An application of matroid theory to the SAT problem

Oliver Kullmann*

Department of Computer Science
University of Toronto
Toronto, Ontario M5S 3G4

e-mail: kullmann@cs.toronto.edu
http://www.cs.utoronto.ca/~kullmann/

February 15, 2000

Abstract

We consider the *deficiency* $\delta(F) := c(F) - n(F)$ and the *maximal deficiency* $\delta^*(F) := \max_{F' \subseteq F} \delta(F)$ of a set F of clauses (a conjunctive normal form), where $c(F)$ is the number of clauses in F and $n(F)$ is the number of variables.

Combining ideas from *matching* and *matroid theory* with techniques from the area of *resolution refutations*, we prove that for clause-sets F with $\delta^*(F) \leq k$, where k is considered as a constant, the *SAT* problem, the *minimally unsatisfiability* problem and the *MAXSAT* problem are decidable in *polynomial time* (previously, only poly-time decidability of the minimally unsatisfiability problem was known, and that only for $k = 1$).

1   Introduction

Minimally unsatisfiable clause-sets

A conjunctive normal form F (represented as a *clause-set* in this paper) is called *minimally unsatisfiable* iff F is unsatisfiable, and any strict subset is satisfiable. Let $\mathcal{MUSAT}$ denote the class of all minimally unsatisfiable clause-sets. It has been shown in [13] that $\mathcal{MUSAT}$ is D^P-*complete*, where $D^P = \{L_1 \cap L_2 : L_1 \in \mathrm{NP} \wedge L_2 \in \text{co-NP}\}$ is the class of languages representable as the intersection of a language in NP and a language in co-NP.

A fundamental property for minimally unsatisfiable clause-sets is proven in [1]. Applying Hall's theorem they show $c(F) \geq n(F)+1$ for $F \in \mathcal{MUSAT}$ (attributing this lemma to M. Tarsi), where $c(F)$ is the number of clauses of F, and $n(F)$ is the number of variables in F.

Now two *special classes* of minimally unsatisfiable clause-sets are known, where the decision problem is in polynomial time:

- $\mathcal{SMUSAT}(1)$, the class of *strongly* minimally unsatisfiable clause-sets F with $c(F) = n(F) + 1$, where "strongly" additionally requires, that adding any literal to any clause of F renders F satisfiable (this class has been introduced in [1], and from their characterisation of this class poly-time decidability follows);

- $\mathcal{MUSAT}(1)$, the class of all minimally unsatisfiable F with $c(F) = n(F) + 1$ (characterised and shown to be poly-time decidable in [4] (using a quite complicated proof)).

At the SAT'98 workshop ([6]) Hans Kleine Büning posed the question whether *all classes* $\mathcal{MUSAT}(k) := \{F \in \mathcal{MUSAT} : c(F) \leq n(F) + k\}$ for constant $k \geq 1$ are *poly-time decidable* (answered positively in this article).

Furthermore, in [3] for $F \in \mathcal{MUSAT}(k)$ the upper bound $2^{k-1} \cdot n(F)^2$ for the minimal number of resolution steps in a *resolution tree refutation* of F has been established (a strengthening of this fact is basic for our considerations).

In essence, most techniques from the work mentioned above is included in our approach. Additionally, the following notions have been of importance.

Autarkies and the deficiency

Straightening and generalising prior work, the use of the notion of an *autarky* for clause-sets F (partial assignments satisfying each clause of F they "touch") and of the *deficiency* $\delta(F) := c(F) - n(F)$ are basic for this paper.

Being unaware of [1], in [5] an alternative (inductive) proof of "Tarsi's lemma" is given, based on the (valuable) observation that substituting a truth value for any (single) variable in $F \in \mathcal{SMUSAT}$ results in a clause-set still in $\mathcal{MUSAT}$. They notice that a clause-set F with $1 \leq c(F) \leq n(F)$ must have a non-empty *autark subset* $F' \subseteq F$, that is, F' is satisfiable by an autarky (not involving any variable appearing in $F \setminus F'$).

Also not using [1], in [14] the property $c(F) \geq n(F) + 1$ has been shown for the more general class $F \in \mathcal{UMUSAT}$, the closure of $\mathcal{MUSAT}$ under finite union. They also show for an arbitrary clause-set F and and a subset $F' \subseteq F$ with the property that $c(F') - n(F')$ is maximal, that for every subset $F'' \subseteq F \setminus F'$ one has $c(\mathrm{var}(F') * F'') - n(\mathrm{var}(F') * F'') \leq 0$, where $\mathrm{var}(F') * F''$ is obtained from F'' by crossing out all variables appearing in F'. *Actually* this says that $F \setminus F'$ is an *autark subset* of F, and thus for $F \in \mathcal{UMUSAT}$ we must have $F' = F$, and property (1) given in the next paragraph follows. Unfortunately the argumentation in [14] is very lengthy and does not arrive at such conceptual insights.

Independent of [14], the notion of *deficiency* $\delta(F) := c(F) - n(F)$ has been introduced in [7] and made fruitful for lucid argumentation. The class of *matched clause-sets* F, given by the property, that for all $F' \subseteq F$ we have $\delta(F') \leq 0$, is introduced, and it is shown that this class is poly-time decidable and contains only satisfiable clause-sets. A short proof of the statement

$$\forall\, F' \subset F : \delta(F') < \delta(F) \tag{1}$$

for $F \in \mathcal{MUSAT}$ is given (since $\delta(\top) = 0$, where $\top$ is the empty clause-set, Tarsi's lemma follows).

These attempts have been unified and strengthened in [9], starting a systematic investigation of the notion of autarkies and autark subsets. *Generalising* $\mathcal{MUSAT}$ (and $\mathcal{UMUSAT}$), the class $\mathcal{LEAN}$ of *lean clause-sets* F has been introduced, where all clauses of F *can* be used in a resolution tree refutation of F. It is shown, that a clause-set F is lean if and only if $\top := \emptyset$ is the *only autark subset* of F.

3

In order to obtain special cases of autark subsets, which are computable in polynomial time, the notion of *linearly autark subsets* is introduced, and the class $\mathcal{LEAN}$ has been enlarged to $\mathcal{LLEAN}$, the class of *linearly lean* clause-sets where $\top$ is the only linearly autark subset. $\mathcal{LLEAN}$ is poly-time decidable (using Linear Programming), and property (1) in fact holds true for all clause-sets $F \in \mathcal{LLEAN}$.

