Proceedings of the International MultiConference of Engineers and Computer Scientists 2012 Vol I,
IMECS 2012, March 14 - 16, 2012, Hong Kong

Maximal Satisfiable CNF Formulas

Stefan Porschen

***Abstract*—For a fixed unsatisfiable formula C, we introduce the class of (C,r)-maximal satisfiable subformulas S and study its basic properties, for every positive integer $r \geq 1$. The parameter r means the smallest number such that the addition of arbitrary r further clauses from C to S always yields an unsatisfiable subformula of C. Regarding the base case $r = 1$ we give a full characterization of the collection of all $(C,1)$-maximal satisfiable subformulas of C. We also clarify the connection to the class of minimal unsatisfiable formulas. Finally, we provide some connections to matroids of clause sets.**

***Index Terms*—propositional satisfiability, hypergraph, minimal unsatisfiable formula, fibre-transversal, matroid**

I. Introduction

The classical propositional satisfiability problem (SAT) of conjunctive normal form (CNF) formulas is a central combinatorial problem, namely one of the first problems that have been proven to be NP-complete [2]. More precisely, it is the natural NP-complete problem and thus lies at the heart of computational complexity theory. Moreover SAT plays a fundamental role in the theory of designing exact algorithms, and it has a wide range of applications because many problems can be encoded as a SAT problem via reduction [6] due to the rich expressiveness of the CNF language. The applicational area is pushed by the fact that meanwhile several powerful solvers for SAT have been developed (cf. e.g. [7], [10] and references therein).

The focus of this paper is the structural investigation of certain CNF subformula classes regarding SAT. Concretely, we introduce the notion of C-maximal satisfiable formulas with respect to a given *unsatisfiable* formula C. A subformula S of C is called C-maximal satisfiable if the addition of any further clause of C to S yields an unsatisfiable formula. We show the connection to the well known class of minimal unsatisfiable formulas, and give a full characterization of the set of all C-maximal satisfiable clause sets, if an arbitrary unsatisfiable formula C is given. Moreover, we clarify how an unsatisfiable formula C can be constructed such that a given satisfiable formula S becomes a C-maximal satisfiable formula. We also prove that there always exists a maximal satisfiable subformula for each unsatisfiable clause set. Further, we consider the generalization to (C,r)-maximal satisfiable formulas $S \subseteq C$, for fixed positive integer $r \geq 1$. Here, by definition, one always obtains an unsatisfiable formula when adding at least r further clauses from C to S, and there is no number less than r satisfying that property. Some hints are given, how $(C,1)$-maximal and (C,r)-maximal satisfiable subformulas are connected. Finally, some relations to matroids are discussed.

Methodical, we shall make use of the so-called fibre-view on clause sets which is introduced in [9], and a central result established therein. A fibre of a formula simply is the collection of all its clauses that contain the same set of variables.

Mathematics Group, Department 4, HTW Berlin, D-10313 Berlin, Germany, e-mail: porschen@htw-berlin.de.

II. Notation and Preliminaries

Let CNF denote the set of formulas (free of duplicate clauses) in conjunctive normal form over propositional variables $x \in \{0,1\}$. Each formula $C \in \text{CNF}$ is considered as a set of its clauses $C = \{c_1, \ldots, c_{|C|}\}$ having in mind that it is a conjunction of these clauses. A *positive (negative)* literal is a (negated) variable. The *negation (or complement)* of a literal l is $\bar{l}$. A clause $c \in C$ is a disjunction of different literals, and is also represented as a set $c = \{l_1, \ldots, l_{|c|}\}$. For formula C, clause c, by $V(C), V(c)$ we denote the variables contained (neglecting negations), correspondingly. Furthermore, CNF_+ denotes the set of *monotone* formulas, i.e., every variable occurs unnegated, only.

The satisfiability problem (SAT) asks, whether input $C \in$ CNF has a *model*, which is a truth value assignment $t : V(C) \to \{0,1\}$ assigning at least one literal in each clause of C to 1. Let UNSAT denote the set of all unsatisfiable members of CNF, and let SAT denote the set of all satisfiable members of CNF. Clearly, we have $\emptyset \in \text{SAT}$, i.e., the empty clause set is satisfiable. Given a set V of propositional variables, it is convenient to identify an assignment t with a clause of length V in the following way: Let $x^0 := \bar{x}$, $x^1 := x$. Then we can identify $t : V \to \{0,1\}$ with the literal set $\{x^{t(x)} : x \in V\}$. Similarly, for $b \subset V$, we identify the *restriction* $t|_b$ of t to b with the clause $\{x^{t(x)} : x \in b\}$. The collection of all clauses of length $|V|$ and containing literals over all variables in V is denoted as W_V and, by the identification described previously, W_V can be regarded as the set of all total assignments $V \to \{0,1\}$.

For a clause c we denote by c^γ the clause in which all its literals are complemented. In case of an assignment $t \in W_V$, we have the correspondence of t^γ to the assignment $1 - t : V \to \{0,1\}$ complementing all truth values. Similarly, let $C^\gamma = \{c^\gamma : c \in C\}$ denote the complemented clause set version of C.

