Minimal Unsatisfiability and Minimal Strongly Connected Digraphs

Hoda Abbasizanjani and Oliver Kullmann

Swansea University, Swansea, UK
{Hoda.Abbasizanjani,O.Kullmann}@Swansea.ac.uk

Abstract. Minimally unsatisfiable clause-sets (MUs) are the hardest unsatisfiable clause-sets. There are two important but isolated characterisations for nonsingular MUs (every literal occurs at least twice), both with ingenious but complicated proofs: Characterising 2-CNF MUs, and characterising MUs with deficiency 2 (two more clauses than variables). Via a novel connection to Minimal Strong Digraphs (MSDs), we give short and intuitive new proofs of these characterisations, revealing an underlying common structure.

1 Introduction

This paper is about understanding basic classes of minimally unsatisfiable CNFs, short MUs. The most basic MUs are those with only *one more clause* than variables, i.e., with deficiency $\delta = 1$. This whole class is explained by the expansion rule, which replaces a single clause C by two clauses $C' \cup \{v\}, C'' \cup \{\overline{v}\}$ for $C' \cup C'' = C$ and a new variable v, starting with the empty clause. So in a sense only trivial reasoning takes place here. Somewhat surprisingly, this covers all Horn cases in MU ([5]). At the next level, there are two classes, namely *two more clauses* than variables ($\delta = 2$), and *2-CNF*. Characterisations have been provided in the seminal paper [12] for the former class, and in the technical report [15] for the latter. Both proofs are a tour de force. We introduce in this paper a new unifying reasoning scheme, based on graph theory.

This reasoning scheme considers MUs with two parts. The clauses of the "core" represent AllEqual, that is, all variables are equal. The two "full monotone clauses", a disjunction over all positive literals and a disjunction over all negative literals, represent the negation of AllEqual. This is the new class FM ("full monotone") of MUs, which still, though diluted, is as complex as all of MU. So we demand that the reasoning for AllEqual is graph theoretical, arriving at the new class DFM ("D" for digraph).

Establishing AllEqual on the variables happens via SDs, "strong digraphs", where between any two vertices there is a path. For minimal reasoning we use MSDs, minimal SDs, where every arc is necessary. Indeed, just demanding to have an MU with two full monotone clauses, while the rest are binary clauses, is enough to establish precisely MSDs. The two most fundamental classes of MSDs are the *(directed) cycles* C_n and the *dipaths*, the directed versions $D(P_n)$ of the undirected paths P_n, where every undirected edge is replaced by two directed arcs, for both directions. The cycles are at the heart of MUs with $\delta = 2$, while the dipaths are at the heart of MUs in 2-CNF.

To connect MSDs (that is, DFMs) with more general MUs, two transformations are used. First, *singular variables*, occurring in one sign only once, are eliminated by singular DP-reduction, yielding *nonsingular* MUs. This main (polytime) reduction removes "trivialities", and indeed deficiency 1 consists purely of these trivialities (as the above generation process shows). Second we need to add "missing" literal occurrences, non-deterministically, to clauses, as long as one stays still in MU. This process is called *saturation*, yielding saturated MUs. As it turns out, the nonsingular MUs of deficiency 2 are already saturated and are already of the form DFM, while the nonsingular 2-CNFs have to be (partially) saturated to reach the form DFM.

Before continuing with the overview, we introduce a few basic notations. The class of MUs as clause-sets is formally denoted by $\mathcal{MU}$, while the nonsingular elements are denoted by $\mathcal{MU}' \subset \mathcal{MU}$ (every variable occurs at least twice positively and twice negatively). The number of clauses of a clause-set F is $c(F)$, the number of (occurring) variables is $n(F)$, and the deficiency is $\delta(F) := c(F) - n(F)$. The basic fact is $\delta(F) \geq 1$ for $F \in \mathcal{MU}$ ([1]), and that deficiency is a good complexity parameter ([6]). We use indices for subclassing in the obvious way, e.g., $\mathcal{MU}_{\delta=1} = \{F \in \mathcal{MU} : \delta(F) = 1\}$. Furthermore, like in the DIMACS file format for clause-sets, we use natural numbers in $\mathbb{N} = \{1, 2, \dots\}$ for variables, and the non-zero integers for literals. So the clause $\{-1, 2\}$ stands for the usual clause $\{\overline{v_1}, v_2\}$, where we just got rid off the superfluous variable-symbol "v". In propositional calculus, this would mean $\neg v_1 \vee v_2$, or, equivalently, $v_1 \to v_2$.

The two fundamental examples After this general overview, we now state the central two families of MUs for this paper, for deficiency 2 and 2-CNF. The MUs $\boldsymbol{\mathcal{F}_n} := \{\{1, \dots, n\}, \{-1, \dots, -n\}, \{-1, 2\}, \dots, \{-(n-1), n\}, \{-n, 1\}\}$ of deficiency 2 have been introduced in [12]. It is known, and we give a proof in Lemma 7, that the $\mathcal{F}_n$ are saturated. As shown in [12], the elements of $\mathcal{MU}'_{\delta=2}$ are exactly (up to isomorphism, of course) the formulas $\mathcal{F}_n$. The elimination of singular variables by singular DP-reduction is not confluent in general for MUs. However in [18] it is shown, that we have confluence up to isomorphism for deficiency 2. These two facts reveal that the elements of $\mathcal{MU}_{\delta=2}$ contain a unique "unadorned reason" for unsatisfiability, namely the presence of a complete cycle over some variables (of unique length) together with the requirement that these variables do not have the same value. In the report [15], as in [20] (called "$F^{(2)}$"), the 2-CNF MUs $\boldsymbol{\mathcal{B}_n} := \{\{-1, 2\}, \{1, -2\}, \dots, \{-(n-1), n\}, \{n-1, -n\}, \{-1, -n\}, \{1, n\}\}$ have been introduced, which are 2-uniform (all clauses have length 2). In [15] it is shown that the nonsingular MUs in 2-CNF are exactly the $\mathcal{B}_n$. By [18] it follows again that we have confluence modulo isomorphism of singular DP-reduction on 2-CNF-MUs. Thus a 2-CNF-MU contains, up to renaming, a complete path of equivalences of variables, where the length of the path is unique; this path establishes the equivalence of all these variables, and then there is the equivalence of the starting point and the negated end point, which yields the contradiction.

Background We have referred above to the fundamental result about singular DP-reduction (sDP) in [18], that for $F \in \mathcal{MU}$ and any $F', F'' \in \mathcal{MU}'$ obtained from F by sDP we have $n(F') = n(F'')$. So we can define $\mathbf{nst(\boldsymbol{F})} := n(F') \in \mathbb{N}_0$ (generalising [18, Definition 75]). We have $0 \leq \operatorname{nst}(F) \leq n(F)$, with $\operatorname{nst}(F) = 0$ iff $\delta(F) = 1$, and $\operatorname{nst}(F) = n(F)$ iff F is nonsingular. The "nonsingularity type" $\operatorname{nst}(F)$ provides basic information about the isomorphism type of MUs, after (completed) sDP-reduction, and suffices for deficiency 2 and 2-CNF.

We understand a class $\mathcal{C} \subseteq \mathcal{MU}$ "fully" if we have a full grasp on its elements, which should include a complete understanding of the isomorphism types involved, that is, an easily accessible catalogue of the essentially different elements of $\mathcal{C}$. The main conjecture is that the nonsingular cases of fixed deficiency have polytime isomorphism decision, and this should be extended to "all basic classes". Singular DP-reduction is essential here, since already Horn-MU, which has deficiency one, is GI-complete (graph-isomorphism complete; [14]).

Before giving an overview on the main proof ideas, we survey the relevant literature on 2-CNF. Irredundant 2-CNF is studied in [21], mostly concentrating on satisfiable cases, while we are considering only unsatisfiable cases. As mentioned, the technical report [15] contains the proof of the characterisation of 2-CNF-MU, while [20] has some bounds, and some technical details are in [2]. MUSs (MU-sub-clause-sets) of 2-CNF are considered in [3], showing how to compute shortest MUSs in polytime, while in [4] MUSs with shortest resolution proofs are determined in polytime. It seems that enumeration of all MUSs of a 2-CNF has not been studied in the literature. However, in the light of the strong connection to MSDs established in this paper, for the future [11] should become important, which enumerates all MSDs of an SD in incremental polynomial time.