Finally, in [9] the *maximal deficiency* $\delta^*(F) := \max_{F' \subseteq F} \delta(F')$ has been investigated. This notion in fact is of basic importance (see at the end of the following subsection). Note that F is a matched clause-set iff $\delta^*(F) = 0$, and that for all minimally unsatisfiable F we have $\delta^*(F) = \delta(F)$ by (1).

The contributions of this paper

Using $\mathcal{C}(k) := \{F \in \mathcal{C} : \delta^*(F) \leq k\}$ for any class $\mathcal{C}$ contained in the set $\mathcal{CLS}$ of all clause-sets (note the consistency with the prior use of this notation), we prove for any $k \geq 0$ (see Section 4):

- $\mathcal{SAT}(k)$ as well as the classes

$$\mathcal{SMUSAT}(k) \subset \mathcal{MUSAT}(k) \subset \mathcal{UMUSAT}(k) \subset$$
$$\subset \mathcal{LEAN}(k) \subset \mathcal{USAT}(k) \cup \{\top\}$$

  are *poly-time decidable* ($\mathcal{SAT}$ and $\mathcal{USAT}$ are the satisfiable resp. unsatisfiable clause-sets).

- For $F \in \mathcal{CLS}(k)$ the two *decompositions*

$$F = F_1^i \cup F_2^i, \ F_1^i \cap F_2^i = \emptyset$$

  ($i = 1, 2$) given by

  - F_1^1 is the *largest lean subset* of F, F_2^1 is the *largest autark subset* of F
  - F_1^2 is the union of all *minimally unsatisfiable* sub-clause-sets of F, F_2^2 the intersection of all *maximally satisfiable* sub-clause-sets

  are computable in *polynomial time*, as well as a *largest satisfiable subset*.

4

Furthermore the improved (and now tight) upper bound $2^{\delta^*(F)-1} \cdot n(F)$ for the minimal number of resolution steps needed in a *resolution tree refutation* is derived.

The basic new insights towards these results are:

1. In Theorem 3.2 lean sub-clause-sets of $F \in \mathcal{CLS}$ are shown to be (special) *circuits* of some *matroids*, derived from the *transversal matroid* $T(F)$ given by the (natural) bipartite graph associated with F. The maximal deficiency $\delta^*(F')$ for subsets $F' \subseteq F$ is the *nullity* $c(F') - \mathrm{rank}(F')$ in $T(F)$ (using the rank function of $T(F)$).

   It follows that for fixed maximal deficiency of F a set of sub-clause-sets containing all lean sub-clause-sets can be computed in *polynomial time*. (See Appendix A, where also some basic notions from matroid theory are explained.)

2. In Theorem 3.11 it is shown that unsatisfiable clause-sets with fixed maximal deficiency have *small* "DPL-like refutations" (or "(generalised) semantic tree refutations") of a *special form*, which can be searched for in polynomial time.

2   Notation

$\mathcal{CLS}$ is the set of all clause-sets (conjunctive normal forms treated as sets), where we require clauses to be complement-free, $\mathcal{SAT}$ is the set of all satisfiable clause-sets and $\mathcal{USAT}$ the set of all unsatisfiable clause-sets. The set of minimally unsatisfiable clause-sets is denoted by $\mathcal{MUSAT}$ (unsatisfiable clause-sets such that every strict subset is satisfiable), while $\mathcal{SMUSAT}$ (called "strongly minimal" in [1] and "saturated" in [5]) is the set of minimally unsatisfiable clause-sets such that adding any literal to any clause renders them satisfiable. Finally $\mathcal{UMUSAT}$ is the closure of $\mathcal{MUSAT}$ under finite union (introduced in [14]). The complement of a literal x is denoted by $\overline{x}$, while for a clause C we set $\overline{C} := \{\overline{x} : x \in C\}$. Variables appearing in a literal x, a clause C or a clause-set F are denoted by $\mathbf{var}(x)$, $\mathbf{var}(C)$ and $\mathbf{var}(F)$ respectively (the latter two being sets).

For $F \in \mathcal{CLS}$ let $\boldsymbol{n(F)}$ be the number of variables and $\boldsymbol{c(F)}$ the number of clauses, while $\boldsymbol{\#_x(F)}$ is the number of occurrences of literal x; $\boldsymbol{\delta(F)} := c(F) - n(F)$ and $\boldsymbol{\delta^* F} := \max_{F' \subseteq F} \delta(F)$.

5

For any $\mathcal{C} \subseteq \mathcal{CLS}$ and $k \geq 0$ let $\boldsymbol{\mathcal{C}(k)} := \{F \in \mathcal{C} : \delta^*(F) \leq k\}$.

For any partial assignment φ (assigning truth values to variables) we denote by $\boldsymbol{\varphi * F}$ the result of applying φ to F, that is, all clauses satisfied by φ are removed, and from the remaining clauses all literals falsified by φ are cancelled. $\boldsymbol{\langle x \to \varepsilon \rangle}$ denotes the partial assignment setting literal x to ε (and the complement $\overline{x}$ to $\overline{\varepsilon}$). φ is called an **autarky** for F if each clause of F either is satisfied by φ or has no variables given a value by φ.

And for a set V of variables let $\boldsymbol{V * F}$ denote the clause-set resulting when crossing out all literals with variables in V. ($V * F$ is equal to the union of all $\varphi * F$ with $\mathrm{var}(\varphi) = V$.)

3 The main lemmas

3.1 Circuits of the transversal matroid $T(F)$

Definition 3.1 *For $F \in \mathcal{CLS}$ and $k \geq 0$ let $\boldsymbol{\mathfrak{C}_k(F)}$ be the set of all sub-clause-sets $F' \subseteq F$ with $\delta(F') = k + 1$ such that for all $F'' \subset F'$ one has $\delta(F'') \leq k$. Furthermore $\boldsymbol{\mathfrak{C}(F)} := \bigcup_{0 \leq k < \delta^*(F)} \mathfrak{C}_k(F)$.* ∎

Theorem 3.2 *Consider $F \in \mathcal{CLS}$ and $k \geq 0$. Then $\mathfrak{C}_{k+1}$ is the set of circuits of a matroid $\boldsymbol{T_k(F)}$ on F (the elements are the clauses of F), where the independent subsets of $T_k(F)$ are those $F' \subseteq F$ with $\delta^*(F') \leq k$.*

Proof: First we consider the case $k = 0$. The set $\{F' \subseteq F : \delta^*(F') = 0\}$ is the set of independent subsets of the *transversal matroid* $T(F) = T(B(F))$ induced by the bipartite graph $B(F) = (F, R, \mathrm{var}(F))$, where the relation $(C, v) \in R \subseteq F \times \mathrm{var}(F)$ holds iff $v \in \mathrm{var}(C)$ (see Appendix A).

