

SATISFIABILITY IN MULTI-VALUED CIRCUITS

PAWEŁ M. IDZIAK AND JACEK KRZACZKOWSKI

*Jagiellonian University,
Faculty of Mathematics and Computer Science,
Department of Theoretical Computer Science,
ul. Prof. S. Łojasiewicza 6,
30-348, Kraków, Poland*

ABSTRACT. Satisfiability of Boolean circuits is among the most known and important problems in theoretical computer science. This problem is NP-complete in general but becomes polynomial time when restricted either to monotone gates or linear gates. We go outside Boolean realm and consider circuits built of any fixed set of gates on an arbitrary large finite domain. From the complexity point of view this is strictly connected with the problems of solving equations (or systems of equations) over finite algebras.

The research reported in this work was motivated by a desire to know for which finite algebras \mathbf{A} there is a polynomial time algorithm that decides if an equation over \mathbf{A} has a solution. We are also looking for polynomial time algorithms that decide if two circuits over a finite algebra compute the same function. Although we have not managed to solve these problems in the most general setting we have obtained such a characterization for a very broad class of algebras from congruence modular varieties. This class includes most known and well-studied algebras such as groups, rings, modules (and their generalizations like quasi-groups, loops, near-rings, nonassociative rings, Lie algebras), lattices (and their extensions like Boolean algebras, Heyting algebras or other algebras connected with multi-valued logics including MV-algebras).

This paper seems to be the first systematic study of the computational complexity of satisfiability of non-Boolean circuits and solving equations over finite algebras. The characterization results provided by the paper is given in terms of nice structural properties of algebras for which the problems are solvable in polynomial time.

E-mail address: idziak@tcs.uj.edu.pl, jacek.krzaczkowski@uj.edu.pl.

Date: July 23, 2017.

The project is partially supported by Polish NCN Grant # 2014/14/A/ST6/00138.

1. INTRODUCTION

One of the most celebrated NP-complete problem is SAT – the problem that takes on a Boolean expression and decides whether there is a $\{0, 1\}$ -valuation of variables that satisfies this expression.

The most popular variant of this problem is CNF-SAT (often called SAT as well) in which the input is in Conjunctive Normal Form. A formula in CNF is a conjunction of clauses each of which is a disjunction of (e.g. at most 3) literals. These clauses (if ternary) can be treated as (ternary) relations on the set $\{0, 1\}$ and the SAT problem simply asks whether a conjunction of atomic formulas (in this new relational language) is satisfiable. This generalizes to any (finite) relational structure, say \mathbb{D} , where the problem lies in answering whether a conjunction of atomic formulas (in the language of \mathbb{D}) is satisfiable in \mathbb{D} . This is now known under the name of Constraint Satisfaction Problem, or CSP for short. A characterization of relational structures over $\{0, 1\}$ for which CSP is solvable in a polynomial time has been done in [35]. The structures for which a polynomial time algorithm is not provided in [35] have been shown there to be NP-complete with respect to CSP. The similar dichotomy conjecture for CSP over arbitrary finite domains has been stated by Feder and Vardi in [8]. With the help of deep algebraic tools two algorithmic paradigms have been shown to be fruitful in establishing polynomial time complexity of a wide range of relational structures. One of these paradigms generalizes Gaussian elimination method to the realm of algebras with few subpowers [25]. The other generalizes DATALOG programming to local consistency checking method [2]. Both of those methods were explored to their limits, so that they cannot be extended any further and a new approach is needed. Very recently three independent proofs (one by D. Zhuk, another one by A. Rafiey, J. Kinne and T. Feder and the third one by A. Bulatov) of the CSP dichotomy conjecture have been announced.

In contrast to CNF-SAT the problem of satisfiability of general Boolean expression is often called CIRCUITS SAT or CSAT for short. After restricting this NP-complete problem for example to the circuits that are either monotone (only AND and OR gates) or linear (only XOR gates) the problem becomes solvable in a polynomial time. Thus it is natural to isolate those collections of 2-valued gates that lead to circuits with polynomially solvable satisfiability problem. Actually such characterization of tractable families of 2-valued gates can be inferred from the results of [13].

In general, different collections of admissible gates (on a given set) give rise to algebras (in the universal algebraic sense). Thus we will talk about circuits over a fixed algebra \mathbf{A} . In this language the output gates of such circuits can be represented by terms of an algebra \mathbf{A} (or polynomials of \mathbf{A} , if values on some input gates are fixed). We also relax the notion of satisfiability of such circuits to be read:

CSAT(\mathbf{A})

given a circuit over \mathbf{A} with two output gates $\mathbf{g}_1, \mathbf{g}_2$ is there a valuation of input gates \bar{x} that gives the same output on $\mathbf{g}_1, \mathbf{g}_2$, i.e. $\mathbf{g}_1(\bar{x}) = \mathbf{g}_2(\bar{x})$.

Note here, that in some cases (including 2-element Boolean algebra) the satisfiability of $\mathbf{g}_1(\bar{x}) = \mathbf{g}_2(\bar{x})$ can be replaced by satisfiability of $\mathbf{g}(\bar{x}) = c$, where c is a constant and \mathbf{g} is a new output gate that combines \mathbf{g}_1 and \mathbf{g}_2 .

In a circuit that has more than two output gates it is also natural to state the following question. We will see that this very similar question has different taste.

MCSAT(\mathbf{A})

given a circuit over \mathbf{A} with output gates $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_k$ is there a valuation of input gates \bar{x} that gives the same output on $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_k$, i.e. $\mathbf{g}_1(\bar{x}) = \mathbf{g}_2(\bar{x}) = \dots = \mathbf{g}_k(\bar{x})$.

From algebraic point of view problem **CSAT(\mathbf{A})** asks for the solutions of an equation over \mathbf{A} . The problem **MCSAT(\mathbf{A})** asks for solutions of a special system of equations over \mathbf{A} . But we can also ask for solutions of arbitrary systems of equations. This however has a more natural wording in purely algebraic terms.

SCSAT(\mathbf{A})

Given polynomials

$$\mathbf{g}_1(\bar{x}), \mathbf{h}_1(\bar{x}), \dots, \mathbf{g}_k(\bar{x}), \mathbf{h}_k(\bar{x})$$

of an algebra \mathbf{A} , is there a valuation of the variables \bar{x} in A such that

$$\begin{aligned} \mathbf{g}_1(x_1, \dots, x_n) &= \mathbf{h}_1(x_1, \dots, x_n) \\ &\vdots \\ \mathbf{g}_k(x_1, \dots, x_n) &= \mathbf{h}_k(x_1, \dots, x_n), \end{aligned}$$

With this natural approach via multi valued circuits also the problem **TAUTOLOGY** has its natural generalization:

CEQV(\mathbf{A})

given a circuit over \mathbf{A} is it true that for all inputs \bar{x} we have the same values on given two output gates $\mathbf{g}_1, \mathbf{g}_2$, i.e. $\mathbf{g}_1(\bar{x}) = \mathbf{g}_2(\bar{x})$.

In the algebraic setting this is simply the question of equivalence of two terms or polynomials. Here equivalence of k pairs of terms/polynomials reduces to k independent **CEQV** queries.

In Boolean realm the problem **CEQV** can be treated as the complement of **CSAT** and therefore is **co-NP**-complete. In general the closely related problem **CEQV(\mathbf{A})** is somehow independent from **CSAT(\mathbf{A})**. This independence means that all four possibilities of tractability/intractability can be witnessed by some finite algebras. For example for the 2-element lattice \mathbf{L} the problem **CSAT(\mathbf{L})** is in **P** while **CEQV(\mathbf{L})** is **co-NP**-complete. An example

of a finite semigroup \mathbf{S} with $\text{CEQV}(\mathbf{S}) \in \text{P}$ and $\text{CSAT}(\mathbf{S})$ being NP-complete can be inferred from [28].

It is worth to note that solving equations (or systems of equations) is one of the oldest and well known mathematical problems which for centuries was the driving force of research in algebra. Let us only mention Galois theory, Gaussian elimination or Diophantine Equations.

In the decision version of these problems one asks if an equation (or system of such equations) expressed in the language of a fixed algebra \mathbf{A} , has a solution in \mathbf{A} . In fact, for \mathbf{A} being the ring of integers this is the famous 10th Hilbert Problem on Diophantine Equations, which has been shown to be undecidable [32]. In finite realms such problems are obviously decidable in nondeterministic polynomial time. There are numerous results related to problems connected with solving equations and systems of equations over fixed finite algebras. Most of them concerns well known algebraic structures as groups [6], [11], [20], [22] rings [18], [7] or lattices [36] but there are also some more general results [1], [31].

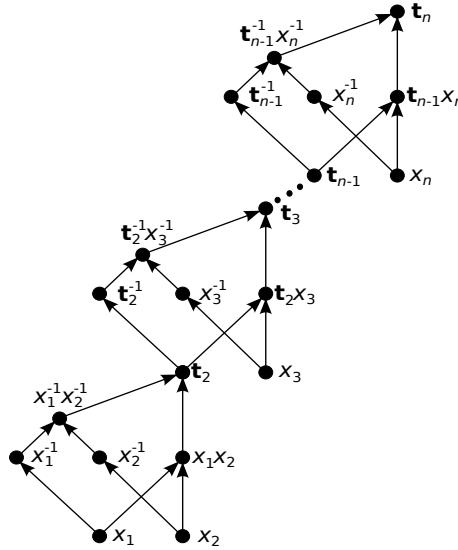
The main goal of this paper is to attack the classification problems of the form: for which finite algebras \mathbf{A} there is an algorithm that answers one of the problems $\text{CSAT}(\mathbf{A})$, $\text{MCSAT}(\mathbf{A})$, $\text{SCSAT}(\mathbf{A})$ or $\text{CEQV}(\mathbf{A})$ in polynomial time with respect to the size of the circuit, i.e. the size of the underlying graph of the circuit. It seems that the most natural way to look at these problems is to treat circuits over \mathbf{A} (or in fact output gates of such circuits) as terms/polynomials of the algebra \mathbf{A} . This obvious translation makes our attack fruitful, as we can apply deep results and techniques developed by universal algebra such as *modular commutator theory* and *tame congruence theory*. These tools are especially useful in case of algebras generating congruence modular variety. This assumption covers many well known structures as groups, rings, modules or lattices. Our attempt to attack the classification problems has resulted in partial characterization of computational complexity of CSAT , MCSAT and CEQV for algebras generating congruence modular varieties. This partial characterization leaves some room to be filled before establishing a dichotomy.

2. THE RESULTS

In this section we present the state of the art in more details and discuss our results and tools.

The first thing in which our research differs from what has been already considered is that we concentrate on circuits rather than on syntactic form of terms or polynomials. This difference is visible in how the size of the input is measured. We have seen how an output gate can be treated as a term or a polynomial. On the other hand, in an obvious way, every term over \mathbf{A} can be treated as a circuit in which each gate is used as an input to at most one other gate. This leads to a circuit whose underlying graph is a tree. However circuits can have more compact representation than terms. For

example, in groups the terms $\mathbf{t}_n(x_1, x_2, \dots, x_n) = [\dots [[x_1, x_2], x_3] \dots x_n]$, (where $[x, y] = x^{-1}y^{-1}xy$ is the group commutator) expressed in the pure group language of $(\cdot, {}^{-1})$ have an exponential size in n , as the number of occurrences of variables doubles whenever we pass from n to $n + 1$. On the other hand the size of a circuit realizing \mathbf{t}_n has $6n - 5$ vertices as can be seen from the picture below.



The consequences of this (exponential) disproportion in measuring the input size for terms and circuits are illustrated by the following example.

Example 2.1. *There are finite groups \mathbf{A} such that $\text{CSAT}(\mathbf{A})$ is NP-complete, while there are polynomial time algorithms for solving equations over \mathbf{A} .*

There are also finite groups \mathbf{B} such that $\text{CEQV}(\mathbf{A})$ is co-NP-complete, while there are polynomial time algorithms for checking the identities in \mathbf{B} .

Proof. The first such example for CSAT was the symmetric group \mathbf{S}_3 for which polynomial time algorithm was shown in [21], while the first author’s observation on the NP-completeness is included in [12].

The papers [19, 22, 23] contain many other examples of solvable non-nilpotent groups which witness both statements in our example. □

Note that in case of SCSAT there is no such disproportion in the size as every polynomial equation $z = \mathbf{t}(\bar{x})$ can be replaced by a system of equations of the form $y = \mathbf{f}(x_1, \dots, x_k)$ or $y = c$, where \mathbf{f} is one of the basic operations and c is a constant. This replacement has linear size with respect to the circuit representing $\mathbf{t}(\bar{x})$. For example for the above term $\mathbf{t}_n(x_1, x_2, \dots, x_n) = [\dots [[x_1, x_2], x_3] \dots x_n]$, slightly abusing our conditions,

we can use the following representation

$$\begin{aligned} t_2 &= x_1^{-1}x_2^{-1}x_1x_2 \\ t_3 &= t_2^{-1}x_3^{-1}t_2x_3 \\ &\vdots \\ t_n &= t_{n-1}^{-1}x_n^{-1}t_{n-1}x_n, \end{aligned}$$

in which t_2, \dots, t_n are treated as variables.

However, even in the setting of a single equation, representing a polynomial $\mathbf{t}(\bar{x})$ by its corresponding circuit and looking at the size of this circuit (instead of the syntactic length of \mathbf{t}) allows us to harmlessly expand the original language of the algebra \mathbf{A} by finitely many polynomials. In fact in our intractability proofs we will often expand the language of the original algebra \mathbf{A} by finitely many polynomials of \mathbf{A} . This will allow us to code NP-complete problems in much more smooth way. Note that the possibility of such expansions show that the characterizations we are looking for can be done up to polynomial equivalence of algebras. Two algebras are polynomially equivalent if they have the same universes and each polynomial of one of them can be defined by composing the polynomials of the other one.

It turns out that quite a few results on the complexity of the problems CSAT, MCSAT, SCSAT and CEQV are already known for particular kinds of (finite) algebras.

Example 2.2. *Finite Groups:*

- If \mathbf{A} is Abelian then $\text{SCSAT}(\mathbf{A}) \in P$ (by Gaussian elimination), and for all other groups $\text{SCSAT}(\mathbf{A})$ is NP-complete [11].
- $\text{CSAT}(\mathbf{A})$ is in P , whenever \mathbf{A} is nilpotent [11] and NP-complete otherwise [11, 22].
- $\text{CEQV}(\mathbf{A})$ is in P , whenever \mathbf{A} is nilpotent [6] and co-NP-complete otherwise [20, 22].

Example 2.3. *Finite Rings:*

- If \mathbf{A} is essentially an Abelian group (i.e. satisfies the identity $xy = 0$) then $\text{SCSAT}(\mathbf{A}) \in P$ (by Gaussian elimination), and for all other rings $\text{SCSAT}(\mathbf{A})$ is NP-complete [31].
- $\text{CSAT}(\mathbf{A})$ is in P , whenever \mathbf{A} is nilpotent [18] and NP-complete otherwise [7].
- $\text{CEQV}(\mathbf{A})$ is in P , whenever \mathbf{A} is nilpotent and NP-complete otherwise (see [24] for commutative rings and [7] for general case).

Example 2.4. *Finite Lattices:*

- $\text{CSAT}(\mathbf{A}) \in P$ if \mathbf{A} is distributive and NP-complete otherwise [36].
- For all nontrivial lattices \mathbf{A} , $\text{SCSAT}(\mathbf{A})$ is NP-complete while $\text{CEQV}(\mathbf{A})$ is co-NP-complete (easy to see).

The examples given above suggest that the existence of polynomial time algorithms for the considered circuits problems go hand in hand with nice structure theory of the underlying algebras. However there are only two results that can be considered general enough to be expressed in structural terms. These results are stated in the following two theorems.

First note that E. Aichinger and N. Mudrinski [1] have shown the following theorem, a partial converse of which is our Theorem 2.11.

Theorem 2.5. *If \mathbf{A} is a finite supernilpotent algebra from a congruence variety then $\text{CEQV}(\mathbf{A})$ is in \mathcal{P} .*

The second general result is that of B. Larose and L. Zádori [31]. After observing that SCSAT has exactly the same expressive power as CSP they used mutual translation between SCSAT and CSP to prove the first part of the next characterization, while the second one is a form of Gaussian elimination.

Theorem 2.6. *For a finite algebra \mathbf{A} from a congruence modular variety:*

- *if $\text{SCSAT}(\mathbf{A})$ is not NP -complete then \mathbf{A} is affine (i.e. \mathbf{A} is polynomially equivalent to a module over a finite ring),*
- *if \mathbf{A} is affine then $\text{SCSAT}(\mathbf{A}) \in \mathcal{P}$.*

Not as much is known when one leaves the congruence modularity realm. It is worth to note however that an important extension of Theorem 2.6 to finite algebras from varieties omitting $\mathbf{1}$ (in the sense of [17]) can be found in [39].

Also a number of results on semigroups do not fall in congruence modular setting but these results are still about particular type of algebras. The paper [29] gives a nice, but somewhat technical, characterization of finite monoids \mathbf{A} for which $\text{SCSAT}(\mathbf{A}) \in \mathcal{P}$. There are also several results on the complexity of $\text{SCSAT}(\mathbf{A})$ for particular semigroups or classes of semigroups, but we are far from having a full characterization similar to that for monoids. This is because the paper [29] contains a proof that the expressive power of $\text{SCSAT}(\mathbf{A})$ over semigroups is equivalent to the expressive power of CSP . Surprisingly another class of algebras with the same expressive power is the class of algebras with unary operations only [4, 9].

In Section 4 we will prove that the expressive power of CSAT is no weaker than this of CSP , as expressed below.

Proposition 2.7. *For every finite relational structure \mathbb{D} (with finitely many relations) there is a finite algebra $\mathbf{A}[\mathbb{D}]$ such that the problem $\text{CSP}(\mathbb{D})$ is polynomially equivalent to $\text{CSAT}(\mathbf{A}[\mathbb{D}])$.*

Unlike in the SCSAT setting we do not know whether the expressive power of CSAT is not bigger than the one of CSP .

Problem 1. *Is it true that for every finite algebra \mathbf{A} there exists a relational structure $\mathbb{D}[\mathbf{A}]$ such that the problems $\text{CSAT}(\mathbf{A})$ and $\text{CSP}(\mathbb{D}[\mathbf{A}])$ are polynomially equivalent?*

The above difference between a single equation and a system of equations is probably a consequence of the presence of an external conjunction in systems of equations. Intuitively, to replace a system of equations by a single equation, one needs to squeeze many terms (or polynomials) into a single one. This requires an analogue of an internal conjunction present in Boolean algebras. Since such a squeeze is not always possible, more algebras may have polynomial time algorithms for CSAT than for SCSAT. Actually our work is going to confirm this claim.

One of the main difficulties in characterizing finite algebras with $\text{SCSAT}(\mathbf{A}) \in \mathbf{P}$ is that this property does not carry over quotient algebras (unless $\mathbf{P} = \text{NP}$). The paper [29] contains an example of a finite semigroup \mathbf{A} and its congruence θ with $\text{SCSAT}(\mathbf{A}/\theta)$ being NP-complete while $\text{SCSAT}(\mathbf{A}) \in \mathbf{P}$. The example below (an easy proof of which is postponed to Section 4) shows that this unwanted phenomena occurs for the CSAT problem, as well.

Example 2.8. *There is a finite algebra \mathbf{A} and its congruence θ such that $\text{CSAT}(\mathbf{A}) \in \mathbf{P}$ while $\text{CSAT}(\mathbf{A}/\theta)$ is NP-complete.*

Since passing to quotient algebras may not preserve polynomial time complexity for CSAT, it is natural to work under the stronger assumption that not only $\text{CSAT}(\mathbf{A}) \in \mathbf{P}$, but $\text{CSAT}(\mathbf{A}/\theta) \in \mathbf{P}$ for all congruences θ of \mathbf{A} . Such assumption has also a natural interpretation. Given \mathbf{A} we want a fast method to solve equations over \mathbf{A} , or at least decide if such equations have solutions. However such solutions may not exist in the original algebra \mathbf{A} . They obviously do exist in $\mathbf{A}/1_{\mathbf{A}}$, where $1_{\mathbf{A}}$ is the congruence collapsing everything. Thus the best we can do, is to determine (existence of) the solutions with best possible precision, i.e. modulo the smallest congruences possible. This however requires \mathbf{A} to be regular enough so that $\text{CSAT}(\mathbf{A}')$ is in \mathbf{P} for all quotients \mathbf{A}' of \mathbf{A} .

After fixing the setting we are working in, we can state our main result in the next theorem which in fact summarizes Theorems 9.1 and 9.2.

