and $h(k) := \mathrm{nM}(i'(k))$ for $k \in \mathbb{N}$, $k \geq 2$. Thus $\Delta i'(k) \in \{0, 1\}$ and $\Delta i'(k) = 1 - \Delta i'(k)$. Furthermore we have $\mathrm{nM}(k) = h(k) + i(k)$, thus $\Delta \mathrm{nM}(k) = \Delta h(k) + \Delta i(k)$, and $i(k) - h(k) \in \{0, 1, 2\}$. Consider $k \geq 2$.*

1. *If $\Delta i(k) = 0$, then:*

   (a) *$\Delta i(k + 1) = 1$*
   (b) *$i(k) \neq h(k)$.*
   (c) *$i(k + 1) = h(k + 1)$.*

2. *If $\Delta i(k) = 1$, then:*

   (a) *$\Delta h(k) = 0$, and so $k \notin J$*
   (b) *$i(k) \neq h(k) + 2$.*

3. *The following conditions are equivalent:*

   (a) *$k \in J$*
   (b) *$\Delta h(k) = 2$*
   (c) *$i(k) = h(k) + 2$*
   (d) *$\Delta i(k - 1) = 1$ and $i(k - 1) = h(k - 1) + 1$*
   (e) *$\Delta i(k - 2) = \Delta i(k - 1) = 1$*
   (f) *$i'(k) = i'(k - 1) = i'(k - 2)$ and $i'(k) \in J$.*

4. *If $k \in J$, then $i'(k) = \max(k' \in J : k' < k)$.*

Proof Part 1a follows by definition. For Part 1b note $i(k + 1) = i(k)$ while $h(k + 1) \geq h(k) + 1$. For Part 1c assume $i(k + 1) > h(k + 1)$. Then we have $i(k) = h(k) + 2$ and $h(k + 1) = h(k) + 1$. However then $i(k) - 1 = h(k) + 1 = h(k + 1) = \mathrm{nM}(k - (i(k) - 1) + 1)$ contradicting the definition of $i(k)$. For Part 2a assume $i(k) = i(k + 1) = i(k + 2)$. We have $i(k) \geq h(k + 2) = \mathrm{nM}(k - i(k) + 3)$, while $i(k) - 1 < \mathrm{nM}(k - (i(k) - 1) + 1) = \mathrm{nM}(k - i(k) + 2)$, i.e., $i(k) \leq \mathrm{nM}(k - i(k) + 2)$, contradicting the strict monotonicity of nM. Part 2b follows by $i(k + 1) \leq h(k + 1) + 2$ and $i(k + 1) = i(k) + 1$, $h(k + 1) = h(k)$. Now consider Part 3.

Condition 3a implies condition 3b due to $\Delta i(k) = 0$ in case of $k \in J$ by Part 2a. Condition 3b implies condition 3c, since $\Delta h(k) = 2$ implies $\Delta i(k) = 0$ (otherwise we had $\Delta \mathrm{nM}(k) = 3$), and so by Part 1c we have $i(k) = i(k + 1) = h(k + 1)$, while the assumption says $h(k + 1) = h(k) + 2$. In turn condition 3c implies condition 3a, since by Part 2b we get $\Delta i(k) = 0$, and thus $\Delta \mathrm{nM}(k) = \Delta h(k)$, while in case of $\Delta h(k) \leq 1$ we would have $i(k) - 1 \geq \mathrm{nM}(k - (i(k) - 1) + 1)$ contradicting the definition of $i(k)$, due to $\mathrm{nM}(k - (i(k) - 1) + 1) = \mathrm{nM}((k + 1) - i(k + 1) + 1) =$

$h(k+1) \leq h(k) + 1 = i(k) - 1$. So now we can freely use the equivalence of these three conditions.

Condition 3c implies condition 3d, since we have $\Delta i(k) = 0$, and thus $\Delta i(k-1) = 1$ with Part 1a, from which we furthermore get $i(k) = i(k-1)+1$ and $h(k-1) = h(k)$, and so $i(k-1) = i(k)-1 = h(k)+1 = h(k-1)+1$. Condition 3d implies condition 3e, since in case of $\Delta i(k-2) = 0$ we had $i(k-1) = h(k-1)$ with Part 1c. In turn condition 3e implies condition 3c, since $i(k) = i(k-1) + 1 = i(k-2) + 2$, while $h(k) = h(k-1) = h(k-2)$, where by definition $i(k-2) \geq h(k-2)$ holds, whence $i(k) \geq h(k)+2$, which implies $i(k) = h(k)+2$. So now the first five conditions have been shown to be equivalent.

Now condition 3e implies condition 3f, since it only remains to show $i'(k) \in J$, which follows with condition 3b (using $\Delta i(k) = 0$). In turn condition 3f implies immediately condition 3e.