Two full clauses The basic new class is $\boldsymbol{\mathcal{FM}} \subset \mathcal{MU}$, which consists of all $F \in \mathcal{MU}$ containing the full positive clause (all variables) and the full negative clause (all complemented variables). Using "monotone clauses" for positive and negative clauses, "$\mathcal{FM}$" reminds of "full monotone". Let A_n be the basic MUs with n variables and 2^n full clauses; so we have $A_n \in \mathcal{FM}$ for all $n \geq 0$. The trivial cases of $\mathcal{FM}$ are A_0 and A_1, while a basic insight is that $\boldsymbol{\mathcal{FM}'} := \mathcal{FM} \cap \mathcal{MU}'$ besides $\{\perp\}$ contains precisely all the nontrivial elements of $\mathcal{FM}$. In this sense it can be said that $\mathcal{FM}$ studies only nonsingular MUs. We expect the class $\mathcal{FM}_{\delta=k}$ at least for $\delta = 3$ to be a stepping stone towards understanding $\mathcal{MU}_{\delta=3}$ (the current main frontier). The most important new class for this paper is $\boldsymbol{\mathcal{DFM}} \subset \mathcal{FM}$, which consists of all $F \in \mathcal{FM}$ such that besides the monotone clauses all other clauses are binary. Indeed graph isomorphisms for MSDs is still GI-complete ([23]), and thus so is isomorphism for $\mathcal{DFM}$.

After having now DFM at our disposal, we gain a deeper understanding how the seminal characterisations of the basic nonsingular F, that is, $F \in \mathcal{MU}'$, $n(F) > 0$, work: [I] From $\delta(F) = 2$ follows $F \cong \mathcal{F}_{n(F)}$ ([12]; see Corollary 1). [II] From $F \in 2\text{–}\mathcal{CLS}$ follows $F \cong \mathcal{B}_{n(F)}$ ([15]; see Corollary 2). The main step is to make the connection to the class $\mathcal{DFM}$: [I] In case of $\delta(F) = 2$, up to renaming it actually already holds that $F \in \mathcal{DFM}$. The main step here to show is the existence of the two full monotone clauses — that the rest then is in 2-CNF follows by the minimality of the deficiency. [II] In case of $F \in 2\text{–}\mathcal{CLS}$ there must exist exactly one positive and one negative clause and these can be saturated to full positive resp. full negative clauses, and so we obtain $F' \in \mathcal{DFM}$. Once the connection to $\mathcal{DFM}$ is established, graph-theoretical reasoning does the remaining job: [I] The MSDs of minimal deficiency 0 are the cycles. [II] The only MSDs G such that the corresponding DFMs can be obtained as partial saturations of *nonsingular* 2-CNF are the dipaths, since we can only have two *linear vertices* in G, vertices of in- and out-degree one.

An overview on the main results of this paper is given in Figure 1.

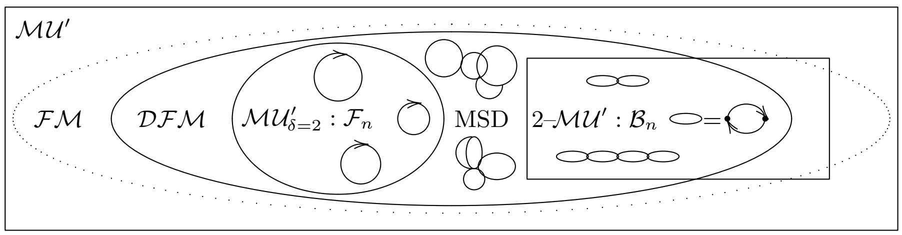

Fig. 1. Directed cycles at the heart of $\mathcal{MU}_{\delta=2}$, and dipaths at the heart of $2\text{–}\mathcal{MU}$.

2 Preliminaries

We use clause-sets F, finite sets of clauses, where a clause is a finite set of literals, and a literal is either a variable or a negated/complemented variable. The set of all variables is denoted by $\mathcal{VA}$ (we use variables also as vertices in graphs), and we assume $\mathbb{N} = \{1, 2, \dots\} \subseteq \mathcal{VA}$. This makes creating certain examples easier, since we can use integers different from zero as literals (as in the DIMACS format). The set of clause-sets is denoted by $\mathcal{CLS}$, the empty clause-set by $\top := \emptyset \in \mathcal{CLS}$ and the empty clause by $\bot := \emptyset$. Clause-sets are interpreted as CNFs, conjunctions of disjunction of literals. A clause-set F is uniform resp. k-uniform, if all clauses of F have the same length resp. length k. This paper is self-contained, if however more background is required, then the Handbook chapter [13] can be consulted.

Clauses C do not contain clashes (conflicts), i.e., they are "non-tautological", which formally is denoted by $C \cap \overline{C} = \emptyset$, where for a set L of literals by $\overline{L}$ we denoted elementwise complementation. With $\mathrm{var}(F)$ we denote the set of variables occurring in F, while by $\mathrm{lit}(F) := \mathrm{var}(F) \cup \overline{\mathrm{var}(F)}$ we denote the possible literals of F (one of the two polarities of a literal in $\mathrm{lit}(F)$ must occur in F). Since the union $\bigcup F$ is the set of occurring literals, we have $\mathrm{lit}(F) = (\bigcup F) \cup \overline{\bigcup F}$, while $\mathrm{var}(F) = \mathrm{lit}(F) \cap \mathcal{VA}$. A clause C is *positive* if $C \subset \mathcal{VA}$, while C is *negative* if $C \subset \overline{\mathcal{VA}}$, and C is *mixed* otherwise; a non-mixed clause is called *monotone*. A **full clause** of a clause-set F is some $C \in F$ with $\mathrm{var}(C) = \mathrm{var}(F)$. A *full clause-set* is an $F \in \mathcal{CLS}$ where all $C \in F$ are full. By $\boldsymbol{A_n}$ we denote the full clause-set consisting of the 2^n full clauses over variables $1, \dots, n$ for $n \in \mathbb{N}_0$. So $A_0 = \{\bot\}$, $A_1 = \{\{-1\}, \{1\}\}$, and $A_2 = \{\{-1, -2\}, \{1, 2\}, \{-1, 2\}, \{1, -2\}\}$. For $F \in \mathcal{CLS}$ we use $\boldsymbol{n(F)} := |\mathrm{var}(F)| \in \mathbb{N}_0$ for the number of (occurring) variables, $\boldsymbol{c(F)} := |F| \in \mathbb{N}_0$ for the number of clauses, and $\boldsymbol{\delta(F)} := c(F) - n(F) \in \mathbb{Z}$ for the **deficiency**. p–$\mathcal{CLS}$ is the set of $F \in \mathcal{CLS}$ such that for all $C \in F$ holds $|C| \le p$. The application of partial assignments φ to $F \in \mathcal{CLS}$, denoted by $\varphi * F$, yields the clause-set obtained from F by removing clauses satisfied by φ, and removing falsified literals from the remaining clauses. Contractions can occur, since we are dealing with clause-*sets*, i.e., previously unequal clauses may become equal, and so more clauses might disappear than expected. Also more variables than just those in φ might disappear, since we consider only occurring variables. $\mathcal{SAT}$ is the set of satisfiable clause-sets, those $F \in \mathcal{CLS}$ where there is a partial assignment φ with $\varphi * F = \top$. $\mathcal{CLS}$ is partitioned into $\mathcal{SAT}$ and the set of unsatisfiable clause-sets. A clause-set F is irredundant iff for every $C \in F$ there exists a total assignment φ which satisfies $F \setminus \{C\}$ (i.e., $\varphi * (F \setminus \{C\}) = \top$) while falsifying C (i.e., $\varphi * \{C\} = \{\bot\}$). Every full clause-set is irredundant.