The set of circuits of $T(F)$ is $\mathfrak{C}_0(F)$. For the rank in $T(F) = T_0(F)$ we have $\mathrm{rank}(F') = c(F') - \delta^*(F')$ for all $F' \subseteq F$. The function $f(F') := \mathrm{rank}(F') + k$ is submodular and increasing, and thus the set $\mathfrak{C}'_k$ consisting of all $F' \subseteq F$ which are minimal w.r.t. the property "$f(F') < c(F')$" is the set of circuits of a matroid $T'_k(F)$, whose set of independent subsets is

$$\mathfrak{I}'_k = \big\{ F' \subseteq F : \forall\, F'' \subseteq F' \, [\, f(F'') \geq c(F'') \,] \big\}$$

(see [12], page 380, 12.1.1 and 12.1.2). Since for $C \in F$ and $F' \subseteq F$ we have $f(F' \cup \{C\}) \leq f(F') + 1$, in fact $\mathfrak{I}'_k = \{F' \subseteq F : f(F') \geq c(F')\} = \{F' \subseteq F : \delta^*(F') \leq k\}$ and $\mathfrak{C}'_k = \mathfrak{C}_k$ holds. ∎

6

Using the enumeration of the circuits of $T_k(F)$ in Appendix A, we get

Corollary 3.3 *For $k \geq 0$ and $F \in \mathcal{CLS}(k)$ the sets $\mathfrak{C}_0(F), \ldots, \mathfrak{C}_{k-1}(F)$ can be computed in polynomial time (k is treated as constant).* ∎

3.2 Lean clause-sets

For $F \in \mathcal{CLS}$ we call $F' \subseteq F$ an **autark subset** of F iff there is an autarky φ for F with $F' = \{C \in F : \varphi(C) = 1\}$. We make frequent use of the following reformulation:

$$F' \text{ is an autark subset of } F \Leftrightarrow \operatorname{var}(F \setminus F') * F' \in \mathcal{SAT}. \tag{2}$$

Generalising minimally unsatisfiable clause-sets, in [9] the notion of *lean clause-sets* has been introduced, which are clause-sets such that every clause *can* be used in a resolution refutation. It has been shown, that a clause-set is lean if and only if it has no non-trivial autarky. In our context, we take this as the definition of lean clause-sets.

Definition 3.4 *Let $\mathcal{LEAN}$ be the set of autarky-free clause-set $F \in \mathcal{CLS}$, that is, the only autark subset of F is $\top$.* ∎

Simple properties (for more information see [9]):

1. $\mathcal{UMUSAT} \subset \mathcal{LEAN} \subset \mathcal{USAT} \cup \{\top\}$.

2. $\mathcal{LEAN}$ is closed under crossing out variables, that is, for $F \in \mathcal{LEAN}$ and $V \subseteq \mathcal{VA}$ we have $V * F \in \mathcal{LEAN}$.

3. $\mathcal{LEAN}$ is closed under union: $F_1, F_2 \in \mathcal{LEAN} \Rightarrow F_1 \cup F_2 \in \mathcal{LEAN}$.

4. $F \in \mathcal{LEAN} \Leftrightarrow F \setminus \{\bot\} \in \mathcal{LEAN}$.

For detecting autarkies in our context, the following inequality (with straightforward proof) from [9] is basic.

Lemma 3.5 *For any clause-set $F \in \mathcal{CLS}$ and any subset $F' \subseteq F$ we have $\delta^*(\operatorname{var}(F') * (F \setminus F')) + \delta(F') \leq \delta^*(F)$.* ∎

7

Since $\delta^*(F) = 0$ implies $F \in \mathcal{SAT}$ (every maximum matching in $B(F)$ induces a satisfying assignment; clause-sets F with $\delta^*(F) = 0$ are called "matched clause-sets" in [7]), by (2) we get the following sufficient condition for a subset being autark (also included in [14], while a forerunner can be found in [5]; for a refinement using linear programming see [9]).

Corollary 3.6 *For $F \in \mathcal{CLS}$ and $F' \subseteq F$ with $\delta(F') = \delta^*(F)$ the set $F \setminus F'$ is an autark subset of F.* ∎

From Corollary 3.6 and Definition 3.4 we obtain the following fundamental lemma (see [9] for a generalisation; in [7] Lemma 3.7 is proved for the special case of minimally unsatisfiable clause-sets).

Lemma 3.7 *For $F \in \mathcal{LEAN}$ we have $\forall F' \subset F : \delta(F') < \delta(F)$. Since $\delta(\top) = 0$ it follows $\delta(F) \geq 1$.* ∎

Hence all lean sub-clause-sets of F (and thus all minimally unsatisfiable sub-clause-sets) are contained in $\mathfrak{C}(F)$:

$$\forall F \in \mathcal{CLS} \; \forall F' \subseteq F : F' \in \mathcal{LEAN} \Rightarrow F' \in \mathfrak{C}(F). \tag{3}$$

Corollary 3.8 $\mathcal{MUSAT}(1) = \mathcal{LEAN}(1) \setminus \{\top\}$. ∎

3.3 Splitting up lean clause-sets

Now we turn to the problem of investigating the deficiency of clause-sets when performing splitting, which is, besides Corollary 3.3, the basis for our decision algorithms, and also establishes the connection to tree-like resolution. For the proofs of Lemmas 3.9 and 3.10 see Appendix D.

In general for any clause-set $F \in \mathcal{CLS}$ and any literal x we have

$$\delta^*(\langle x \to 1 \rangle * F) \leq \delta^*(F) + 1 \tag{4}$$

since for every $F' \subseteq \langle x \to 1 \rangle * F$ there is a $F'' \subseteq F$ with $c(F'') = c(F')$ and $n(F'') \leq n(F') + 1$. Our problem is to actually *reduce* the deficiency by splitting.