Finally, we prove Part 4 by induction on k (regarding the enumeration of J). We have $i'(4) = 1$, and so the induction holds for $k = 4$, the smallest jump position $k \geq 2$. Now assume that the assertion holds for all elements of $J \cap \{1, \ldots, k-1\}$, where $k > 4$, and we have to show the assertion for k. By Part 3f we know $i'(k) \in J$, where $2 \leq i'(k) < k$. Assume there is $k' \in J$ with $i'(k) < k' < k$. Now by induction hypothesis we get $i'(k) \leq i'(k') < k'$. However by Part 1 we get $\Delta i'(k') = 1$, and thus $i'(k) > i'(k')$ (since $k > k'$). ∎

Corollary 3.7 *We have $J = \{2^{m+1} - m - 2 : m \in \mathbb{N}\}$.*

Proof Let k_m for $m \in \mathbb{N}$ be the mth element of J; so the assertion is $k_m = 2^{m+1} - m - 2$. We have $k_1 = 4 - 1 - 2 = 1 = \min J$; in the remainder assume $m \geq 2$. We prove the assertion by induction, in parallel with $i(k_m) = 2^{m+1} - 2^m$. For $m = 2$ we have $k_2 = 8 - 2 - 2 = 4 = \min J \setminus \{1\}$, while $i(4)$ is the smallest $i \in \{2, 3, 4\}$ with $i \geq \mathrm{nM}(5 - i)$, which yields $i(4) = 4 = 2^3 - 2^2$. Now we consider the induction step, from $m-1$ to m. The induction hypothesis yields $k_{m-1} = 2^m - m - 1$ and $i(k_{m-1}) = 2^m - 2^{m-1}$. Lemma 3.6, Part 4 yields $i'(k_m) = k_{m-1}$, from which by $i'(k_m) = k_m - i(k_m) + 1$ follows $k_m = 2^m - m - 2 + i(k_m)$. By definition we get $i(k_m) = \Delta i(k_m - 1) + \cdots + \Delta i(k_{m-1}) + i(k_{m-1})$. By Lemma 3.6, Parts 1 - 3 the sequence of Δ-values has the form (starting with the lowest index) $0, 1, 0, 1, \ldots, 0, 1, 1$, and thus their sum has the value $\frac{1}{2}(k_m - k_{m-1} - 1) + 1$. So we get $i(k_m) = \frac{1}{2}(k_m - k_{m-1} - 1) + 1 + i(k_{m-1}) = \frac{1}{2}(2^m - m - 2 + i(k_m) - 2^m + m + 1 - 1) + 1 + 2^m - 2^{m-1} = \frac{1}{2}i(k_m) - 1 + 1 + 2^m - 2^{m-1}$, from which $i(k_m) = 2^{m+1} - 2^m$ follows. Finally $k_m = 2^m - m - 2 + 2^{m+1} - 2^m = 2^{m+1} - m - 2$. ∎

Now the closed formula for $\mathrm{nM}(k)$ can be proven (using $\mathrm{ld}(x) := \log_2(x)$):

Lemma 3.8 *For $k \in \mathbb{N}$ let $\mathrm{fld}(k) := \lfloor \mathrm{ld}(k) \rfloor$ ("floor of logarithm dualis"). Then we have for $k \in \mathbb{N}$ the equality $\mathrm{nM}(k) = k + \mathrm{fld}(k + 1 + \mathrm{fld}(k+1))$.*

Proof Let $g(k) := \mathrm{fld}(k+1+\mathrm{fld}(k+1))$ and $f(k) := k + g(k)$ (so $\mathrm{nM}(k) = f(k)$ is to be shown, for $k \geq 1$). We have $f(1) = 1 + \mathrm{fld}(2 + \mathrm{fld}(2)) = 1 + \mathrm{fld}(3) = 2 = \mathrm{nM}(1)$. We will now prove that the function $g(k)$ changes values exactly at the transitions $k \mapsto k + 1$ for $k \in J$, that is, for indices $k = k_m := 2^{m+1} - m - 2$ (using Corollary 3.7) with $m \in \mathbb{N}$ we have $\Delta g(k_m) = 1$, while otherwise we have $\Delta g(k_m) = 0$, from which the assertion follows (by the definition of J).