Isomorphism of clause-sets $F, G \in \mathcal{CLS}$ is denoted by $\boldsymbol{F \cong G}$, that is, there exists a complement-preserving bijection from $\operatorname{lit}(F)$ to $\operatorname{lit}(G)$ which induces a bijection from the clauses of F to the clauses of G. For example for an unsatisfiable full clause-set F we have $F \cong A_{n(F)}$. $\boldsymbol{\mathcal{RHO}}$ is the set of renamable Horn clause-sets, i.e., $F \in \mathcal{CLS}$ with $F \cong G$ for some Horn clause-set G (where every clause contains at most one positive literal, i.e., $\forall\, C \in G : |C \cap \mathcal{VA}| \le 1$).

The **DP-operation** (sometimes also called "variable elimination") for $F \in \mathcal{CLS}$ and a variable v results in $\mathbf{DP}_{\boldsymbol{v}}\boldsymbol{(F)} \in \mathcal{CLS}$, which replaces all clauses in F containing variable v (positively or negatively) by their resolvents on v. Here for clauses C, D with $C \cap \overline{D} = \{x\}$ the resolvent of C, D on $\operatorname{var}(x)$ is $(C \setminus \{x\}) \cup (D \setminus \{\overline{x}\})$ (note that clauses can only be resolved if they contain *exactly* one clashing literal, since clauses are non-tautological).

We conclude by recalling some notions from graph theory: A graph/digraph G is a pair (V, E), with $V(G) := V$ a finite set of "vertices", while $E(G) := E$ is the set of "edges" resp. "arcs", which are two-element subsets $\{a, b\} \subseteq V$ resp. pairs $(a, b) \in V^2$ with $a \ne b$. An isomorphism between two (di)graphs is a bijection between the vertex sets, which induces a bijection on the edges/arcs. Isomorphism of clause-sets can be naturally reduced in polytime to graph isomorphism, and *GI-completeness* of such isomorphism problems means additionally that also the graph isomorphism problem can be reduced to it.

3 Review on Minimal Unsatisfiability (MU)

$\boldsymbol{\mathcal{MU}}$ is the set of unsatisfiable clause-sets such that every strict sub-clause-set is satisfiable. For $F \in \mathcal{CLS}$ holds $F \in \mathcal{MU}$ iff F is unsatisfiable and irredundant. We note here that "$\mathcal{MU}$" is the *class* of MUs, while "MU" is used in text in a substantival role. $\boldsymbol{\mathcal{MU}'} \subset \mathcal{MU}$ is the set of **nonsingular** MUs, that is, $F \in \mathcal{MU}$ such that every literal occurs at least twice. We use $\mathbf{2}$–$\boldsymbol{\mathcal{MU}} := \mathcal{MU} \cap 2$–$\mathcal{CLS}$ and $\mathbf{2}$–$\boldsymbol{\mathcal{MU}'} := \mathcal{MU}' \cap 2$–$\mathcal{CLS}$. **Saturated** MUs are those unsatisfiable $F \in \mathcal{CLS}$, such that for every $C \in F$ and every clause $D \supset C$ we have $(F \setminus \{C\}) \cup \{D\} \in \mathcal{SAT}$. For $F \in \mathcal{MU}$ a **saturation** is some saturated $F' \in \mathcal{MU}$ where there exists a bijection $\alpha : F \to F'$ with $\forall\, C \in F : C \subseteq \alpha(C)$; by definition, every MU can be saturated. If we just add a few literal occurrences (possibly zero), staying with each step within $\mathcal{MU}$, then we speak of a **partial saturation** (this includes saturations); we note that the additions of a partial saturation can be arbitrarily permuted. Dually there is the notion of **marginal** MUs, those $F \in \mathcal{MU}$ where removing any literal from any clause creates a redundancy, that is, some clause following from the others. For $F \in \mathcal{MU}$ a **marginalisation** is some marginal $F' \in \mathcal{MU}$ such that there is a bijection $\alpha : F' \to F$ with $\forall\, C \in F' : C \subseteq \alpha(C)$; again, every MU can be marginalised, and more generally we speak of a **partial marginalisation**. As an example all $A_n \in \mathcal{MU}$, $n \in \mathbb{N}_0$, are saturated and marginal, while A_n is nonsingular iff $n \neq 1$.

For $F \in \mathcal{MU}$ and a variable $v \in \mathrm{var}(F)$, we define **local saturation** as the process of adding literals $v, \overline{v}$ to some clauses in F (not already containing $v, \overline{v}$), until adding any additional v or $\overline{v}$ yields a satisfiable clause-set. Then the result is **locally saturated on** v. For a saturated $F \in \mathcal{MU}$, as shown in [17, Lemma C.1], assigning any (single) variable in F (called "splitting") yields MUs (for more information see [19, Subsection 3.4]). The same proof yields in fact, that for a locally saturated $F \in \mathcal{MU}$ on a variable v, splitting on v maintains minimal unsatisfiability:

Lemma 1. *Consider $F \in \mathcal{MU}$ and a variable $v \in \mathrm{var}(F)$. If F is locally saturated for variable v, then we have $\langle v \to \varepsilon \rangle * F \in \mathcal{MU}$ for both $\varepsilon \in \{0, 1\}$.*

In general, application of the DP-operation to some MU may or may not yield another MU. A positive example for $n \in \mathbb{N}$ and $v \in \{1, \dots, n\}$ is $\mathrm{DP}_v(A_n) \cong A_{n-1}$. A special case of DP-reduction, guaranteed to stay inside MU, is **singular DP-reduction**, where v is a singular variable in F. In this case, as shown in [18, Lemma 9], no tautological resolvents can occur and no contractions can take place (recall that we are using clause-*sets*, where as a result of some operations previously different clause can become equal – a "contraction"). So even each $\mathcal{MU}_{\delta=k}$ is stable under singular DP-reduction. We use $\mathbf{sDP(\mathit{F})} \subset \mathcal{MU}'_{\delta=\delta(F)}$, $F \in \mathcal{MU}$, for the set of all clause-sets obtained from F by singular DP-reduction. By [18, Corollary 64] for any $F', F'' \in \mathrm{sDP}(F)$ holds $n(F') = n(F'')$. So we can define for $F \in \mathcal{MU}$ the **nonsingularity type** $\mathbf{nst(\mathit{F})} := n(F') \in \mathbb{N}_0$ via any $F' \in \mathrm{sDP}(F)$. Thus $\mathrm{nst}(F) = n(F)$ iff F is nonsingular.

MU(1) A basic fact is that $F \in \mathcal{MU} \setminus \{\{\bot\}\}$ contains a variable occurring positively and negatively each at most $\delta(F)$ times ([17, Lemma C.2]). So the minimum variable degree (the number of occurrences) is $2\delta(F)$ (sharper bounds are given in [19]). This implies that $F \in \mathcal{MU}_{\delta=1}$ has a **1-singular** variable (i.e., degree 2). It is well-known that for $F \in \mathcal{RHO}$ there exists an input-resolution tree T yielding $\{\bot\}$ ([10]); in the general framework of [9], these are those T with the Horton-Strahler number $\mathrm{hs}(T)$ at most 1. W.l.o.g. we can assume all these trees to be regular, that is, along any path no resolution variable is repeated. This implies that for $F \in \mathcal{MU} \cap \mathcal{RHO}$ holds $\delta(F) = 1$, and all variables in F are singular. By [17] all of $\mathcal{MU}_{\delta=1}$ is described by a binary tree T, which just describes the expansion process as mentioned in the Introduction, and which is basically the same as a resolution tree refuting F (T is not unique). Since the variables in the tree are all unique (the creation process does not reuse variables), any two clauses clash in at most one variable. For $F \in \mathcal{MU}_{\delta=1}$ with exactly one 1-singular variable holds $\mathrm{hs}(T) = 1$ (and so $F \in \mathcal{RHO}$), since $\mathrm{hs}(T) \geq 2$ implies that there would be two nodes whose both children are leaves, and so F would have two 1-singular variables. Furthermore if $F \in \mathcal{MU} \cap \mathcal{RHO}$ has a full clause C, then C is on top of T and so the complement of its literals occur only once:

Lemma 2. *Consider $F \in \mathcal{MU}$. If $\delta(F) = 1$, and F has only one 1-singular variable, then $F \in \mathcal{RHO}$. If $F \in \mathcal{RHO}$, then all variables are singular, and if F has a full clause C, then for every $x \in C$ the literal $\overline{x}$ occurs only once in F.*