Part 3 of the following lemma generalises (and simplifies) the "splitting theorem" from [3].

Lemma 3.9 *Consider $F \in \mathcal{LEAN}$ with $\forall\, v \in \mathrm{var}(F) : (\#_v + \#_{\overline{v}})(F) \geq 3$.*

1. *For $\emptyset \neq V \subseteq \mathrm{var}(F)$: $c(\{C \in F : \mathrm{var}(C) \cap V \neq \emptyset\}) \geq |V| + 2$.*

2. *For $F' \subseteq F$ with $\mathrm{var}(F') \subset \mathrm{var}(F)$: $\delta(F') \leq \delta(F) - 2$.*

3. *For all literals x with $\#_x(F) \geq 2$ we have*

$$\delta^*(\langle x \to 1 \rangle * F) \leq \delta(F) - 1.$$

Using "Davis-Putnam resolution" (see Appendix B) we get

Lemma 3.10 *Consider $F \in \mathcal{USAT}$ and $F' \subseteq F$ with $F' \in \mathcal{MUSAT}$ and $\delta(F') \geq 2$. Then there is a variable $v \in \mathrm{var}(F)$ such that for both $\varepsilon \in \{0, 1\}$ the clause-set $\langle v \to \varepsilon \rangle * F$ contains some $F'_\varepsilon \in \mathcal{MUSAT}(\delta(F') - 1)$.*

The following theorem now immediately follows.

Theorem 3.11 *Consider $F \in \mathcal{USAT}(k)$ for some $k \geq 1$. Then there is a binary tree T (directed from the root to the leaves) of height at most $k - 1$ such that*

- *the root of T is labelled by F;*

- *the two direct successors of an inner node labelled by F' are labelled by $\langle v \to 0 \rangle * F'$ resp. $\langle v \to 1 \rangle * F'$ for some variable $v \in \mathrm{var}(F')$;*

- *every leaf label G contains some $G' \subseteq G$ with $G' \in \mathcal{MUSAT}(1)$.* ∎

From the results in [10] it follows that the class

$$\mathcal{C} := \{F \in \mathcal{CLS} \mid \exists\, F' \subseteq F : F' \in \mathcal{MUSAT}(1)\}$$

of possible leaf labels of trees T is (properly) contained in the class of clause-set refutable by "read-once" resolution (resolution refutations in tree form, where every input clause can be used at most once), shown to be NP-complete in [8]. It is an interesting question whether also the problem "$F \in \mathcal{C}$?" is NP-complete.

We conclude this section by an application to resolution complexity. Theorem 3.11 together with Lemma C.3 yields immediately

Corollary 3.12 $2^{\delta^*(F)-1} \cdot n(F)$ *is an upper bound for the minimal number of inner nodes in resolution tree refutations of $F \in \mathcal{USAT}$.* ∎

4 The poly-time computability results

Theorem 4.1 *For fixed $k \geq 0$ the class $\mathcal{SAT}(k)$ is decidable in polynomial time. It follows that also $\mathcal{MUSAT}(k)$ and $\mathcal{SMUSAT}(k)$ are poly-time decidable. Furthermore for $F \in \mathcal{CLS}(k)$ a satisfiable $F' \subseteq F$ of largest size (containing at least $c(F) - \delta^*(F)$ clauses) is poly-time computable (that is, the MAXSAT problem is poly-time solvable for $\mathcal{CLS}(k)$).*

Proof: Given $F \in \mathcal{CLS}(k)$, search for a tree T according to Theorem 3.11:

For a leaf label G we have $\delta^*(G) \leq k + (k - 1) = 2k - 1$ by (4). Hence the property "$\exists\, G' \subseteq G : G' \in \mathcal{MUSAT}(1)$?" is decidable in poly-time by Corollary 3.3 and Lemma C.3.

It follows that the existence of such a T can be decided in poly-time, and hence SAT decision for F is poly-time.

W.r.t the properties "minimally unsatisfiable" and "saturated minimally unsatisfiable" now we can simply check whether the definitions hold.

And w.r.t. MAXSAT note that a basis F' of the transversal matroid $T(F)$ is computable in polynomial time. We have $F' \in \mathcal{SAT}$ and $c(F') = c(F) - \delta^*(F)$, and now the question is whether a larger satisfiable subset exists. So we simply search through all (polynomially many) subset $F'' \subseteq F$ with $c(F'') > c(F) - \delta^*(F)$ and decide whether $F'' \in \mathcal{SAT}$ holds. ∎

The unique "autarky-decomposition"

From [9] we get the following unique decomposition of $F \in \mathcal{CLS}$ into the largest lean subset F_{l} and the largest autark subset F_{a}

$$F = F_{\mathrm{l}} \cup F_{\mathrm{a}}, \quad F_{\mathrm{l}} \cap F_{\mathrm{a}} = \emptyset \tag{5}$$

where F_{l} has the following characterisations:

- F_{l} is the union of all lean subsets $F' \in \mathcal{LEAN}$ of F;

- F_{l} is the largest lean subset $F' \in \mathcal{LEAN}$ of F;

- F_{l} is the set of all clauses $C \in F$ such that a resolution tree refutation of F exists using C as an axiom;

on the other side, F_{a} has the following characterisations:

- F_{a} is the union of all autark subsets of F;

- F_{a} is the largest autark subset of F.

Theorem 4.2 *For fixed $k \geq 0$ and $F \in \mathcal{CLS}(k)$ the unique decomposition of F according to equation (5) is computable in polynomial time. Thus the class $\mathcal{LEAN}(k)$ is poly-time decidable.*

Proof: Since F_{a} is the largest autark subset of F, by (2) and (3) we can characterise F_{l} as the smallest subset (w.r.t. inclusion) $F' \subseteq F$ with

$$F' \in \mathfrak{C}(F) \quad \text{and} \quad \mathrm{var}(F') * (F \setminus F') \in \mathcal{SAT}.$$

Lemma 3.5 gives $\delta^*(\mathrm{var}(F') * (F \setminus F')) \leq \delta^*(F) - \delta(F') \leq \delta^*(F) - 1$.