Theorem 2.9. *Let \mathbf{A} be a finite algebra from a congruence modular variety.*

- (1) *If \mathbf{A} has no quotient \mathbf{A}' with $\text{CSAT}(\mathbf{A}')$ being NP-complete then \mathbf{A} is isomorphic to a direct product $\mathbf{N} \times \mathbf{D}$, where \mathbf{N} is a nilpotent algebra and \mathbf{D} is a subdirect product of 2-element algebras each of which is polynomially equivalent to the 2-element lattice.*
- (2) *If \mathbf{A} decomposes into a direct product $\mathbf{N} \times \mathbf{D}$, where \mathbf{N} is a supernilpotent algebra and \mathbf{D} is a subdirect product of 2-element algebras each of which is polynomially equivalent to the 2-element lattice, then for every quotient \mathbf{A}' of \mathbf{A} the problem $\text{CSAT}(\mathbf{A}')$ is solvable in polynomial time.*

To understand the above result first note that the congruence modularity assumption covers most algebraic structures considered in classical mathematics. In particular it includes groups (and their extensions like rings,

fields), and lattices (and their extensions like Boolean algebras or other algebras connected with multi-valued logics). This assumption does not cover however semigroups (or even semilattices) or multiunary algebras.

The conditions (1) and (2) show that the nilpotent groups and rings as well as distributive lattices mentioned in Examples 2.2, 2.3 and 2.4 are in fact paradigms for CSAT tractability in congruence modular realm. In fact the structural conditions described in Theorem 2.9, when specialized to groups, rings or lattices, gives the already known characterizations presented in the Examples.

The decomposition enforced in (1) is a result of almost a dozen constructions interpreting NP-complete problems (mostly SAT and k -COLORABILITY) into CSAT(\mathbf{A}), whenever \mathbf{A} , or some of its quotients, fails to satisfy one of the structural conditions that finally lead to this nice decomposition.

The second factor, \mathbf{D} , of this decomposition is easier to understand than the first one. It essentially behaves like a finite distributive lattice, but the algebra \mathbf{D} does not need to actually have lattice operations. Instead \mathbf{D} is composed of 2-element algebras each of which does have lattice operations expressible by polynomials (and all other operations monotone with respect to this lattice order).

The first factor, \mathbf{N} , of this decomposition requires the general algebraic notion of nilpotency in congruence modular setting that goes back to the late 1970's when Smith [37], Hagemann and Herrmann [16], Gumm [15] and finally Freese and McKenzie [10] developed necessary deep tools of *modular commutator theory*. In fact a notion of the commutator multiplication $[\alpha, \beta]$ of congruences α, β of arbitrary algebras was defined in a way that extends multiplication of ideals in ring theory and commutator multiplication of normal subgroups in group theory. With the help of such commutator one can define solvable and nilpotent congruences and algebras.

Finite nilpotent groups (and rings) behave very nicely. In particular they decompose into direct products of groups (or rings) of prime power order. Unfortunately such nice decomposition of nilpotent algebras in congruence modular varieties does not hold in general. However, in this general setting, nilpotent algebras that have this nice decomposition (and have only finitely many basic operations) are exactly those that are supernilpotent. In fact supernilpotency has been introduced by another universal algebraic generalization of commutator multiplication of congruences.

The nilpotent/supernilpotent gap that occurs in Theorem 2.9 resists to be easily filled. This is because in supernilpotent case there is a bound on the arity of the so called commutator polynomials. These commutator polynomials can imitate the behavior of the long conjunction. In nilpotent (but not supernilpotent) case arbitrary long conjunctions are expressible. But this can be probably done at the expense of exponentially large (with respect to the arity) circuits needed to represent those conjunctions. This expected exponential size probably prevents polynomial time reduction of NP-complete problems to CSAT in nilpotent but not supernilpotent case.

The reductions we have produced to show intractability of the considered problems are based on the local behavior described by the second deep tool of universal algebra known as *tame congruence theory*. This theory, created and described by D. Hobby and R. McKenzie in [17], is a perfect tool for studying the local structure of finite algebras. Instead of considering the whole algebra and all of its operations at once, tame congruence theory allows us to localize to small subsets on which the structure is much simpler to understand and to handle. According to this theory there are only five possible ways a finite algebra can behave locally. The local behavior must be one of the following:

1. a finite set with a group action on it,
2. a finite vector space over a finite field,
3. a two element Boolean algebra,
4. a two element lattice,
5. a two element semilattice.

Now, if from our point of view a local behavior of an algebra is ‘bad’ then we can often show that the algebra itself behaves ‘badly’. For example, since CSAT or CEQV is intractable in 2-element Boolean algebra one can argue that in any finite algebra with tractable CSAT or CEQV type **3** cannot occur (see Theorem 5.1).

On the other hand it is not true that if the local behavior is ‘good’ then the global one is good as well. Several kinds of interactions between these small sets can produce a fairly messy global behavior. Such interactions often contribute to NP-completeness of the considered problems (see for example Lemma 7.1). Also the relative ‘geographical layout’ of those small sets can result in unpredictable phenomena, as in Theorems 6.3 and 6.4.

Combining Theorems 2.9 and 2.6 we are able to infer the following corollary.

Corollary 2.10. *Let \mathbf{A} be a finite algebra from a congruence modular variety.*

- (1) *If \mathbf{A} has no quotient \mathbf{A}' with $\text{MCSAT}(\mathbf{A}')$ being NP-complete then \mathbf{A} is isomorphic to a direct product $\mathbf{M} \times \mathbf{D}$, where \mathbf{M} is an affine algebra and \mathbf{D} is a subdirect product of 2-element algebras each of which is polynomially equivalent to the 2-element lattice.*
- (2) *If \mathbf{A} decomposes into a direct product $\mathbf{M} \times \mathbf{D}$, where \mathbf{M} is an affine algebra and \mathbf{D} is a subdirect product of 2-element algebras each of which is polynomially equivalent to the 2-element lattice, then for every quotient \mathbf{A}' of \mathbf{A} the problem $\text{CSAT}(\mathbf{A}')$ is solvable in polynomial time.*

Our constructions used to show that lack of nice structure of the algebra \mathbf{A} leads to intractability of $\text{CSAT}(\mathbf{A})$ can be also modified to work for intractability of $\text{CEQV}(\mathbf{A})$ so that we are able to prove a partial converse to Theorem 2.5.

Theorem 2.11. *Let \mathbf{A} be a finite algebra from a congruence modular variety. If \mathbf{A} has no quotient \mathbf{A}' with $\text{CEQV}(\mathbf{A}')$ being co-NP-complete then \mathbf{A} is nilpotent.*

A short informal summary of these results is completed in the following table, where ‘DL-like’ stays for being a subdirect product of algebras polynomially equivalent to 2-element lattices.

	tractable (polynomial time)	open	intractable (co-NP- or NP-complete)
CEQV	supernilpotent Aichinger & Mudrinski [1]	nilpotent but not supernilpotent	non nilpotent Thm 2.11
CSAT	supernilpotent \times DL-like Thm 2.9 (2)	nilpotent but not supernilpotent	non (nilpotent \times DL-like) Thm 2.9 (1)
MCSAT	affine \times DL-like Cor 2.10 (2)	—	otherwise Cor 2.10 (1)
SCSAT	affine Gaussian elimination	—	otherwise Larose & Zádori [31]

An obvious open question is the following:

Problem 2. *Determine the computational complexity of CEQV and CSAT for nilpotent, but not supernilpotent finite algebras from congruence modular varieties.*

Another question that arises naturally is the role of quotient algebras in the proofs of NP-completeness of considered problems. Note that the result of B. Larose and L. Zádori [31] for SCSAT makes no use of quotient algebras. This is because a quotient of an affine algebra is affine itself.

Example 2.8 shows that in general it is not enough to establish NP-completeness for a quotient algebra to conclude it for the original one. However it may suffice in some more restricted setting like for example congruence modularity. In concrete algebraic structures where basic operations are described explicitly it might be much easier. In fact in structures described in Examples 2.2, 2.3 and 2.4, passing to quotients is hidden in the hardness proofs and (implicitly) replaced by an involved control over congruences in groups, rings or lattices, respectively.

Problem 3. *Is it true that NP-completeness of CSAT for some quotient of a finite algebra \mathbf{A} from a congruence modular variety implies NP-completeness of CSAT for \mathbf{A} itself.*

Even if the answer to Problem 3 would be negative the next one remains open.

Problem 4. *Do the characterizations of Theorems 2.9 (1), 2.11 and Corollary 2.10 (1) remain true without passing to quotient algebras.*

Note here that when restricting to equations of the form

$$\mathbf{t}(x_1, \dots, x_n) = c$$

where \mathbf{t} is a polynomial but c is a constant, the satisfiability in the quotient \mathbf{A}/θ reduces to the satisfiability of at least one of the equations in the following disjunction

$$\mathbf{t}(x_1, \dots, x_n) = c_1 \vee \dots \vee \mathbf{t}(x_1, \dots, x_n) = c_s,$$

where $\{c_1, \dots, c_s\}$ is the equivalence class of c modulo θ . This Cook style reduction gives the hope to attack the following problem.

Problem 5. *Characterize finite algebras \mathbf{A} for which determining the existence of a solution to the equations of the form $\mathbf{t}(\bar{x}) = c$ can be done in polynomial time.*

In view of Problem 1, the natural conjecture about dichotomy for CSAT is not so evident. However there is a slightly bigger hope for such dichotomy after:

- restricting CSAT to the equations of the form $\mathbf{t}(\bar{x}) = c$, and
- relaxing many-to-one reductions to Cook reductions.

Problem 6. *Prove the dichotomy in the above settings.*

3. BACKGROUND MATERIAL

In general we use the terminology and notation of [33]. Our brief introduction to this terminology, notation and the facts that we are using in this paper is modelled after that in [3].

An algebra $\mathbf{A} = \langle A, \mathbf{f}_i(i \in I) \rangle$ is a nonvoid set A together with a collection of finitary operations \mathbf{f}_i on A indexed by a set I . The set A is called the *universe* of the algebra and the \mathbf{f}_i 's are the *fundamental operations* of \mathbf{A} . For $i \in I$ the operation \mathbf{f}_i maps A^{n_i} to A , that is, \mathbf{f}_i is n_i -ary. The function from I to the integers given by $i \mapsto n_i$ is the *similarity type* of the algebra \mathbf{A} . If I is finite, then the algebra is said to be of *finite similarity type*. For algebras of finite similarity type we often just list the operations, e.g., a Boolean algebra might be given as $\mathbf{B} = \langle B, \wedge, \vee, \neg, 0, 1 \rangle$. An algebra is *finite* if its universe is finite and is *trivial* if its universe has only one element.

An algebra $\mathbf{A} = \langle A, \mathbf{f}_i(i \in I) \rangle$ may also be viewed as a model in the language L where L consists of all the function symbols \mathbf{f}_i for $i \in I$. When necessary, we distinguish the function symbol \mathbf{f}_i in L from the fundamental operation \mathbf{f}_i on A by writing $\mathbf{f}_i^{\mathbf{A}}$ to denote the n_i -ary operation on the algebra \mathbf{A} . A *term* for L over a set of variables $X = \{x_1, x_2, \dots\}$ is defined inductively by letting every $x_j \in X$ be a term and if $i \in I$ and $\mathbf{t}_1, \dots, \mathbf{t}_{n_i}$ are terms, then $\mathbf{f}_i(\mathbf{t}_1, \dots, \mathbf{t}_{n_i})$ is also a term. If the variables that appear in a term \mathbf{t} are in the set $\{x_1, \dots, x_n\}$, then we say \mathbf{t} is n -ary and denote

this by writing $\mathbf{t}(x_1, \dots, x_n)$. If $\mathbf{t}(x_1, \dots, x_n)$ is an n -ary term for L over X and \mathbf{A} is an algebra in the language L , then the *term operation* $\mathbf{t}^{\mathbf{A}}$ on \mathbf{A} corresponding to \mathbf{t} is defined by letting $x_i^{\mathbf{A}}$ be the projection on the i -th coordinate and if

$$\mathbf{t}(x_1, \dots, x_n) = \mathbf{f}_i(\mathbf{t}_1(x_1, \dots, x_n), \dots, \mathbf{t}_{n_i}(x_1, \dots, x_n)),$$

then

$$\mathbf{t}^{\mathbf{A}}(a_1, \dots, a_n) = \mathbf{f}_i^{\mathbf{A}}(\mathbf{t}_1^{\mathbf{A}}(a_1, \dots, a_n), \dots, \mathbf{t}_{n_i}^{\mathbf{A}}(a_1, \dots, a_n))$$

for all $(a_1, \dots, a_n) \in A^n$. To simplify notation we often suppress the subscript on fundamental operations and just write \mathbf{f} or $\mathbf{f}(x_1, \dots, x_r)$. Likewise, we often omit the algebra superscript on term operations. We also use the bar convention by writing \bar{a} for (a_1, \dots, a_n) .

The collection of all term operations on an algebra forms a *clone*, that is, a family of operations on a set that contains all the projection operations and is closed under composition. Thus the set of term operations of \mathbf{A} is called the *clone of \mathbf{A}* and is denoted $\text{Clo } \mathbf{A}$. The clone of all operations on A that can be obtained from the term operations of \mathbf{A} and all the constant operations is called the *clone of polynomial operations* of \mathbf{A} and is denoted $\text{Pol } \mathbf{A}$. The set of n -ary polynomial operations of \mathbf{A} is written $\text{Pol}_n \mathbf{A}$. Two algebras \mathbf{A} and \mathbf{B} are said to be *polynomially equivalent* if they have the same universe and $\text{Pol } \mathbf{A} = \text{Pol } \mathbf{B}$. A unary polynomial $\mathbf{e} \in \text{Pol}_1 \mathbf{A}$ is said to be idempotent if $\mathbf{e}(\mathbf{e}(a)) = \mathbf{e}(a)$ for all $a \in A$.

A *subuniverse* of an algebra \mathbf{A} is a set $S \subseteq A$ that is closed under the fundamental operations of \mathbf{A} , that is, $\mathbf{f}(\bar{a}) \in S$ for every fundamental operation \mathbf{f} of \mathbf{A} and every $\bar{a} \in S^r$. An algebra \mathbf{B} is a *subalgebra* of \mathbf{A} if \mathbf{B} and \mathbf{A} have the same similarity type, the universe of \mathbf{B} is a subuniverse of \mathbf{A} , and for every operation symbol \mathbf{f}_i , the operation $\mathbf{f}_i^{\mathbf{B}}$ is the restriction to B of the operation $\mathbf{f}_i^{\mathbf{A}}$. Since the intersection of an arbitrary family of subuniverses of an algebra \mathbf{A} is a subuniverse it follows that the set of all subuniverses of \mathbf{A} , denoted $\text{Sub } \mathbf{A}$, forms a complete lattice when ordered by inclusion. For a subset X of A , the *subuniverse of \mathbf{A} generated by X* , denoted $\text{Sg}^{\mathbf{A}}(X)$, is the intersection of all subuniverses of \mathbf{A} that contain X . Another way to describe $\text{Sg}^{\mathbf{A}}(X)$ is to observe that the subuniverse generated by X consists of all elements of the form $\mathbf{t}^{\mathbf{A}}(\bar{x})$ where \mathbf{t} ranges over all terms for the language of \mathbf{A} and the \bar{x} are tuples from X .

Given two algebras \mathbf{A} and \mathbf{B} of the same similarity type, a function $h : A \rightarrow B$ is called a *homomorphism* if $h(\mathbf{f}(a_1, \dots, a_r)) = \mathbf{f}(h(a_1), \dots, h(a_r))$ for every fundamental operation \mathbf{f} of \mathbf{A} and all $a_i \in A$. A homomorphism is an *isomorphism* if it is both one-to-one and onto. If h is a homomorphism from \mathbf{A} to \mathbf{B} , then $h(A)$ is a subuniverse of \mathbf{B} and if $A = \text{Sg}^{\mathbf{A}}(X)$, then the subuniverse $h(A)$ is generated by $h(X)$.

A congruence relation on an algebra \mathbf{A} is an equivalence relation θ on A that is preserved by the fundamental operations of \mathbf{A} , that is, if f is an r -ary fundamental operation and $(a_1, b_1), \dots, (a_r, b_r) \in \theta$, then $(\mathbf{f}(\bar{a}), \mathbf{f}(\bar{b})) \in \theta$. Notation that is often used to express that (a, b) is in the congruence relation

θ includes $a\theta b$ and $a \stackrel{\theta}{\equiv} b$. For a congruence relation θ on \mathbf{A} the congruence class containing an element a is denoted a/θ and A/θ is the set of all congruence classes of θ . The intersection of a family of congruence relations of an algebra is again a congruence relation so the set of all congruence relations of \mathbf{A} , when ordered by inclusion, forms a complete lattice. The lattice of congruence relations of \mathbf{A} is denoted $\text{Con } \mathbf{A}$. The top element of this lattice is $A \times A$ and is written as 1_A ; the bottom element is the diagonal 0_A , which consists of all pairs (a, a) for $a \in A$. We frequently omit the subscripts in 0_A and 1_A .

For a set $Z \subseteq A \times A$ the *congruence relation on \mathbf{A} generated by Z* is the intersection of all $\theta \in \text{Con } \mathbf{A}$ for which $Z \subseteq \theta$. We write $\text{Cg}^{\mathbf{A}}(Z)$ for this congruence relation but in the case that $Z = \{(a, b)\}$ we write $\text{Cg}^{\mathbf{A}}(a, b)$. Like in the case of subuniverses $\text{Sg}^{\mathbf{A}}(X)$ there is an intrinsic way to describe the congruence $\text{Cg}^{\mathbf{A}}(Z)$.

Lemma 3.1. *Suppose that in an algebra \mathbf{A} we have $(a, b) \in \text{Cg}^{\mathbf{A}}(Z)$ for some $Z \subseteq A^2$. Then*

- *there is a natural number n , a sequence $(y_1, z_1), \dots, (y_n, z_n)$ of pairs in Z , a sequence of unary polynomials $\mathbf{p}_1, \dots, \mathbf{p}_n$ of \mathbf{A} and a sequence x_0, \dots, x_n of elements of \mathbf{A} such that*

$$a = x_0, \quad x_n = b \quad \text{and} \\ \{x_{i-1}, x_i\} = \{\mathbf{p}_i(y_i), \mathbf{p}_i(z_i)\} \text{ for all } 1 \leq i \leq n,$$

- *if additionally \mathbf{A} is finite and has a ternary polynomial \mathbf{d} that behaves like a Malcev operation on a subset $B \subseteq A$ (i.e., $\mathbf{d}(x, x, y) = y = \mathbf{d}(y, x, x)$ for all $x, y \in B$) which is the range of a unary idempotent polynomial \mathbf{e}_B of \mathbf{A} , then for $a, b \in B$ and $Z = \{(c, d)\}$ there is a single unary polynomial \mathbf{p} with $\mathbf{p}(c) = a$ and $\mathbf{p}(d) = b$.*

Proof. The first item is due to Malcev. The second item is also a part of folklore, but we will include its proof for the reader convenience.

To see the second item note that from the first one we know that \mathbf{A} has the unary polynomials $\mathbf{p}_1, \dots, \mathbf{p}_n$ and elements x_0, \dots, x_n such that

$$a = x_0, b = x_n \text{ and } \{x_{i-1}, x_i\} = \{\mathbf{p}_i(c), \mathbf{p}_i(d)\} \text{ for all } 1 \leq i \leq n$$

and applying \mathbf{e}_B to this chain we may assume that the ranges of the \mathbf{p}_i 's are contained in B , so that the entire chain of the x_i 's lives in B . First look at $\{x_{i-1}, x_i\} = \{\mathbf{p}_i(c), \mathbf{p}_i(d)\}$. If $(x_{i-1}, x_i) = (\mathbf{p}_i(d), \mathbf{p}_i(c))$, replace $\mathbf{p}_i(x)$ by $\mathbf{d}_B(\mathbf{p}_i(c), \mathbf{p}_i(x), \mathbf{p}_i(d))$, so that after this replacement we have $(x_{i-1}, x_i) = (\mathbf{p}_i(c), \mathbf{p}_i(d))$ for all i . Now we will show that if $n > 1$ such sequence can be shortened and this additional requirements are kept. Indeed, for $\mathbf{p}_{1,2}(x) = \mathbf{d}_B(\mathbf{p}_1(x), \mathbf{p}_1(d), \mathbf{p}_2(x))$ we have $(\mathbf{p}_{1,2}(c), \mathbf{p}_{1,2}(d)) = (x_0, x_2)$. \square

Some terminology from lattice theory is used in describing $\text{Con } \mathbf{A}$. For $a \leq b$ in a lattice \mathbf{L} the ordered pair (a, b) is called a *quotient* in \mathbf{L} and the *interval* from a to b , written $I[a, b]$, is the subuniverse of \mathbf{L} consisting

of $\{c \in L : a \leq c \leq b\}$. The element a is *covered* by b if $a < b$ and $I[a, b] = \{a, b\}$. If a is covered by b , then we write $a \prec b$ and call $I[a, b]$ a *prime interval* or a *prime quotient*. A *subcover* of an element b is any element covered by b . An *atom* in a lattice with least element 0 is any element that covers 0 and a *coatom* or *dual atom* in a lattice with largest element 1 is any element covered by 1. If $I[a, b]$ and $I[c, d]$ are intervals such that $b \wedge c = a$ and $b \vee c = d$, then $I[a, b]$ is said to *transpose up* to $I[c, d]$, written $I[a, b] \nearrow I[c, d]$; and $I[c, d]$ is said to *transpose down* to $I[a, b]$, written $I[c, d] \searrow I[a, b]$; and the two intervals are called *transposes* of one another. Two intervals are said to be *projective* if one can be obtained from the other by a finite sequence of transposes. A fundamental fact in lattice theory is that a lattice is modular if and only if its projective intervals are isomorphic. Another equivalent condition for modularity is that the lattice has no elements a, b, c satisfying $a < b, a \vee c = b \vee c$ and $a \wedge c = b \wedge c$. Such a 5-element sublattice generated by a, b, c will be called an $[a, b, c]$ -*pentagon*.