The Splitting Ansatz The main method for analysing $F \in \mathcal{MU}$ is "splitting": choose an appropriate variable v in $F \in \mathcal{MU}$, apply the partial assignments $\langle v \to 0 \rangle$ and $\langle v \to 1 \rangle$ to F, obtain F_0, F_1, analyse them, and lift the information obtained back to F. An essential point here is to have $F_0, F_1 \in \mathcal{MU}$. In general this does not hold. The approach of Kleine Büning and Zhao, as outlined in [16, Section 3], is to remove clauses appropriately in F_0, F_1, and study various conditions. Our method is based on the observation, that if a clause say in F_0 became redundant, then $\overline{v}$ can be added to this clause in F, while still remaining MU, and so the assignment $v \to 0$ then takes care of the removal. This is the essence of *saturation*, with the advantage that we are dealing again with MUs. A saturated MU is characterised by the property, that for any variable, splitting yields two MUs. For classes like 2–$\mathcal{CLS}$, which are not stable under saturation, we introduced *local saturation*, which only saturates the variable we want to split on. In our application, the local saturation uses all clauses, and this is equivalent to a "disjunctive splitting" as surveyed [2, Definition 8]. On the other hand, for deficiency 2 the method of saturation is more powerful, since we have stability under saturation, and the existence of a variable occurring twice positively and twice negatively holds *after* saturation. Splitting needs to be done on nonsingular variables, so that the deficiency becomes strictly smaller in F_0, F_1 — we want these instances 'to be 'easy", to know them well. In both our cases we obtain indeed renamable Horn clause-sets. For deficiency 2 we exploit, that the splitting involves the minimal number of clauses, while for 2-CNF we exploit that the splitting involves the maximal number of clauses after local saturation. In order to get say F_0 "easy", while F is "not easy", the part which gets removed, which is related to F_1, must have special properties.

4 MU with Full Monotone Clauses (FM)

We now introduce formally the main classes of this paper, $\mathcal{FM} \subset \mathcal{MU}$ (Definition 1) and $\mathcal{DFM} \subset \mathcal{FM}$ (Definition 4). Examples for these classes showed up in the literature, but these natural classes haven't been studied yet.

Definition 1. *Let $\boldsymbol{\mathcal{FM}}$ be the set of $F \in \mathcal{MU}$ such that there is a full positive clause $P \in F$ and a full negative clause $N \in F$ (that is, $\mathrm{var}(P) = \mathrm{var}(N) = \mathrm{var}(F)$, $P \subset \mathcal{VA}$, $N \subset \overline{\mathcal{VA}}$). More generally, let $\boldsymbol{\mathcal{FC}}$ be the set of $F \in \mathcal{MU}$ such that there are full clauses $C, D \in F$ with $D = \overline{C}$.*

The closure of $\mathcal{FM}$ under isomorphism is $\mathcal{FC}$. In the other direction, for any $F \in \mathcal{FC}$ and any pair $C, D \in F$ of full clauses with $D = \overline{C}$ (note that in general such a pair is not unique), flip the signs so that C becomes a positive clause (and thus D becomes a negative clause), and we obtain an element of $\mathcal{FM}$. As usual we call the subsets of nonsingular elements $\boldsymbol{\mathcal{FM}'}$ resp. $\boldsymbol{\mathcal{FC}'}$. The *trivial elements* of $\mathcal{FM}$ and $\mathcal{FC}$ are the MUs with at most one variable: $\mathcal{FM}_{n \le 1} = \mathcal{FM}_{\delta=1} = \mathcal{FC}_{n \le 1} = \mathcal{FC}_{\delta=1} = \{\{\bot\}\} \cup \{\{v\}, \{\overline{v}\} : v \in \mathcal{VA}\}$. The singular cases in $\mathcal{FM}$ and $\mathcal{FC}$ are just these cases with only one variable:

Lemma 3. $\mathcal{FM}' = \mathcal{FM}_{\delta \ge 2} \cup \{\{\bot\}\}$, $\mathcal{FC}' = \mathcal{FC}_{\delta \ge 2} \cup \{\{\bot\}\}$.

Proof. Assume that there is a singular $F \in \mathcal{FC}$ with $n(F) \ge 2$. Let C, D be full complementary clauses in F. W.l.o.g. we can assume that there is $x \in C$ (so $\overline{x} \in D$) such that literal x only occurs in C. Consider now some $y \in D \setminus \{\overline{x}\}$ (exists due to $n(F) \ge 2$). There exists a satisfying assignment φ for $F' := F \setminus \{D\}$, and it must hold $\varphi(x) = 1$ and $\varphi(y) = 0$ (otherwise F would be satisfiable). Obtain φ' by flipping the value of x. Now φ' still satisfies F', since the only occurrence of literal x is C, and this clause contains $\overline{y}$ — but now φ' satisfies F. □

So the study of $\mathcal{FM}$ is about special nonsingular MUs. In general we prefer to study $\mathcal{FM}$ over $\mathcal{FC}$, since here we can define the "core" as a sub-clause-set:

Definition 2. *For $F \in \mathcal{FM}$ there is exactly one positive clause $P \in F$, and exactly one negative clause $N \in F$ (otherwise there would be subsumptions in F), and we call $F \setminus \{P, N\}$ the* ***core*** *of F.*

We note that cores consist only of mixed clauses, and in general any mixed clause-set (consisting only of mixed clauses) has always at least two satisfying assignments, the all-0 and the all-1 assignments. The decision complexity of $\mathcal{FM}$ is the same as that of $\mathcal{MU}$ (which is the same as $\mathcal{MU}'$), which has been determined in [22, Theorem 1] as complete for the class D^P, whose elements are differences of NP-classes (for example "MU = Irredundant minus SAT"):

Theorem 1. *For $F \in \mathcal{CLS}$, the decision whether "$F \in \mathcal{FM}$?" is D^P-complete.*

Proof. The decision problem is in D^P, since $F \in \mathcal{FM}$ iff F is irredundant with full monotone clauses and $F \notin \mathcal{SAT}$. For the reduction of $\mathcal{MU}$ to $\mathcal{FM}$, we consider $F \in \mathcal{CLS}$ with $n := n(F) \ge 2$, and first extend F to F', forcing a full positive clause, by taking a new variable v, adding literal v to all clauses of $\mathcal{F}_n$ and adding literal $\overline{v}$ to all clauses of F. Then we force additionally a full negative clause, extending F' to F'' in the same way, now using new variable w, and adding w to all clauses of F' and adding $\overline{w}$ to all clauses of $\mathcal{F}_{n+1}$. We have $F \in \mathcal{MU}$ iff $F'' \in \mathcal{MU}$. □

We now turn to the semantics of the core:

Definition 3. *For $V \subset \mathcal{VA}$ the* ***AllEqual*** *function on V is the boolean function which is true for a total assignment of V if all variables are assigned the same value, and false otherwise. A* ***CNF-realisation*** *of AllEqual on V is a clause-set F with $\mathrm{var}(F) \subseteq V$, which is as a boolean function the AllEqual function on V.*

The core of every FM F realises AllEqual on $\mathrm{var}(F)$ irredundantly, and this characterises $\mathcal{FM}$, yielding the AllEqual Theorem:

Theorem 2. *Consider $F \in \mathcal{CLS}$ with a full positive clause $P \in F$ and a full negative clause $N \in F$, and let $F' := F \setminus \{P, N\}$. Then $F \in \mathcal{FM}$ if and only if F' realises AllEqual on $\mathrm{var}(F)$, and F' is irredundant.*

5 FM with Binary Clauses (DFM)

Definition 4. ***$\mathcal{DFM}$*** *is the subset of $\mathcal{FM}$ where the core is in $2\text{–}\mathcal{CLS}$, while* ***$\mathcal{DFC}$*** *is the set of $F \in \mathcal{FC}$, such that there are full complementary clauses $C, D \in F$ with $F \setminus \{C, D\} \in 2\text{–}\mathcal{CLS}$.*

The core of DFMs consists of clauses of length exactly 2. $\mathcal{DFC}$ is the closure of $\mathcal{DFM}$ under isomorphism.