We have $g(1) = 1$ and $g(2) = 2$. Now consider $m \in \mathbb{N}$ and $k_m + 1 \leq k \leq k_{m+1}$. We show $g(k) = m + 1$, which proves the claim. Note that $g(k)$ is monotonically increasing. Now $g(k) \geq g(k_m + 1) = \lfloor \mathrm{ld}(2^{m+1} - m + \lfloor \mathrm{ld}(2^{m+1} - m) \rfloor) \rfloor = \lfloor \mathrm{ld}(2^{m+1} - m + m) \rfloor = m + 1$ and $g(k) \leq g(k_{m+1}) = \lfloor \mathrm{ld}(2^{m+2} - m - 2 + \lfloor \mathrm{ld}(2^{m+2} - m - 2) \rfloor) \rfloor \leq \lfloor \mathrm{ld}(2^{m+2} - m - 2 + m + 1) \rfloor = \lfloor \mathrm{ld}(2^{m+2} - 1) \rfloor = m + 1$. ∎

As a result, we obtain very precise bounds:

Corollary 3.9 $k + \mathrm{fld}(k+1) \leq \mathrm{nM}(k) \leq k + 1 + \mathrm{fld}(k)$ *holds for* $k \in \mathbb{N}$.

Proof   The lower bound follows trivially. The upper bound holds (with equality) for $k \leq 2$, so assume $k \geq 3$. We have to show $g(k) = \mathrm{fld}(k+1+\mathrm{fld}(k+1)) \leq 1+\mathrm{fld}(k)$, which follows from $\mathrm{ld}(k + 1 + \mathrm{fld}(k + 1)) \leq 1 + \mathrm{ld}(k)$. Now $\mathrm{ld}(k + 1 + \mathrm{fld}(k + 1)) \leq \mathrm{ld}(k + 1 + \mathrm{ld}(k + 1)) \leq \mathrm{ld}(k + k) = 1 + \mathrm{ld}(k)$. ∎

4   Lean clause-sets and the surplus

In this section we prove the main result of this paper, Theorem 4.1. The proof consists in first handling a special case, minimally unsatisfiable clause-sets instead of lean clause-sets, in Subsection 4.1, and then lifting the result to the general case in Subsection 4.2. In Subsection 4.3 we consider the algorithmic implications of this result.

Theorem 4.1 *We have* $\mu\mathrm{vd}(F) \leq \mathrm{nM}(\sigma(F))$ *for a lean multi-clause-set* F *with* $n(F) > 0$. *More precisely, there exists a variable* $v \in \mathrm{var}(F)$ *with* $\mathrm{vd}_F(v) \leq \mathrm{nM}(\sigma(F))$ *and* $\mathrm{ld}_F(v), \mathrm{ld}_F(\overline{v}) \leq \sigma(F)$.

We obtain a sufficient criterion for the existence of a non-trivial autarky.

Corollary 4.2 *Consider a multi-clause-set* F *with* $n(F) > 0$. *If* $\sigma(F) \leq 0$, *then* F *has a non-trivial matching autarky. So assume* $\sigma(F) \geq 1$. *If we have* $\mu\mathrm{vd}(F) > \mathrm{nM}(\sigma(F))$, *then for every* $\emptyset \neq V \subseteq \mathrm{var}(F)$ *with* $\delta(F[V]) = \sigma(F)$ *we have an autarky* φ *for* F *with* $\mathrm{var}(\varphi) = V$ *(and thus* F *has a non-trivial autarky).*

The quantities $\mu\mathrm{vd}(F)$ and $\mathrm{nM}(\sigma(F))$ (resp. $\mathrm{nM}(\delta(F))$) are computable in polynomial time, and so the applicability of Corollary 4.2 can be checked in polynomial time. We conjecture that also "constructivisation" of Corollary 4.2 can be done in polynomial time:

Conjecture 4.3 *There is a poly-time algorithm for computing a non-trivial autarky in case of* $\mu\mathrm{vd}(F) > \mathrm{nM}(\sigma(F))$ *(or* $\mu\mathrm{vd}(F) > \mathrm{nM}(\delta(F))$*) for matching-lean clause-sets* F.

See Subsection 4.3 for more discussion on Conjecture 4.3 (there also the remaining details of Corollary 4.2 are proven).

4.1   The special case of minimally unsatisfiable clause-sets

The main auxiliary lemma is the following statement, which receives its importance from the fact that every minimally unsatisfiable clause-set can be saturated (this method was first applied in [6]).

Lemma 4.4 *Consider* $F \in \mathcal{SMU}_{\delta=k}$ *for* $k \in \mathbb{N}$ *and a variable* $v \in \mathrm{var}(F)$ *realising the minimal var-degree (i.e.,* $\mathrm{vd}_F(v) = \mu\mathrm{vd}(F)$*). Using* $m_0 := \mathrm{ld}_F(\overline{v})$ *and* $m_1 := \mathrm{ld}_F(v)$ *we have* $\langle v \to \varepsilon \rangle * F \in \mathcal{MU}_{k-m_\varepsilon+1}$ *for* $\varepsilon \in \{0, 1\}$, *where* $n(\langle v \to \varepsilon \rangle * F) = n(F) - 1$. *Since minimally unsatisfiable clause-sets have deficiency at least one, we get* $m_\varepsilon \leq k$.