An algebra \mathbf{A} is *simple* if it is nontrivial and $\text{Con } \mathbf{A}$ consists solely of 1_A and 0_A . An algebra is called *congruence distributive* or *congruence modular* if its congruence lattice satisfies the distributive identity or the modular identity. Two congruence relations $\theta, \tau \in \text{Con } \mathbf{A}$ *permute* if $\theta \circ \tau = \tau \circ \theta$. If θ and τ permute, then $\theta \vee \tau = \theta \circ \tau$ in $\text{Con } \mathbf{A}$. An algebra is *congruence permutable* if every pair of its congruence relations permute.

Homomorphisms and congruence relations are naturally linked: If h is a homomorphism on \mathbf{A} , then the *kernel of h* , denoted $\ker(h)$, is the set of all $(a_1, a_2) \in A^2$ for which $h(a_1) = h(a_2)$. For every homomorphism h the relation $\ker(h)$ is a congruence on \mathbf{A} . On the other hand, if $\theta \in \text{Con } \mathbf{A}$, then the congruence classes of θ form the elements of an algebra \mathbf{A}/θ and the map $a \mapsto a/\theta$ is a homomorphism from \mathbf{A} onto \mathbf{A}/θ with kernel θ .

We next consider direct products of algebras. Suppose \mathbf{A}_j , for $j \in J$, are algebras of the same similarity type indexed by a set J . The *direct product* of these algebras, denoted $\prod_{j \in J} \mathbf{A}_j$, is an algebra of the same similarity type as the \mathbf{A}_j with universe $\prod_{j \in J} A_j$ and fundamental operations defined coordinatewise: $\mathbf{f}(\bar{a}, \bar{b}, \bar{c}, \dots)_j = \mathbf{f}(a_j, b_j, c_j, \dots)$ for all $j \in J$. Often the index set J is finite, say $J = \{1, \dots, n\}$, and we write $\mathbf{A}_1 \times \dots \times \mathbf{A}_n$ for the direct product in this situation. If J is the empty set, then $\prod_{j \in J} \mathbf{A}_j$ is a trivial algebra. A direct product of copies of a single algebra \mathbf{A} is called a *direct power* of \mathbf{A} . We write \mathbf{A}^J for a direct power of \mathbf{A} indexed by a set J and we often view the elements of this algebra as functions from J to A .

If $\mathbf{A} = \prod_{j \in J} \mathbf{A}_j$, then the j -th *projection* map π_j is a homomorphism of \mathbf{A} onto \mathbf{A}_j . The kernel of π_j is usually written as η_j and thus for $a, b \in A$ we have $(a, b) \in \eta_j$ if and only if $a(j) = b(j)$. The η_j are called *projection kernels*. It is easily checked that if J_1 and J_2 are nonvoid complementary subsets of J and $\alpha_i = \bigwedge_{j \in J_i} \eta_j$ for $i = 1, 2$, then in $\text{Con } \mathbf{A}$ we have:

- (1) the congruences α_1 and α_2 permute,

- (2) $\alpha_1 \vee \alpha_2 = 1_A$,
(3) $\alpha_1 \wedge \alpha_2 = 0_A$.

Conversely, if \mathbf{A} is any algebra and α_1 and α_2 are any two congruences for which these three conditions hold, then α_1 and α_2 are each called *factor congruences* of \mathbf{A} and $\mathbf{A} \simeq \mathbf{A}/\alpha_1 \times \mathbf{A}/\alpha_2$. An algebra is called *directly indecomposable* if it is nontrivial and is not isomorphic to the direct product of two nontrivial algebras. Every finite algebra is isomorphic to the direct product of directly indecomposable algebras but this is not necessarily the case for infinite algebras.

Certain subalgebras of a direct product called subdirect products play an important role in our work. An algebra \mathbf{A} is a *subdirect product* of the algebras \mathbf{A}_j , for $j \in J$, if \mathbf{A} is a subalgebra of $\prod_{j \in J} \mathbf{A}_j$ and for each $j \in J$ the projection map from \mathbf{A} to \mathbf{A}_j is onto. Thus, if \mathbf{A} is a subdirect product of the \mathbf{A}_j for $j \in J$ and γ_j is the kernel of the j -th projection homomorphism from \mathbf{A} to \mathbf{A}_j , then $\bigcap_{j \in J} \gamma_j = 0_A$ and each \mathbf{A}_j is isomorphic to \mathbf{A}/γ_j . Conversely, if a family of congruence relations, γ_j for $j \in J$, on an algebra \mathbf{A} has the property that $\bigcap_{j \in J} \gamma_j = 0_A$, then \mathbf{A} is isomorphic to an algebra that is the subdirect product of \mathbf{A}/γ_j for $j \in J$. A *subdirect representation* of \mathbf{A} with subdirect factors \mathbf{A}_j is a homomorphic embedding h of \mathbf{A} into $\prod_{j \in J} \mathbf{A}_j$ for which $h(A)$ is the universe of an algebra that is a subdirect product of the \mathbf{A}_j .

An algebra \mathbf{A} is *subdirectly irreducible* if it is nontrivial and in any subdirect representation of \mathbf{A} at least one of the projection maps is an isomorphism. We use the following internal characterization of a subdirectly irreducible algebra: An algebra \mathbf{A} is subdirectly irreducible if and only if there is a $\mu \in \text{Con } \mathbf{A}$ such that $0_A < \mu$ and $\mu \leq \theta$ for all $0_A < \theta \in \text{Con } \mathbf{A}$. The congruence relation μ is called the *monolith* of the subdirectly irreducible algebra \mathbf{A} . Thus, \mathbf{A} is subdirectly irreducible if and only if in the lattice $\text{Con } \mathbf{A}$ the element 0_A is strictly meet irreducible. A theorem of Birkhoff states that every algebra is a subdirect product of subdirectly irreducible algebras. This theorem is equivalent to the statement that in any algebra \mathbf{A} the congruence relation 0_A is the intersection of all strictly meet irreducible members of $\text{Con } \mathbf{A}$. If \mathbf{A} is a subdirectly irreducible algebra with monolith μ , then $\mu = \text{Cg}^{\mathbf{A}}(a, b)$ for every $(a, b) \in \mu - 0$.

The class of algebras is a *variety* if it is closed under taking homomorphic images, subalgebras and products of algebras. A classic preservation theorem of Birkhoff states that a class of algebras is a variety if and only if it is an equational class, i.e. a class consisting of all algebras satisfying a certain set of identities.

A variety is *congruence distributive*, *congruence modular*, or *congruence permutable* if every algebra in the variety is congruence distributive, congruence modular, or has permuting congruence relations, respectively. Obviously congruence distributive varieties are congruence modular. But also congruence permutable varieties are known to be congruence modular. A

classic result of Malcev states that a variety \mathcal{V} is congruence permutable if and only if \mathcal{V} has a ternary term \mathbf{d} for which $\mathcal{V} \models \mathbf{d}(x, x, y) \approx \mathbf{d}(y, x, x) \approx y$. Such a term is called a *Malcev term* for \mathcal{V} .

There are also several characterizations (due to B. Jónsson, A. Day or H.P. Gumm) of congruence modular or distributive varieties in terms of identities they have to satisfy. We will use one such characterization via so called *directed Gumm terms* which is described in [26].

Theorem 3.2. *A variety \mathcal{V} is congruence modular if and only if \mathcal{V} has ternary terms*

$$\mathbf{D}_1(x, y, z), \dots, \mathbf{D}_n(x, y, z), \mathbf{Q}(x, y, z)$$

satisfying the following equalities:

$$\begin{aligned} x &= \mathbf{D}_i(x, y, x), & \text{for all } i = 1, \dots, n, \\ x &= \mathbf{D}_1(x, x, y), \\ \mathbf{D}_i(x, y, y) &= \mathbf{D}_{i+1}(x, x, y), & \text{for all } i = 1, \dots, n-1, \\ \mathbf{D}_n(x, y, y) &= \mathbf{Q}(x, y, y), \\ & \mathbf{Q}(x, x, y) &= y. \end{aligned}$$

The material on universal algebra presented so far is “classical” and was all known by the mid 1960s. In our work we will require some deep results that have come out of two more recent developments: generalized commutator theory and tame congruence theory. We now present the basics of these two topics.

Fuller discussions of the generalized commutator may be found in [10], [33, Section 4.13] and [17, Chapter 3]. The main reference for tame congruence theory is [17].

We begin with the theory of the commutator. Let \mathbf{A} be an algebra, $\gamma \in \text{Con } \mathbf{A}$, and $R, S \subseteq A^2$. We say R *centralizes* S *modulo* γ , denoted $C(R, S; \gamma)$, if for every $n \geq 1$, every $(n+1)$ -ary term \mathbf{t} , every $(a, b) \in R$, and every $(c_1, d_1), \dots, (c_n, d_n) \in S$ we have

$$\mathbf{t}(a, \bar{c}) \stackrel{\gamma}{\cong} \mathbf{t}(a, \bar{d}) \quad \text{iff} \quad \mathbf{t}(b, \bar{c}) \stackrel{\gamma}{\cong} \mathbf{t}(b, \bar{d}).$$

The following facts are easily verified.

Proposition 3.3. *For binary relations that are congruence relations on \mathbf{A} :*

- (1) *If $\alpha' \subseteq \alpha$ and $\beta' \subseteq \beta$, then $C(\alpha, \beta; \gamma)$ implies $C(\alpha', \beta'; \gamma)$.*
- (2) *If $C(\alpha, \beta, \gamma_i)$ for all $i \in I$, then $C(\alpha, \beta; \bigcap_{i \in I} \gamma_i)$.*
- (3) *$C(\alpha, \beta; \alpha)$ and $C(\alpha, \beta; \beta)$.*
- (4) *If $C(\alpha_i, \beta, \gamma)$ for all $i \in I$, then $C(\bigvee_{i \in I} \alpha_i, \beta; \gamma)$.*
- (5) *If $\theta \subseteq \alpha, \beta, \gamma$ then $C(\alpha, \beta; \gamma)$ holds in \mathbf{A} iff $C(\alpha/\theta, \beta/\theta; \gamma/\theta)$ holds in the quotient \mathbf{A}/θ .*

Moreover, (1) and (2) hold for arbitrary binary relations $\alpha, \alpha', \beta, \beta'$, and (3) holds if α and β are binary relations that are preserved by the fundamental operations of \mathbf{A} .

An algebra \mathbf{A} is *Abelian*, or is said to satisfy the *term condition*, if $C(1_A, 1_A; 0_A)$ holds. Note that if $C(1_A, 1_A; \gamma)$, then \mathbf{A}/γ is Abelian.

If α and β are congruence relations on an algebra \mathbf{A} , then the *commutator* of α and β , denoted $[\alpha, \beta]$, is the least congruence γ for which $C(\alpha, \beta; \gamma)$. The *centralizer* of β modulo α , denoted $(\alpha : \beta)$, is the largest congruence δ for which $C(\delta, \beta; \alpha)$.

We will appeal, often without reference, to the following facts about the centralizer and the commutator:

Proposition 3.4. *For congruence relations in an arbitrary algebra \mathbf{A}*

- (1) $C(\alpha, \beta; \gamma)$ if and only if $\alpha \leq (\gamma : \beta)$,
- (2) $(\alpha : 0_A) = 1_A$,
- (3) $\alpha \leq (\alpha : \beta)$.

If \mathbf{A} belongs to a congruence modular variety then we additionally have (see [10])

- (4) $[\alpha, \beta] = [\beta, \alpha]$,
- (5) $C(\alpha, \beta; \gamma)$ if and only if $[\alpha, \beta] \leq \gamma$,
- (6) $[\alpha, \bigvee_i \beta_i] = \bigvee_i [\alpha, \beta_i]$,
- (7) $(\alpha : \bigvee_i \beta_i) = \bigwedge_i (\alpha : \beta_i)$,
- (8) $(\bigwedge_i \alpha_i : \beta) = \bigwedge_i (\alpha_i : \beta)$,
- (9) if the intervals $I[\alpha_1, \beta_1]$ and $I[\alpha_2, \beta_2]$ are projective in the lattice $\text{Con } \mathbf{A}$, then $(\alpha_1 : \beta_1) = (\alpha_2 : \beta_2)$.

A consequence of items (2) and (3) in Proposition 3.3 is that $[\alpha, \beta] \leq \alpha \cap \beta$ for all congruence relations in an arbitrary algebra, however, for algebras in congruence distributive varieties it is known that $[\alpha, \beta] = \alpha \cap \beta$, see e.g., [33, p. 258].

By means of the commutator it is possible to define notions of Abelian, solvable and nilpotence for arbitrary algebras. Let $\alpha \leq \beta$ be congruence relations of an algebra \mathbf{A} . The congruence relation β is *Abelian over α* if $C(\beta, \beta; \alpha)$ and β is *Abelian* if $C(\beta, \beta; 0_A)$. We say β is *solvable over α* if there exists a finite chain of congruence relations $\beta = \gamma_0 \geq \gamma_1 \geq \dots \geq \gamma_m = \alpha$ such that γ_i is Abelian over γ_{i+1} for all $i < m$. A congruence relation β is *solvable* if it is solvable over 0_A . An algebra \mathbf{A} is *solvable* if 1_A , and hence every congruence relation of \mathbf{A} , is solvable. An algebra \mathbf{A} is *locally solvable* if every finitely generated subalgebra of \mathbf{A} is solvable. It can be argued that in the congruence lattice of a finite algebra \mathbf{A} the join of all the solvable congruence relations is itself solvable. This largest solvable congruence relation is called the *solvable radical* of \mathbf{A} .

For a congruence θ and $i = 1, 2, \dots$ we write

$$\begin{aligned} \theta^{(1)} &= \theta & \theta^{[1]} &= \theta \\ \theta^{(i+1)} &= [\theta, \theta^{(i)}] & \theta^{[i+1]} &= [\theta^{[i]}, \theta^{[i]}]. \end{aligned}$$

A congruence relation θ on \mathbf{A} is called *k-step left nilpotent* if $\theta^{(k+1)} = 0_A$ and the algebra \mathbf{A} is *left nilpotent* if 1_A is *k-step left nilpotent* for some

finite k . In the congruence modular varieties we use the word nilpotent rather than left nilpotent. Note that θ is solvable if $\theta^{[k]} = 0_A$ for some k .

The following strengthening of the nilpotency is also relevant in our setting. First, for a bunch of congruences $\alpha_1, \dots, \alpha_k, \beta, \gamma \in \mathbf{Con} \mathbf{A}$ we say that $\alpha_1, \dots, \alpha_k$ centralize β modulo γ , and write $C(\alpha_1, \dots, \alpha_k, \beta; \gamma)$, if for all polynomials $\mathbf{f} \in \mathbf{Pol} \mathbf{A}$ and all tuples $\bar{a}_1 \stackrel{\alpha_1}{\equiv} \bar{b}_1, \dots, \bar{a}_k \stackrel{\alpha_k}{\equiv} \bar{b}_k$ and $\bar{u} \stackrel{\beta}{\equiv} \bar{v}$ such that

$$\mathbf{f}(\bar{x}_1, \dots, \bar{x}_k, \bar{u}) \stackrel{\gamma}{\equiv} \mathbf{f}(\bar{x}_1, \dots, \bar{x}_k, \bar{v})$$

for all possible choices of $(\bar{x}_1, \dots, \bar{x}_k)$ in $\{\bar{a}_1, \bar{b}_1\} \times \dots \times \{\bar{a}_k, \bar{b}_k\}$ but $(\bar{b}_1, \dots, \bar{b}_k)$, we also have

$$\mathbf{f}(\bar{b}_1, \dots, \bar{b}_k, \bar{u}) \stackrel{\gamma}{\equiv} \mathbf{f}(\bar{b}_1, \dots, \bar{b}_k, \bar{v}).$$

This notion was introduced by A. Bulatov [5] and further developed by E. Aichinger and N. Mudrinski [1]. In particular they have shown that for all $\alpha_1, \dots, \alpha_k \in \mathbf{Con} \mathbf{A}$ there is the smallest congruence γ with $C(\alpha_1, \dots, \alpha_k; \gamma)$ called the k -ary commutator and denoted by $[\alpha_1, \dots, \alpha_k]$. Such generalized commutator behaves especially well in algebras from congruence modular varieties. In particular this commutator is monotone, join-distributive and we have

$$[\alpha_1, [\alpha_2, \dots, \alpha_k]] \leq [\alpha_1, \dots, \alpha_k]$$

Thus every k -supernilpotent algebra, i.e. algebra satisfying $[\overbrace{1, \dots, 1}^{k+1 \text{ times}}] = 0$, is k -nilpotent. The following properties, that can be easily inferred from the deep work of R. Freese and R. McKenzie [10] and K. Kearnes [27], have been summarized in [1].

Theorem 3.5. *For a finite algebra \mathbf{A} from a congruence modular variety the following conditions are equivalent:*

- (1) \mathbf{A} is k -supernilpotent,
- (2) \mathbf{A} is k -nilpotent, decomposes into a direct product of algebras of prime power order and the clone $\mathbf{Clo} \mathbf{A}$ is generated by finitely many operations,
- (3) \mathbf{A} is k -nilpotent and all commutator polynomials have rank at most k .

The commutator polynomials mentioned in condition (3) of Theorem 3.5 are the paradigms for the failure of supernilpotency. We say that $\mathbf{t}(x_1, \dots, x_{k-1}, z) \in \mathbf{Pol}_k \mathbf{A}$ is a commutator polynomial of rank k if

- $\mathbf{t}(a_1, \dots, a_{k-1}, b) = b$ whenever $b \in \{a_1, \dots, a_{k-1}\} \subseteq A$,
- $\mathbf{t}(a_1, \dots, a_{k-1}, b) \neq b$ for some $a_1, \dots, a_{k-1}, b \in A$.

We next sketch the material on tame congruence theory that we will need.

For a nonvoid subset U of an algebra \mathbf{A} the *algebra induced by \mathbf{A} on U* is the algebra $\mathbf{A}|_U$ whose universe is U and whose fundamental operations are all polynomials $\mathbf{p} \in \mathbf{Pol}_m \mathbf{A}$ for which $\mathbf{p}|_{U^m}$ maps U^m into U . The algebra $\mathbf{A}|_U$ is nonindexed, that is, there is no index set specified for the

set of fundamental operations. Note that every polynomial operation of $\mathbf{A}|_U$ is its fundamental operation. Two nonvoid subsets U and V of \mathbf{A} are called *polynomially isomorphic* if there exist $\mathbf{f}, \mathbf{g} \in \text{Pol}_1 \mathbf{A}$ such that $\mathbf{f}(U) = V$, $\mathbf{g}(V) = U$, \mathbf{fg} is the identity on V , and \mathbf{gf} is the identity on U . If U and V are polynomially isomorphic, then the algebras $\mathbf{A}|_U$ and $\mathbf{A}|_V$ are isomorphic as nonindexed algebras, that is, it is possible to index the fundamental operations of each with one index set so that the resulting algebras are isomorphic in the usual sense.

An idempotent polynomial for an algebra \mathbf{A} is any $\mathbf{e} \in \text{Pol}_1 \mathbf{A}$ such that $\mathbf{e}^2(x) = \mathbf{e}(x)$ for all $x \in A$. For an idempotent polynomial \mathbf{e} the restriction $\mathbf{e}|_{\mathbf{e}(A)}$ is the identity map on $\mathbf{e}(A)$. Algebras induced by \mathbf{A} on the range of an idempotent polynomial have a particularly simple characterization for their fundamental operations. Namely, if \mathbf{e} is idempotent for \mathbf{A} and $U = \mathbf{e}(A)$, then the fundamental operations of $\mathbf{A}|_U$ consist of all polynomials of the form $\mathbf{ep}|_U$ where \mathbf{p} ranges over all polynomials of \mathbf{A} . The collection of all idempotent polynomials for \mathbf{A} is denoted $E(\mathbf{A})$.