Definition 5. *For $F \in \mathcal{DFM}$ the* ***positive implication digraph*** $\mathbf{pdg}(\boldsymbol{F})$ *has vertex set $\mathrm{var}(F)$, i.e., $V(\mathrm{pdg}(F)) := \mathrm{var}(F)$, while the arcs are the implications on the variables as given by the core F' of F, i.e., $E(\mathrm{pdg}(F)) := \{(a, b) : \{\overline{a}, b\} \in F', a, b \in \mathrm{var}(F)\}$. This can also be applied to any mixed binary clause-set F (note that the core F' is such a mixed binary clause-set).*

The essential feature of mixed clause-sets $F \in 2\text{–}\mathcal{CLS}$ is that for a clause $\{\overline{v}, w\} \in F$ we only need to consider the "positive interpretation" $v \to w$, not the "negative interpretation" $\overline{w} \to \overline{v}$, since the positive literals and the negative literals do not interact. So we do not need the (full) implication digraph. Via the positive implication digraphs we can understand when a mixed clause-set realises AllEqual. We recall that digraph G is a **strong digraph** (SD), if G is strongly connected, i.e., for every two vertices a, b there is a path from a to b. A **minimal strong digraph** (MSD) is an SD G, such that for every arc $e \in E(G)$ holds that $(V(G), E(G) \setminus \{e\})$ is not strongly connected. Every digraph G with $|V(G)| \le 1$ is an MSD. We are ready to formulate the Correspondence Lemma:

Lemma 4. *A mixed binary clause-set F is a CNF-realisation of AllEqual iff $\mathrm{pdg}(F)$ is an SD, where F is irredundant iff $\mathrm{pdg}(F)$ is an MSD.*

Proof. The main point here is that the resolution operation for mixed binary clauses $\{\overline{a}, b\}, \{\overline{b}, c\}$, resulting in $\{\overline{a}, c\}$, corresponds exactly to the formation of transitive arcs, i.e., from $(a, b), (b, c)$ we obtain (a, c). So the two statements of the lemma are just easier variations on the standard treatment of logical reasoning for 2-CNFs via "path reasoning". $\square$

As explained before, $F \mapsto \mathrm{pdg}(F)$ converts mixed binary clause-sets with full monotone clauses to a digraph. Also the reverse direction is easy:

Definition 6. *For a finite digraph G with $V(G) \subset \mathcal{VA}$, the clause-set* $\mathbf{mcs}(\boldsymbol{G}) \in \mathcal{CLS}$ *("m" like "monotone") is obtained by interpreting the arcs $(a, b) \in E(G)$ as binary clauses $\{\overline{a}, b\} \in \mathrm{mcs}(G)$, and adding the two full monotone clauses $\{V(G), \overline{V(G)}\} \subseteq \mathrm{mcs}(G)$.*

For the map $G \mapsto \mathrm{mcs}(G)$, we use the vertices of G as the variables of $\mathrm{mcs}(G)$. An arc (a,b) naturally becomes a mixed binary clause $\{\overline{a}, b\}$, and we obtain the set F' of mixed binary clauses, where by definition we have $\mathrm{pdg}(F') = G$. This yields a bijection between the set of finite digraphs G with $V(G) \subset \mathcal{VA}$ and the set of mixed binary clause-sets. By the Correspondence Lemma 4, minimal strong connectivity of G is equivalent to F' being an irredundant AllEqual-representation. So there is a bijection between MSDs and the set of mixed binary clause-sets which are irredundant AllEqual-representation. We "complete" the AllEqual-representations to MUs, by adding the full monotone clauses, and we get the DFM $\mathrm{mcs}(G)$. We see, that DFMs and MSDs are basically the "same thing", only using different languages, which is now formulated as the Correspondence Theorem (with obvious proofs left out):

Theorem 3. *The two formations $F \mapsto \mathrm{pdg}(F)$ and $G \mapsto \mathrm{mcs}(G)$ are inverse to each other, i.e., $\mathrm{mcs}(\mathrm{pdg}(F)) = F$ and $\mathrm{pdg}(\mathrm{mcs}(G)) = G$, and they yield inverse bijections between DFMs and MSDs: For every $F \in \mathcal{DFM}$ the digraph $\mathrm{pdg}(F)$ is an MSD, and for every MSD G with $V(G) \subset \mathcal{VA}$ we have $\mathrm{mcs}(G) \in \mathcal{DFM}$.*

The Correspondence Theorem 3 can be considerably strengthened, by including other close relations, but here we formulated only what we need. For a DFM $F \neq \{\bot\}$ and an MSD $G \neq (\emptyset, \emptyset)$ we obtain $\delta(\mathrm{pdg}(F)) = \delta(F) - 2$ and $\delta(\mathrm{mcs}(G)) = \delta(G) + 2$, where we define the **deficiency** of a digraph G as $\boldsymbol{\delta(G)} := |E(G)| - |V(G)|$. Concerning isomorphisms there is a small difference between the two domains, since the notion of clause-set isomorphism includes flipping of variables, which for DFMs can be done all at once (flipping "positive" and "negative") — this corresponds in $\mathrm{pdg}(F)$ to the reversal of the direction of all arcs. For our two main examples, cycles and dipaths, this yields an isomorphic digraph, but this is not the case in general.

Marginalisation of DFMs concerns only the full monotone clauses and not the binary clauses, formulated as the Marginalisation Lemma:

Lemma 5. *Consider a clause-set F obtained by partial marginalisation of a non-trivial DFM F'. Then F has no unit-clause and its formation did not touch binary clauses but only shortened its monotone clauses.*

Proof. By definition, partial marginalisation can be arbitrarily reordered. If some binary clause would be shortened, then, put first, this would yield unit-clauses, subsuming some full monotone clauses. □

Deciding $F \in \mathcal{DFM}$ can be done in polynomial time: Check whether we have the two full monotone clauses, while the rest are binary clauses, if yes, translate the binary clauses to a digraph and decide whether this digraph is an MSD (which can be done in quadratic time; recall that deciding the SD property can be done in linear time) — if yes, then $F \in \mathcal{DFM}$, otherwise $F \notin \mathcal{DFM}$. We now come to the two simplest example classes, cycles and "di-paths". Let $C_n := (\{1,\dots,n\}, \{(1,2),\dots,(n-1,n),(n,1)\})$ for $n \geq 2$ be the directed cycle of length n. The directed cycles C_n have the minimum deficiency zero among MSDs. We obtain the basic class $\mathcal{F}_n$, as already explained in the Introduction:

Definition 7. *Let* $\boldsymbol{\mathcal{F}_n} := \mathrm{mcs}(C_n) \in \mathcal{DFM}$ *for* $n \geq 2$ *(Definition 6).*

Lemma 6. *For* $F \in \mathcal{DFM}_{\delta=2}$ *holds* $F \cong \mathcal{F}_{n(F)}$*.*

Proof. By the Correspondence Theorem 3, $\mathrm{pdg}(F)$ is an MSD with the deficiency $\delta(F) - 2 = 0$, and thus is a directed cycle of length $n(F)$. □

Lemma 7. *For every* $n \geq 2$*,* $\mathcal{F}_n$ *is saturated.*

Proof. We show that adding a literal x to any $C \in \mathcal{F}_n$ introduces a satisfying assignment, i.e, $\mathcal{F}_n$ is saturated. The monotone clauses are full, and saturation can only touch the mixed clauses. Recall $\mathrm{var}(\mathcal{F}_n) = \{1, \ldots, n\}$. Due to symmetry assume $C = \{-n, 1\}$, and we add $x \in \{2, \ldots, n-1\}$ to C. Let φ be the total assignment setting all variables $2, \ldots, n$ to true and 1 to false. Then φ satisfies the monotone clauses and the new clause $\{-n, 1, x\}$. Recall that every literal occurs only once in the core of $\mathcal{F}_n$. So literal 1 occurs only in C. Thus φ satisfies also every mixed clause in $F \setminus \{C\}$ (which has a positive literal other than 1). □

For a tree G (a finite connected acyclic graph with at least one vertex) we denote by $\boldsymbol{D(G)} := (V(G), \{(a,b),(b,a) : \{a,b\} \in E(G)\})$ the directed version of G, converting every edge $\{a,b\}$ into two arcs $(a,b),(b,a)$; in [7] these are called "directed trees", and we use **ditree** here. For every tree G the ditree $D(G)$ is an MSD. Let $P_n := (\{1, \ldots, n\}, \{\{1,2\}, \ldots, \{n-1,n\}\}$, $n \in \mathbb{N}_0$, be the pathgraph.