The *fibre-view* as introduced in [9] regards a clause set C as composed of *fibres* over a hypergraph: All clauses c of C *projecting* onto the same variable set $b = V(c)$, when negations are eliminated, form the *fibre* C_b *over* b, namely $C_b = \{c \in C : V(c) = b\}$. The collection of these *base elements* b determines the edge set of a hypergraph, the *base hypergraph* $\mathcal{H}(C) = (V(C), B(C))$ of C, i.e., $B(C) = \{V(c) : c \in C\} \in \text{CNF}_+$ which can be regarded as a monotone clause set. Hence, C is the disjoint union of all its fibres: $C = \bigcup_{b \in B(C)} C_b$. Conversely, we can start with a given arbitrary hypergraph $\mathcal{H} = (V, B)$ serving as a base hypergraph if its vertices $x \in V$ are regarded as Boolean variables such that for each $x \in V$ there is a (hyper)edge $b \in B$ containing x. Recalling that W_b collects

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

all possible clauses over b, the set of all clauses over $\mathcal{H}$ is $K_{\mathcal{H}} := \bigcup_{b\in B} W_b$, also called the *total clause set over* $\mathcal{H}$. W_b is the fibre of $K_{\mathcal{H}}$ over b that sometimes is refered to as the *complete* fibre over b. For example, given $V = \{x_1, x_2, x_3\}$, and $B = \{b_1 := x_1x_2, b_2 := x_1x_3\}$, we have $K_{\mathcal{H}} = W_{b_1} \cup W_{b_2}$ where $W_{b_1} = \{x_1x_2, \bar{x}_1x_2, x_1\bar{x}_2, \bar{x}_1\bar{x}_2\}$, $W_{b_2} = \{x_1x_3, \bar{x}_1x_3, x_1\bar{x}_3, \bar{x}_1\bar{x}_3\}$ are the complete fibres over b_1, b_2.

A *formula over* $\mathcal{H}$ (or a $\mathcal{H}$*-based)* formula is a subset $C \subseteq K_{\mathcal{H}}$ such that $C_b := C \cap W_b \neq \emptyset$, for each $b \in B$. Given a $\mathcal{H}$-based formula $C \subseteq K_{\mathcal{H}}$ with the additional property that $\bar{C}_b := W_b \setminus C_b \neq \emptyset$ holds, for each $b \in B$, then we can define its $\mathcal{H}$-based *complement formula* $\bar{C} := \bigcup_{b\in B} \bar{C}_b = K_{\mathcal{H}} \setminus C$ with fibres $\bar{C}_b$. By construction $\bar{C}$ has the same base hypergraph as C. For example, given $\mathcal{H} = (V, B)$ with $V = \{x_1, x_2, x_3\}$, and $B = \{x_1x_2, x_1x_3\}$, let $C = \{x_1x_2, x_1\bar{x}_2, x_1\bar{x}_3, \bar{x}_1\bar{x}_3\}$ then $K_{\mathcal{H}} = C \cup \bar{C}$ where $\bar{C} = \{\bar{x}_1x_2, \bar{x}_1\bar{x}_2, x_1x_3, \bar{x}_1x_3\}$.

Let $\mathcal{H} = (V, B)$ be a fixed base hypergraph with total clause set $K_{\mathcal{H}}$. Recall that $C \subseteq K_{\mathcal{H}}$ is called *minimal unsatisfiable* if C is unsatisfiable but $\forall c \in C : C \setminus \{c\}$ is satisfiable, see for instance [1].

A *fibre-transversal* of $K_{\mathcal{H}}$ (not to be confused with a hitting set [5]) is a $\mathcal{H}$-based formula $F \subset K_{\mathcal{H}}$ such that $|F \cap W_b| = 1$, for each $b \in B$. Hence F is a formula containing exactly one clause of each fibre W_b of $K_{\mathcal{H}}$; let that clause be refered to as $F(b)$. An important type of fibre-transversals F are those containing each variable of V as a pure literal, that is, occurring in F with a single polarity only. Such fibre-transversals are called *compatible* and have the property that $\bigcup_{b\in B} F(b) \in W_V$. As a simple example for a compatible fibre-transversal, consider the base hypergraph with variable set $V := \{x_1, x_2, x_3\}$ and $B := \{b_1 := x_1x_2, b_2 := x_1x_3, b_3 := x_2x_3\}$. Then, e.g., the clauses $c_1 := \bar{x}_1x_2 \in W_{b_1}$, $c_2 := \bar{x}_1\bar{x}_3 \in W_{b_2}$ and $c_3 := x_2\bar{x}_3 \in W_{b_3}$, denoted as literal strings, form a compatible fibre-transversal of the corresponding $K_{\mathcal{H}}$, because $c_1 \cup c_2 \cup c_3 = \bar{x}_1x_2\bar{x}_3 \in W_V$.

We can define a (compatible) fibre-transversal of a $\mathcal{H}$-based formula $C \subset K_{\mathcal{H}}$ as a (compatible) fibre-transversal $K_{\mathcal{H}}$ that is contained in C.

The following result proved in [9] characterizes satisfiability of a formula C in terms of compatible fibre-transversals in its based complement formula $\bar{C}$.

Theorem 1: [9] For $\mathcal{H} = (V, B)$, let $C \subset K_{\mathcal{H}}$ be a $\mathcal{H}$-based formula such that $\bar{C}$ is $\mathcal{H}$-based, too. Then C is satisfiable if and only if $\bar{C}$ admits a compatible fibre-transversal F. Moreover, the union of all clauses in F^{γ} is a model of C.

III. The Concept of r-Maximal Satisfiability

In this section we introduce the notion of (C, r)*-maximal satisfiable clause sets* and provide some basic examples and facts. Let us begin with the formal definition.