Let $\alpha < \beta$ in the congruence lattice of a finite algebra \mathbf{A} . By $U_{\mathbf{A}}(\alpha, \beta)$ we denote all sets of the form $\mathbf{f}(A)$, with at least two elements, where $\mathbf{f} \in \text{Pol}_1 \mathbf{A}$ and $\mathbf{f}(\beta) \not\subseteq \alpha$. Minimal members of $U_{\mathbf{A}}(\alpha, \beta)$, that is, minimal when ordered by inclusion, are called *(α, β) -minimal sets of \mathbf{A}* . The set of all (α, β) -minimal sets of \mathbf{A} is denoted $M_{\mathbf{A}}(\alpha, \beta)$.

In a finite algebra \mathbf{A} a quotient (α, β) in $\text{Con } \mathbf{A}$ is called *tame* if there exist $V \in M_{\mathbf{A}}(\alpha, \beta)$ and $\mathbf{e} \in E(\mathbf{A})$ such that $\mathbf{e}(A) = V$ and for all $\gamma \in \text{Con } \mathbf{A}$ if $\alpha < \gamma < \beta$, then $\gamma|_V \neq \alpha|_V$ and $\gamma|_V \neq \beta|_V$. Note that every prime quotient is tame. A basic result in tame congruence theory is that if (α, β) is a tame quotient, then all (α, β) -minimal sets of \mathbf{A} are polynomially isomorphic. If (α, β) is tame and $U \in M_{\mathbf{A}}(\alpha, \beta)$, then any set of the form $a/\beta \cap U$ that is not of the form $a/\alpha \cap U$ is called a *trace* of U and an *(α, β) -trace of \mathbf{A}* . The union of all (α, β) -traces of U is called the *body* of U and those elements of U not in the body of U form the *tail* of U . If N is a trace for U , then $\alpha|_N$ denotes $\alpha \cap N^2$, and $\alpha|_N$ is a congruence on the nonindexed algebra $\mathbf{A}|_N$.

The interest in tame congruence theory in tame quotients and their minimal sets and traces arises from the fact that the local behavior of a tame quotient falls into one of five distinct situations. More specifically, for any finite algebra \mathbf{A} , for any tame quotient (α, β) , and for any trace N of $U \in M_{\mathbf{A}}(\alpha, \beta)$, the quotient algebra $(\mathbf{A}|_N)/(\alpha|_N)$ must be polynomially equivalent to one of the following five types of algebras:

1. a G-set,
2. a finite dimensional vector space over a finite field,
3. a 2-element Boolean algebra,
4. a 2-element distributive lattice,
5. a 2-element semilattice.

Moreover, the particular type **1**, **2**, **3**, **4**, or **5** is independent of the choice of U and N . This is called the *type* of the tame quotient (α, β) and is denoted $\text{typ}(\alpha, \beta)$.

The type of a tame quotient in a finite algebra has significant consequences for local behavior and for the algebraic structure of the algebra and the quotient. For example, it is known that for a tame quotient (α, β) , $\text{typ}(\alpha, \beta) \in \{\mathbf{1}, \mathbf{2}\}$ if and only if β is Abelian over α . Because of this, types **1** and **2** are referred to as the *Abelian types* and types **3**, **4**, and **5** are the *non-Abelian types*.

In our work the tame quotients that we consider are usually prime quotients. The following terminology is used in connection with the set of types of prime quotients in finite algebra.

For $\alpha \prec \beta$ the fact $\text{typ}(\alpha, \beta) = \mathbf{i}$ will be sometimes denoted by $\alpha \prec_{\mathbf{i}} \beta$. For $\gamma < \delta$ in $\text{Con } \mathbf{A}$ the set of all types $\text{typ}(\alpha, \beta)$ for $\gamma \leq \alpha \prec \beta \leq \delta$ is denoted $\text{typ}\{\gamma, \delta\}$. The *type set of a finite algebra* \mathbf{A} , denoted $\text{typ}\{\mathbf{A}\}$, is $\text{typ}\{0_{\mathbf{A}}, 1_{\mathbf{A}}\}$. The *type set of a class* \mathcal{K} of algebras consists of the union of the type sets of the finite algebras in \mathcal{K} and is denoted $\text{typ}\{\mathcal{K}\}$.

Two preservation theorems involving the calculus of types that we will frequently use are that type is preserved under homomorphism and that projective prime quotients have the same type, that is, for a finite algebra \mathbf{A} ,

- if $\delta \leq \alpha \prec \beta$ in $\text{Con } \mathbf{A}$, then $\text{typ}(\alpha, \beta) = \text{typ}(\alpha/\delta, \beta/\delta)$ in \mathbf{A}/δ ,
- if $\alpha_1 \prec \beta_1$ and $\alpha_2 \prec \beta_2$ are projective prime quotients in $\text{Con } \mathbf{A}$, then $\text{typ}(\alpha_1, \beta_1) = \text{typ}(\alpha_2, \beta_2)$.

These two results show that if $\mathbf{i} \in \text{typ}\{\mathbf{A}\}$, then there is a subdirectly irreducible algebra \mathbf{A}' with monolith μ such that \mathbf{A}' is a homomorphic image of \mathbf{A} and $\text{typ}(0_{\mathbf{A}'}, \mu) = \mathbf{i}$. Many of our arguments involve an analysis of $(0, \mu)$ -minimal sets and traces for such a monolith μ .

We next summarize some of the algebraic properties that are consequences of a prime quotient having a particular type. Consider an arbitrary finite algebra \mathbf{A} with $\alpha \prec \beta$. Let $U \in M_{\mathbf{A}}(\alpha, \beta)$ and let N be an (α, β) -trace contained in U . Suppose $\text{typ}(\alpha, \beta) = \mathbf{3}$ or $\mathbf{4}$. For these two types, it is known that N is the unique (α, β) -trace contained in U , $\alpha|_N = 0$ and the algebra $\mathbf{A}|_N$ is polynomially equivalent to a 2-element Boolean algebra or 2-element distributive lattice. In both cases there are two binary polynomial $\wedge, \vee \in \text{Pol}_2 \mathbf{A}$ such that $\wedge|_U$ is a *pseudo-meet* and $\vee|_U$ is a *pseudo-join*, [17, Definition 4.18]. This in particular means that we can label the two elements of N with 0 and 1 so that $\langle \{0, 1\}, \wedge|_N, \vee|_N \rangle$ is a distributive lattice with $0 < 1$. Thus, every n -ary operation on $N = \{0, 1\}$ that preserves this order is of the form $\mathbf{p}|_N$ for some $\mathbf{p} \in \text{Pol}_n \mathbf{A}$. If $\text{typ}(\alpha, \beta) = \mathbf{3}$, then in addition to the binary polynomials \wedge and \vee that we have in the type **4** case, there is also a unary polynomial $'$ such that $0' = 1, 1' = 0$ and $A' = U$. The algebra $\langle \{0, 1\}, \wedge|_N, \vee|_N, '|_N \rangle$ is a Boolean algebra and thus every n -ary

operation on N is the restriction to N of some n -ary polynomial on \mathbf{A} that can be built using \wedge, \vee , and $'$.

If $\text{typ}(\alpha, \beta) = \mathbf{2}$, then there may be more than one trace contained in U . Let B be the body of U . A useful result (see [17, Definition 4.22]) that applies to this type $\mathbf{2}$ case is that there is a $\mathbf{d} \in \text{Pol}_3 \mathbf{A}$ such that

- (1) $\mathbf{d}(x, x, x) = x$ for all $x \in U$.
- (2) $\mathbf{d}(x, x, y) = y = \mathbf{d}(y, x, x)$ for all $x \in B$ and $y \in U$.
- (3) For every $a, b \in B$, the unary polynomials given by $\mathbf{d}(x, a, b)$, $\mathbf{d}(a, x, b)$, and $\mathbf{d}(a, b, x)$ are permutations of U
- (4) B is closed under \mathbf{d} , that is, $\mathbf{d}(a, b, c) \in B$ for all $a, b, c \in B$.

The polynomial \mathbf{d} is called a *pseudo-Malcev operation* for U .

Since we are particularly interested in finite algebras from congruence modular varieties we conclude our discussion of tame congruence theory by citing some results that connect it with the theory of the generalized commutator in *locally finite varieties*, i.e. varieties in which finitely generated algebras are finite.

Theorem 3.6. *Let \mathcal{V} be a locally finite variety.*

- (1) \mathcal{V} is congruence modular if and only if $\text{typ}\{\mathcal{V}\} \subseteq \{\mathbf{2}, \mathbf{3}, \mathbf{4}\}$ and minimal sets in finite algebras of \mathcal{V} have empty tails.
- (2) $\text{typ}\{\mathcal{V}\} \subseteq \{\mathbf{2}\}$ if and only if \mathcal{V} is congruence permutable and every algebra in \mathcal{V} is locally solvable.

The varietal conditions given in item (2) of Theorem 3.6 will be of special interest in our work. A variety \mathcal{V} is called *affine* if it is congruence modular and Abelian. It can be argued that if \mathcal{V} is affine then it is also congruence permutable. The properties of affine varieties are developed in [10]. Each affine variety \mathcal{V} has a corresponding ring \mathbf{R} with unit such that every algebra in \mathcal{V} is polynomially equivalent to an \mathbf{R} -module and conversely every \mathbf{R} -module is polynomially equivalent to an algebra in \mathcal{V} .

4. SOME EASY OBSERVATION

Except canonical NP-complete problems (like SAT or k -colorability of graphs) used in our proofs of NP-completeness we will also need the following easy observation, a straightforward proof of which can be found in [12].

Proposition 4.1. *It is NP-complete to decide whether the following systems of two equations of the form*

$$\bigwedge_{i=1}^m x_1^i \vee x_2^i \vee x_3^i = 1,$$

$$\bigvee_{i=1}^n y_1^i \vee y_2^i \vee y_3^i = 0,$$

where x_j^i and y_j^i are variables, have solutions in the 2-element lattice. \square

We continue this section with the proofs of Proposition 2.7 and Example 2.8.

PROPOSITION 2.7. *For every finite relational structure \mathbb{D} (with finitely many relations) there is a finite algebra $\mathbf{A}[\mathbb{D}]$ such that the problem $\text{CSP}(\mathbb{D})$ is polynomially equivalent to $\text{CSAT}(\mathbf{A}[\mathbb{D}])$.*

Proof. Without loss of generality we may assume that \mathbb{D} has both satisfiable and unsatisfiable instances, say \top and \perp , respectively. Now, for the relational structure $\mathbb{D} = (D, \mathcal{R})$ put $\mathbf{A}[\mathbb{D}]$ to be $(A; \wedge, \{f_R\}_{R \in \mathcal{R}})$, where

- $A = D \cup \{0, 1\}$ with $0, 1 \notin D$,
- the binary operation \wedge is defined by:

$$a \wedge b = \begin{cases} 1, & \text{if } a = 1 = b, \\ 0, & \text{otherwise,} \end{cases}$$

- f_R is the $\{0, 1\}$ -characteristic function of the relation R , i.e.

$$f_R(a_1, \dots, a_k) = \begin{cases} 1, & \text{if } (a_1, \dots, a_k) \in R, \\ 0, & \text{otherwise.} \end{cases}$$

It should be obvious that the instance

$$(1) \quad R_1(x_1^1, \dots, x_{k_1}^1) \wedge \dots \wedge R_s(x_1^s, \dots, x_{k_s}^s)$$

of $\text{CSP}(\mathbb{D})$ transforms equivalently to the following instance of $\text{CSAT}(\mathbf{A}[\mathbb{D}])$

$$f_{R_1}(x_1^1, \dots, x_{k_1}^1) \wedge \dots \wedge f_{R_s}(x_1^s, \dots, x_{k_s}^s) = 1.$$

On the other hand the only polynomials of $\mathbf{A}[\mathbb{D}]$ that are non constant are among those that have the following form:

$$(2) \quad x_1^0 \wedge \dots \wedge x_{k_0}^0 \wedge f_{R_1}(x_1^1, \dots, x_{k_1}^1) \wedge \dots \wedge f_{R_s}(x_1^s, \dots, x_{k_s}^s) \wedge 1,$$

where the last conjunct (namely 1) may be absent and the x_j^0 's are not among the x_j^i 's with $i \geq 1$. Moreover the range of such polynomials is contained $\{0, 1\}$ where the value 1 is obtained by sending all the x_j^i 's to 1 and the other variable to the values in D satisfying (1). Thus the only nontrivial instances (i.e. the ones that do not transform to \top or \perp) of $\text{CSAT}(\mathbf{A}[\mathbb{D}])$ have the form $\mathbf{t}(\bar{x}) = 1$, with $\mathbf{t}(\bar{x})$ being described in (2). Such an equation obviously translates to the equivalent instance (1) of $\text{CSP}(\mathbb{D})$. \square

EXAMPLE 2.8. *There is a finite algebra \mathbf{A} and its congruence θ such that $\text{CSAT}(\mathbf{A}) \in P$ while $\text{CSAT}(\mathbf{A}/\theta)$ is NP-complete.*

Proof. The operations of the algebra \mathbf{A} will be defined in such a way that the satisfiability of the polynomial equation $\mathbf{t}(\bar{x}) = \mathbf{s}(\bar{x})$ easily reduces to the one over the 2-element lattice whenever the ranges of the polynomials \mathbf{t} and \mathbf{s} are not disjoint. On the other hand it will be possible to define the congruence θ and lattice-like polynomials \mathbf{l} and \mathbf{r} so that the one equation of the form $\mathbf{l}(\bar{x}) = \mathbf{r}(\bar{x})$ encodes modulo θ a system of two lattice equations.

For the underlying set of our algebra we put

$$A = \{0, 1, 0', 1', 0_{\wedge,l}, 1_{\wedge,l}, s_{\wedge,l}, 0_{\wedge,r}, 1_{\wedge,r}, s_{\wedge,r}, 0_{\vee,l}, 1_{\vee,l}, s_{\vee,l}, 0_{\vee,r}, 1_{\vee,r}, s_{\vee,r}\}.$$

Our basic operations come in two sorts:

- left: ternary ‘disjunction’ D_l , binary ‘conjunction’ \wedge_l and unary f_l
- right: ternary ‘conjunction’ C_r , binary ‘disjunction’ \vee_r and unary f_r ,

To define these operations we will refer to an external two element lattice $(\{\perp, \top\}; \wedge, \vee)$ in which $\perp < \top$. This reference is done by passing from $x \in \{0, 0_{\wedge,l}, 0_{\wedge,r}, 0_{\vee,l}, 0_{\vee,r}\}$ to $\hat{x} = \perp$ and for $x \in \{1, 1_{\wedge,l}, 1_{\wedge,r}, 1_{\vee,l}, 1_{\vee,r}\}$ to $\hat{x} = \top$ and putting

$$\begin{aligned} D_l(x, y, z) &= \begin{cases} 1_{\vee,l}, & \text{if } x, y, z \in \{0, 1\} \text{ and } \hat{x} \vee \hat{y} \vee \hat{z} = \top, \\ 0_{\vee,l}, & \text{if } x, y, z \in \{0, 1\} \text{ and } \hat{x} \vee \hat{y} \vee \hat{z} = \perp, \\ s_{\vee,l}, & \text{otherwise,} \end{cases} \\ x \wedge_l y &= \begin{cases} 1_{\wedge,l}, & \text{if } x, y \in \{0_{\wedge,l}, 1_{\wedge,l}, 0_{\vee,l}, 1_{\vee,l}\} \text{ and } \hat{x} \wedge \hat{y} = \top, \\ 0_{\wedge,l}, & \text{if } x, y \in \{0_{\wedge,l}, 1_{\wedge,l}, 0_{\vee,l}, 1_{\vee,l}\} \text{ and } \hat{x} \wedge \hat{y} = \perp, \\ s_{\wedge,l}, & \text{otherwise,} \end{cases} \\ C_r(x, y, z) &= \begin{cases} 1_{\wedge,r}, & \text{if } x, y, z \in \{0, 1\} \text{ and } \hat{x} \wedge \hat{y} \wedge \hat{z} = \top, \\ 0_{\wedge,r}, & \text{if } x, y, z \in \{0, 1\} \text{ and } \hat{x} \wedge \hat{y} \wedge \hat{z} = \perp, \\ s_{\wedge,r}, & \text{otherwise,} \end{cases} \\ x \vee_r y &= \begin{cases} 1_{\vee,r}, & \text{if } x, y \in \{0_{\wedge,r}, 1_{\wedge,r}, 0_{\vee,r}, 1_{\vee,r}\} \text{ and } \hat{x} \vee \hat{y} = \top, \\ 0_{\vee,r}, & \text{if } x, y \in \{0_{\wedge,r}, 1_{\wedge,r}, 0_{\vee,r}, 1_{\vee,r}\} \text{ and } \hat{x} \vee \hat{y} = \perp, \\ s_{\vee,r}, & \text{otherwise,} \end{cases} \\ f_l(x) &= \begin{cases} 1', & \text{if } x = 1_{\wedge,l}, \\ x, & \text{otherwise,} \end{cases} \\ f_r(x) &= \begin{cases} 0', & \text{if } x = 0_{\vee,r}, \\ x, & \text{otherwise.} \end{cases} \end{aligned}$$

A careful inspection of the above definitions shows that there not many ways to compose operations of $\mathbf{A} = (A; D_l, C_r, \wedge_l, \vee_r, f_l, f_r)$ in a meaningful way, i.e. to get polynomials that have essential arity at least 4. Moreover for two such polynomials \mathbf{t}, \mathbf{s} either they have disjoint ranges or the equation $\mathbf{t}(\bar{x}) = \mathbf{s}(\bar{x})$ has a solution. However disjointness of the ranges can be checked by inspecting how the polynomials \mathbf{t}, \mathbf{s} are built from the basic operations. This shows that $\text{CSAT}(\mathbf{A}) \in P$.

Now let Θ be the congruence with one nontrivial block $\{0', 1'\}$. This opens the way to transform the system of two lattice equations

$$\begin{aligned} (x_1^1 \vee x_2^1 \vee x_3^1) \wedge \dots \wedge (x_1^m \vee x_2^m \vee x_3^m) &= 1 \\ (y_1^1 \wedge y_2^1 \wedge y_3^1) \vee \dots \vee (y_1^n \wedge y_2^n \wedge y_3^n) &= 0 \end{aligned}$$

into a single equation

$$f_l(D_l(x_1^1, x_2^1, x_3^1) \wedge_l \dots \wedge_l D_l(x_1^m, x_2^m, x_3^m)) = f_r(C_r(y_1^1, y_2^1, y_3^1) \vee_r \dots \vee_r C_r(y_1^n, y_2^n, y_3^n)),$$

so that the system is solvable in two element lattice if and only if this single equation is satisfied in \mathbf{A}/Θ by the very same $\{0, 1\}$ -values for all the variables. Together with Proposition 4.1 this allows us to conclude that $\text{CSAT}(\mathbf{A}/\theta)$ is NP-complete. \square

5. TYPE **3** NEED NOT APPLY

The most classical problem of solving equation is satisfiability of Boolean formulas which is actually a paradigm for NP-complete problems. The presence of Boolean behavior inside a finite algebra is in fact ruled out by the following theorem.

Theorem 5.1. *If \mathbf{A} is finite algebra from a congruence modular variety such that $\mathbf{3} \in \text{typ}\{\mathbf{A}\}$, then $\text{CSAT}(\mathbf{A})$ is NP-complete.*

Proof. Suppose that \mathbf{A} is a finite algebra containing a type **3** minimal set $U = \{0, 1\}$ with respect to some covering pair $\alpha \prec \beta$ of its congruences. Then $\mathbf{A}|_U$ is polynomially equivalent to a 2-element Boolean algebra, so that there are polynomials \wedge, \vee, \neg of \mathbf{A} that behave on U like meet, join and negation, respectively. Moreover, there is a unary idempotent polynomial \mathbf{e}_U of \mathbf{A} with the range U .

Now the 3-SAT instance:

$$\Phi \equiv \bigwedge_{i=1}^m \ell_1^i \vee \ell_2^i \vee \ell_3^i,$$

where $\ell_j^i \in \{x_j^i, \neg x_j^i\}$, can be easily translated to the equation

$$(3) \quad \bigwedge_{i=1}^m \delta_1^i \mathbf{e}_U(z_1^i) \vee \delta_2^i \mathbf{e}_U(z_2^i) \vee \delta_3^i \mathbf{e}_U(z_3^i) = 1,$$

where

$$\delta_j^i \mathbf{e}_U(z_j^i) = \begin{cases} \mathbf{e}_U(z_j^i), & \text{if the literal } \ell_j^i \text{ is the variable, i.e., } \ell_j^i = x_j^i, \\ \neg \mathbf{e}_U(z_j^i), & \text{if } \ell_j^i \text{ is the negated variable, i.e., } \ell_j^i = \neg x_j^i. \end{cases}$$

It should be obvious that the formula Φ is satisfiable if and only if the equation (3) has a solution. \square

Combining Theorems 3.6 and 5.1 we get the following corollary.