Definition 8. *Let* $\mathbf{DB}_{\boldsymbol{n}} := \mathrm{mcs}(D(P_n)) \in \mathcal{DFM}$ *(*$n \in \mathbb{N}_0$*) (Definition 6).*

So $\mathrm{DB}_n = A_n$ for $n \leq 2$, while in general $n(\mathrm{DB}_n) = n$, and for $n \geq 1$ holds $c(\mathrm{DB}_n) = 2 + 2(n-1) = 2n$, and $\delta(\mathrm{DB}_n) = n$. DB_n for $n \neq 1$ is nonsingular, and every variable in $\mathrm{var}(\mathrm{DB}_n) \setminus \{1, n\}$ is of degree 6 for $n \geq 2$, while the variables $1, n$ (the endpoints of the dipath) have degree 4. Among ditrees, only dipaths can be marginalised to nonsingular 2-uniform MUs, since dipaths are the only ditrees with exactly two *linear vertices* (i.e., vertices with indegree and outdegree equal to 1). The unique marginal MUs obtained from dipaths are as follows:

Definition 9. *For* $n \geq 1$ *obtain the uniform* $\boldsymbol{\mathcal{B}_n} \in 2\text{-}\mathcal{MU}$ *from* DB_n *by replacing the full positive/negative clause with* $\{1, n\}$ *resp.* $\{-1, -n\}$*, i.e.,* $\mathcal{B}_n = \{\{-1, -n\}, \{1, n\}, \{-1, 2\}, \{1, -2\}, \ldots, \{-(n-1), n\}, \{n-1, -n\}\}$*;* $\mathcal{B}_0 := \mathrm{DB}_0$*.*

6 Deficiency 2 Revisited

We now come to the first main application of the new class $\mathcal{DFM}$, and we give a new and relatively short proof, that the $\mathcal{F}_n$ are precisely the nonsingular MUs of deficiency 2. The core combinatorial-logical argument is to show $\mathcal{MU}'_{\delta=2} \subseteq \mathcal{FC}_{\delta=2}$, i.e., every $F \in \mathcal{MU}'_{\delta=2}$ must have two full complementary clauses $C, D \in F$. The connection to the "geometry" then is established by showing $\mathcal{FM}_{\delta=2} \subseteq \mathcal{DFM}_{\delta=2}$, i.e, if an FM F has deficiency 2, then it must be a DFM, i.e., all clauses besides the full monotone clauses are binary. The pure geometrical argument is the characterisation of $\mathcal{DFM}_{\delta=2}$, which has already been done in Lemma 6.

The proof of the existence of full clauses $D = \overline{C}$ in F is based on the Splitting Ansatz, as explained in Section 3. Since $\mathcal{MU}_{\delta=2}$ is stable under saturation, we can start with a saturated F, and can split on any variable (though later an argument is needed to undo saturation). There must be a variable v occurring at most twice positively as well as negatively (otherwise the basic lemma $\delta(F) \geq 1$ for any MU F would be violated), and due to nonsingularity v occurs exactly twice positively and negatively. The splitting instances F_0, F_1 have deficiency 1. So they have at least one 1-singular variable. There is very little "space" to reduce a nonsingular variable in F to a 1-singular variable in F_0 resp. F_1, and indeed those two clauses whose vanishing in F_0 do this, are included in F_1, and vice versa. Since clauses in $\mathcal{MU}_{\delta=1}$ have at most one clash, F_0, F_1 have exactly one 1-singular variable. And so by the geometry of the structure trees (resp. their Horton-Strahler numbers), both F_0, F_1 are in fact renamable Horn! Thus every variable in F_0, F_1 is singular, and F_0, F_1 must contain a unit-clause. Again considering both sides, it follows that the (two) positive occurrences of v must be a binary clause (yielding the unit-clause) and a full clause C (whose vanishing yields the capping of all variables to singular variables), and the same for the (two) negative occurrences, yielding D. So $F_0, F_1 \in \mathcal{RHO}$ both contain a full clause and we know that the complements of the literals in the full clause occur exactly once in F_0 resp. F_1. Thus in fact C resp. D have the "duty" of removing each others complement, and we get $D = \overline{C}$.

Now consider $F \in \mathcal{FM}_{\delta=2}$ with monotone full clauses $C, D \in F$. Transform the core F' within F into an equivalent F'', by replacing each clause in F' by a contained prime implicate of F', which, since the core means that all variables are equal (semantically), is binary. So we arrive in principle in $\mathcal{DFM}$, but we could have created redundancy — and this can not happen, since an MSD has minimum deficiency 0. The details are as follows:

Theorem 4. $\mathcal{DFC}_{\delta=2} = \mathcal{FC}_{\delta=2} = \mathcal{MU}'_{\delta=2}$.

Proof. By definition and Lemma 3 we have $\mathcal{DFC}_{\delta=2} \subseteq \mathcal{FC}_{\delta=2} \subseteq \mathcal{MU}'_{\delta=2}$. First we show $\mathcal{MU}'_{\delta=2} \subseteq \mathcal{FC}_{\delta=2}$, i.e., every $F \in \mathcal{MU}'_{\delta=2}$ has two full complementary clauses. Recall that F has a variable $v \in \mathrm{var}(F)$ of degree 4, which by nonsingularity is the minimum variable degree. So v has two positive occurrences in clauses $C_1, C_2 \in F$ and two negative occurrences in clauses $D_1, D_2 \in F$. We assume that F is saturated (note that saturation maintains minimal unsatisfiability and deficiency). By the Splitting Ansatz, $F_0 := \langle v \to 0 \rangle * F \in \mathcal{MU}_{\delta=1}$ and $F_1 := \langle v \to 1 \rangle * F \in \mathcal{MU}_{\delta=1}$. So F_0 removes D_1, D_2 and shortens C_1, C_2, while F_1 removes C_1, C_2 and shortens D_1, D_2. Both F_0, F_1 contain a 1-singular variable (i.e., of degree 2), called a resp. b. We obtain $\{a, \overline{a}\} \subseteq D_1 \cup D_2$, since F has no singular variable and only by removing D_1, D_2 the degree of a decreased to 2. Similarly $\{b, \overline{b}\} \subseteq C_1 \cup C_2$. In $\mathcal{MU}_{\delta=1}$ any two clauses have at most one clash, and thus indeed F_0, F_1 have each exactly one 1-singular variable. Now $F_0, F_1 \in \mathcal{MU}_{\delta=1}$ with exactly one 1-singular variable are renamable Horn clause-sets (Lemma 2). Since $F_0, F_1 \in \mathcal{RHO} \cap \mathcal{MU}$ contain unit-clauses, created by clause-shortening, one of C_1, C_2 and one of D_1, D_2 are binary. W.l.o.g. assume C_1, D_1 are binary. Furthermore by Lemma 2 all variables in F_0, F_1 are singular, while F has no singular variable. So in F_0 all singularity is created by the removal of D_1, D_2, and in F_1 all singularity is created by the removal of C_1, C_2. Thus C_2, D_2 are full clauses. For a full clause in F_0, F_1, the complement of its literals occur only once (recall Lemma 2). Thus C_2 and D_2 have the duty of eliminating each others complements, and so we obtain $C_2 = \overline{D_2}$. To finish the first part, we note that the literals in clauses C, D each occurs exactly twice by the previous argumentation, and thus, since F was nonsingular to start with, indeed the initial saturation did nothing.