Definition 1: Let $\mathcal{H} = (V, B)$ be a fixed base hypergraph. Given a fixed unsatisfiable formula $C \subseteq K_{\mathcal{H}}$ with $B(C) = B$, we call a (proper) subset $S \subseteq C$ a (C, r)-maximal satisfiable clause set *if S is satisfiable and r is the smallest positive integer such that for each r-subset $\hat{S} \subseteq C \setminus S$ the clause set $S \cup \hat{S}$ is unsatisfiable. S simply is called C*-maximal satisfiable *if $r = 1$.*

Addressing the natural question whether there are C-maximal satisfiable formulas at all, we can provide a direct answer on basis of minimal unsatisfiable formulas [1]. By definition such a formula becomes satisfiable if any of its clauses is removed. Therefore, if C is minimal unsatisfiable, and c is a clause of C then obviously $S = C \setminus \{c\}$ is C-maximal satisfiable. Here arise two questions. First, does the converse of the previous argument also hold? And closely related, are all maximal satisfiable formulas as trivial as those corresponding to minimal unsatisfiable ones? We give answers to both questions subsequently by the next example that can be obtained in view of Theorem 1 as follows. Consider the total clause set $K_{\mathcal{H}}$, for any base hypergraph $\mathcal{H} = (V, B)$, then clearly $K_{\mathcal{H}} \in$ UNSAT. Moreover, let $t \in W_V$ be chosen arbitrarily. Then $t_B := \{t|_b : b \in B\}$ is a compatible fibre-transversal of $K_{\mathcal{H}}$. Therefore $S = K_{\mathcal{H}} \setminus t_B$ is a satisfiable formula with model t^{γ} according to Theorem 1 as $\bar{S}$ exists. Furthermore, adding any clause of t_B to S provides a complete fibre in the resulting formula which cannot be satisfiable, therefore S is $K_{\mathcal{H}}$-maximal satisfiable.

Lemma 1: For any $t \in W_V$, t^{γ} is the unique model of $K_{\mathcal{H}} \setminus t_B$.

Proof. By construction we have $\overline{K_{\mathcal{H}} \setminus t_B} = t_B$ which is the only compatible fibre-transversal therein, hence t^{γ} is the unique model of $K_{\mathcal{H}} \setminus t_B$. □

In the next section we shall provide a precise characterization of the class of all maximal satisfiable clause sets of a given unsatisfiable formula C.

Let us briefly return to the first question stated above, namely whether every C-maximal satisfiable formula S has the property that $S \cup \{c\}$ is minimal unsatisfiable, for arbitrary $c \in C \setminus S$. A counterexample already is provided by the last example: Since the clause set $(K_{\mathcal{H}} \setminus t_B) \cup \{c\}$ contains the complete fibre $W_{V(c)}$, it is unsatisfiable. However, removing a clause of any other fibre of the formula leaves the resulting clause set unsatisfiable.

Addressing the case $r > 1$, let us first investigate whether there exists a pair S, C, and a positive integer $r \geq 2$ such that S is (C, r)-maximal satisfiable. Assume that the edge set B of $\mathcal{H}$ contains edges b with $|b| > \log r$ and again consider $K_{\mathcal{H}}$ and $t \in W_V$ as above. Let S be the formula obtained from $K_{\mathcal{H}} \setminus t_B$ by removing $r - 1$ clauses from any of its fibres whose base point b satisfies $|b| > \log r$. Then S is an example for a $(K_{\mathcal{H}}, r)$-maximal satisfiable formula. Indeed, we have to verify that each $T \subset K_{\mathcal{H}} \setminus S$ with $|T| = r$ yields an unsatisfiable formula $S \cup T$. This is clear because there is only one such set namely $T = K_{\mathcal{H}} \setminus S$. To see that r is the smallest integer with that property, let $T \subset (K_{\mathcal{H}} \setminus S)_b$ with $|T| = r - 1$ and such that $t|_b \notin T$, then $S \cup T$ is satisfiable with model t^{γ} according to Theorem 1.

The next result which can easily be established takes the converse perspective.

Proposition 1: Assume $C \in$ UNSAT. Let S be a (C, r)-maximal satisfiable formula, where $r \geq 2$. Then there exists a set $T \subseteq C \setminus S$ of $r-1$ elements such that $S \cup T$ is C-maximal satisfiable.

Proof. Let S be (C, r)-maximal satisfiable. Since r is the smallest integer for which this property holds, there must exist a subformula T of $C \setminus S$ consisting of $r-1$ clauses such that $S \cup T$ remains satisfiable. However adding an arbitrary further clause from $C \setminus (S \cup T)$ to $S \cup T$ yields an unsatisfiable

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

formula. Thus $S \cup T$ is C-maximal satisfiable as asserted. □

It is far less easy to clarify whether the converse of the previous result also holds.

IV. Characterization of the Case $r = 1$

Next, let us concentrate on the base case $r = 1$ addressing the following topics:

(1) Given $C \in$ UNSAT; does there always exist a C-maximal satisfiable subformula?
(2) How can C-maximal satisfiable clause sets be fully characterized?
(3) What is the structural relationship between maximal satisfiability and minimal unsatisfiability?

Regarding question (1), we already saw that it can be answered positively whenever C is minimal unsatisfiable, or is the total clause set. Below we shall provide a positive answer in the general case also exploring that the minimal unsatisfiable clause sets in a certain sense are an extreme subclass. Before, let us address the second topic. To that end, we provide a useful operation as follows. Given $C \in$ CNF and $t \in W_{V(C)}$, we set

$$C \ominus t := C \setminus t_{B(C)}$$

where $t_{B(C)} := \{t|_b : b \in B(C)\}$. So, from C we obtain $C \ominus t$ by substracting the clause $t|_b$ from every fibre C_b of C, where b runs through all edges of the base hypergraph $B(C)$ of C. Let $C \in$ UNSAT, and $S = C \ominus t$, for any $t \in W_{V(C)}$. Observe that in case of $B(C) = B(S)$ we have, in fact, that t^γ is a model of S by Theorem 1. Observe that, since $C \in$ UNSAT we always have $C \cap t_{B(C)} \neq \emptyset$, for every $t \in W_{V(C)}$. Otherwise there is a t with $t_{B(C)} \subseteq \bar{C}$ meaning that C is satisfiable according to Theorem 1. We state an useful Lemma the proof of which is obvious.