Proof   We have $n(\langle v \to \varepsilon \rangle * F) = n(F) - 1$ since F contains no pure variable, while v realises the minimum of var-degrees. Thus $\delta(\langle v \to \varepsilon \rangle * F) = \delta(F) - m_\varepsilon + 1$, while $\langle v \to \varepsilon \rangle * F \in \mathcal{MU}$ by Lemma 2.1, Part 1. ∎

Theorem 4.5 *For all $k \in \mathbb{N}$ and $F \in \mathcal{MU}_{\delta \leq k}$ we have $\mu\mathrm{vd}(F) \leq \mathrm{nM}(k)$. More precisely, for $n(F) > 0$ there exists a variable $v \in \mathrm{var}(F)$ with $\mathrm{vd}_F(v) \leq \mathrm{nM}(k)$ and $\mathrm{ld}_F(v), \mathrm{ld}_F(\overline{v}) \leq k$.*

Proof The assertion is known for $k = 1$, so assume $k > 1$, and we apply induction on k. Assume $\delta(F) = k$ (due to $k > 1$ we have $n(F) > 1$). Saturate F and obtain F'. Consider a variable $v \in \mathrm{var}(F')$ realising the min-var-degree of F'. If $\mathrm{vd}_{F'}(v) = 2$ then we are done, so assume $\mathrm{vd}_{F'}(v) \geq 3$. Let $i := \max(\mathrm{ld}_{F'}(v), \mathrm{ld}_{F'}(\overline{v}))$; so $\mathrm{vd}_{F'}(v) \leq 2i$. W.l.o.g. assume that $i = \mathrm{ld}_{F'}(v)$. By Lemma 4.4 we get $2 \leq i \leq k$. Applying the induction hypothesis and Lemma 4.4 we obtain a variable $w \in \mathrm{var}(G)$ for $G := \langle v \to 1 \rangle * F$ with $\mathrm{vd}_G(w) \leq \mathrm{nM}(k-i+1)$. By definition we have $\mathrm{vd}_{F'}(w) \leq \mathrm{vd}_G(w) + \mathrm{ld}_{F'}(v)$. Altogether we get $\mu\mathrm{vd}(F) \leq \min(2i, \mathrm{nM}(k-i+1)+i) \leq \mathrm{nM}(k)$. ∎

It is interesting to generalise Theorem 4.5 for generalised clause-sets (see [11, 12] for a systematic study, and [10] for the underlying report). Generalised clause-sets have literals "$v \neq \varepsilon$" for variables v with domains D_v and values $\varepsilon \in D_v$, and the deficiency is generalised by giving every variable a weight $|D_v| - 1$ (which is 1 in the boolean case). The base case of deficiency $k = 1$ is handled in Lemma 5.4 in [12], showing that for generalised clause-sets we have here $\mu\mathrm{vd}(F) \leq \max_{v \in \mathrm{var}(F)} |D_v|$. But $k \geq 2$ requires more work:

1. The basic method of saturation is not available for generalised clause-sets, as discussed in Subsection 5.1 in [12]. Thus the proofs for the boolean case seem not to be generalisable.

2. Stipulating the effects of saturation via the "substitution stability parameter regarding irredundancy", in Corollary 5.10 in [12] one finds a first approach towards generalising the basic bound $\mu\mathrm{vd}(F) \leq 2\delta(F)$ (for the boolean case) by $\mu\mathrm{vd}(F) \leq \max_{v \in \mathrm{var}(F)} |D_v| \cdot \delta(F)$.

3. Another approach uses translations to boolean clause-sets. The "generic translation scheme" (see [9, 12]) allows (for certain instances) to preserve the deficiency and the other structures relevant here. So we get general upper bounds for the minimum number of occurrences of variables in generalised clause-sets from the boolean case. But further investigations are needed in these bounds.

4.2   Proof of the general case

Now consider an arbitrary (multi-)clause-set F. Consider a set of variables $\emptyset \neq V \subseteq \mathrm{var}(F)$ realising the surplus of F, i.e., such that $\delta(F[V])$ is minimal. If $F[V]$ would be satisfiable, then a satisfying assignment would give a non-trivial autarky for F. Assuming that F is lean thus yields that $F[V]$ must be unsatisfiable. So there exists a minimally unsatisfiable $F' \subseteq F[V]$. If now $\mathrm{var}(F') \neq \mathrm{var}(F[V]) = V$ would be the case, then we would loose control over the deficiency of F'. Fortunately this can not happen, as the following lemma shows.