Corollary 5.2. *If \mathbf{A} is finite algebra from a congruence modular variety such that $\text{CSAT}(\mathbf{A})$ is not NP-complete then $\text{typ}\{\mathbf{A}\} \subseteq \{\mathbf{2}, \mathbf{4}\}$* \square

6. TRANSFER PRINCIPLES AND DECOMPOSITION

In this section we prove that every finite algebra \mathbf{A} from a congruence modular variety for which $\text{CSAT}(\mathbf{A})$ is not NP-complete decomposes into a direct product of a solvable algebra and an algebra that has only **4** in its typeset. In order to obtain such a nice decomposition we will first establish so called transfer principles introduced by Matthew Valeriote in [38].

Definition 6.1. We say that a finite algebra \mathbf{A} satisfies the (\mathbf{i}, \mathbf{j}) -transfer principle if whenever $\alpha \prec_{\mathbf{i}} \beta \prec_{\mathbf{j}} \gamma$ are congruences of \mathbf{A} then there exists a congruence β' with $\alpha \prec_{\mathbf{j}} \beta' \leq \gamma$.

The next Lemma helps us in a better localizing unwanted failures of the transfer principles.

Lemma 6.2. If an algebra \mathbf{A} fails to have (\mathbf{i}, \mathbf{j}) -transfer principle and $\text{Con } \mathbf{A}$ is modular then

- (1) \mathbf{A} has congruences $\alpha' \prec_{\mathbf{i}} \beta' \prec_{\mathbf{j}} \gamma'$ with α' being meet irreducible,
- (2) \mathbf{A} has congruences $\alpha' \prec_{\mathbf{i}} \beta' \prec_{\mathbf{j}} \gamma'$ with γ' being join irreducible.

Proof. To see (1) suppose that the failure of (\mathbf{i}, \mathbf{j}) -transfer principle is witnessed by the three element chain $\alpha \prec_{\mathbf{i}} \beta \prec_{\mathbf{j}} \gamma$. Pick α' to be a maximal congruence that is over α but not over β . Then obviously α' is meet irreducible, as otherwise $\alpha' = \alpha_1 \cap \alpha_2$ with $\alpha_i > \alpha'$ would give $\alpha_i \geq \beta$ so that $\alpha' = \alpha_1 \cap \alpha_2 \geq \beta$. One can easily check that $I[\alpha, \beta] \nearrow I[\alpha', \beta \vee \alpha']$. Moreover modularity of the lattice $\text{Con } \mathbf{A}$ gives $I[\beta, \gamma] \nearrow I[\beta \vee \alpha', \gamma \vee \alpha']$. Summing up we get $\alpha' \prec_{\mathbf{i}} \beta' \prec_{\mathbf{j}} \gamma'$ for $\beta' = \beta \vee \alpha'$ and $\gamma' = \gamma \vee \alpha'$.

The item (2) can be shown in a dual way, by replacing γ with a minimal congruence γ' that is below γ but not below β . \square

The next two Theorems establish both possible transfer principles, as the typeset $\text{typ}\{\mathbf{A}\}$ is restricted in Corollary 5.2.

Theorem 6.3. If \mathbf{A} is finite algebra from a congruence modular variety in which $(2, 4)$ -transfer principle fails, then $\text{CSAT}(\mathbf{A})$ is NP-complete.

Proof. Suppose that $(2, 4)$ -transfer principle fails in \mathbf{A} . By Lemma 6.2.(2) this failure can be witnessed with a three element chain of congruences $\theta \prec_2 \alpha \prec_4 \beta$ with β being a join irreducible. Let $U = \{0, 1\}$ be an (α, β) -minimal set and V be an (θ, α) -minimal set in \mathbf{A} . Moreover let \mathbf{e}_U and \mathbf{e}_V be unary idempotent polynomials of \mathbf{A} with the range U and V , respectively. Taking into account the types of minimal sets U and V we know that \mathbf{A} has the polynomials \wedge, \vee that serve as the lattice operations on $\mathbf{A}|_U$ (with respect to the lattice order $0 < 1$) and a polynomial $\mathbf{d}_V(x, y, z)$ that has the range contained in V and is a Malcev operation on V .

$$(6.1) \text{ For every } (a, b) \in \alpha|_V \text{ there is a unary polynomial } \mathbf{f}_{ab}(x) \text{ of } \mathbf{A} \text{ such that } \mathbf{f}_{ab} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

To produce such a polynomial \mathbf{f}_{ab} note that $(a, b) \in \alpha|_V \subseteq \beta$ and $\beta = \text{Cg}^{\mathbf{A}}(0, 1)$, as β is join irreducible and $(0, 1) \notin \alpha$. Now simply recall Lemma 3.1.

Now with the help of (6.1) we will transform the system of two lattice equations

$$(6.2) \begin{cases} \bigwedge_{i=1}^m x_1^i \vee x_2^i \vee x_3^i = 1 \\ \bigvee_{i=1}^n y_1^i \vee y_2^i \vee y_3^i = 0 \end{cases}$$

into a single equation of the algebra \mathbf{A} . In view of Proposition 4.1 this will establish NP-completeness of $\text{CSAT}(\mathbf{A})$.

We start with picking $(a, b) \in \alpha|_V - \theta$ to put

$$(6.3) \quad \mathbf{d}_V(\mathbf{f}_{ab}(\nu(\bar{y})), a, \mathbf{f}_{ab}(\nu(\bar{y}) \vee \pi(\bar{x}))) = b, \text{ where}$$

$$\begin{aligned} \pi(\bar{x}) &= \bigwedge_{i=1}^m \mathbf{e}_U(x_1^i) \vee \mathbf{e}_U(x_2^i) \vee \mathbf{e}_U(x_3^i), \\ \nu(\bar{y}) &= \bigvee_{i=1}^n \mathbf{e}_U(y_1^i) \vee \mathbf{e}_U(y_2^i) \vee \mathbf{e}_U(y_3^i). \end{aligned}$$

First note that if \bar{x}, \bar{y} is the $\{0, 1\}$ -lattice solution to (6.2) then keeping the values for the x 's and y 's we have $\mathbf{f}_{ab}(\nu(\bar{y})) = a$ and $\mathbf{f}_{ab}(\nu(\bar{y}) \vee \pi(\bar{x})) = b$ so that

$$\mathbf{d}_V(\mathbf{f}_{ab}(\nu(\bar{y})), a, \mathbf{f}_{ab}(\nu(\bar{y}) \vee \pi(\bar{x}))) = \mathbf{d}_V(a, a, b) = b,$$

as required.

Conversely, if (6.3) has a solution \bar{x}, \bar{y} in \mathbf{A} then

- $\mathbf{f}_{ab}(\nu(\bar{y})) = a = \mathbf{f}_{ab}(\nu(\bar{y}) \vee \pi(\bar{x}))$ is not possible, as $\mathbf{d}_V(a, a, a) = a$,
- $\mathbf{f}_{ab}(\nu(\bar{y})) = b$ and $\mathbf{f}_{ab}(\nu(\bar{y}) \vee \pi(\bar{x})) = a$ is not possible, as then we would have $1 = \nu(\bar{y}) \leq \nu(\bar{y}) \vee \pi(\bar{x}) = 0$ in the set U , contrary to our choice of $0 < 1$,
- $\mathbf{f}_{ab}(\nu(\bar{y})) = b = \mathbf{f}_{ab}(\nu(\bar{y}) \vee \pi(\bar{x}))$ is not possible, as then $\mathbf{d}_V(b, a, b) = b$ contrary to the fact that $\mathbf{d}_V(b, b, b) = b$ and $x \mapsto \mathbf{d}_V(b, x, b)$ is a permutation of V .

Thus the only possibility for a solution \bar{x}, \bar{y} to (6.3), is to satisfy $\mathbf{f}_{ab}(\nu(\bar{y})) = a$ and $\mathbf{f}_{ab}(\nu(\bar{y}) \vee \pi(\bar{x})) = b$, or equivalently that $\nu(\bar{y}) = 0$ and $\pi(\bar{x}) = 1$. Therefore evaluating the x_j^i 's and y_j^i 's by $\mathbf{e}_U(x_j^i)$ and $\mathbf{e}_U(y_j^i)$, respectively, we get a solution to the system of lattice equations (6.2). \square

Theorem 6.4. *If \mathbf{A} is finite algebra from a congruence modular variety in which $(4, 2)$ -transfer principle fails, then $\text{CSAT}(\mathbf{A})$ is NP-complete.*

Proof. Suppose that $(4, 2)$ -transfer principle fails in \mathbf{A} . By Lemma 6.2.(2) this failure can be witnessed with a three element chain $\theta \prec_4 \alpha \prec_2 \beta$ with θ being a meet irreducible congruence. Let U be an (α, β) -minimal set and $V = \{0, 1\}$ be an (θ, α) -minimal set in \mathbf{A} . Moreover let \mathbf{e}_U and \mathbf{e}_V be unary idempotent polynomials of \mathbf{A} with the range U and V , respectively. Taking into account the types of minimal sets U and V we know that \mathbf{A} has the polynomials \wedge, \vee that serve as the lattice operations on $\mathbf{A}|_V$ (with respect to the lattice order $0 < 1$) and a polynomial $\mathbf{d}_U(x, y, z)$ that has the range contained in U and is a Malcev operation on U .

We start with the following claims:

$$(6.4) \quad (0, 1) \in \text{Cg}^{\mathbf{A}}(c, d) \text{ for each } (c, d) \notin \theta,$$

$$(6.5) \quad \text{for every } (c, d) \in U^2 - \theta \text{ there is a unary polynomial } \mathbf{f}_{cd}(x) \text{ such that } \mathbf{f}_{cd}\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$(6.6) \quad \alpha|_U \subseteq \theta.$$

Note that $(0, 1) \in \alpha \leq \theta \vee \text{Cg}^{\mathbf{A}}(c, d)$ which together with Lemma 3.1 gives a Malcev chain that connects 0 with 1 via projections by unary polynomials pairs from θ or the pair (c, d) . Applying \mathbf{e}_V to this chain we get that it entirely lives in $\{0, 1\}$. Since $\theta|_V = 0$ we get that at some link in this chain the pair 0, 1 occurs as a projection by a unary polynomial \mathbf{f} applied to the pair (c, d) . This obviously gives $(0, 1) \in \text{Cg}^{\mathbf{A}}(c, d)$.

If, as in the assumptions of (6.5), $(c, d) \in U^2 - \theta$ then either \mathbf{f} is good enough to serve as \mathbf{f}_{cd} or we put $\mathbf{f}_{cd}(x) = \mathbf{f}\mathbf{d}_U(c, x, d)$.

To see (6.6) suppose to the contrary that $(0', 1') \in \alpha|_U - \theta$. In particular $(0', 1') \in \alpha = \theta \vee \text{Cg}^{\mathbf{A}}(0, 1)$. Thus there is a Malcev chain connecting $0'$ and $1'$ via links of the form $\{\mathbf{f}(c), \mathbf{f}(d)\}$, with \mathbf{f} being unary polynomials of \mathbf{A} and $(c, d) \in \theta \cup \{0, 1\}$. By applying \mathbf{e}_U to this chain we may assume that it is fully contained in U . Since $(0', 1') \notin \theta$ at least one link must be obtained by projecting the set $\{0, 1\}$ onto a pair $(0'', 1'') \in \alpha|_U - \theta$. Therefore $\{0'', 1''\}$ is a (θ, α) -minimal set of type 4 lying inside a minimal set U of type 2, which is not possible.

Now we pick a transversal $\{c_0, c_1, \dots, c_k\}$ of U/α and define a unary polynomial

$$\mathbf{s}(x) = \bigvee_{i=1}^k \mathbf{f}_{c_0 c_i}(x)$$

so that $\mathbf{s}(c_0) = 0$ and $\mathbf{s}(c_i) = 1$ for $i = 1, \dots, k$. In fact

$$(6.7) \quad \text{for every } a \in U \text{ we have } \mathbf{s}(a) = \begin{cases} 0, & \text{if } a \in c_0/\alpha, \\ 1, & \text{otherwise,} \end{cases}$$

as for every $a \in U \cap c_i/\alpha$ we have $(\mathbf{s}(a), \mathbf{s}(c_i)) \in \alpha|_U \subseteq \theta$, which together with $\theta|_{\{0,1\}} = 0$ gives $\mathbf{s}(a) = \mathbf{s}(c_i)$.

Now our proof of NP-complete ss of $\text{CSAT}(\mathbf{A})$ splits into two cases depending on the size of U/α .

CASE 1. $|U/\alpha| = 2$, i.e., $U = \{c_0, c_1\}$.

In this case we will transform each 3-SAT instance:

$$\Phi \equiv \bigwedge_{i=1}^m \ell_1^i \vee \ell_2^i \vee \ell_3^i,$$

where $\ell_j^i \in \{x_j^i, \neg x_j^i\}$, to an equation

$$(6.8) \quad \bigwedge_{i=1}^m z_1^i \vee z_2^i \vee z_3^i = 1, \text{ where}$$

$$z_j^i = \begin{cases} \mathbf{se}_U(x_j^i), & \text{if } \ell_j^i = x_j^i, \\ \mathbf{sd}_U(\mathbf{e}_U(x_j^i), c_0, c_1), & \text{if } \ell_j^i = \neg x_j^i. \end{cases}$$

It should be obvious that each evaluation satisfying Φ can be transformed into a solution of (6.8) by sending the 0's to c_0 and the 1's to c_1 .

Conversely, if the a_j^i 's form a solution to (6.8) then putting

$$x_j^i = \begin{cases} 0, & \text{if } \mathbf{e}_U(a_j^i) \in c_0/\alpha \cap U, \\ 1, & \text{if } \mathbf{e}_U(a_j^i) \in c_1/\alpha \cap U, \end{cases}$$

we get a valuation satisfying Φ .

CASE 2. $|U/\alpha| \geq 3$.

In this case with each graph $G = (V, E)$ with $V = \{v_1, \dots, v_n\}$ we associate a polynomial $\mathbf{t}_G(x_1, \dots, x_n) \in \text{Pol } \mathbf{A}$ in such a way that G is $|U/\alpha|$ -colorable iff the equation $\mathbf{t}_G(\bar{x}) = 1$ has a solution in \mathbf{A} .

First observe that for the polynomial $\mathbf{s}'(x, y) = \mathbf{sd}_U(x, y, c_0)$ and $a, b \in U$ we have

$$(6.9) \quad \mathbf{s}'(a, b) = \begin{cases} 0, & \text{if } (a, b) \in \alpha, \\ 1, & \text{otherwise.} \end{cases}$$

This is due to the fact that the polynomial $\mathbf{d}_U(x, y, z)$, after fixing values of any two of its variables (by values in U), is a permutation of U with respect to the remaining variable.

Now we put

$$\mathbf{t}_G(x_1, \dots, x_n) = \bigwedge_{\{v_i, v_j\} \in E} \mathbf{s}'(\mathbf{e}_U(x_i), \mathbf{e}_U(x_j)).$$

From (6.9) we know that $\mathbf{t}_G(x_1, \dots, x_n) = 1$ iff for all i, j such that v_i and v_j are connected in G by an edge, $\mathbf{e}_U(x_i)$ and $\mathbf{e}_U(x_j)$ are in different $\alpha|_U$ -classes. This means that $\mathbf{t}_G(\bar{x}) = 1$ has a solution in \mathbf{A} iff G can be properly colored by $\alpha|_U$ -classes. \square

With the help of Theorems 6.3 and 6.4 we are ready to prove the promised decomposition.

Corollary 6.5. *If \mathbf{A} is finite algebra from a congruence modular variety then either \mathbf{A} is isomorphic to a direct product $\mathbf{A}_2 \times \mathbf{A}_4$, where $\text{typ}\{\mathbf{A}_2\} \subseteq \{\mathbf{2}\}$ and $\text{typ}\{\mathbf{A}_4\} \subseteq \{\mathbf{4}\}$ or $\text{CSAT}(\mathbf{A})$ is NP-complete.*

Proof. Suppose that $\text{CSAT}(\mathbf{A})$ is not NP-complete. From Theorem 5.1 we know that $\text{typ}\{A\} \subseteq \{\mathbf{2}, \mathbf{4}\}$. To get the required decomposition we start with the following easy claim:

$$(6.10) \quad \text{For } \alpha, \beta \in \text{Con } \mathbf{A} \text{ we have } \text{typ}\{\alpha \cap \beta, \alpha\} = \text{typ}\{\beta, \alpha \vee \beta\}.$$

In a modular lattice the intervals $I[\alpha \cap \beta, \alpha]$ and $I[\beta, \alpha \vee \beta]$ are isomorphic under the mutually converse mappings $\gamma \mapsto \gamma \vee \beta$ and $\alpha \cap \delta \longleftarrow \delta$. In particular every covering pair $\alpha \cap \beta \leq \gamma \prec \gamma' \leq \alpha$ is mapped onto a covering pair $\beta \leq \gamma \vee \beta \prec \gamma' \vee \beta \leq \alpha \vee \beta$. In fact $I[\gamma, \gamma'] \nearrow I[\gamma \vee \beta, \gamma' \vee \beta]$ so that $\text{typ}(\gamma, \gamma') = \text{typ}(\gamma \vee \beta, \gamma' \vee \beta)$.

(6.11) For $i \in \{2, 4\}$ there is the largest congruence $\rho_i \in \text{Con } \mathbf{A}$ with $\text{typ}\{0, \rho_i\} \subseteq \{\mathbf{i}\}$.

To prove this it suffices to show that for $\alpha, \beta \in \text{Con } \mathbf{A}$ with $\text{typ}\{0, \alpha\} \subseteq \{\mathbf{i}\} \supseteq \text{typ}\{0, \beta\}$ we have $\text{typ}\{0, \alpha \vee \beta\} \subseteq \{\mathbf{i}\}$. Thus let $\alpha \cap \beta \leq \delta \prec \delta' \leq \alpha \vee \beta$. If $\alpha \cap \delta < \alpha \cap \delta'$ or $\alpha \vee \delta < \alpha \vee \delta'$ that such an inequality is actually a covering so that, by our assumptions and (6.10), it has type \mathbf{i} . Therefore $\delta \prec \delta'$ inherits type either from $\alpha \cap \delta \prec_{\mathbf{i}} \alpha \cap \delta'$ or from $\alpha \vee \delta \prec_{\mathbf{i}} \alpha \vee \delta'$. On the other hand at least one of those strong inequalities has to hold, as otherwise the congruences $\alpha \cap \delta \leq \alpha, \delta, \delta' \leq \alpha \vee \delta$ would form a pentagon, contradicting the modularity of $\text{Con } \mathbf{A}$.

(6.12) For $\{i, j\} = \{2, 4\}$ we have $\text{typ}\{\rho_i, 1\} \subseteq \{\mathbf{j}\}$.

Suppose that (6.12) fails, and α is a minimal congruence above ρ_i that has a cover, say β , of type \mathbf{i} . Since, by the definition of ρ_i all its covers are of type \mathbf{j} we know that $\rho < \alpha \prec_{\mathbf{i}} \beta$. Therefore there is $\theta \in \text{Con } \mathbf{A}$ with $\rho \leq \theta \prec_{\mathbf{j}} \alpha \prec_{\mathbf{i}} \beta$. By Theorems 6.3 and 6.4 we know that there is $\theta' \in \text{Con } \mathbf{A}$ with $\theta \prec_{\mathbf{i}} \theta' \leq \beta$. This contradicts the minimality of α as $\theta < \alpha$ and θ has a cover of type \mathbf{i} .

From (6.12) we know that for $\mathbf{A}_2 = \mathbf{A}/\rho_4$ and $\mathbf{A}_4 = \mathbf{A}/\rho_2$ we have $\text{typ}\{\mathbf{A}_i\} \subseteq \{\mathbf{i}\}$. To show that \mathbf{A} is isomorphic with the product $\mathbf{A}_2 \times \mathbf{A}_4$ first note that $\rho_2 \cap \rho_4 = 0$, by the definitions of ρ_i , and $\rho_2 \vee \rho_4 = 1$, by (6.12). Finally, by Theorem 6.2 of [10], ρ_2 permutes with all congruences of \mathbf{A} , as ρ_2 is solvable. This gives that ρ_2, ρ_4 gives a factorization of \mathbf{A} , as required. \square

The decomposition established in Corollary 6.5 together with the possibility of passing to the quotients allows us to separately consider solvable algebras, i.e. algebras with typeset contained in $\{\mathbf{2}\}$ and entirely lattice type algebras, i.e. algebras with typeset contained in $\{\mathbf{4}\}$.

7. RESTRICTING SOLVABLE BEHAVIOR

The aim of this section is to show that every finite solvable algebra \mathbf{A} from a congruence modular variety is in fact nilpotent or \mathbf{A} has a homomorphic image \mathbf{A}' with $\text{CSAT}(\mathbf{A}')$ being NP-complete. We start with the following construction.