Lemma 2: For a formula C and $t, t' \in W_{V(C)}$, it holds that $C \ominus t = C \ominus t'$ if and only if

$$(*) \quad t_{B(C)} \cap C = t'_{B(C)} \cap C$$

Moreover, it holds that $C \cap t_{B(C)} \neq \emptyset$, for every $t \in W_{V(C)}$, iff $C \in$ UNSAT.

For $C \in$ CNF and $V := V(C)$, defining

$$t_1 \sim t_2 \quad \text{iff} \quad C \ominus t_1 = C \ominus t_2$$

obviously yields an equivalence relation on W_V, with pairwise disjoint classes $[\tilde{t}_j] \subseteq W_V$, for arbitrary fixed representatives $\tilde{t}_j$, where the index j runs through an appropriate finite index set J satisfying $W_V = \bigcup_{j \in J} [\tilde{t}_j]$. We clearly have that t^γ is a model of $S_j := C \ominus \tilde{t}_j$, for every $t \in [\tilde{t}_j]$. But there might exist further models of S_j not contained therein. The next result provides a precise criterion for the existence of a C-maximal satisfiable subformula S in case of equal base hypergraphs.

Theorem 2: Let $C \in$ UNSAT be fixed and let $S \subseteq C$ be a satisfiable subformula with $B(S) = B(C)$. Then S is a C-maximal satisfiable formula if and only if each model t of S satisfies $S = C \ominus t^\gamma$, i.e., $[t]^\gamma$ equals the set of models of S.

Proof. Fix $C \in$ UNSAT, and let $S \subseteq C$ be satisfiable. For the if-direction we assume that each model t of S satisfies $S = C \ominus t^\gamma$. Observe that $B(S) = B(C) =: B$ implies $V(S) = V(C) =: V$. Let $t \in W_V$ be a fixed model of S, and t^γ be its corresponding compatible fibre-transversal of $\bar{S}$ according to Theorem 1. For an arbitrary clause $c \in C \setminus S$, there is a $b \in B$ such that $c = t^\gamma|_b$. Otherwise we have $c \in S$, because $B(S) = B(C)$. For verifying $S \cup \{c\} \in$ UNSAT, we first note that $B(S \cup \{c\}) = B(C) = B$, thus we have to distinguish two cases.

Case (1): $B(\overline{S \cup \{c\}}) \subset B$. Then both the fibres C_b containing c and also $[S \cup \{c\}]_b$ are complete, i.e., they both equal W_b yielding an unsatisfiable fibre of $S \cup \{c\}$.

Case (2): $B(\overline{S \cup \{c\}}) = B$. Theorem 1 directly implies that t^γ cannot be a compatible fibre-transversal in $\overline{S \cup \{c\}}$ meaning that t is no model of $S \cup \{c\}$. Moreover, assume that there is another assignment $\tilde{t} \in V$ which is a model of $S \cup \{c\}$. Then $\tilde{t}$ specifically is a model of S, too; hence it satisfies $S = C \ominus \tilde{t}^\gamma$ where $\tilde{t}^\gamma$ is the corresponding compatible fibre-transversal of $\overline{S \cup \{c\}}$. Obviously, we therefore have $\tilde{t}^\gamma|_b \neq c = t^\gamma|_b$. Because of $c \in C$ we obtain a contradiction to the fact that $t_B^\gamma \cap C = \tilde{t}_B^\gamma \cap C \neq \emptyset$ according to $(*)$ in Lemma 2. Thus $S \cup \{c\}$ cannot have a model finishing the argumentation.

For the converse direction, suppose that S is a C-maximal satisfiable subformula of C hence $\bar{S}$ exists. Let $t \in W_V$ be an arbitrary model of S, and let t^γ be the corresponding compatible fibre-transversal of $\bar{S}$ according to Theorem 1. We claim that $S = C \ominus t^\gamma$. Indeed, otherwise there is a base point $b \in B$ such that $c \in C_b \setminus S_b$ and $c \neq t|_b$. This however implies that $S \cup \{c\}$ also is satisfiable by t contradicting that S is C-maximal satisfiable. On the other hand, there can be no clause from S removed when calculating $C \ominus t^\gamma$, because this requires $b \in B$ such that $t^\gamma|_b \in S_b$ which is impossible as t^γ is a compatible fibre-transversal of $\bar{S}$, hence $t^\gamma|_b \in \bar{S}_b$ verifying the claim. □

However, it is by no means guaranteed that t is the unique model of $C \ominus t^\gamma$ as it was the case for $C = K_{\mathcal{H}}$. Actually, we obtain the following connection between maximal satisfiable formulas and those possessing a unique model.

Corollary 1: For $C \in$ UNSAT and a satisfiable subformula S with $B(S) = B(C) = B$, it holds that S is a C-maximal satisfiable formula with unique model $t \in W_{V(C)}$ if and only if $S = C \ominus t^\gamma$ and $t_B^\gamma \cap C$ can be extended to a compatible fibre-transversal of $\bar{S}$ only by $t_B^\gamma \setminus C$, which specifically holds if $t_B^\gamma \subseteq C$.

Proof. Assume that $S = C \ominus t^\gamma$ and that $t_B^\gamma \cap C$ can be extended to a compatible fibre-transversal of $\bar{S}$ only by $t_B^\gamma \setminus C$. Then clearly t is a model of S. Since $t_B^\gamma \cap C$ must be fixed and $B(S) = B(C)$, there can be no other $\tilde{t} \in W_{V(C)}$ such that $S = C \ominus \tilde{t}^\gamma$ according to Lemma 2. Thus t is the unique model of S and Theorem 2 tells us that S is a C-maximal satisfiable formula.