Lemma 4.6 *Consider a multi-clause-set F with $\sigma(F) = \delta(F)$. Then for every unsatisfiable sub-multi-clause-set $F' \leq F$ we have $\mathrm{var}(F') = \mathrm{var}(F)$.*

Proof Assume $\mathrm{var}(F') \subset \mathrm{var}(F)$, and consider a minimally unsatisfiable sub-clause-set $F'' \subseteq F'$. By definition we have $\delta(F'') + \delta(F[\mathrm{var}(F) \setminus \mathrm{var}(F'')]) \leq \delta(F)$, where $\delta(F[\mathrm{var}(F) \setminus \mathrm{var}(F'')]) \geq \sigma(F) = \delta(F)$, from which we conclude $\delta(F'') \leq 0$, but $\delta(F'') \geq 1$ must hold since F'' is minimally unsatisfiable. ∎

Finally we are able to prove Theorem 4.1. Recall that F is a lean multi-clause-set with $n(F) > 0$, and we have to show the existence of a variable v with $\mathrm{vd}_F(v) \leq \mathrm{nM}(\sigma(F))$ and $\mathrm{ld}_F(v), \mathrm{ld}_F(\overline{v}) \leq \sigma(F)$.

Consider $\emptyset \neq V \subseteq \text{var}(F)$ with $\delta(F[V]) = \sigma(F)$, and let $F' := F[V]$. F' is unsatisfiable, since F is lean. Because of $\delta(F') = \sigma(F)$ we have $\delta(F') = \sigma(F')$. Consider some minimally unsatisfiable $F'' \subseteq F'$. By Lemma 4.6 we have $\text{var}(F'') = \text{var}(F')$. So we get $\delta(F'') = \delta(F') - (c(F') - c(F''))$. By Theorem 4.5 there is $v \in \text{var}(F'')$ with $\text{vd}_{F''}(v) \leq \text{nM}(\delta(F'')) = \text{nM}(\delta(F') - (c(F') - c(F''))) \leq \text{nM}(\delta(F')) - (c(F') - c(F''))$ and $\text{ld}_{F''}(v), \text{ld}_{F''}(\overline{v}) \leq \delta(F'') = \delta(F') - (c(F') - c(F''))$. Finally we have $\text{vd}_F(v) \leq \text{vd}_{F''}(v) + (c(F') - c(F''))$ (note that all occurrences of v in F are also in F'), and similarly for the literal degrees. QED

Corollary 4.7 *For a lean multi-clause-set F with $n(F) > 0$ we have $\mu\text{vd}(F) \leq \text{nM}(\delta(F))$.*

Corollary 4.8 *Consider a lean multi-clause-set F.*

1. *$\sigma(F) = 1$ holds if and only if $\mu\text{vd}(F) = 2$ holds.*

2. *$\mu\text{vd}(F) = 3$ implies $\sigma(F) = 2$.*

Proof First consider Part 1. If $\sigma(F) = 1$ (so $n(F) > 0$), then by Theorem 4.1 we have $\mu\text{vd}(F) \leq \text{nM}(1) = 2$, while in case of $\mu\text{vd}(F) = 1$ there would be a matching autarky for F. If on the other hand $\mu\text{vd}(F) = 2$ holds, then by definition $\sigma(F) \leq 2 - 1 = 1$, while $\sigma(F) \geq 1$ holds since F is matching lean. For Part 2 note that due to $\sigma(F) + 1 \leq \mu\text{vd}(F)$ we have $\sigma(F) \leq 2$, and then the assertion follows by Part 1. ∎

Remarks:

1. If F is lean, then $\sigma(F) = 2$ implies $\mu\text{vd}(F) \in \{3, 4\}$. An example for $\mu\text{vd}(F) = 4$ is given by the full unsatisfiable clause-set with 2 variables.

2. Is there a minimally unsatisfiable F with $\mu\text{vd}(F) = 4$ and $\sigma(F) = 3$?

3. More generally, is there for every $k \in \mathbb{N}$ a minimally unsatisfiable F with $\sigma(F) = k$ and $\mu\text{vd}(F) = k + 1$?

4.3 On finding the autarky

The following lemma (with Theorem 4.1) yields the proof of Corollary 4.2:

Lemma 4.9 *Consider a matching-lean multi-clause-set F with $n(F) > 0$. If we have $\mu\text{vd}(F) > \text{nM}(\sigma(F))$, then all $F[V]$ for $\emptyset \subset V \subseteq \text{var}(F)$ with $\delta(F[V]) = \sigma(F)$ are satisfiable.*

Proof If some $F[V]$ would be unsatisfiable, then by the proof of Theorem 4.1 in Subsection 4.2 there would be a variable v with $\text{vd}_F(v) \leq \text{nM}(\sigma(F))$. ∎

Now consider a matching-lean multi-clause-set F with $n(F) > 0$, where Corollary 4.2 is applicable (recall that we have $\sigma(F) \geq 1$), that is, we have $\mu\text{vd}(F) > \text{nM}(\sigma(F))$. So we know that F has a non-trivial autarky. Conjecture 4.3 states that finding such a non-trivial autarky in this case can be done in polynomial time (recall that finding a non-trivial autarky in general is NP-complete, which was shown in [7]).