We turn to the second part of the proof, showing $\mathcal{FM}_{\delta=2} \subseteq \mathcal{DFM}_{\delta=2}$, i.e., the core F' of every $F \in \mathcal{FM}_{\delta=2}$ contains only binary clauses. By the characterisation of FMs, the AllEqual Theorem 2, F' realises AllEqual over the variables of F. The deficiency of F' is $\delta(F') = \delta(F) - 2 = 0$. Obtain F'' by replacing each $C \in F'$ by a prime implicate $C'' \subseteq C$ of F', where every prime implicate is binary. Now F'' is logically equivalent to F', and we can apply the Correspondence Lemma 4 to F'', obtaining an MSD $G := \mathrm{pdg}(F'')$ with $\delta(G) = \delta(F'')$. Due to the functional characterisation of F' we have $\mathrm{var}(F'') = \mathrm{var}(F') = \mathrm{var}(F)$. Using that MSDs have minimal deficiency 0, thus $\delta(G) = 0$, and so G is the cycle of length $n(F)$, and thus F'' is isomorphic to $\mathcal{F}_{n(F)}$. Now $\mathcal{F}_{n(F)}$ is saturated (Lemma 7), and thus indeed $F'' = F'$. □

Corollary 1 ([12]). *For $F \in \mathcal{MU}'_{\delta=2}$ holds $F \cong \mathcal{F}_{n(F)}$.*

7 MU for 2-CNF

Lemma 8. *$F \in 2\text{-}\mathcal{MU}$ with a unit-clause is in $\mathcal{RHO}$, and has at most two unit-clauses ([2]). In every $F \in 2\text{-}\mathcal{MU}$ each literal occurs at most twice ([15]).*

Lemma 9. *In $F \in 2\text{-}\mathcal{MU}$ with exactly two unit-clauses, every literal occurs exactly once. Both unit-clauses can be partially saturated to a full clause (yielding two saturations), and these two full clauses are complementary.*

Proof. By Lemma 8, $F \in \mathcal{RHO}$ with $\delta(F) = 1$. Since F is uniform except of two unit-clauses, the number of literal occurrences is $2c(F) - 2 = 2n(F)$, and so every literal in F occurs only once (F is marginal). Consider an underlying tree T according to Section 3, in the form of an input-resolution tree T. The key is that any of the two unit-clauses can be placed at the top, and thus can be saturated (alone) to a full clause. Since in an input-resolution tree, at least one unit-clause is needed at the bottom, to derive the empty clause, we see that the two possible saturations yield complementary clauses. □

We now come to the main results of this section, characterising the nonsingular MUs in 2-CNF. First the combinatorial part of the characterisation: the goal is to show that $F \in 2\text{-}\mathcal{MU}'$ can be saturated to a DFM, up to renaming, i.e., there exist a positive clause and a negative clause which can be partially saturated to full positive and negative clauses. The proof is based on the Splitting Ansatz. Unlike $\mathcal{MU}'_{\delta=2}$, 2-CNF MUs are not stable under saturation. So we use *local* saturation on a variable $v \in \mathrm{var}(F)$, where we get splitting instances $F_0, F_1 \in 2\text{–}\mathcal{MU}$. Then we show F_0, F_1 are indeed in $\mathcal{RHO}$ with exactly 2 unit-clauses, and we apply that any of these unit-clauses can be saturated to a full clause. W.l.o.g. we saturate any of the two unit-clauses in F_0 to a full positive clause. Now one of the two unit-clauses in F_1 can be saturated to a full negative clause, and the two full monotone clauses can be lifted to F. This yields a DFM which is a partial saturation of F. The details are as follows:

Theorem 5. *Every $2\text{–}\mathcal{MU}'$ can be partially saturated to some $\mathcal{DFC}$.*

Proof. We show $F \in 2\text{–}\mathcal{MU}'$ contains, up to flipping of signs, exactly one positive and one negative clause, and these can be saturated to full monotone clauses. F has no unit-clause and is 2-uniform. By Lemma 8 every literal in F has degree 2. Let $F' \in \mathcal{MU}$ be a clause-set obtained from F by locally saturating $v \in \mathrm{var}(F)$. So $F_0 := \langle v \to 0 \rangle * F'$ and $F_1 := \langle v \to 1 \rangle * F'$ are in $2\text{–}\mathcal{MU}$ (Lemma 1) and each has exactly two unit-clauses (obtained precisely from the clauses in F containing $v, \overline{v}$). By Lemma 8 holds $F_0, F_1 \in \mathcal{RHO} \cap \mathcal{MU}_{\delta=1}$. And by Lemma 9 all variables are 1-singular and in each of F_0, F_1, both unit-clauses can be partially saturated to a full clause. These full clauses can be lifted to the original F (by adding v resp. $\overline{v}$) while maintaining minimal unsatisfiability (if both splitting results are MU, so is the original clause-set; see [19, Lemma 3.15, Part 1]). Now we show that for a full clause in F_0, F_1 adding v or $\overline{v}$ yields a full clause in F, i.e., only v vanished by splitting. All variables in F_0, F_1 are 1-singular, while F has no singular variable. If there would be a variable w in F_0 but not in F_1, then the variable degree of w would be 2 in F, a contradiction. Thus $\mathrm{var}(F_0) \subseteq \mathrm{var}(F_1)$. Similarly we obtain $\mathrm{var}(F_1) \subseteq \mathrm{var}(F_0)$. So $\mathrm{var}(F_0) = \mathrm{var}(F_1) = \mathrm{var}(F) \setminus \{v\}$.

It remains to show that we can lift w.l.o.g. a full positive clause from F_0 and a full negative clause from F_1. Let $C_1, C_2 \in F$ be the clauses containing v and $D_1, D_2 \in F$ be the clauses containing $\overline{v}$. Assume the unit-clause $C_1 \setminus \{v\} \in F_0$ can be saturated to a full positive clause. This implies that every $C \in F \setminus \{C_1\}$ has a negative literal (since $F \setminus \{C_1\}$ is satisfied by setting all variables to false). Then by Lemma 9 the unit-clause $C_2 \setminus \{v\}$ can be saturated to a full negative clause in F_0. Similarly we obtain that every clause in $F \setminus \{C_2, D_1, D_2\}$ has a positive literal. So F has exactly one positive clause C_1 and all binary clauses in F_0, F_1 are mixed. Since $c(F_1) = n(F_1) + 1 = (n(F) - 1) + 1 = n(F)$ and there are $n(F) - 1$ occurrences of each literal in F_1, w.l.o.g. D_1 is a negative clause and D_2 is mixed. Recall that in $\mathcal{MU}_{\delta=1}$ every two clauses have at most one clash, and so $D_1 \setminus \{\overline{v}\} \in F_1$ can be saturated to a full negative clause (otherwise there would be a clause with more than one clash with the full clause). So we obtain a DFM which is a partial saturation of F. □

By [7, Theorem 4], every MSD with at least two vertices has at least two linear vertices. We need to characterise a special case of MSDs with exactly two linear vertices. This could be derived from the general characterisation by [8, Theorem 7], but proving it directly is useful and not harder than to derive it:

Lemma 10. *An MSD G with exactly two linear vertices, where every other vertex has indegree and outdegree both at least 2, is a dipath.*

Proof. We show that G is a dipath by induction on $n := |V(G)|$. For $n = 2$ clearly G is MSD iff G is a dipath. So assume $n \geq 3$. Consider a linear vertex $v \in V(G)$ with arcs (w, v) and (v, w'), where $w, w' \in V(G)$. If $w \neq w'$ would be the case, then the MSD obtained by removing v and adding the arc (w, w') had only one linear vertex (since the indegree/outdegree of other vertices are unchanged). So we have $w = w'$. Let G' be the MSD obtained by removing v. Now w is a linear vertex in G' (since every MSD has at least two linear vertices). By induction hypothesis G' is a dipath, and the assertion follows. $\square$

By definition, for a mixed binary clause-set F a 1-singular variable is a linear vertex in $\mathrm{pdg}(F)$. So by the Correspondence Theorem, a variable v in a DFM F has degree 4 (i.e., degree 2 in the core) iff v is a linear vertex in $\mathrm{pdg}(F)$.