For the converse direction, let S be a C-maximal satisfiable formula with unique model t then Theorem 2 directly yields $S = C \ominus t^\gamma$. Suppose there is a further $t \neq \tilde{t} \in W_{V(C)}$ such that $(t_B^\gamma \cap C) \cup (\tilde{t}_B^\gamma \setminus C)$ is a compatible fibre-transversal of $\bar{S}$. Then, according to Theorem 1, this yields another model $[(t_{B(C)}^\gamma \cap C) \cup (\tilde{t}_B^\gamma \setminus C)]^\gamma$ of S which cannot exist by assumption. □

Let us consider the question under which conditions a satisfiable formula S with a unique model admits an enlargement C such that S is C-maximal satisfiable. The next result states the answer including the case $B(S) \subset B(C)$; by 2^M we denote the power set of a (finite) set M. Moreover,

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

recall that $t_{B(S)} := \{t|_b : b \in B(S)\}$.

Theorem 3: For $S \in$ SAT with the unique model $t \in W_{V(S)}$, let $T = A \cup D$ be a non-empty clause set such that

$$A \subseteq t^\gamma_{B(S)} \quad \text{and} \quad D \subseteq \{t^\gamma|_b : b \in 2^{V(S)} \setminus B(S)\}$$

Then S is $(S \cup T)$-maximal satisfiable, and moreover, there exists no other superset C of S such that S is C-maximal satisfiable.

PROOF. First, we have to verify that $C = S \cup T \in$ UNSAT whenever $T = A \cup D \neq \emptyset$. First suppose that $c \in A \neq \emptyset$. Then over $B(S)$ there can be no compatible fibre-transversal since t, by assumption, is unique. Therefore no partial fibre-transversal over $2^{V(S)}$ can be enlarged to a complete compatible fibre-transversal in $\bar{C}$. So, we conclude that $C \in$ UNSAT. Next assume $D \neq \emptyset$. Observe that for each $c \in D$ there is $b \in 2^{V(S)} \setminus B(S)$ such that $c = t^\gamma|_{V(c)}$. Thus from $c \in D$ it follows that there can be no enlargement of the unique partial transversal t^γ over W_b yielding $C \in$ UNSAT also in this case. Obviously, the preceding argumentation also shows that $S \cup \{c\} \in$ UNSAT for every $c \in T$.

It remains to prove the last statement of the assertion. To that end, observe that any other superformula C' of S either properly satisfies $V(C') = V(S)$ and $b \in 2^{V(S)} \setminus B(S)$ contains one of the superformulas C as described above. Or it contains a clause c with at least one new variable $x \notin V(S)$. In the latter case, we consider $S \cup \{c\}$ where $c = c_n \cup c_S$ such that $V(c_n) \cap V(S) = \emptyset$ and $V(c_S) \subseteq V(S)$. Then t^γ can be extended to a compatible fibre-transversal of $\overline{S \cup \{c\}}$ because there always is the clause $d = t^\gamma|_{V(c_S)} \cup c_n^\gamma$ in $W_{V(c)}$, and $t^\gamma \cup d$ is compatible. In the first case, we can add a clause $c \in C' \setminus (t^\gamma_{B(S)} \cup \{t^\gamma|_b : b \in 2^V \setminus B(S)\})$ to S yielding a satisfiable formula, as we can provide a compatible fibre-transversal using the clauses in $t^\gamma_{B(S)} \cup \{t^\gamma|_b : b \in 2^V \setminus B(S)\} \subset \overline{S \cup \{c\}}$. So, S cannot be C'-maximal satisfiable. □

To finish the structural characterization of maximal satisfiable formulas, we have to generalize the last result to formulas S with more than one model. This requires that each enlargement of S disturbs all models of S simultaneously. Again let us include the case $B(S) \subset B(C)$ and profit by the fibre-view on clause sets. The proof proceeds similar to the previous one and is omitted therefore.

Theorem 4: For $S \in$ SAT with the set of models $\phi(S) \subseteq W_{V(S)}$, let $T = A \cup D$ be a non-empty clause set with $A = \emptyset$ or

$$A \subseteq \bigcup_{t \in \phi(S)} t^\gamma_{B(S)} \quad \text{s.t.} \quad \forall t \in \phi(S):\ A \cap t^\gamma_{B(S)} \neq \emptyset$$

and $D = \emptyset$ or

$$D \subseteq \bigcup_{t \in \phi(S)} t^\gamma_{B'(S)} \quad \text{s.t.} \quad \forall t \in \phi(S):\ D \cap t^\gamma_{B'(S)} \neq \emptyset$$

where $B'(S) := 2^{V(S)} \setminus B(S)$. Then S is $(S \cup T)$-maximal satisfiable, and moreover, there exists no other superset C of S such that S is C-maximal satisfiable.

Finally, let us reformulate without proof the last result from the perspective of a given $C \in$ UNSAT, i.e., generalizing Theorem 2 for the case $B(S) \subseteq B(C)$.

Theorem 5: Let $C \in$ UNSAT be fixed, let $S \subseteq C$ be a satisfiable subformula and let $\phi(S) \subseteq W_{V(S)}$ be the set of its models. Then S is a C-maximal satisfiable formula if and only if $V(S) = V(C) =: V$ and
(1) $\forall t \in \phi(S):\ S = C_{B(S)} \ominus t^\gamma$,
(2) $\forall t \in \phi(S), \forall c \in C_{B'(S)} : c \in t^\gamma_{B'(C)}$,
where $B'(S) := B(C) \setminus B(S)$ and $C_B := \bigcup_{b \in B} C_b$, for every $B \subseteq B(C)$.