Hence by Corollary 3.3 and Theorem 4.1 we can find F_{l} in polynomial time. ∎

"Minimally unsatisfiable" vs. "maximally satisfiable"

Lemma 4.3 *For a clause-set F let*

$$\begin{aligned}
\min(F) &:= \{F' \subseteq F : F' \text{ minimally unsatisfiable}\} \\
\max(F) &:= \{F' \subseteq F : F' \text{ maximally satisfiable}\},
\end{aligned}$$

where $F' \subseteq F$ is maximally satisfiable iff F' is satisfiable and every $F'' \subseteq F$ with $F' \subset F''$ is unsatisfiable. Then

$$\bigcup \min(F) \cup \bigcap \max(F) = F, \quad \bigcup \min(F) \cap \bigcap \max(F) = \emptyset.$$

Proof: Consider $C \in F$:

If there is a minimally unsatisfiable $F' \subseteq F$ with $C \in F'$, then consider a maximally satisfiable $F'' \subseteq F$ with $F' \setminus \{C\} \subseteq F''$ — here we have $C \notin F''$.

If there is a maximally satisfiable $F' \subseteq F$ with $C \notin F'$, consider a minimally unsatisfiable $F'' \subseteq F$ with $F'' \subseteq F' \cup \{C\}$ — now $C \in F''$. ∎

Note that the largest autark subset F_{a} from equation (5) is contained in $\bigcap \max(F)$ (since for an autark subset F_1 of F and a satisfiable subset F_2 of F also $F_1 \cup F_2$ is satisfiable).

Theorem 4.4 *For fixed $k \geq 0$ and $F \in \mathcal{CLS}(k)$ the unique decomposition of F according to Lemma 4.3 is computable in polynomial time. Thus the class $\mathcal{UMUSAT}(k)$ is poly-time decidable.*

Proof: By Corollary 3.3, (3) in Subsection 3.2 and Theorem 4.1 all minimally unsatisfiable subsets of F can be enumerated in poly-time. ∎

To conclude, it seems to me an interesting observation that $\mathcal{UMUSAT}$ is the closure of $\mathcal{MUSAT}$ under crossing out variables (stated without a proof here, since it is not the subject of this extended abstract), that is

$$\mathcal{UMUSAT} = \big\{\, V * F : V \subseteq \mathcal{VA} \wedge F \in \mathcal{MUSAT} \,\big\}.$$

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] Martin Aigner. *Kombinatorik; II. Matroide und Transversaltheorie.* Springer-Verlag, 1976.

[3] Hans Kleine Büning. An upper bound for minimal resolution refutations. In Georg Gottlob, Etienne Grandjean, and Katrin Seyr, editors, *Computer Science Logic 12th International Workshop, CSL '98*, volume 1584 of *Lecture Notes in Computer Science*, pages 171–178. Springer, 1999.

[4] 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.

[5] T. Fliti and G. Reynaud. Sizes of minimally unsatisfiable conjunctive normal forms. Faculté des Sciences de Luminy, Dpt. Mathematique-Informatique, 13288 Marseille, France, November 1994.

[6] John Franco, Giorgio Gallo, Hans Kleine Büning, and Ewald Speckenmeyer, editors. *Second Workshop On The Satisfiability Problem*, 1998. Eringerfeld (Germany), May 10 - 14 (1998).

[7] John Franco and Allen Van Gelder. A perspective on certain polynomial time solvable classes of satisfiability. To appear in Discrete Applied Mathematics, 2000.

[8] Kazuo Iwama and Eiji Miyano. Intractability of read-once resolution. In *Proceedings Structure in Complexity Theory, Tenth Annual Conference*, pages 29–36. IEEE, 1995.

[9] Oliver Kullmann. Investigations on autark assignments. To appear in Discrete Applied Mathematics, 2000.

[10] Oliver Kullmann. On structural properties of minimally unsatisfiable clause-sets. In preparation, January 2000.

[11] László Lovász and Michael D. Plummer. *Matching Theory*, volume 121 of *Mathematics Studies*. North-Holland, 1986.

[12] James G. Oxley. *Matroid Theory*. Oxford University Press, 1992.

[13] Christos H. Papadimitriou and David Wolfe. The complexity of facets resolved. *Journal of Computer and System Sciences*, 37:2–13, 1988.

[14] Ma Shaohan and Liang Dongmin. A polynomial-time algorithm for reducing the number of variables in MAX SAT problem. *Science in China (Series E)*, 40(3):301–311, June 1997.

[15] Hassler Whitney. On the abstract properties of linear dependence. *American Journal of Mathematics*, 57:509–533, 1935.

APPENDIX

A   The transversal matroid of a bipartite graph

We recall some basic matroid theory (see any introductory text on the subject, like [12], Chapter 1). Let X be a finite set. A set $\mathfrak{I} \subseteq \mathbb{P}(X)$ of subsets of X is the set of *independent subsets* of a matroid $M(X)$ on X iff the following "independence axioms" are fulfilled:

I1 $\emptyset \in \mathfrak{I}$

I2 $\forall A \in \mathfrak{I} \; \forall A' \subseteq A : A' \in \mathfrak{I}$

I3 $\forall A_1, A_2 \in \mathfrak{I} : |A_1| > |A_2| \Rightarrow \exists a \in A_1 \, [\, A_2 \cup \{a\} \in \mathfrak{I} \,]$.

$M(X)$ is uniquely determined by $\mathfrak{I}$ fulfilling (I1) - (I3). A *circuit* of $M(X)$ is a minimal *dependent* (that is, not independent) subset of X. The set $\mathfrak{C}$ of all circuits fulfils the "circuit axioms"

C1 $\emptyset \notin \mathfrak{C}$

C2 $\forall C_1, C_2 \in \mathfrak{C} : C_1 \neq C_2 \Rightarrow C_1 \not\subseteq C_2$

C3 $\forall C_1, C_2 \in \mathfrak{C} \; \forall x \in C_1 \cap C_2 \; \exists C \in \mathfrak{C} : C \subseteq (C_1 \cup C_2) \setminus \{x\}$.

$M(X)$ is (also) uniquely determined by $\mathfrak{C}$ fulfilling (C1) - (C3). The *rank* of $A \subseteq X$ is the size of a maximal independent subset of A (called a *basis* of A — all basis' have the same cardinality).