Lemma 7.1. *Let \mathbf{A} be a finite solvable subdirectly irreducible algebra from a congruence modular variety. If $[1, \mu] > 0$, where μ is the monolith of \mathbf{A} then $\text{CSAT}(\mathbf{A})$ is NP-complete.*

Proof. Put $\alpha = (0 : \mu)$. If $[1, \mu] > 0$ then there is $\beta \in \text{Con } \mathbf{A}$ such that $\mu \leq \alpha \prec \beta$ and obviously $\alpha \prec_{\mathbf{2}} \beta$ by the solvability of \mathbf{A} . Moreover we have $[\alpha, \mu] = 0$ while $[\beta, \mu] = \mu$. Pick:

- an (α, β) -minimal set U ,
- a transversal $\{d_0, d_1, \dots, d_k\}$ of U/α ,
- a $(0, \mu)$ -minimal set V ,

- a pair $(e, a) \in \mu|_V - 0$ and let $N = e/\mu \cap V$ be the trace of V containing both e and a .

We know that $\mathbf{A}|_N$ is polynomially equivalent to a (one dimensional) vector space and we may assume that e is its zero element with respect to the vectors addition $+$ which has to be a polynomial of \mathbf{A} .

Now note that by the choice of the d_i 's we know that $\alpha < \alpha \vee \text{Cg}^{\mathbf{A}}(d_i, d_j)$ for $i \neq j$ which gives $[\text{Cg}^{\mathbf{A}}(d_i, d_j), \mu] = \mu$. This has to be witnessed by a polynomial $\mathbf{s}_{ij}(x, y_1, \dots, y_m)$ and elements $(c, d) \in \text{Cg}^{\mathbf{A}}(d_i, d_j)$ and $(a_1, b_1), \dots, (a_m, b_m) \in \mu = \text{Cg}^{\mathbf{A}}(e, a)$ satisfying

$$\begin{aligned} s'_{ij}(c, a_1, \dots, a_m) &= s'_{ij}(c, b_1, \dots, b_m), \\ s'_{ij}(d, a_1, \dots, a_m) &\neq s'_{ij}(d, b_1, \dots, b_m). \end{aligned}$$

Since \mathbf{A} is a solvable algebra in a congruence modular variety, the variety generated by \mathbf{A} is solvable and therefore congruence permutable (by Theorem 6.3 of [10]). Thus \mathbf{A} has a Malcev term, say \mathbf{d} , and by Lemma 3.1 we get unary polynomials $\mathbf{q}, \mathbf{p}_1, \dots, \mathbf{p}_m$ of \mathbf{A} with $\mathbf{q}(d_i) = c, \mathbf{q}(d_j) = d$ and $\mathbf{p}_k(e) = a_k, \mathbf{p}_k(a) = b_k$ for all $1 \leq k \leq m$. Thus for the polynomial $\mathbf{s}_{ij}(x, y) = e_V \mathbf{s}'_{ij}(\mathbf{q}(y), \mathbf{p}_1(x), \dots, \mathbf{p}_m(x))$ we have

$$\begin{aligned} \mathbf{s}_{ij}(e, d_i) &= \mathbf{s}_{ij}(a, d_i), \\ \mathbf{s}_{ij}(e, d_j) &\neq \mathbf{s}_{ij}(a, d_j). \end{aligned}$$

Again referring to Lemma 3.1 and using $(e, a) \in \text{Cg}^{\mathbf{A}}(\mathbf{s}_{ij}(e, d_j), \mathbf{s}_{ij}(a, d_j))$ we get a unary polynomial \mathbf{p} of \mathbf{A} that takes the pair $(\mathbf{s}_{ij}(e, d_j), \mathbf{s}_{ij}(a, d_j))$ to (e, a) . Now, replacing $\mathbf{s}_{ij}(x, y)$ by $\mathbf{d}(\mathbf{s}_{ij}(x, y), \mathbf{s}_{ij}(e, y), e)$ we get that

$$\begin{aligned} \mathbf{s}_{ij}(e, d_i) &= \mathbf{s}_{ij}(a, d_i) = e, \\ e &= \mathbf{s}_{ij}(e, d_j) \neq \mathbf{s}_{ij}(a, d_j) = a. \end{aligned}$$

In fact we know that $\mathbf{s}_{ij}(e, y) = e$ for all $y \in A$.

Now, for each fixed $y \in A$ the unary polynomial

$$V \ni v \longmapsto \mathbf{s}_{ij}(v, y) \in V$$

is either a permutation of V or collapses $\mu|_V$ to 0, i.e., it is constant on $\mu|_V$ -classes. Thus, iterating $\mathbf{s}_{ij}(v, y)$ in the first variable a sufficient number of times we can modify \mathbf{s}_{ij} to additionally satisfy that for each fixed $y \in A$ the new polynomial $\mathbf{s}_{ij}(v, y)$ is either the identity map on V or it is constant on $\mu|_V$ -classes. Actually, in the second case, i.e. if $\mathbf{s}_{ij}(v, y)$ collapses $\mu|_V$ to 0 then it collapses the trace N to $\mathbf{s}_{ij}(e, y) = e$. Summing up, we produced polynomials \mathbf{s}_{ij} satisfying

$$\begin{aligned} \mathbf{s}_{ij}(e, y) &= e, \quad \text{for each } y \in A, \\ \mathbf{s}_{ij}(v, d_i) &= e, \quad \text{for each } v \in N, \\ \mathbf{s}_{ij}(v, d_j) &= v, \quad \text{for each } v \in V. \end{aligned}$$

Now, using the fact that $[\mu, \alpha] = 0$ we can keep the above equalities by varying the second variable modulo α

$$\begin{aligned} \mathbf{s}_{ij}(e, y) &= e, \quad \text{for each } y \in A, \\ \mathbf{s}_{ij}(v, y) &= e, \quad \text{for each } v \in N \text{ and } y \in d_i/\alpha, \\ \mathbf{s}_{ij}(v, y) &= v, \quad \text{for each } v \in V \text{ and } y \in d_j/\alpha. \end{aligned}$$

Now for $j = 0, \dots, k$ define

$$\mathbf{s}_j(x, y) = \mathbf{s}_{i_1 j}(\dots \mathbf{s}_{i_{k-1} j}(\mathbf{s}_{i_k j}(x, y), y) \dots, y),$$

where $\{j, i_1, \dots, i_k\} = \{0, 1, \dots, k\}$. It is easy to observe that \mathbf{s}_j has the range contained in V and

$$\begin{aligned} \mathbf{s}_j(e, y) &= e, \quad \text{for each } y \in A, \\ \mathbf{s}_j(v, y) &= e, \quad \text{for each } v \in N \text{ and } y \notin d_j/\alpha, \\ \mathbf{s}_j(v, y) &= v, \quad \text{for each } v \in V \text{ and } y \in d_j/\alpha. \end{aligned}$$

Indeed, the first and the last item follows directly from the definition of \mathbf{s}_j . To see the middle one note that for $v \in N$ and $y \in d_{i_\ell}/\alpha$ we have

$$\begin{aligned} v' &= \mathbf{s}_{i_{\ell+1}}(\dots \mathbf{s}_{i_{k-1} j}(\mathbf{s}_{i_k j}(v, y), y) \dots, y) \\ &\stackrel{\mu}{=} \mathbf{s}_{i_{\ell+1}}(\dots \mathbf{s}_{i_{k-1} j}(\mathbf{s}_{i_k j}(e, y), y) \dots, y) \\ &= e, \end{aligned}$$

i.e. $v' \in N$ so that $\mathbf{s}_{i_\ell j}(v', y) = e$, and consequently

$$\begin{aligned} \mathbf{s}_j(v, y) &= \mathbf{s}_{i_1 j}(\dots \mathbf{s}_{i_{\ell-1} j}(\mathbf{s}_{i_\ell j}(v', y), y) \dots, y) \\ &= \mathbf{s}_{i_1 j}(\dots \mathbf{s}_{i_{\ell-1} j}(e, y) \dots, y) \\ &= e. \end{aligned}$$

As $\mathbf{A}|_N$ is polynomially equivalent to a vector space (with e being its neutral element) and for $v \in N$ and $y \in U$ the elements $\mathbf{s}_j(v, y)$ are in $\mathbf{A}|_N$ then it makes sense to sum them up and define

$$\mathbf{s}(x, y) = \sum_{j=1}^k \mathbf{s}_j(x, y)$$

to get

$$\begin{aligned} \mathbf{s}(e, y) &= e, \quad \text{for each } y \in A, \\ \mathbf{s}(v, y) &= e, \quad \text{for each } v \in N \text{ and } y \in U \cap d_0/\alpha, \\ \mathbf{s}(v, y) &= v, \quad \text{for each } v \in V \text{ and } y \in U - d_0/\alpha. \end{aligned}$$

Indeed, in the sum defining \mathbf{s} , at most one summand differs from e , namely $\mathbf{s}_j(v, y)$ for the unique j such that $y \in U \cap d_j/\alpha$.

Now we are ready to code each instance of an NP-complete problem in a single equation of \mathbf{A} endowed with some polynomials. As for Theorem 6.4 our proof splits into two cases depending on the size of U/α .

CASE 1. $|U/\alpha| \geq 3$.

In this case with each graph $G = (V, E)$ with $V = \{v_1, \dots, v_n\}$ we associate a polynomial $\mathbf{t}_G(x_1, \dots, x_n) \in \text{Pol}_n \mathbf{A}$ in such a way that G is $|U/\alpha|$ -colorable iff the equation $\mathbf{t}_G(\bar{x}) = 1$ has a solution in \mathbf{A} .

For more readability we define polynomials of the form $v \&_{\mathbf{s}} Y$ acting on $N \times U^m$ as follows

$$v \&_{\mathbf{s}} \{y_1, \dots, y_m\} = \mathbf{s}(\dots \mathbf{s}(\mathbf{s}(v, y_1), y_2) \dots, y_m).$$

Note that if $(v, \bar{y}) \in N \times U^m$ then the value of $v \&_{\mathbf{s}} Y$ does not depend on the order of the y_i 's and in fact we have

$$a \&_{\mathbf{s}} Y = \begin{cases} a, & \text{if } Y \cap d_0/\alpha = \emptyset, \\ e, & \text{otherwise.} \end{cases}$$

Moreover note that for the polynomial $x - y = \mathbf{d}_U(x, y, d_0)$ and $x, y \in U$ we have

$$x - y \in d_0/\alpha \text{ iff } (x, y) \in \alpha,$$

as $\mathbf{d}_U(x, y, d_0)$ is a permutation of U whenever the value for one of the variables x, y is fixed.

Now for a graph G define $\mathbf{t}_G(\bar{x})$ by putting

$$\mathbf{t}_G(x_1, \dots, x_n) = a \&_{\mathbf{s}} \{\mathbf{e}_U(x_i) - \mathbf{e}_U(x_j) : \{v_i, v_j\} \in E\}.$$

From what it was said about $a \&_{\mathbf{s}} Y$ and the difference $-$ on $\mathbf{A}|_N$ it should be clear that the equation $\mathbf{t}_G(\bar{x}) = a$ has a solution in \mathbf{A} iff the elements $\mathbf{e}_U(x_i)$ and $\mathbf{e}_U(x_j)$, corresponding to the edge $\{v_i, v_j\}$, are evaluated in different $\alpha|_U$ -classes, i.e., if G is $(k+1)$ -colorable.

CASE 2. $|U/\alpha| = 2$, i.e., $U = \{d_0, d_1\}$.

Being in this case we start with the following polynomial $\mathbf{w}(v, y_1, y_2, y_3)$ of \mathbf{A} acting on $N \times U^3$ as follows

$$\begin{aligned} \mathbf{w}(v, y_1, y_2, y_3) &= \mathbf{s}(\mathbf{s}(\mathbf{s}(v, y_1), y_2), y_3) \\ &\quad - \mathbf{s}(\mathbf{s}(v, y_1), y_2) - \mathbf{s}(\mathbf{s}(v, y_1), y_3) - \mathbf{s}(\mathbf{s}(v, y_2), y_3) \\ &\quad + \mathbf{s}(v, y_1) + \mathbf{s}(v, y_2) + \mathbf{s}(v, y_3), \end{aligned}$$

where the addition $+$ and the subtraction $-$ is taken in the vector space $\mathbf{A}|_N$. One can easily check that

$$\begin{aligned} \mathbf{w}(e, y_1, y_2, y_3) &= e \text{ for all } y_1, y_2, y_3 \in A, \\ \mathbf{w}(v, y_1, y_2, y_3) &= e \text{ for } v \in N \text{ and } \{y_1, y_2, y_3\} \subseteq U \cap d_0/\alpha, \\ \mathbf{w}(v, y_1, y_2, y_3) &= v \text{ for } v \in N \text{ and } \{y_1, y_2, y_3\} \cap (U - d_0/\alpha) \neq \emptyset. \end{aligned}$$

Analogously to Case 1 we define a polynomials of the form $v \star_{\mathbf{w}} T$ acting on $N \times U^{3m}$, where now $T = \{(y_1^i, y_2^i, y_3^i) : i = 1, \dots, m\}$ is a set of triples of variables, by putting

$$v \star_{\mathbf{w}} T = \mathbf{w}(\dots (\mathbf{w}(\mathbf{w}(v, y_1^1, y_2^1, y_3^1), y_1^2, y_2^2, y_3^2) \dots, y_1^m, y_2^m, y_3^m)).$$

Again, if the variable v is evaluated in N and all the y_j^i 's in U then the value of $v \star_{\mathbf{w}} T$ does not depend on the order of triples in T neither on the order inside the triples. In fact we have

$$a \star_{\mathbf{w}} T = \begin{cases} e, & \text{if there is } j = 1, \dots, m \text{ with } \{y_1, y_2, y_3\} \subseteq U \cap d_0/\alpha, \\ a, & \text{otherwise,} \end{cases}$$

i.e., $a \star_{\mathbf{w}} T$ acts like a conjunction of disjunction of triples. Indeed, a 3-SAT instance

$$\Phi \equiv \bigwedge_{i=1}^m \ell_1^i \vee \ell_2^i \vee \ell_3^i,$$

where $\ell_j^i \in \{x_j^i, \neg x_j^i\}$, can be translated to a polynomial

$$\mathbf{t}_{\Phi}(\bar{x}) = a \star_{\mathbf{w}} \{ \{z_1^i, z_2^i, z_3^i\} : i = 1, \dots, m \},$$

where

$$z_j^i = \begin{cases} \mathbf{e}_U(x_j^i), & \text{if the literal } \ell_j^i \text{ is the variable, i.e., } \ell_j^i = x_j^i, \\ \mathbf{d}_U(d_1, \mathbf{e}_U(x_j^i), d_0), & \text{if } \ell_j^i \text{ is the negated variable, i.e., } \ell_j^i = \neg x_j^i. \end{cases}$$

First note that for $z_j^i = \mathbf{d}_U(d_1, \mathbf{e}_U(x_j^i), d_0)$ we have $z_j^i \in U \cap d_{1-\ell}/\alpha$ whenever $\mathbf{e}_U(x_j^i) \in U \cap d_{\ell}/\alpha$, i.e. $\mathbf{e}_U(x) \mapsto \mathbf{d}_U(d_1, \mathbf{e}_U(x), d_0)$ acts as a negation on the set $\{d_0/\alpha, d_1/\alpha\}$. Moreover, from what has been already said about $a \star_{\mathbf{w}} T$, it should be clear that the equation $\mathbf{t}_{\Phi}(\bar{x}) = a$ has a solution in \mathbf{A} iff Φ is satisfiable. Indeed, it suffices to evaluate the x 's in Φ by the boolean value ℓ iff in the corresponding solution of $\mathbf{t}_{\Phi}(\bar{x}) = a$ they are evaluated in a way that $\mathbf{e}_U(x) \in d_{\ell}/\alpha$. \square

Corollary 7.2. *If a finite algebra \mathbf{A} from a congruence modular variety is solvable but not nilpotent then \mathbf{A} has a homomorphic image \mathbf{A}' with $\text{CSAT}(\mathbf{A}')$ being NP-complete.*

Proof. If \mathbf{A} is solvable but not nilpotent then there is a natural number k such that

$$1 > 1^{(2)} > \dots > 1^{(k)} = 1^{(k+1)} > 0.$$

Now, picking a maximal congruence φ which is not above $1^{(k)}$ we know that φ is meet-irreducible (with the unique cover φ^+) and that the quotient $\mathbf{A}' = \mathbf{A}/\varphi$ is solvable but not nilpotent, as in \mathbf{A}' we have

$$1^{(k)} = 1^{(k+1)} = \varphi^+/\varphi.$$

Now we are in a position to apply Lemma 7.1. \square

8. RESTRICTING LATTICE BEHAVIOR

In this section we study finite algebras from congruence modular varieties such that all prime quotients of its congruences are of lattice type, i.e., of type 4. We will show that if such an algebra \mathbf{A} is not a subdirect product of algebras each of which is polynomially equivalent to the 2-element lattice, then \mathbf{A} has a homomorphic image \mathbf{A}' with $\text{CSAT}(\mathbf{A}')$ being NP-complete.

In the first Lemma of this section we collect some configurations that lead to NP-completeness in type 4 algebras.

Lemma 8.1. *Let \mathbf{A} be a finite algebra from a congruence modular variety such that $\text{typ}\{\mathbf{A}\} \subseteq \{4\}$. If one of the following configuration*

- (1) $\{0, 1\}$ is a minimal set and there are three different elements $a, b, c \in A$ and $\mathbf{f}_a, \mathbf{f}_c \in \text{Pol}_1 \mathbf{A}$ such that $\mathbf{f}_a \binom{1}{0} = \binom{b}{a}$ and $\mathbf{f}_c \binom{1}{0} = \binom{c}{b}$,
- (2) $\{0, 1\}$ is the range of a polynomial $\mathbf{p} \in \text{Pol}_1 \mathbf{A}$, $\{a, b\}$ is a minimal set and there are polynomials $\mathbf{f}, \mathbf{g} \in \text{Pol}_1 \mathbf{A}$ such that $\mathbf{f} \binom{1}{0} = \binom{b}{a}$ and $\mathbf{g} \binom{1}{0} = \binom{a}{b}$,
- (3) $\{0, 1\}$ is a minimal set and one of the sets

$$\begin{aligned} F_1 &= \bigcap \{ \mathbf{f}^{-1}(1) : \mathbf{f} \in \text{Pol}_1 \mathbf{A} \text{ and } \mathbf{f}(A) = \{0, 1\} \}, \\ F_0 &= \bigcap \{ \mathbf{f}^{-1}(0) : \mathbf{f} \in \text{Pol}_1 \mathbf{A} \text{ and } \mathbf{f}(A) = \{0, 1\} \}, \end{aligned}$$

is empty

can be found in \mathbf{A} then $\text{CSAT}(\mathbf{A})$ is NP-complete.

Proof. Suppose we have a configuration described in (1) and let \mathbf{e} be a unary idempotent polynomial with the range $\{0, 1\}$. In this case the required NP-completeness follows from Proposition 4.1 by transforming the system of the two lattice equations:

$$\begin{aligned} \bigwedge_{i=1}^m x_1^i \vee x_2^i \vee x_3^i &= 1, \\ \bigvee_{i=1}^n y_1^i \vee y_2^i \vee y_3^i &= 0 \end{aligned}$$

into a single equation

$$\mathbf{f}_a \left(\bigwedge_{i=1}^m \mathbf{e}(x_1^i) \vee \mathbf{e}(x_2^i) \vee \mathbf{e}(x_3^i) \right) = \mathbf{f}_c \left(\bigvee_{i=1}^n \mathbf{e}(y_1^i) \vee \mathbf{e}(y_2^i) \vee \mathbf{e}(y_3^i) \right)$$

of the algebra \mathbf{A} , where meets and joins are performed in the minimal set $\{0 < 1\}$.

In case (2) we code the 3-SAT instance:

$$\Phi \equiv \bigwedge_{i=1}^m \ell_1^i \vee \ell_2^i \vee \ell_3^i,$$

by the equation

$$\bigwedge_{i=1}^m \delta_1^i \mathbf{p}(z_1^i) \vee \delta_2^i \mathbf{p}(z_2^i) \vee \delta_3^i \mathbf{p}(z_3^i) = b,$$

where

$$\delta_j^i \mathbf{p}(z_j^i) = \begin{cases} \mathbf{fp}(z_j^i), & \text{if the literal } \ell_j^i \text{ is the variable, i.e., } \ell_j^i = x_j^i, \\ \mathbf{gp}(z_j^i), & \text{if } \ell_j^i \text{ is the negated variable, i.e., } \ell_j^i = \neg x_j^i. \end{cases}$$

and meets and joins are taken in the minimal set $\{a < b\}$.