Observe that $t^\gamma_{B'(C)}$ in condition (2) is well defined because $V(S) = V(C)$ must hold. Moreover, Theorem 5 becomes Theorem 2 in case that $B'(S) = \emptyset$ as then condition (2) becomes a tautology.

In view of the preceding discussion we realize that minimal unsatisfiable formulas, that we collect in $\mathcal{I} \subset$ UNSAT, are the most easy candidates to provide maximal satisfiable subformulas. First notice that for $C \in \mathcal{I}$ the complement formula $\bar{C}$ always is defined as long as $|B| > 1$; the case $|B| = 1$ is trivial and therefore omitted in what follows. Moreover, let $C \in \mathcal{I}$ with $\mathcal{H}(C) = (V, B)$ then we have $|C \cap t_B| = 1$, for every $t \in W_V$. Conversely, for fixed $t \in W_V$ there is exactly one $b \in B$ such that $t|_b \notin \bar{C}_b$. Meaning that C and $S := C \ominus t$ must differ in exactly one clause $c = t|_b$, which also holds for all $\tilde{t} \in \phi(S)$. In that sense $\mathcal{I}$ admit maximal satisfiable subformulas with minimal complement. In a natural generalization, we define the classes $\mathcal{I}_s$ within UNSAT of *s-minimal unsatisfiable formulas*, for every positive integer s: $C \in \mathcal{I}_s$ iff for each $T \subset C$ with $|T| = s - 1$, we have $C \setminus T \in \mathcal{I}$. Specifically it is $\mathcal{I}_1 = \mathcal{I}$. The next assertion is easy to verify and states the relationship to maximal satisfiability.

Theorem 6: Let $s > 0$ be any integer, then for every $C \in \mathcal{I}_s$ and $T \subset C$ with $|T| = s$ we have that $S = C \setminus T$ is C-maximal satisfiable.

PROOF. Case $s = 1$ is clear. Fix $s \geq 2$, $C \in \mathcal{I}_s$ and let $T \subset C$ with $|T| = s$ be chosen arbitrarily. By definition of $\mathcal{I}_s$, we have $S = C \setminus T \in$ SAT. Suppose there is $c \in T$ such that $S \cup \{c\}$ is satisfiable, it follows that $C \setminus (T \setminus \{c\}) \in \mathcal{I}$ is satisfiable yielding a contradiction. □

The question, for which values of $s > 1$ the family $\mathcal{I}_s \neq \emptyset$ is worth to be addressed in the future. The other extreme is given by the class of formulas $\mathcal{N} \subset$ UNSAT the members of which admit no maximal satisfiable subformulas at all. Meaning that for each $C \in \mathcal{N}$ there is no $T \subset C$ such that $C \setminus T$ is a maximal satisfiable subformula. The answer to question (3) posed in the beginning of this section is open so far. If $\mathcal{N} \neq \emptyset$ it must be answered negative, but we finally have a positive answer.

Theorem 7: It holds that $\mathcal{N} = \emptyset$, i.e., for each $C \in$ UNSAT there is a C-maximal satisfiable subformula.

PROOF. Let $C \in$ UNSAT and set $V := V(C), B := B(C)$. Define

$$\mu(C) := \min\{|t_B \cap C| : t \in W_V\}$$

as the minimum intersection cardinality of assignments with C. Collect all assignments fulfilling that cardinality in the subset $W_V^{\mu(C)}$. Since $C \in$ UNSAT we always have $\mu(C) \geq 1$. Now, we claim that $S_t := C \ominus t$ is a C-maximal satisfiable subformula if $t \in W_V^{\mu(C)}$. Clearly, S_t is satisfiable because $t^\gamma \in \phi(S_t)$. Assume there is $c \in C \setminus S_t$ such that $S_t \cup \{c\}$ is satisfiable with model $\tilde{t}^\gamma$. Then $\tilde{t}^\gamma$ also is a model of S_t. Since $c \in t_B$ we have $c \notin \tilde{t}$ and therefore $\tilde{t}_B \cap C \neq t_B \cap C$. However, since $\tilde{t}^\gamma$ is a model of S_t which is no model of C it must have been created by the operation

$C \ominus t$. Thus $\tilde{t}_B \cap C \subseteq t_B \cap C$; in summary we obtain $|\tilde{t}_B \cap C| < |t_B \cap C| = \mu(C)$ yielding a contradiction. Hence S_t is maximal satisfiable. Since $C \in$ UNSAT, it always holds that $W_V^{\mu(C)} \neq \emptyset$ completing the proof. □

Note that $\mu(I) = 1$ whenever $I \in \mathcal{I}$. Moreover for the class $\mathcal{I}$ we also have the converse of the claim in the preceding proof. Indeed, assume for $I \in \mathcal{I}$ there is $t \in W_{V(I)}$ such that $|t_{B(I)} \cap I| \geq 2$ and $S_t = I \ominus t$ is I-maximal satisfiable. Let c, c' be two of the clauses in the intersection. Then clearly $S \cup \{c\} = I \setminus \{c'\}$ is satisfiable yielding a contradiction.

In that respect minimal unsatisfiable formulas are extreme candidates for providing maximal satisfiable subformulas. In view of Lemma 1 we see that the total clause set provides the extreme coming from the other side, because the minimum intersection cardinality then equals the number of edges of the base hypergraph. In general the converse of the claim in the last proof does not hold and we refer to Theorems 2 and 5 regarding the full characterization.