The task of actually finding the autarky can be considered as finding a satisfying assignment for the following class $\mathcal{MLCR} \subset \mathcal{SAT} \cap \mathcal{MLEAN}$ of satisfiable(!) clause-sets F, obtained by considering all $F[V]$ for minimal sets of variables V with $\delta(F[V]) = \sigma(F)$ (where "CR" stands for "critical"):

Definition 4.10 *Let $\mathcal{MLCR}$ be the class of clause-sets F fulfilling the following three conditions:*

1. *F is matching-lean, has at least one variable, and does not contain the empty clause.*

2. *The only $\emptyset \neq V \subseteq \mathrm{var}(F)$ with $\delta(F[V]) = \sigma(F)$ is $V = \mathrm{var}(F)$ (and thus we have $\delta(F) = \sigma(F)$).*

3. *$\mu\mathrm{vd}(F) > \mathrm{nM}(\sigma(F))$.*

It is sufficient to find a non-trivial autarky for this class of satisfiable clause-sets.

Lemma 4.11 *Conjecture 4.3 is equivalent to the statement, that finding a non-trivial autarky for clause-sets in $\mathcal{MLCR}$ can be achieved in polynomial time.*

At the time of writing this article, we are not aware of elements of $\mathcal{MLCR}$ with a deficiency at least 2.

5 On strengthening the bound

For a class $\mathcal{C}$ of clause-sets let $\mu\mathrm{vd}(\mathcal{C})$ be the supremum of $\mu\mathrm{vd}(F)$ for $F \in \mathcal{C}$ with $n(F) > 0$. So by Theorem 4.5 we have $\mu\mathrm{vd}(\mathcal{MU}_{\delta=k}) \leq \mathrm{nM}(k)$ for all $k \in \mathbb{N}$. The task of precisely determining $\mu\mathrm{vd}(\mathcal{MU}_{\delta=k})$ for all k will be pursued in the forthcoming [13]; we need more theory for minimally unsatisfiable clause-sets (especially for unsatisfiable hitting clause-sets), and so here we can only mention some results connected with this article.

- We can show for infinitely many k that $\mu\mathrm{vd}(\mathcal{MU}_{\delta=k}) = \mathrm{nM}(k)$.

- We can also show that the smallest k where we don't have equality is $k = 6$, namely $\mu\mathrm{vd}(\mathcal{MU}_{\delta=6}) = 8 = \mathrm{nM}(6) - 1$.

- Let $\mathrm{nM}_1 : \mathbb{N} \to \mathbb{N}$ be defined by the recursion as in Definition 3.1, however with different start values, namely $\mathrm{nM}_1(k) := \mathrm{nM}(k)$ for $1 \leq k \leq 5$, while $\mathrm{nM}_1(6) := \mathrm{nM}(6) - 1 = 8$. We have $\mathrm{nM}_1(k) = \mathrm{nM}(k)$ for $k \notin \{2^m - m + 1 : m \in \mathbb{N}, m \geq 3\}$, while for $k = 2^m - m + 1$ we have $\mathrm{nM}_1(k) = \mathrm{nM}(k) - 1 = 2^m$.

- With the same proof as for Theorem 4.5 we can show $\mu\mathrm{vd}(\mathcal{MU}_{\delta=k}) \leq \mathrm{nM}_1(k)$ for all $k \in \mathbb{N}$.

- It seems that this bound can not be generalised to lean clause-sets (as in Theorem 4.1).

Conjecture 5.1 *For all $k \in \mathbb{N}$ we have $\mu\mathrm{vd}(\mathcal{MU}_{\delta=k}) \geq \mathrm{nM}(k) - 1$.*

Now we consider the question whether the bound holds for a larger class of clause-sets, that is, whether Theorem 4.1 can be generalised further, incorporating non-lean clause-sets. We consider the large class $\mathcal{MLEAN}$ of matching lean clause-sets, as introduced in [7], which is natural, since a basic property of $F \in \mathcal{MU}$ used in the proof of Theorem 4.1 is $\delta(F) \geq 1$ for $F \neq \top$, and this actually holds for all $F \in \mathcal{MLEAN}$. We will construct for arbitrary deficiency $k \in \mathbb{N}$ and $K \in \mathbb{N}$ clause-sets $F \in \mathcal{MLEAN}$ of deficiency k where every variable occurs positively at least K times. Thus neither the upper bound $\max(\mathrm{ld}_F(v), \mathrm{ld}_F(\overline{v})) \leq f(\delta(F))$ nor $\mathrm{ld}_F(v) + \mathrm{ld}_F(\overline{v}) = \mathrm{vd}_F(v) \leq f(\delta(F))$ for some chosen variable v and for any function f does hold for $\mathcal{MLEAN}$.