Finally, in case (3) suppose that $F_1 = \emptyset$. This in particular means that the set

$$P = \{\mathbf{f} \in \text{Pol}_1 \mathbf{A} \text{ and } \mathbf{f}(A) = \{0, 1\}\}$$

contains at least two different polynomials. Note that if $\mathbf{g}_1, \mathbf{g}_2 \in P$ then for $\mathbf{g}(x) = \mathbf{g}_1(x) \wedge \mathbf{g}_2(x)$ we have $\mathbf{g}^{-1}(1) = \mathbf{g}_1^{-1}(1) \cap \mathbf{g}_2^{-1}(1)$, and if this intersection is nonempty then also $\mathbf{g} \in P$ as then $\mathbf{g}(A) = \{0, 1\}$. Thus $F_1 = \emptyset$ gives that there are $\mathbf{f}, \mathbf{g} \in P$ with $\mathbf{f}^{-1}(1) \cap \mathbf{g}^{-1}(1) = \emptyset$. Now we can transform the 3-SAT instance Φ into the equation

$$\bigwedge_{i=1}^m \delta_1^i(z_1^i) \vee \delta_2^i(z_2^i) \vee \delta_3^i(z_3^i) = 1$$

where

$$\delta_j^i(z_j^i) = \begin{cases} \mathbf{f}(z_j^i), & \text{if the literal } \ell_j^i \text{ is the variable, i.e., } \ell_j^i = x_j^i, \\ \mathbf{g}(z_j^i), & \text{if } \ell_j^i \text{ is the negated variable, i.e., } \ell_j^i = \neg x_j^i. \end{cases}$$

and meets and joins are taken in the minimal set $\{0 < 1\}$.

The case $F_0 = \emptyset$ can be treated similarly. \square

Endowed with the tools provided by Lemma 8.1 we start enforcing nice lattice behavior of an algebra \mathbf{A} with $\text{typ}\{\mathbf{A}\} = \{\mathbf{4}\}$ by associating with every join irreducible congruence α of \mathbf{A} a binary relation \leq_α which will turn to be a partial order on A whenever $\text{CSAT}(\mathbf{A})$ is not NP-complete. If α is join irreducible then by α^- we denote its unique subcover. Moreover pick $\{0, 1\}$ to be an (α^-, α) -minimal set. Since $\text{typ}(\alpha^-, \alpha) = \mathbf{4}$ we know that $\mathbf{A}|_{\{0,1\}}$ is polynomially equivalent with the 2-element lattice, and without loss of generality we assume that $0 < 1$ in this lattice. We are going to denote this choice of order on this minimal set by typing $\{0 < 1\}$.

Now for $a, b \in A$ put:

$$a \leq_\alpha b \text{ iff there is a polynomial } \mathbf{f} \in \text{Pol}_1 \mathbf{A} \text{ with } \mathbf{f}\binom{1}{0} = \binom{b}{a}.$$

Note that this relation is independent of the choice of the (ordered) (α^-, α) -minimal set $\{0 < 1\}$ as all (α^-, α) -minimal set are polynomially equivalent and this equivalence with $\{0 < 1\}$ propagates the order in the unique way.

Moreover, for further simplicity, we will use the following notation:

- $a <_\alpha b$ iff $a \leq_\alpha b$ and $a \neq b$,
- $a \diamond_\alpha b$ iff $a \leq_\alpha b$ or $b \leq_\alpha a$,
- $a \triangleleft_\alpha b$ iff $a \diamond_\alpha b$ and $a \neq b$,

Lemma 8.2. *Let \mathbf{A} be a finite algebra from a congruence modular variety with $\text{typ}\{\mathbf{A}\} \subseteq \{\mathbf{4}\}$ and let α be a join irreducible congruence of \mathbf{A} . Then*

- (1) $a \triangleleft_\alpha b$, whenever $\{a, b\} \in M_{\mathbf{A}}(\delta, \delta')$ for some $\delta \prec \delta' \leq \alpha$,
- (2) for every $a \in A$ the graph $(a/\alpha, \triangleleft_\alpha)$ is connected,

- (3) *all unary polynomials of \mathbf{A} preserve the relation \leq_α , and all polynomials of \mathbf{A} preserve the transitive closure of \leq_α .*

Proof. To see (1) apply Lemma 3.1 to $(a, b) \in \delta' \leq \alpha = \text{Cg}^{\mathbf{A}}(0, 1)$ to get a chain connecting a with b , where each link in this chain is obtained by projecting the set $\{0, 1\}$ by a unary polynomial. Since the ends of this chain lie in the minimal set $\{a, b\}$ we can apply the unary idempotent polynomial with the range $\{a, b\}$ to put the chain into the set $\{a, b\}$. Obviously at least one link in this chain has to be $\{a, b\}$, which finishes the proof.

To see (2) we recall Lemma 2.17 of [17] which gives that for every prime quotient (δ, δ') each pair $(a, b) \in \delta'$ can be connected via (δ, δ') -traces and δ -links. Now, each link of the form $(c, d) \in \delta$ can be decomposed into a chain of links modulo join irreducible congruences below δ . Thus recursively we get that any pair $(a, b) \in \alpha$ can be connected via traces with respect to the prime quotients of the form (β^-, β) , where β ranges over join irreducible congruences below α . Now, by (1), each such trace is an edge in the graph $(a/\alpha, \diamond_\alpha)$.

Finally, for (3), assume that $a \leq_\alpha b$, i.e. $\mathbf{f} \binom{1}{0} = \binom{b}{a}$ for some $\mathbf{f} \in \text{Pol}_1 \mathbf{A}$. Then obviously for $\mathbf{p} \in \text{Pol}_1 \mathbf{A}$ we have $\binom{\mathbf{p}(b)}{\mathbf{p}(a)} = \mathbf{p} \mathbf{f} \binom{1}{0}$, so that $\mathbf{p}(a) \leq_\alpha \mathbf{p}(b)$. Now if $\mathbf{p} \in \text{Pol}_s \mathbf{A}$ and $a_i \leq_\alpha b_i$ for all $i = 1, \dots, s$ then

$$\begin{aligned} \mathbf{p}(a_1, a_2, a_3, \dots, a_s) &\leq_\alpha \mathbf{p}(b_1, a_2, a_3, \dots, a_s) \\ &\leq_\alpha \mathbf{p}(b_1, b_2, a_3, \dots, a_s) \\ &\leq_\alpha \dots \\ &\leq_\alpha \mathbf{p}(b_1, b_2, b_3, \dots, b_s), \end{aligned}$$

so that $(\mathbf{p}(a_1, a_2, \dots, a_s), \mathbf{p}(b_1, b_2, \dots, b_s))$ lies in the transitive closure of \leq_α . \square

Lemma 8.3. *Let \mathbf{A} be a finite algebra from a congruence modular variety, α be a join irreducible congruence of \mathbf{A} and $\text{typ}\{\mathbf{A}\} \subseteq \{\mathbf{4}\}$. Then either $\text{CSAT}(\mathbf{A})$ is NP-complete or all the following hold:*

- (1) \leq_α is a partial order on A without 3-element chains,
- (2) \leq_α is preserved by all polynomials of \mathbf{A} ,
- (3) for every $a \in A$ the graph $(a/\alpha, \diamond_\alpha)$ is connected and acyclic,
- (4) $a \diamond_\alpha b$ if and only if $\{a, b\} \in M_{\mathbf{A}}(\delta, \delta')$ for some $\delta \prec \delta' \leq \alpha$.

Proof. To see (1) suppose that $a <_\alpha b <_\alpha c$, and that this is witnessed by unary polynomials \mathbf{f}_a and \mathbf{f}_c , i.e. $\mathbf{f}_a \binom{1}{0} = \binom{b}{a}$ and $\mathbf{f}_c \binom{1}{0} = \binom{c}{b}$. Now, if $c \neq a$ refer to Lemma 8.1.(1). If $c = a$ and $\{a, b\}$ is a minimal set in \mathbf{A} , Lemma 8.1.(2) does the job. Now suppose that $c = a$ but $\{a, b\}$ is not a minimal set in \mathbf{A} . Since $(a, b) \in \alpha$ then arguing as in the proof of Lemma 8.2.(2), a and b can be connected via (β^-, β) -minimal sets where β ranges over join irreducible congruences below α . In particular there is $d \in A$ so that $\{a, d\}$ is minimal and therefore, by Lemma 8.2.(1), $\{a, d\} = \{\mathbf{f}(0), \mathbf{f}(1)\}$ for some

$\mathbf{f} \in \text{Pol}_1 \mathbf{A}$. Now either \mathbf{f}, \mathbf{f}_a or \mathbf{f}, \mathbf{f}_c put us into the setting of Lemma 8.1.(1). Summing up, this shows that \leq_α is a partial order without 3 element chains.

For (2) use transitivity of \leq_α and refer to Lemma 8.2.(3).

The last part of the Lemma would follow from (3) and the fact that the binary relation T defined by

$$aTb \text{ iff } \{a, b\} \text{ is a trace with respect to some prime quotient } \delta \prec \delta' \leq \alpha$$

is connected on α -classes and is contained in \diamond_α (see Lemma 8.2). Since \diamond_α is acyclic \diamond_α has to coincide with T .

The hardest part of the Lemma is to show (3). It can be inferred from Theorem 3.6 in [30] but we decided to include our proof which seems to be more direct.

Thus suppose to the contrary that C is a cycle in the bipartite graph (A, \diamond_α) . Thus $|C|$ is even and $|C| \geq 4$.

Moreover let $\mathbf{D}_1(x, y, z), \dots, \mathbf{D}_n(x, y, z), \mathbf{Q}(x, y, z)$ be the directed Gumm terms with the properties described in Theorem 3.2. Our goal is to show that

$$(8.1) \quad x = \mathbf{D}_1(x, y, z) = \dots = \mathbf{D}_n(x, y, z) \text{ for all } x, y, z \in C.$$

Given (8.1) we know that $\mathbf{Q}(x, y, y) = x$ whenever $x, y \in C$. This together with the last equality in Theorem 3.2 gives that \mathbf{Q} is a polynomial that behaves like a Malcev operation on cycle C . Now, picking $a, b \in C$ with $a <_\alpha b$ we get $b = \mathbf{Q}(b, a, a) \leq_\alpha \mathbf{Q}(b, b, a) = a$, a contradiction.

In order to prove (8.1) we will simplify notation by omitting the subscript α in $\leq_\alpha, <_\alpha, \diamond_\alpha, \diamond_\alpha$. Instead we will introduce the notation \leq^ℓ for $\ell \in \{1, -1\}$ where \leq^1 stays for \leq and \leq^{-1} for \geq and we will use the notation $a \diamond_C b$ to denote that $a \diamond b$ and $a, b \in C$. Moreover for a, b in the same α -class we define:

- $\text{dist}(a, b)$ to be the distance of a and b in the graph $(a/\alpha, \diamond)$.

If $a, b, c \in C$ we put

- $\text{dist}_C(a, b)$ to be the distance of a and b in the graph (C, \diamond_C) ,
- $\text{dist}_c(a, b)$ to be the length of the shortest path between a and b fully contained in C and containing the vertex c ,

Suppose to the contrary with (8.1) that there are $a, b, c \in C$ with $\mathbf{D}_1(a, b, c) \neq a$. This configuration will allow us to construct the sequence of triples $(a_i, b_i, c_i)_{i=0}^k$ of vertices in C , so that after putting $d_i = \mathbf{D}_1(a_i, b_i, c_i)$, the following invariants will be kept:

$$(8.2) \quad \text{dist}(b_i, c_i) \text{ is even,}$$

$$(8.3) \quad \text{dist}_{a_{i+2}}(b_{i+2}, c_{i+2}) < \text{dist}_{a_i}(b_i, c_i),$$

$$(8.4) \quad d_i \neq a_i,$$

$$(8.5) \quad a_k \in \{b_k, c_k\}.$$

The last item gives $d_k = \mathbf{D}_1(a_k, b_k, c_k) = a_k$ contrary to (8.4).

First we will define the triple (a_0, b_0, c_0) . If $\text{dist}(b, c)$ is even then we simply put $(a_0, b_0, c_0) = (a, b, c)$ so that (8.2) and (8.4) hold. If $\text{dist}(b, c)$ is odd, then exactly one of the distances $\text{dist}(a, b)$, $\text{dist}(a, c)$ is even. Suppose this is $\text{dist}(a, b)$. We then move b to its neighbor $b_0 \diamond_C b$ while (a_0, c_0) is set to (a, b) . Obviously $\text{dist}(b_0, c_0)$ is now even. Moreover $b_0 <^\ell b$ for some $\ell \in \{1, -1\}$. In particular both b and a are $<^\ell$ -maximal. Applying (2) to $b_0 <^\ell b$ we get $\mathbf{D}_1(a, b_0, c) \leq^\ell \mathbf{D}_1(a, b, c) \neq a$. Thus either $d_0 = \mathbf{D}_1(a, b_0, c) = \mathbf{D}_1(a, b, c) \neq a = a_0$ or $d_0 = \mathbf{D}_1(a, b_0, c) <^\ell \mathbf{D}_1(a, b, c)$ is not $<^\ell$ -maximal, while $a_0 = a$ is. This shows (8.4).

Now, as long as $a_i \in \{b_i, c_i\}$ fails, we create the next triple $(a_{i+1}, b_{i+1}, c_{i+1})$ by either moving a_i along an edge and keeping b_i, c_i untouched, or by moving simultaneously b_i and c_i towards a_i which stays unchanged. More formally:

- Case 1:** if $\text{dist}(a_i, d_i) = 1$ then put
- $a_{i+1} \diamond_C a_i$ such that $\text{dist}(a_{i+1}, d_i) = 2$,
 - $b_{i+1} = b_i$ and $c_{i+1} = c_i$.
- Case 2:** if $\text{dist}(a_i, d_i) \geq 2$ then put
- $a_{i+1} = a_i$,
 - $b_{i+1} \diamond_C b_i$ and $c_{i+1} \diamond_C c_i$ such that $\text{dist}_{a_i}(b_{i+1}, c_{i+1}) \leq \text{dist}_{a_i}(b_i, c_i) - 2$.

Note that the very last inequality is strong only if the initial situation is $a_i \diamond_C b_i = c_i$ and results either in $a_i = b_{i+1} \diamond_C b_i \diamond_C c_{i+1}$ or in $a_i = c_{i+1} \diamond_C c_i \diamond_C b_{i+1}$. Indeed, in this case $\text{dist}_{a_i}(b_i, c_i) = |C| \geq 4$ while $\text{dist}_{a_i}(b_{i+1}, c_{i+1}) \in \{0, 2\}$.

Being in CASE 1 we know that $d_i <^\ell a_i$ for some $\ell \in \{1, -1\}$. Then $a_{i+1} \in C$ is chosen so that $a_i >^\ell a_{i+1} \neq d_i$. Thus we have $d_{i+1} \leq^\ell d_i$ and this inequality cannot be strong as then we would have a 3-element path $d_{i+1} <^\ell d_i <^\ell a_i$, contrary to (1).

Moreover, after CASE 1 is performed we fall into CASE 2. Thus in each two consecutive rounds the distance between b_i and c_i through a_i decreases by at least 2, so that the invariant (8.3) is kept. Since in CASE 2 we move both b_i and c_i to their neighbors, (8.2) holds as well. To see (8.4) note that $b_{i+1} <^\ell b_i$ and $c_{i+1} <^\ell c_i$ for the very same $\ell \in \{1, -1\}$. Thus $d_{i+1} \leq^\ell d_i$ while $a_{i+1} = a_i$, which together with $\text{dist}(a_i, d_i) \geq 2$ gives $d_{i+1} \neq a_{i+1}$.

Finally note that (8.3) gives that there is $k \leq \text{dist}_{a_0}(b_0, c_0)$ for which (8.5) holds.

This finishes the proof that $\mathbf{D}_1(a, b, c) = a$ whenever $a, b, c \in C$. Now we can repeat recursively this argument for $\mathbf{D}_2, \dots, \mathbf{D}_n$, so that (8.1), and therefore (3) is shown. \square

Now we are ready to establish quite strong property enforced by tractability of CSAT.

Theorem 8.4. *If \mathbf{A} is a finite subdirectly irreducible algebra from a congruence modular variety and $\text{typ}\{\mathbf{A}\} = \{4\}$ then either $|A| = 2$ or $\text{CSAT}(\mathbf{A})$ is NP-complete.*

Proof. To be able to use the properties of the partial orders \leq_α established in Lemma 8.3 we assume that $\text{CSAT}(\mathbf{A})$ is not NP-complete.

Let μ be the monolith of \mathbf{A} and Υ be the set of all join irreducible congruences of \mathbf{A} . Pick a $(0, \mu)$ -minimal set $N = \{0, 1\}$. Then for every $\alpha \in \Upsilon$ and each (α^-, α) -minimal set U there is a unary polynomial \mathbf{f}_α such that $\mathbf{f}_\alpha(U) = \{0, 1\}$. Note that any unary polynomial \mathbf{g} for which $\mathbf{g}(U) = \{0, 1\}$ we have $\mathbf{g}|_U = \mathbf{f}_\alpha|_U$, as otherwise we would have $0 <_\alpha 1 <_\alpha 0$, contrary to Lemma 8.3. This allows us to name the elements of an (α^-, α) -minimal set by $0_\alpha, 1_\alpha$ by requiring $\mathbf{f}_\alpha \begin{pmatrix} 1_\alpha \\ 0_\alpha \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Note that this naming is independent of the choice of polynomials \mathbf{f}_α . Now the ordering $0_\alpha < 1_\alpha$ determines the (bipartite) order \leq_α with the properties described in Lemma 8.3. Moreover the orderings of the form \leq_α are coherent in the following sense:

$$(8.6) \text{ for } \alpha, \beta \in \Upsilon \text{ and } (a, b) \in \alpha \cap \beta \text{ we have } a \leq_\alpha b \text{ iff } a \leq_\beta b.$$

Indeed suppose that $a <_\alpha b$ and $a >_\beta b$. This gives that $\mathbf{g}_\alpha \begin{pmatrix} 1_\alpha \\ 0_\alpha \end{pmatrix} = \begin{pmatrix} b \\ a \end{pmatrix} = \mathbf{g}_\beta \begin{pmatrix} 0_\beta \\ 1_\beta \end{pmatrix}$ for some unary polynomials $\mathbf{g}_\alpha, \mathbf{g}_\beta$. On the other hand the pair $\begin{pmatrix} b \\ a \end{pmatrix}$ can be polynomially mapped, by say \mathbf{f} , onto $\{0, 1\}$. Now either $\mathbf{f} \mathbf{g}_\alpha \begin{pmatrix} 1_\alpha \\ 0_\alpha \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ or $\mathbf{f} \mathbf{g}_\beta \begin{pmatrix} 0_\beta \\ 1_\beta \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, contrary to our previous choices of orders in minimal sets.

Our first goal is to show that

$$(8.7) \text{ the transitive closure } \leq \text{ of the sum } \bigcup_{\alpha \in \Upsilon} \leq_\alpha \text{ is a connected partial order } A \text{ and it is preserved by all polynomials of } \mathbf{A}.$$

The only obstacle for \leq to be a partial order is the existence of a cycle of the form

$$a_0 <_{\alpha_1} a_1 <_{\alpha_2} a_2 \dots a_{k-1} <_{\alpha_k} a_k <_{\alpha_0} a_0.$$

We know that $U = \{a_0, a_k\}$ is a (β^-, β) -minimal set for some join irreducible congruence $\beta \leq \alpha_0$. Applying unary idempotent polynomial \mathbf{e}_U , with the range U , to such a cycle we get

$$\mathbf{e}_U(a_0) \leq_{\alpha_1} \mathbf{e}_U(a_1) \leq_{\alpha_2} \mathbf{e}_U(a_2) \dots \mathbf{e}_U(a_{k-1}) \leq_{\alpha_k} \mathbf{e}_U(a_k) \leq_{\alpha_0} \mathbf{e}_U(a_0).$$

We induct on $j = 0, 1, \dots, k$ to show that $\mathbf{e}_U(a_j) = a_0$. First note that (8.6) applied to $a_k <_{\alpha_0} a_0$ gives $a_0 \not\leq_{\alpha_{j+1}} a_k$. However, by the induction hypothesis $a_0 = \mathbf{e}_U(a_j) \leq_{\alpha_{j+1}} \mathbf{e}_U(a_{j+1}) \in \{a_0, a_k\}$, so that we must have $\mathbf{e}_U(a_{j+1}) = a_0$, as required. But now we have $a_0 = \mathbf{e}_U(a_k) = a_k$, an obvious contradiction.

To see that the partial order \leq is connected and preserved by the polynomials simply recall the arguments used in the proof of Lemma 8.2.(2) and Lemma 8.3.(2).

We are working under the assumption that $\text{CSAT}(\mathbf{A})$ is not NP-complete. Thus, defining the sets F_v with $v \in N = \{0, 1\}$ by putting

$$F_v = \bigcap \{ \mathbf{f}^{-1}(v) : \mathbf{f} \in \text{Pol}_1 \mathbf{A} \text{ and } \mathbf{f}(A) = N \},$$

Lemma 8.1.(3) allows us to assume that both F_0 and F_1 are nonempty. Pick an ‘upper’ element $u \in F_1$ and a ‘lower’ element $d \in F_0$. The names for them are justified by the following observation.

(8.8) For every $\mathbf{f} \in \text{Pol}_1 \mathbf{A}$ the element $\mathbf{f}(u)$ is maximal in $\mathbf{f}(A)$, while $\mathbf{f}(d)$ is minimal in $\mathbf{f}(A)$. In particular, u is maximal, while d is minimal in the poset (A, \leq) .