V. Some Observations Regarding Matroids

This section is devoted to reveal some connections of maximal satisfiable clause sets to matroids. Matroids are a well-known concept in combinatorics and algebra [8]. From an algorithmic point of view, matroids are closely related to greedy algorithms, a prominent example is the matroid of forests in the context of minimum spanning trees in edge-weighted simple graphs [3]. Here, we consider some clause sets with matroid structure. For given $\mathcal{H} = (V, B)$ and $C \subseteq K_{\mathcal{H}}$, denote by $\mathcal{S}(C)$ the collection of all C-maximal satisfiable clause sets.

Definition 2: A matroid $\mathcal{M} = (C, \mathcal{T}), \mathcal{T} \subseteq 2^C$ is called a matroid of satisfiable clause sets (mscs) *if $\mathcal{T} \subset$ SAT. The set of bases of $\mathcal{M}$ is denoted as $\mathcal{U}(\mathcal{M})$ and let $\phi(\mathcal{M}) := \bigcap_{U \in \mathcal{U}(\mathcal{M})} \phi(U)$. An mscs $\mathcal{M}$ is called* satisfiable *if $\phi(\mathcal{M}) \neq \emptyset$.*

Not any matroid of satisfiable clause sets is satisfiable. For instance consider $\mathcal{M} = \{\emptyset, \{c_1\}, \{\bar{c}_1\}\}$, which trivially is a matroid of satisfiable clause sets but is not a satisfiable matroid.

A first obvious example providing the connection to maximal satisfiable clause sets is given by the class $\mathcal{I}$ as follows.

Proposition 2: For each $C \in \mathcal{I}$ we have that $(C, 2^C \setminus \{C\})$ is a mscs and $\mathcal{S}(C) = \{C \setminus \{c\} : c \in C\}$ is the set of its bases. The matroid is unsatisfiable.

Proof. We only verify the second assertion, namely that the mscs defined in the Proposition is unsatisfiable. This can be seen easily relying on Theorem 5. Indeed, let $c_1, c_2 \in C$ and suppose $t \in \phi(S_1) \cap \phi(S_2)$, where $S_i := C \setminus \{c_i\}$, $i = 1, 2$. Since S_1, S_2 both are C-maximal satisfiable, we have $V(S_1) = V(S_2) = V(C) =: V$, hence $t \in W_V$. It is $c_2 \in S_1$ therefore $c_2 \not\subset t^\gamma$. Either $B(S_2) = B(C)$, then $c_2 \subseteq t^\gamma$ because $c_2 \notin S_2$. Or it holds that $B(S_2) \subset B(C)$ then $c \in t^\gamma_{B'(S_2)}$ also implying $c_2 \subseteq t^\gamma$. So, we obtain a contradiction in either case. □

Relying on the total clause set over $\mathcal{H} = (V, B)$, we obtain the next obvious assertion.

Proposition 3: For each $S \in \mathcal{S}(K_{\mathcal{H}})$, $(K_{\mathcal{H}}, 2^S)$ is a satisfiable matroid.

However, we cannot enlarge the previous matroid by the power set of a further $K_{\mathcal{H}}$-maximal satisfiable subformula without disturbing its matroid structure.

Theorem 8: There is no matroid over $K_{\mathcal{H}}$ containing more than one $K_{\mathcal{H}}$-maximal satisfiable formula.

Proof. Using the results of the previous section it is not hard to see that every maximal satisfiable subformula of $K_{\mathcal{H}}$ is of the form $S_t := K_{\mathcal{H}} \ominus t$, for a $t \in W_V$. So, let $t, s \in W_V, t \neq s$ and assume that both S_t, S_s are members of the same matroid set system $\mathcal{T}$, then clearly the power sets of both also are in $\mathcal{T}$. It suffices to verify that then there are $T, U \in \mathcal{T}$ with $|U| = |T| + 1$ and there is no $c \in U \setminus T$ such that $T \cup \{c\} \in \mathcal{T}$. Since $t \neq s$ there is $b \in B$ such that $c_t := t|_b \neq s|_b =: c_s$. Moreover $c_s \in S_t$, and $c_t \in S_s$. We claim that $U := S_t$ and $T := S_S \setminus \{c_t\}$ is a pair having the desired property. Choose an arbitrary $c \in U \setminus T$. Since $c \in s_B$ we have $T \cup \{c\} \in$ UNSAT. Indeed, let $V(c) = b' \neq b$, then $T \cup \{c\}$ contains the complete fibre $W_{b'}$. If $V(c) = b$ then $c = c_s$ and $T \cup \{c_s\}$ has at b the fibre $W_b \setminus \{c_t\}$. Thus we also obtain an unsatisfiable formula because $(s \setminus c_s) \cup c_t$ is the only candidate for a compatible fibre-transversal in the based complement formula $\overline{T \cup \{c_s\}}$. However, that cannot work as, by assumption, we have $c_t \neq c_s$ meaning, e.g., $x \in c_s$ and $\bar{x} \in c_t$ implying $\{x, \bar{x}\} \subset (s \setminus c_s) \cup c_t$. It follows that there is no $t' \in W_V$ with $T \cup \{c\} \subset S_{t'}$, which is valid for every $c \in U \setminus T$. Thus we cannot enlarge $\mathcal{T}$ by further $K_{\mathcal{H}}$-maximal satisfiable formulas. □

The last assertion can easily be reformulated as follows.