A *submodular* function on X is a map $f : \mathbb{P}(X) \to \mathbb{R}$ from subsets of X to real numbers fulfilling

$$\forall A, B \in \mathbb{P}(X) : f(A \cap B) + f(A \cup B) \leq f(A) + f(B).$$

Consider a bipartite graph $G = (X, R, Y)$ $(R \subseteq X \times Y)$. The function

$$A \subseteq X \mapsto |R(A)|$$

is submodular (on X), increasing, integer-valued, and fulfils $f(\emptyset) = 0$, and thus

$$\mathfrak{I}_G := \big\{ A \subseteq X \mid \forall A' \subseteq A : |A'| \leq |R(A')| \big\}$$

is the set of independent sets of a matroid on X, called the *transversal matroid* $T(G)$ (w.r.t. X). For a proof of this basic property of submodular functions see for example [12], Chapter 12.1 or [2], Chapter VI.1.

By Hall's theorem, the independent subsets of $T(G)$ are exactly the "partial transversals" $A \subseteq X$, that is, subsets A of X which can be covered by a matching in G, which in turn means, that there is an injective mapping $m : A \to Y$ with $\forall a \in A : (a, m(a)) \in R$.

14

For the rank in $T(G)$ we have the formula

$$\text{rank}(A) = |A| - \delta^*(A)$$

for $A \subseteq X$, where $\delta(A)^* := \max_{A' \subseteq A} \delta(A)$, and $\delta(A) := |A| - |R(A)|$ is the *deficiency* of A (see also [11], Chapter 1).[1] $\text{rank}(A)$ is the size of a largest matching in the subgraph $(A, R|(A \times Y), Y)$, and thus $\text{rank}(A)$ as well as $\delta^*(A)$ are computable in polynomial time. Note that $A \subseteq X$ is independent iff $\text{rank}(A) = |A|$ iff $\delta^*(A) = 0$.

For a clause-set $F \in \mathcal{CLS}$ the bipartite graph $B(F)$ associated with F is $(F, R, \text{var}(F))$ with $(C, v) \in R \Leftrightarrow v \in \text{var}(C)$, that is, the "left side" of the bipartite graph are the clauses of F, the "right side" are the variables of F, and an edge joins clause C and variable v if v is contained in C (positively or negatively).

Computing the set of all circuits of a matroid

The set of all circuits of a matroid can be computed in polynomial time as follows (pointed out to me by Günter M. Ziegler).

Given a basis B of X and an element $x \in X \setminus B$, there is a unique circuit $C_x(B) \subseteq B \cup \{x\}$ (see [12], page 18, 1.2.6), and since every independent subset can be extended to a basis, it is easy to see that for every circuit C there is a basis B and an element $x \in X \setminus B$ with $C = C_x(B)$.

Consider a matroid $M(X)$ such that the property "$A \subseteq X$ is independent" is decidable in polynomial time (w.r.t. $|X|$). (Now for a given dependent subset A a circuit $C \subseteq A$ can be found in poly-time, and for $B \subseteq X$ we can decide in poly-time whether B is a basis of $M(X)$ or not.)

Assume furthermore that the corank of $M(X)$ is bounded by a constant k, that is $|X| - \text{rank}(X) \leq k$. Then by running through all subsets of X of size $\text{rank}(X)$ we can enumerate all bases' B in polynomial time, and then considering all $x \in X \setminus B$ yields all circuits $C_x(B)$ of $M(X)$ (in poly-time).

---

[1] For general matroids $M(X)$ the quantity $|A| - \text{rank}(A)$ for $A \subseteq X$ has been called the "nullity" of A in the fundamental paper [15], which generalises the notion of deficiency.

B Reduction by "Davis-Putnam resolution"

For clauses C, D with exactly one "clashing literal" x, that is $C \cap \overline{D} = \{x\}$, let $C \diamond D := (C \setminus \{x\}) \cup (D \setminus \{\overline{x}\})$ denote their *resolvent*.

For $a, b \in \mathbb{N} \cup \{\infty\}$ and clause-sets $F, F' \in \mathcal{CLS}$ the relation $F \xrightarrow{a,b} F'$ holds if there is a variable v with

$$\#_v(F) \leq a, \#_{\overline{v}}(F) \leq b \text{ or } \#_v(F) \leq b, \#_{\overline{v}}(F) \leq a$$

such that F' is obtained from F by *Davis-Putnam resolution* on v without any *"blockings" or contractions*, that is for

$$P := \{(C, D) \in F^2 : v \in C \wedge \overline{v} \in D\}$$

the following conditions hold:

- $P \neq \emptyset$ (neither v or $\overline{v}$ is pure)

- $(C, D) \in P \implies C \cap \overline{D} = \{v\}$ and $C \diamond D \notin F$ (all pairs $(C, D) \in P$ are resolvable, and none of their resolvents is already contained in F)

- $(C, D), (C', D') \in P$ and $(C, D) \neq (C', D') \implies C \diamond D \neq C' \diamond D'$ (any two different pairs from P yield different resolvents)

and we have

$$F' = \{C \in F : v \notin \mathrm{var}(C)\} \cup \{C \diamond D : (C, D) \in P\}.$$

It is well known that $\mathrm{DP}_v(F)$ and F are satisfiability equivalent. The following invariance properties are basic.

Lemma B.1 *For clause-sets F, F' with $F \xrightarrow{1,\infty} F'$ we have*

1. *$\delta(F) = \delta(F')$ (and $n(F') = n(F) - 1$)*

2. *$F \in \mathcal{MUSAT} \Leftrightarrow F' \in \mathcal{MUSAT}$*

3. *$F \in \mathcal{LEAN} \Leftrightarrow F' \in \mathcal{LEAN}$.*

16

Proof: Part 1 and 2 follow directly from the definition of $\xrightarrow{1,\infty}$. For part 3 consider v and P according to the above definition and assume w.l.o.g. that $P = \{C\} \times \{D_1, \dots, D_m\}$.

If φ is a (non-trivial) autarky for F (satisfying at least one clause) then φ is also a (non-trivial) autarky for F'.