An example for $F \in \mathcal{MLEAN}_{\delta=1}$ with $\mu\mathrm{ld}(F) \geq 2$ (and thus $\mu\mathrm{vd}(F) \geq 4$) is given in Section 5 in [8], displaying a "star-free" (thus satisfiable) clause-set F with deficiency 1. In Subsection 9.3 in [11] it is shown that this clause-set is matching lean. "Star-freeness" in our context means, that there are no singular variables (occurring in one sign only once). Our simpler construction pushes the number of positive occurrences arbitrary high, but there are variables with only one negative occurrence (i.e., there are singular variables).

For a finite set V of variables let $M(V) \subseteq A(V)$ be the full clause-set over V containing all full clauses with at most one complementation. Obviously $\delta(F) = 1$ holds, and it is easy to see that $M(V) \in \mathcal{MLEAN}$ (for every $\emptyset \neq F' \subset F \subseteq A(V)$ we have $\delta(F') < \delta(F)$, and thus a full clause-set F is matching lean iff $\delta(F) \geq 1$). Furthermore by definition we have $\mathrm{ld}_{M(V)}(v) = |V|$ and $\mathrm{ld}_{M(V)}(\overline{v}) = 1$ for $v \in V$.

Lemma 5.2 *For $k \in \mathbb{N}$ and $K \in \mathbb{N}$ there are clause-sets $F \in \mathcal{MLEAN}_{\delta=k}$ such that for all variables $v \in \mathrm{var}(F)$ we have $\mathrm{ld}_F(v) \geq K$.*

Proof For $k = 1$ we can set $F := M(\{v_1, \ldots, v_K\})$; so assume $k \geq 2$. Consider any clause-set $G \in \mathcal{MLEAN}_{\delta=k-1}$ with $n := n(G) \geq K$ (for example we could use $F \in \mathcal{MU}_{\delta=k-1}$), and let $V := \mathrm{var}(G)$. Consider a disjoint copy of V, that is a set V' of variables with $V' \cap V = \emptyset$ and $|V'| = |V|$, and consider two enumerations of the clauses $M(V) = \{C_1, \ldots, C_{n+1}\}$, $M(V') = \{C'_1, \ldots, C'_{n+1}\}$. Now

$$F := G \cup \big\{\, C_i \cup C'_i : i \in \{1, \ldots, n+1\} \,\big\}$$

has no matching autarky: If φ is a matching autarky for F, then $\mathrm{var}(\varphi) \cap V = \emptyset$ since G is matching lean, whence $\mathrm{var}(\varphi) \cap V' = \emptyset$ since $M(V')$ is matching lean, and thus φ must be trivial. Furthermore we have $n(F) = 2n$ and $c(F) = c(G) + n + 1$, and thus $\delta(F) = c(G) + n + 1 - 2n = \delta(G) + 1 = k$. By definition for all variables $v \in \mathrm{var}(F)$ we have $\mathrm{ld}_F(v) \geq n$. ∎

Remarks:

1. It remains open whether for deficiency $k \in \mathbb{N}$ we find examples $F \in \mathcal{MLEAN}_{\delta=k}$ with $\mu\mathrm{ld}(F) \geq k+1$ (the above mentioned star-free clause-sets shows that this is the case for $k = 1$), or stronger, $\mu\mathrm{ld}(F) \geq K$ for arbitrary $K \in \mathbb{N}$.

2. The clause-sets F constructed in Lemma 5.2 are not elements of $\mathcal{MLCR}_{\delta=k}$ for $k \geq 2$, since $\delta(F[V']) = n + 1 - n = 1$, thus $\sigma(F) = 1$, and so Condition 2 of Definition 4.10 is not fulfilled. The corresponding autarky is a satisfying assignment of $M(V')$, which is easy to find.

6 Conclusion and open problems

We have shown the upper bound $\mu\mathrm{vd}(F) \leq \mathrm{nM}(\sigma(F))$ for lean clause-sets (Theorem 4.1). The function $\mathrm{nM}(k)$ has been characterised in Lemma 3.8 and Corollary 3.9. We presented first initial results regarding the sharpness of the bound and regarding the constructive aspects of the bound (i.e., what happens if the bound is violated). There remain several open problems:

1. Prove Conjecture 4.3, which says that such an autarky, which must exist if a clause-set does not fulfil the upper bound on the minimum variable degree of Theorem 4.1, can be found in polynomial time. See Subsection 4.3 for more information on this topic.