Lemma 3: An mscs $\mathcal{M}$ over $K_{\mathcal{H}}$ is satisfiable if and only if there is a $K_{\mathcal{H}}$-maximal satisfiable clause set containing all bases of $\mathcal{M}$. □

So, we have to search for satisfiable matroids over $K_{\mathcal{H}}$ appearing as the intersection of those matroids for different $K_{\mathcal{H}}$-maximal satisfiable subformulas. Observe that for every $t^\gamma \in \phi(\mathcal{M})$ no base $U \in \mathcal{U}(\mathcal{M})$ is allowed to contain a clause being a subset of t. For a given $t^\gamma \in \phi(\mathcal{M})$ we have $S_t = K_{\mathcal{H}} \setminus t_B$. Hence we obtain $\bigcap_{t^\gamma \in \phi(\mathcal{M})} S_t = K_{\mathcal{H}} \setminus [\bigcup_{t^\gamma \in \phi(\mathcal{M})} t_B]$.

The previous Lemma directly yields the following statement, because if $\mathcal{M} = (K_{\mathcal{H}}, \mathcal{T})$ is a satisfiable matroid, then for each $t^\gamma \in \phi(\mathcal{M})$ it holds that $U \subset S_t, \forall U \in \mathcal{U}(\mathcal{M})$.

Lemma 4: Let $\mathcal{M} = (K_{\mathcal{H}}, \mathcal{T})$ be a satisfiable matroid with $\mathcal{U}(\mathcal{M})$ the set of its bases, then $U \subseteq K_{\mathcal{H}} \setminus [\bigcup_{t^\gamma \in \phi(\mathcal{M})} t_B]$, for each $U \in \mathcal{U}(\mathcal{M})$. □

VI. Concluding Remarks and Open Problems

We introduced the concept of (C, r)-maximal satisfiable clause sets and studied its basic properties. Specifically, we provided a full characterization of the set of all C-maximal satisfiable subformulas of a given unsatisfiable formula C. We also investigated the relationship to minimal unsatisfiable clause sets. Concretely, we proved that for the class $\mathcal{I}$ it holds that $I \ominus t$ is an I-maximal satisfiable subformula iff $t \in W_V^{\mu(I)}$. In that respect minimal unsatisfiable formulas are extreme candidates for providing maximal satisfiable subformulas. In general, only the if-direction of the assertion holds true implying that every unsatisfiable formula C admits a C-maximal satisfiable subformula.

There remain several open problems for future research. First of all to investigate the existence problems and the structural features of (C, r)-maximal satisfiable clause sets

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

more deeply. Observe that the argumentation providing a $(K_{\mathcal{H}}, r)$-maximal subformula, in general cannot be transfered when the total clause set $K_{\mathcal{H}}$ is replaced with an arbitrary unsatisfiable $C \subset K_{\mathcal{H}}$. From the structural point of view, the main topic that remained open is whether one can proof the converse of Prop. 1, namely the conjecture that S is a (C, r)-maximal satisfiable formula, for fixed $r \geq 2$, if and only if there exists a set $T \subseteq C \setminus S$ of $r - 1$ elements such that $S \cup T$ is C-maximal satisfiable. A deeper structural reasearch with respect to matroids might help here; in that direction we only made some first steps.

The criteria for (C, r)-maximal satisfiable subformulas, $r \geq 1$, presented so far are inadequate for an efficient algorithmic construction of maximal satisfiable subformulas of $C \in \text{UNSAT}$. Perhaps Theorem 7 might provide some efficient approach. However, notice that already in the case of C-maximal satisfiable subformulas with a unique model a recognization is algorithmically hard, even if one knows that only one model exists [11]. Closely related is the question whether one can identify fixed-parameter tractable classes with respect to the parameter r [4]. However for gaining a progress in that direction, a deeper understanding of r-maximal satisfiability is required.

Finally one should study in more detail how the concept of maximal satisfiability could be applied in the area of combinatorial optimization.

References

[1] Aharoni, R., Linial, N., Minimal non-two-colorable hypergraphs and minimal unsatisfiable formulas, Journal of Combinatorial Theory, Series A 43, (1986) 196-204.

[2] Cook, S.A., The Complexity of Theorem Proving Procedures. In: Proceedings of the 3rd ACM Symposium on Theory of Computing, pp. 151-158, 1971.

[3] Cook, W.J., Cunningham, W.H., Pulleyblank, W.R., Shrijver, A., Combinatorial Optimization, Wiley Interscience, New York, 1998.

[4] Downey, R.G., Fellows, M.R., Parameterized Complexity, Springer-Verlag, New York, 1999.

[5] Garey M.R., Johnson, D.S., Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman and Company, San Francisco, 1979.

[6] Gu, J., Purdom, P.W., Franco, J., Wah, B.W., Algorithms for the Satisfiability (SAT) Problem: A Survey, in: D. Du, J. Gu, P. M. Pardalos (Eds.), Satisfiability Problem: Theory and Applications, DIMACS Workshop, March 11-13, 1996, DIMACS Series, vol. 35, pp. 19-151, American Mathematical Society, Providence, Rhode Island, 1997.

[7] Le Berre, D., Simon, L., The Essentials of the SAT 2003 Competition, Proc. SAT 2003, LNCS, vol. 2919, pp. 172-187, Springer-Verlag, 2004.

[8] Oxley, J., Matroid Theory, Oxford University Press, Oxford, 2006.

[9] Porschen, S., A CNF Formula Hierarchy over the Hypercube, Proc. AI 2007, LNAI, vol. 4830, pp. 234-243, 2007.

[10] Speckenmeyer, E., Min Li, C., Manquinho V., Tacchella, A., (Eds.), Special Issue on the 2007 Competitions, Journal on Satisfiability, Boolean Modeling and Computation, 4 (2008).

[11] Valiant, L.G., Vazirani, V.V., NP is as easy as detecting unique solutions, Theoret Comput. Sci. 47 (1986) 85-93.