If on the other side φ is a non-trivial autarky for F', then, assuming w.l.o.g. $v \notin \mathrm{var}(\varphi)$, we construct a non-trivial autarky φ' for F as follows:

- if $\varphi(C) = 1$, then $\varphi' := \varphi \cup \langle \overline{v} \to 1 \rangle$;

- if $\mathrm{var}(C) \cap \mathrm{var}(\varphi) = \emptyset$, then $\varphi' := \varphi$;

- otherwise $\varphi(D_i) = 1$ for all D_i, and we set $\varphi' := \varphi \cup \langle v \to 1 \rangle$. $\blacksquare$

Note that if for some variable v, occurring in one sign at most once in F, the additional conditions on P in the definition of $\xrightarrow{1,\infty}$ are violated, then F is not minimally unsatisfiable.

The proof of Lemma 3.10 uses the following compatibility properties (here the proofs are somewhat technical, and the reader may skip them).

Lemma B.2 *To unify notations, for clause-sets F, F' the relation $F \xrightarrow{\mathrm{del}} F'$ holds iff there is a clause $C \in F$ with $F' = F \setminus \{C\}$. (As usual, by "$\to^*$" we denote the reflexive-transitive closure of the relation "$\to$".)*

For all clause-sets F, F_1, F_2 and partial assignments φ we have

1. *$F \xrightarrow{\mathrm{del}}^* F_1 \implies \varphi * F \xrightarrow{\mathrm{del}}^* \varphi * F_1$.*

2. *If $\mathrm{var}(\varphi) \subseteq \mathrm{var}(F_1)$, then there is $F' \in \mathcal{CLS}$ with $F \xrightarrow{1,\infty}^* F_1 \implies \varphi * F \xrightarrow{\mathrm{del}}^* F' \xrightarrow{1,\infty}^* \varphi * F_1$.*

3. *There is $F' \in \mathcal{CLS}$ with $F \xrightarrow{1,\infty}^* F_1 \xrightarrow{\mathrm{del}} F_2 \implies F \xrightarrow{\mathrm{del}}^* F' \xrightarrow{1,\infty}^* F_2$.*

Proof: We only have to consider $\varphi = \langle x \to 1 \rangle$ for some literal x, and also only single applications of the reductions.

Part 1: For $F_1 = F \setminus \{C_0\}$ either $\langle x \to 1 \rangle * F = \langle x \to 1 \rangle * (F \setminus \{C_0\})$ holds or

$$(\langle x \to 1 \rangle * F) \setminus \{C_0 \setminus \{\overline{x}\}\} = \langle x \to 1 \rangle * (F \setminus \{C_0\}).$$

For parts 2 and 3 let

$$F_1 = \{C \in F : v \notin \text{var}(C)\} \cup \{C \diamond D : (C, D) \in P\}$$

according to the definition of $\xrightarrow{1,\infty}$ from the beginning of this section. We have to take care that also for the modified application of $\xrightarrow{1,\infty}$ the three conditions on P (applied accordingly) are fulfilled. W.l.o.g. we may assume $\#_v(F) = 1$. So there is exactly one $C_0 \in F$ with $v \in C_0$.

Part 2: Since $\text{var}(x) \in \text{var}(F_1)$, we have $\text{var}(x) \neq v$. Let

$$P' := \big\{ (C, D) \in (\langle x \to 1 \rangle * F)^2 : v \in C \wedge \overline{v} \in D \big\}.$$

If $P' = \emptyset$ then we have

$$(\langle x \to 1 \rangle * F) \setminus \{C \in \langle x \to 1 \rangle * F : v \in \text{var}(C)\} = \langle x \to 1 \rangle * F_1.$$

So assume $P' \neq \emptyset$. Let

$$
\begin{aligned}
G \quad := \quad & \big(\{C \diamond D : (C, D) \in P'\} \cap \langle x \to 1 \rangle * F\big) \\
& \cup \big\{ D \mid \exists D' : (C_0, D), (C_0, D') \in P' \wedge D \neq D' \wedge C_0 \diamond D = C_0 \diamond D' \\
& \qquad \wedge (C_0, D \cup \{\overline{x}\}) \in P \big\}.
\end{aligned}
$$

Using Davis-Putnam resolution on v we get

$$(\langle x \to 1 \rangle * F) \setminus G \xrightarrow{1,\infty} \langle x \to 1 \rangle * F_1.$$

Part 3: Let $F_2 = F_1 \setminus \{E\}$. If $E \in F$ then $F \setminus \{E\} \xrightarrow{1,\infty} F_2$. Otherwise let $E = C_0 \diamond D$ for $(C_0, D) \in P$. In case of $\#_{\overline{v}}(F) = 1$ we have $F \setminus \{C_0, D\} = F_2$. So assume $\#_{\overline{v}}(F) \geq 2$. Now $F \setminus \{D\} \xrightarrow{1,\infty} F_2$ holds. $\blacksquare$

C   Minimally unsatisfiable clause-sets with $\delta = 1$

In this subsection we give a simplified prove of the result from [4], that the class $\mathcal{MUSAT}(1)$ is poly-time decidable. For further characterisations of $\mathcal{MUSAT}(1)$ as well as of $\mathcal{SMUSAT}(1)$ and related classes see the forthcoming paper [10] (including short proofs of all known characterisations).

The following lemma motivated the use of $\mathcal{SMUSAT}$ in [5] (called "saturated minimally unsatisfiable clause-sets" there).

Lemma C.1 *[5] For any clause-set $F \in \mathcal{SMUSAT}$ and any literal x we have $\langle x \to 1 \rangle * F \in \mathcal{MUSAT}$.*

Proof: Consider $F' \subseteq \langle x \to 1 \rangle * F$, $F' \in \mathcal{MUSAT}$ and assume there is $C \in \langle x \to 1 \rangle * F$, $C \notin F'$. Since F is minimally unsatisfiable we must have $C \in F$ (otherwise for $C_0 \in F$ with $C_0 \setminus \{\overline{x}\} = C$ both $\langle x \to 1 \rangle * (F \setminus \{C_0\})$ and $\langle x \to 0 \rangle * (F \setminus \{C_0\})$ would be unsatisfiable). But now replacing C in F by $C \cup \{x\}$ maintains unsatisfiability (since application of $\langle x \to 1 \rangle$ and $\langle x \to 0 \rangle$ still yield unsatisfiable clause-sets), contradicting the assumption that F is saturated. ∎

As an easy application, in [5] an alternative proof (by induction on $n(F)$) is given for the fact $\delta(F) \geq 1$ if F is minimally unsatisfiable:

If $n(F) = 0$ then $F = \{\bot\}$ and thus $\delta(F) = 1$. Otherwise saturate F and obtain $F' \in \mathcal{SMUSAT}$. Choose $v \in \mathrm{var}(F)$ with $\#_v(F') + \#_{\overline{v}}(F')$ minimal. By induction hypothesis and the previous lemma we have $\delta(\langle x \to 1 \rangle * F') \geq 1$ while $\delta(\langle x \to 1 \rangle * F') \leq (c(F) - 1) - (n(F) - 1) = \delta(F') = \delta(F)$. ✓

For the special case $k = 1$ the next lemma has been shown in [4] (using a quite complicated proof).