Theorem 6. *$F \in \mathcal{DFC}$ can be partially marginalised to some nonsingular element of 2–$\mathcal{CLS}$ if and only if $F \cong \mathrm{DB}_{n(F)}$.*

Proof. Since $\mathcal{B}_n$ is a marginalisation of DB_n (obviously then the unique nonsingular one), it remains to show that a DMF F, which can be partially marginalised as in the assertion, is isomorphic to $\mathrm{DB}_{n(F)}$. We show that $\mathrm{pdg}(F)$ has exactly two linear vertices, while all other vertices have indegree and outdegree at least two, which proves the statement by Lemma 10. Consider a nonsingular $G \in 2$–$\mathcal{MU}$ obtained by marginalisation of F. Recall that by the Marginalisation Lemma 5 the mixed clauses are untouched. $\mathrm{pdg}(F)$ has at least two linear vertices, so the mixed clauses in G have at least two 1-singular variables. Indeed the core of F has exactly two 1-singular variables, since these variables must occur in the positive and negative clauses of G, which are of length two. The other vertices have indegree/outdegree at least two due to nonsingularity. $\square$

By Theorems 5, 6 we obtain a new proof for the characterisation of nonsingular MUs with clauses of length at most two:

Corollary 2 ([15]). *For $F \in 2$–$\mathcal{MU}'$ holds $F \cong \mathcal{B}_{n(F)}$.*

8 Conclusion

We introduced the novel classes $\mathcal{FM}$ and $\mathcal{DFM}$, which offer new conceptual insights into MUs. Fundamental for $\mathcal{FM}$ is the observation, that the easy syntactical criterion of having both full monotone clauses immediately yields the complete understanding of the semantics of the core. Namely that the satisfying assignments of the core are precisely the negations of the full monotone clauses, and so all variables are either all true or all false, i.e., all variables are equivalent. $\mathcal{DFM}$ is the class of FMs where the core is a 2-CNF. This is equivalent to the clauses of the core, which must be mixed binary clauses $\{\overline{v}, w\}$, constituting an MSD via the arcs $v \to w$. Due to the strong correspondence between DFMs and MSDs, once we connect a class of MUs to $\mathcal{DFM}$, we can use the strength of graph-theoretical reasoning. As a first application of this approach, we provided the known characterisations of $\mathcal{MU}'_{\delta=2}$ and 2–$\mathcal{MU}'$ in an accessible manner, unified by revealing the underlying graph-theoretical reasoning.

References

1. Ron Aharoni and Nathan Linial. Minimal non-two-colorable hypergraphs and minimal unsatisfiable formulas. *Journal of Combinatorial Theory, Series A*, 43(2):196–204, November 1986. doi:10.1016/0097-3165(86)90060-9.
2. Hans Kleine Büning, Piotr Wojciechowski, and K. Subramani. On the computational complexity of read once resolution decidability in 2CNF formulas. In T. Gopal, G. Jger, and S. Steila, editors, *International Conference on Theory and Applications of Models of Computation (TAMC 2017)*, volume 10185 of *Lecture Notes in Computer Science (LNCS)*, pages 362–372. Springer, 2017. doi:10.1007/978-3-319-55911-7_26.
3. Joshua Buresh-Oppenheim and David Mitchell. Minimum witnesses for unsatisfiable 2CNFs. In Armin Biere and Carla P. Gomes, editors, *Theory and Applications of Satisfiability Testing - SAT 2006*, volume 4121 of *Lecture Notes in Computer Science*, pages 42–47. Springer, 2006. doi:10.1007/11814948_6.
4. Joshua Buresh-Oppenheim and David Mitchell. Minimum 2CNF resolution refutations in polynomial time. In Joao Marques-Silva and Karem A. Sakallah, editors, *Theory and Applications of Satisfiability Testing - SAT 2007*, volume 4501 of *Lecture Notes in Computer Science*, pages 300–313. Springer, 2007. doi:10.1007/978-3-540-72788-0_29.
5. 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(3-4):229–245, 1998. doi:10.1023/A:1018924526592.
6. Herbert Fleischner, Oliver Kullmann, and Stefan Szeider. Polynomial–time recognition of minimal unsatisfiable formulas with fixed clause–variable difference. *Theoretical Computer Science*, 289(1):503–516, November 2002. doi:10.1016/S0304-3975(01)00337-1.
7. J. García-López and C. Marijuán. Minimal strong digraphs. *Discrete Mathematics*, 312(4):737–744, February 2012. doi:10.1016/j.disc.2011.11.010.
8. J. García-López, C. Marijuán, and L.M. Pozo-Coronado. Structural properties of minimal strong digraphs versus trees. *Linear Algebra and its Applications*, 540:203–220, March 2018. doi:10.1016/j.laa.2017.11.027.
9. Matthew Gwynne and Oliver Kullmann. Generalising unit-refutation completeness and SLUR via nested input resolution. *Journal of Automated Reasoning*, 52(1):31–65, January 2014. doi:10.1007/s10817-013-9275-8.
10. Lawrence J. Henschen and Lawrence Wos. Unit refutations and Horn sets. *Journal of the Association for Computing Machinery*, 21(4):590–605, October 1974. doi:10.1145/321850.321857.
11. Leonid Khachiyan, Endre Boros, Khaled Elbassioni, and Vladimir Gurvich. On enumerating minimal dicuts and strongly connected subgraphs. *Algorithmica*, 50:159–172, January 2008. doi:10.1007/s00453-007-9074-x.
12. Hans Kleine Büning. On subclasses of minimal unsatisfiable formulas. *Discrete Applied Mathematics*, 107(1-3):83–98, 2000. doi:10.1016/S0166-218X(00)00245-6.
13. 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. doi:10.3233/978-1-58603-929-5-339.

14. Hans Kleine Büning and Daoyun Xu. The complexity of homomorphisms and renamings for minimal unsatisfiable formulas. *Annals of Mathematics and Artificial Intelligence*, 43(1-4):113–127, 2005. doi:10.1007/s10472-005-0422-8.
15. Hans Kleine Büning and Xishun Zhao. Minimal unsatisfiability: Results and open questions. Technical Report tr-ri-02-230, Series Computer Science, University of Paderborn, University of Paderborn, Department of Mathematics and Computer Science, 2002. http://wwwcs.uni-paderborn.de/cs/ag-klbue/de/research/MinUnsat/index.html.
16. Hans Kleine Büning and Xishun Zhao. On the structure of some classes of minimal unsatisfiable formulas. *Discrete Applied Mathematics*, 130(2):185–207, 2003. doi:10.1016/S0166-218X(02)00405-5.
17. Oliver Kullmann. An application of matroid theory to the SAT problem. In *Proceedings of the 15th Annual IEEE Conference on Computational Complexity*, pages 116–124, July 2000. See also TR00-018, Electronic Colloquium on Computational Complexity (ECCC), March 2000. doi:10.1109/CCC.2000.856741.
18. Oliver Kullmann and Xishun Zhao. On Davis-Putnam reductions for minimally unsatisfiable clause-sets. *Theoretical Computer Science*, 492:70–87, June 2013. doi:10.1016/j.tcs.2013.04.020.
19. Oliver Kullmann and Xishun Zhao. Bounds for variables with few occurrences in conjunctive normal forms. Technical Report arXiv:1408.0629v5 [math.CO], arXiv, January 2017. Available from: http://arxiv.org/abs/1408.0629.
20. Choongbum Lee. On the size of minimal unsatisfiable formulas. *The Electronic Journal of Combinatorics*, 16(1), 2009. Note #N3. Available from: http://www.combinatorics.org/Volume_16/Abstracts/v16i1n3.html.
21. Paolo Liberatore. Redundancy in logic II: 2CNF and Horn propositional formulae. *Artificial Intelligence*, 172(2-3):265–299, February 2008. doi:10.1016/j.artint.2007.06.003.
22. Christos H. Papadimitriou and David Wolfe. The complexity of facets resolved. *Journal of Computer and System Sciences*, 37(1):2–13, August 1988. doi:10.1016/0022-0000(88)90042-6.
23. Brendan McKay (https://mathoverflow.net/users/9025/brendan-mckay). Answer: Graph isomorphism problem for minimally strongly connected digraphs. MathOverflow, August 2017. URL:https://mathoverflow.net/q/279299 (version: 2017-08-23); question asked by Hoda Abbasizanjani (Swansea University). Available from: https://mathoverflow.net/a/279386.