2. Generalise Theorem 4.5 to clause-sets with non-boolean variables; see the discussion after Theorem 4.5.

3. See the remarks to Corollary 4.8 (an underlying question is to understand better the quantity "surplus").

4. Strengthen the bound on the minimum variable degree for minimally unsatisfiable clause-sets (see the forthcoming [13]).

5. Strengthen the construction of Lemma 5.2 (perhaps completely different constructions are needed).

As mentioned in the introduction, a major motivation for us is the project of the classification of minimally unsatisfiable clause-sets for deficiencies $\delta = 1, 2, \ldots$. Especially the classification of unsatisfiable hitting clause-sets in dependency on the deficiency seems very interesting (recall that a hitting clause-set F is defined by the condition that every two clauses $C, C' \in F$, $C \neq C'$, clash in at least one variable, that is $|C \cap \overline{C'}| \geq 1$). The main conjecture is:

Conjecture 6.1 *For every deficiency $k \in \mathbb{N}$ there are only finitely many isomorphism types of non-singular unsatisfiable hitting clause-sets.*

For $k \leq 2$ this conjecture follows from known results, while recently we were able to prove it for $k = 3$.

References

[1] Ron Aharoni and Nathan Linial. Minimal non-two-colorable hypergraphs and minimal unsatisfiable formulas. *Journal of Combinatorial Theory*, A 43:196–204, 1986.

[2] Gennady Davydov, Inna Davydova, and Hans Kleine Büning. An efficient algorithm for the minimal unsatisfiability problem for a subclass of CNF. *Annals of Mathematics and Artificial Intelligence*, 23:229–245, 1998.

[3] Heidi Gebauer, Tibor Szabo, and Gabor Tardos. The local lemma is tight for SAT. Technical Report arXiv:1006.0744v1 [math.CO], arXiv.org, June 2010.

[4] Hans Kleine Büning. On subclasses of minimal unsatisfiable formulas. *Discrete Applied Mathematics*, 107:83–98, 2000.

[5] Hans Kleine Büning and Oliver Kullmann. Minimal unsatisfiability and autarkies. In Armin Biere, Marijn J.H. Heule, Hans van Maaren, and Toby Walsh, editors, *Handbook of Satisfiability*, volume 185 of *Frontiers in Artificial Intelligence and Applications*, chapter 11, pages 339–401. IOS Press, February 2009.

[6] Oliver Kullmann. An application of matroid theory to the SAT problem. In *Fifteenth Annual IEEE Conference on Computational Complexity (2000)*, pages 116–124. IEEE Computer Society, July 2000.

[7] Oliver Kullmann. Lean clause-sets: Generalizations of minimally unsatisfiable clause-sets. *Discrete Applied Mathematics*, 130:209–249, 2003.

[8] Oliver Kullmann. On some connections between linear algebra and the combinatorics of clause-sets. In John Franco, Enrico Giunchiglia, Henry Kautz, Hans Kleine Büning, Hans van Maaren, Bart Selman, and Ewald Speckenmeyer, editors, *Sixth International Conference on Theory and Applications of Satisfiability Testing*, pages 45–59, May 2003. Santa Margherita Ligure – Portofino (Italy), May 5, 2003 to May 8, 2003.

[9] Oliver Kullmann. Green-Tao numbers and SAT. In Ofer Strichman and Stefan Szeider, editors, *Theory and Applications of Satisfiability Testing - SAT 2010*, volume 6175 of *Lecture Notes in Computer Science*, pages 352–362. Springer, 2010.

[10] Oliver Kullmann. Constraint satisfaction problems in clausal form. Technical Report arXiv:1103.3693v1 [cs.DM], arXiv, March 2011.

[11] Oliver Kullmann. Constraint satisfaction problems in clausal form I: Autarkies and deficiency. *Fundamenta Informaticae*, 109, 2011. To appear.

[12] Oliver Kullmann. Constraint satisfaction problems in clausal form II: Minimal unsatisfiability and conflict structure. *Fundamenta Informaticae*, 109, 2011. To appear.

[13] Oliver Kullmann and Xishun Zhao. On extremal conjunctive normal forms w.r.t. variables with few occurrences. In preparation, 2011.

[14] Craig A. Tovey. A simplified NP-complete satisfiability problem. *Discrete Applied Mathematics*, 8:85–89, 1984.

[15] Xishun Zhao and Ding Decheng. Two tractable subclasses of minimal unsatisfiable formulas. *Science in China (Series A)*, 42(7):720–731, July 1999.