Lemma C.2 *For any clause-set $F \in \mathcal{MUSAT}(k)$ with $F \neq \{\bot\}$ there is a variable $v \in \mathrm{var}(F)$ with $\#_v(F), \#_{\overline{v}}(F) \leq k$.*

Proof: Assume the contrary and saturate F, obtaining $F' \in \mathcal{SMUSAT}$. Consider $v \in \mathrm{var}(F)$ with $\#_v(F') + \#_{\overline{v}}(F')$ minimal. W.l.o.g. we assume $\#_v(F') \geq k+1$. Now $\langle v \to 1 \rangle * F' \in \mathcal{MUSAT}$ and thus $\delta(\langle x \to 1 \rangle * F') \geq 1$, but $c(\langle x \to 1 \rangle * F') \leq c(F) - (k + 1)$ and $n(\langle x \to 1 \rangle * F') = n(F) - 1$. ∎

The following characterisation of $\mathcal{MUSAT}(1)$ (first proven in [4]) is an immediate consequence of Lemma C.2 and Lemma B.1.

Lemma C.3 *For $F \in \mathcal{CLS}$ the following conditions are equivalent:*

1. *$F \in \mathcal{MUSAT}(1)$*

2. *$F \xrightarrow{1,\infty}{}^* \{\bot\}$ (for some or any order of applications of $\xrightarrow{1,\infty}$)*

3. *$F \xrightarrow{1,1}{}^* \{\bot\}$ (for some or any order of applications of $\xrightarrow{1,1}$).*

Thus "$F \in \mathcal{MUSAT}(1)$?" is poly-time decidable. ∎

D The proof of the Splitting theorem

Proof of Lemma 3.9: For *part 1* let $V^* := \mathrm{var}(F) \setminus V$ and

$$F^* := (V^* * F) \setminus \{\bot\}, \ F' := \{C \in F : \mathrm{var}(C) \cap V \neq \emptyset\}.$$

By definition we have $c(F^*) \leq c(F')$ and $\mathrm{var}(F^*) = V$. Lemma 3.7 gives $\delta(F^*) \geq 1$. If there is a variable $v \in V$ with $\#_v(F^*) + \#_{\overline{v}}(F^*) \leq 2$ then in fact $c(F') \geq c(F^*) + 1$, and thus $c(F') - |V| \geq \delta(F^*) + 1 = 2$. Otherwise Corollary 3.8 and Lemma C.2 yield $c(F') - |V| \geq \delta(F^*) \geq 2$.

For *part 2* let

$$V^* := \mathrm{var}(F) \setminus \mathrm{var}(F'), \ F^* := \{C \in F : \mathrm{var}(C) \cap V^* \neq \emptyset\}.$$

Because of $F^* \cap F' = \emptyset$ we have $c(F') \leq c(F) - c(F^*)$, while part 1 gives $c(F^*) \geq |V^*| + 2 = n(F) - n(F') + 2$, which altogether yield $\delta(F') \leq \delta(F) - 2$.

For *part 3* consider $F' \subseteq \langle x \to 1 \rangle * F$. If $n(F') = n(F) - 1$, then $\delta(F') \leq (c(F) - 2) - (n(F) - 1) = \delta(F) - 1$.

Otherwise consider $F'_0 \subseteq F$ with $c(F'_0) = c(F')$ and $\langle x \to 1 \rangle * F'_0 = F'$. Now we have $F'_0 \subset F$, and thus by part 2: $\delta(F') \leq \delta(F'_0) + 1 \leq \delta(F) - 1$. ∎

Proof of Lemma 3.10: Consider F'' such that $F' \xrightarrow{1,\infty} F''$ and every variable occurs in F'' in both signs at least twice.

By Lemma 3.9, part 3, for any variable $v \in \mathrm{var}(F'')$ and both $\varepsilon \in \{0,1\}$ there is F''_ε with

$$\langle v \to \varepsilon \rangle * F'' \xrightarrow{\mathrm{del}}_* F''_\varepsilon \in \mathcal{MUSAT}(\delta(F') - 1)$$

(using the relation $\xrightarrow{\mathrm{del}}$ from Lemma B.2). Altogether:

$$F \xrightarrow{\mathrm{del}}_* F' \xrightarrow{1,\infty}_* F'' \to \langle v \to \varepsilon \rangle * F'' \xrightarrow{\mathrm{del}}_* F''_\varepsilon.$$

By Lemma B.2, part 2 there is G with

$$F \xrightarrow{\mathrm{del}}_* F' \to \langle v \to \varepsilon \rangle * F' \xrightarrow{\mathrm{del}}_* G \xrightarrow{1,\infty}_* \langle v \to \varepsilon \rangle * F'' \xrightarrow{\mathrm{del}}_* F''_\varepsilon,$$

and part 1 of this lemma now yields

$$\langle v \to \varepsilon \rangle * F' \xrightarrow{\mathrm{del}}_* G \xrightarrow{1,\infty}_* \langle v \to \varepsilon \rangle * F'' \xrightarrow{\mathrm{del}}_* F''_\varepsilon.$$

Finally by part 3 of that lemma there is F'_ε with

$$\langle v \to \varepsilon \rangle * F' \xrightarrow{\mathrm{del}}_* F'_\varepsilon \xrightarrow{1,\infty}_* F''_\varepsilon,$$

and Lemma B.1 gives $F'_\varepsilon \in \mathcal{MUSAT}(\delta(F') - 1)$. ∎