Indeed, otherwise there is $w \in A$ such that $\mathbf{f}(u) <_\alpha \mathbf{f}(w)$ for some $\alpha \in \Upsilon$. But then $\{\mathbf{f}(u), \mathbf{f}(w)\}$ is a minimal set and using the consistency of our partial order one can polynomially project, say by \mathbf{g} , the pair $\begin{pmatrix} \mathbf{f}(w) \\ \mathbf{f}(u) \end{pmatrix}$ onto the pair $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Composing this with a unary idempotent polynomial \mathbf{e} with the range N we have $\mathbf{egf}\begin{pmatrix} w \\ u \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, contrary to our choice of u .

By the very same token one shows the properties of d .

We will show that in fact there are no other minimal or maximal elements in (A, \leq) .

(8.9) u is the largest element in the poset (A, \leq) , while d is the smallest one.

By symmetry of our assumptions we can restrict ourselves to show that u is the largest element. Suppose to the contrary that there is another maximal element in (A, \leq) . Since the poset is connected we may assume that there are elements $b, c \in A$ with $u > b < c$ and c being maximal.

We will be using directed Gumm terms $\mathbf{D}_1, \dots, \mathbf{D}_n, \mathbf{Q}$, provided by Theorem 3.2, to define unary polynomials

$$\mathbf{f}_i(x) = \mathbf{D}_i(x, b, c), \quad \text{for all } i = 1, \dots, n$$

and show that they satisfy:

- (i) $\mathbf{f}_i(c) = c$,
- (ii) $\mathbf{f}_i(b) \neq c$,
- (iii) $\mathbf{f}_i(u) \neq c$.

The item (i) follows directly from the properties of the \mathbf{D}_i 's. Moreover, for $i = 1$ the item (ii) is secured by $\mathbf{f}_1(b) = \mathbf{D}_1(b, b, c) = b < c$, while the failure of (iii) would lead to a contradiction

$$c = \mathbf{f}_1(u) = \mathbf{D}_1(u, b, c) \leq \mathbf{D}_1(u, u, c) = u.$$

The failure of any of the items (ii) or (iii) at the level $(i + 1)$ would give one of the following

$$\begin{aligned} c &= \mathbf{f}_{i+1}(b) = \mathbf{D}_{i+1}(b, b, c) \leq \mathbf{D}_{i+1}(u, u, c), \\ c &= \mathbf{f}_{i+1}(u) = \mathbf{D}_{i+1}(u, b, c) \leq \mathbf{D}_{i+1}(u, u, c). \end{aligned}$$

In each case the maximality of c yields

$$c = \mathbf{D}_{i+1}(u, u, c) = \mathbf{D}_i(u, c, c) \geq \mathbf{D}_i(u, b, c) = \mathbf{f}_i(u),$$

and now the maximality of $\mathbf{f}_i(u)$ gives $\mathbf{f}_i(u) = c$, a contradiction with the induction hypothesis.

Now,

$$c = \mathbf{Q}(b, b, c) \leq \mathbf{Q}(u, c, c) = \mathbf{D}_n(u, c, c),$$

together with maximality of c gives

$$c = \mathbf{D}_n(u, c, c) \geq \mathbf{D}_n(u, b, c) = \mathbf{f}_n(u),$$

so that maximality of $\mathbf{f}_n(u)$ gives $\mathbf{f}_n(u) = c$, contrary to (iii). This contradiction shows (8.9).

To conclude our proof we will strengthen (8.9) to:

$$(8.10) \quad A = \{u, d\}.$$

First suppose that $\{0', 1'\}$ is a $(0, \mu)$ -minimal set. Thus for every $\alpha \in \Upsilon$ we have $d \leq 0' <_\alpha 1' \leq u$. Now if $d < 0'$ or $1' < u$ then we can pick an element a such that either $a <_\alpha 0'$ or $1' <_\alpha a$ for some $\alpha \in \Upsilon$. In any case we will have a 3-element directed path in the poset $(A, <_\alpha)$ which is not possible in view of Lemma 8.3.(1). Thus $\{d, u\}$ is the only $(0, \mu)$ -minimal set of \mathbf{A} .

Now, suppose there are elements $a, b \in A$ such that $d \leq a <_\alpha b \leq u$. Then $\{a, b\}$ is a minimal set which has to be the range of some unary idempotent polynomial \mathbf{e} of \mathbf{A} . But then $\mathbf{e}(d) \leq \mathbf{e}(a) < \mathbf{e}(b) \leq \mathbf{e}(u)$ and consequently monotonicity of \mathbf{e} gives $\mathbf{e}\left(\begin{smallmatrix} d \\ u \end{smallmatrix}\right) = \begin{smallmatrix} a \\ b \end{smallmatrix}$. In particular $(a, b) \in \mu$ and in fact the set $\{a, b\}$ is $(0, \mu)$ -minimal, so that $\{a, b\} = \{d, u\}$. This shows that A can not have any other elements than d or u . \square

Directly from Theorem 8.4 we get

Corollary 8.5. *Let \mathbf{A} be a finite algebra from a congruence modular variety and $\text{typ}\{\mathbf{A}\} = \{4\}$. Then either \mathbf{A} is a subdirect product of 2-element algebras each of which is polynomially equivalent to the 2-element lattice, or \mathbf{A} has a subdirectly irreducible homomorphic image \mathbf{A}' such that $\text{CSAT}(\mathbf{A}')$ is NP-complete. \square*

The next example shows that a subdirect product of 2-element algebras each of which is polynomially equivalent to the 2-element lattice need not be polynomially equivalent to a distributive lattice.

Example 8.6. *Let $\mathbf{A} = (A, \mathbf{m})$ be a subreduct of $(\{0, 1\}, \wedge, \vee)^3$, with*

- $A = \{(1, 1, 1), (0, 1, 1), (1, 0, 1), (1, 1, 0)\}$
- and \mathbf{m} being the majority operation $\mathbf{m}(x, y, z) = (x \vee y) \wedge (y \vee z) \wedge (z \vee x)$.

Then

- \mathbf{A} belongs to congruence distributive variety
- \mathbf{A} is a subdirect product of algebras polynomially equivalent to two element lattices,
- \mathbf{A} is not polynomially equivalent to a distributive lattice.

Proof. The first two items are obvious. To see the third one note that, up to isomorphism, there are only two four element lattices:

- the four element chain,
- the four element Boolean lattice.

On the other hand, for three pairwise different elements $a, b, c \in A$ we have $\mathbf{m}(a, b, c) = \bar{1}$, where $\bar{1} = (1, 1, 1)$. Sending isomorphically, say by h , all possible 3-element tuples from A into one of the above 4-element lattices we simply cannot find a room for $h(\bar{1})$ under the assumption that \mathbf{m} preserves lattice order. \square

9. POLYNOMIAL TIME ALGORITHMS

Combining Corollaries 6.5, 7.2 and 8.5 we get the following Theorem.

Theorem 9.1. *Let \mathbf{A} be a finite algebra from a congruence modular variety such that $\text{CSAT}(\mathbf{A}')$ is not NP-complete for every quotient \mathbf{A}' of \mathbf{A} . Then \mathbf{A} is isomorphic to a direct product $\mathbf{N} \times \mathbf{D}$, where \mathbf{N} is a nilpotent algebra and \mathbf{D} is a subdirect product of 2-element algebras each of which is polynomially equivalent to the 2-element lattice.* \square

The aim of this section is to prove a partial converse to Theorem 9.1 where nilpotency is strengthened to supernilpotency.

Theorem 9.2. *Let \mathbf{A} be a finite algebra from a congruence modular variety that decomposes into a direct product $\mathbf{N} \times \mathbf{D}$, where \mathbf{N} is a supernilpotent algebra and \mathbf{D} is a subdirect product of 2-element algebras each of which is polynomially equivalent to the 2-element lattice. Then for every quotient \mathbf{A}' of \mathbf{A} the problem $\text{CSAT}(\mathbf{A}')$ is solvable in polynomial time.*

Before proving this Theorem note that if an algebra \mathbf{A} decomposes into a direct product of the form described above then all its quotients decompose in the very same way. This is an immediate consequence of the fact that the product of the form $\mathbf{N} \times \mathbf{D}$ has no skew congruences which in turn follows from $\text{typ}\{\mathbf{D}\} \subseteq \{\mathbf{4}\}$.

The proof of Theorem 9.2 splits into two parts. We show that for both factors of \mathbf{A} the problem has polynomial time solution. Actually we will show that in both cases if the polynomial equation $\mathbf{t}(x_1, \dots, x_n) = \mathbf{s}(x_1, \dots, x_n)$ has a solution in A^n then it has a solution in a relatively small subset S of A^n , namely in a subset of size bounded by a polynomial in n . The reader should be however warned here that we are not going to show that all solutions are contained in this small set S .

Theorem 9.3. *Let \mathbf{D} be a subdirect product of finitely many 2-element algebras each of which is polynomially equivalent to the 2-element lattice. Then $\text{CSAT}(\mathbf{D})$ is solvable in polynomial time.*

Proof. The basic observation is that for the 2-element lattice \mathbf{D} , and therefore for every algebra polynomially equivalent to the 2-element lattice, the problem $\text{CSAT}(\mathbf{D})$ is solvable in polynomial time by a very special algorithm.

Indeed, if $\mathbf{t}, \mathbf{s} \in \text{Pol } \mathbf{D}$ the equation $\mathbf{t}(\bar{x}) = \mathbf{s}(\bar{x})$ has a solution, say (a_1, \dots, a_n) , then both $\mathbf{t}(a_1, \dots, a_n)$ and $\mathbf{s}(a_1, \dots, a_n)$ have the same value a . But for a polynomial \mathbf{t} over the 2-element lattice one can easily show, that if $\mathbf{t}(a_1, \dots, a_n) = a$ then $\mathbf{t}(a, \dots, a) = a$. Indeed, by the monotonicity of the polynomials of \mathbf{D} we have

$$\mathbf{t}(0, \dots, 0) \leq \mathbf{t}(a_1, \dots, a_n) \leq \mathbf{t}(1, \dots, 1)$$

and if $\mathbf{t}(a_1, \dots, a_n) = 0$ then $\mathbf{t}(0, \dots, 0)$ has to be 0 as well. Similarly $\mathbf{t}(a_1, \dots, a_n) = 1$ implies $\mathbf{t}(1, \dots, 1) = 1$.

Therefore, to determine if $\mathbf{t}(\bar{x}) = \mathbf{s}(\bar{x})$ has a solution over \mathbf{D} it suffices to show whether $\mathbf{t}(a, \dots, a) = \mathbf{s}(a, \dots, a)$ for some $a \in D$.

We say that an algebra \mathbf{A} has Uniform Solution Property, or USP for short, if for every polynomial $\mathbf{t}(\bar{x}) \in \text{Pol } \mathbf{A}$ and $a \in A$

$$(\exists \bar{x} \ \mathbf{t}(x_1, \dots, x_n) = a) \Rightarrow \mathbf{t}(a, \dots, a) = a$$

What we have just shown is that the 2-element lattice has USP, and that $\text{CSAT}(\mathbf{A})$ is polynomially time solvable for every finite algebra \mathbf{A} with USP.

Now we can conclude the proof by noting that a subdirect product of algebras with USP, has USP itself. Actually USP is preserved under forming homomorphic images, subalgebras, products or reducts. \square

The reduction of searching a solution of an equation in supernilpotent realm to a relatively small set is much more involved than in lattice case. Our proof is modeled after the Ramsey type argument introduced by Mikael Goldmann and Alexander Russell in [11] for nilpotent groups, and later cleaned up by Gábor Horváth [18] in the realm of nilpotent groups and nilpotent rings.

Theorem 9.4. *Let \mathbf{A} be a finite supernilpotent algebra from a congruence modular variety. Then $\text{CSAT}(\mathbf{A})$ is solvable in polynomial time.*

Proof. Suppose now that \mathbf{A} is a finite nilpotent algebra from a congruence modular variety. Then \mathbf{A} generates a variety in which every algebra is nilpotent, and therefore congruence permutable. In particular \mathbf{A} has a Malcev term $\mathbf{d}(x, y, z)$.

From now on we fix an element 0 of \mathbf{A} and define a binary operation $+$ by putting:

$$x + y = \mathbf{d}(x, 0, y).$$

Unfortunately the binary operation $+$ does not need to be associative. Thus in longer sums we adopt the convention of associating to the left. More formally, if $\langle a_1, a_2, \dots, a_\ell \rangle$ is an ordered list of elements of \mathbf{A} then by $\sum \langle a_1, a_2, \dots, a_\ell \rangle$ we mean $((a_1 + a_2) + a_3) + \dots + a_\ell$.

According to Corollary 7.4 in [10], we know that for every $b, c \in A$ the function $x \mapsto \mathbf{d}(x, b, c)$ is a permutation of A . In particular

$$\mathbf{d}(x, y, 0) = 0 \quad \text{iff} \quad x = y.$$

Thus, each equation $\mathbf{t}(\bar{x}) = \mathbf{s}(\bar{x})$ can be equivalently replaced by an equation of the form $\mathbf{w}(\bar{x}) = 0$, where $\mathbf{w}(\bar{x}) = \mathbf{d}(\mathbf{t}(\bar{x}), \mathbf{s}(\bar{x}), 0)$ has linear size in terms of the size of the original equation.

Our polynomial time algorithm for checking whether $\mathbf{w}(\bar{x}) = 0$ has a solution is based on the following phenomena of supernilpotent algebras:

- (9.1) For every finite supernilpotent algebra \mathbf{A} there is a positive integer d such that every equation of the form $\mathbf{w}(\bar{x}) = 0$ has a solution iff it has a solution with at most d non-zero values for the x_i 's, i.e. $\#\|\bar{x} \neq 0\| \leq d$.

Given (9.1) we simply check if $\mathbf{w}(x_1, \dots, x_n) = 0$ has a solution among $\binom{n}{d} \cdot |A|^d$ possible evaluations of the x_i 's with $\#\|\bar{x} \neq 0\| \leq d$. Unfortunately the degree d of the polynomial bounding the run time of the algorithm can be really huge, as it is obtained by a Ramsey type argument applied to the numbers:

- k – the degree of supernilpotency of the algebra \mathbf{A} ,
- $C = |A|^{k \cdot |A|}$,
- $m = (k - 1)! \cdot |A|$

to get that:

- (9.2) There is a positive integer d such that for every set S with $|S| \geq d$ and every coloring of all at most $(k - 1)$ -element subsets of S with C colors there exists m -element subset T of S such that all at most $(k - 1)$ -element subsets of T with the same number of elements have the same color.

For a proof of the above statement we refer e.g. to Theorem 2, Chapter 1 in the monograph [14].

Now, to prove (9.1) we will show that

- (9.3) Each solution $\bar{b} = b_1, \dots, b_n$ of $\mathbf{w}(x_1, \dots, x_n) = 0$ with $\#\|\bar{b} \neq 0\| > d$ can be replaced with a solution $\bar{b}' = b'_1, \dots, b'_n$ with $\#\|\bar{b}' \neq 0\| = \#\|\bar{b} \neq 0\| - m$.

For the rest of the proof we fix a solution $\bar{b} = b_1, \dots, b_n$ of $\mathbf{w}(x_1, \dots, x_n) = 0$. Now, a very careful reading of Chapter XIV of [10], especially the proof of Lemma 14.6, allows us to represent $\mathbf{w}(\bar{x})$ as $\sum \langle \mathbf{w}_1(\bar{x}), \dots, \mathbf{w}_k(\bar{x}) \rangle$, where k is the degree of supernilpotency, and therefore nilpotency, of \mathbf{A} and each $\mathbf{w}_\ell(\bar{x})$ has the form

$$\mathbf{w}_\ell(\bar{x}) = \sum \langle c_\ell, \mathbf{t}_{\ell,1}(\bar{x}), \dots, \mathbf{t}_{\ell,n_\ell}(\bar{x}) \rangle,$$

with

- $c_\ell \in A$,
- $\mathbf{t}_{\ell,j}(\bar{x}) = 0$ whenever $x_i = 0$ for at least one $i \in \text{Ess}(\mathbf{t}_{\ell,j})$, where $\text{Ess}(\mathbf{t}_{\ell,j})$ is the set of numbers of variables on which $\mathbf{t}_{\ell,j}$ essentially depends,

- for each $a \in A$ the sublist $J_{\ell,a}$ of $\langle 1, \dots, n_\ell \rangle$ consisting of the j 's for which $\mathbf{t}_{\ell,j}(\bar{b}) = a$ is convex in $\langle 1, \dots, n_\ell \rangle$.

The second item above simply means that the $\mathbf{t}_{\ell,j}$'s are commutator expressions and therefore our assumption that \mathbf{A} is k -supernilpotent gives

- $|\text{Ess}(\mathbf{t}_{\ell,j})| < k$.

Now, suppose that the set $S = \|\bar{b} \neq 0\|$ is too big, i.e. $\#S > d$. Define a coloring φ of (at most $(k-1)$ -element) subsets of S by C colors as follows. For a subset $I \subseteq S$ its color φ_I is set to be a function of the form $\{1, \dots, k\} \times A \rightarrow \{0, 1, \dots, |A| - 1\}$ determined by

$$\varphi_I(\ell, a) = \#\{j \in J_{\ell,a} : I \subseteq \text{Ess}(\mathbf{t}_{\ell,j})\} \pmod{|A|}.$$

Now, (9.2) supplies us with $T \subseteq S$ such that $|T| = m$ and $\varphi_{I_1} = \varphi_{I_2}$ whenever $I_1, I_2 \subseteq T$ and $|I_1| = |I_2| < k$.

We modify \bar{b} to \bar{b}' by zeroing the x_i 's with $i \in T$, i.e.

$$b'_i = \begin{cases} 0, & \text{if } i \in T, \\ b_i, & \text{otherwise.} \end{cases}$$

Obviously $\#\|\bar{b}' \neq 0\| = \#\|\bar{b} \neq 0\| - m$, as required in (9.3). To prove that $\mathbf{w}(b'_1, \dots, b'_n) = 0$ we will show that

(9.4) for each $\ell = 1, \dots, k$ and $a \in A$ we have

$$\sum \langle c, \dots, \mathbf{t}_{\ell,j}(\bar{b}'), \dots \rangle_{j \in J_{\ell,a}} = \sum \langle c, \dots, \mathbf{t}_{\ell,j}(\bar{b}), \dots \rangle_{j \in J_{\ell,a}}.$$

Note that in the sum on the left hand side some of the summands switched from a to 0 (if $a \neq 0$). Let Z collect the numbers j of such summands. Obviously $Z = \{j \in J_{\ell,a} : T \cap \text{Ess}(\mathbf{t}_{\ell,j}) \neq \emptyset\}$. We will show that

(9.5) $|Z|$ is divisible by $|A|$.

Given (9.5) we argue towards (9.4) as follows. We have already noticed that $x \mapsto x + a = \mathbf{d}(x, 0, a)$ is a permutation of A . Let σ be the order of this permutation, so that

$$\sum \langle x, \underbrace{a, \dots, a}_{\sigma \text{ times}} \rangle = x.$$

Moreover the fact that $x + 0 = \mathbf{d}(x, 0, 0) = x$ allows us to omit all the 0's in the lists under the sums. This gives the first equality in the display below.

$$\begin{aligned} \sum \langle c, \dots, \mathbf{t}_{\ell,j}(\bar{b}'), \dots \rangle_{j \in J_{\ell,a}} &= \sum \langle c, \underbrace{a, \dots, a}_{(|J_{\ell,a}| - |Z|) \text{ times}} \rangle \\ &= \sum \langle c, \underbrace{a, \dots, a}_{|J_{\ell,a}| \text{ times}} \rangle \\ &= \sum \langle c, \dots, \mathbf{t}_{\ell,j}(\bar{b}), \dots \rangle_{j \in J_{\ell,a}} \end{aligned}$$

The second equality in this display uses the fact that σ divides $|Z|$, which follows from (9.5) and Lemma 14.7 in [10] telling us that σ divides $|A|$.

Thus, we are left with the proof of (9.5). To do this, for $I \subseteq T$ define

$$Z_I = \{j \in J_{\ell, a} : I \subseteq \text{Ess}(\mathbf{t}_{\ell, j})\}$$

and observe that

$$Z = \bigcup_{i \in T} Z_{\{i\}} \quad \text{and} \quad Z_I = \bigcap_{i \in I} Z_{\{i\}}.$$

Thus the inclusion-exclusion principle, together with $|\text{Ess}(\mathbf{t}_{\ell, j})| < k$, gives

$$|Z| = \sum_{\substack{I \subseteq T \\ 0 < |I| < k}} (-1)^{|I|+1} \cdot |Z_I|.$$

However we know that $|Z_I|$ is (modulo $|A|$) nothing else but $\varphi_I(\ell, a)$. Since φ_I depends only on the size of I we put $\zeta_q = \varphi_I(\ell, a)$ for $|I| = q$ to get that modulo $|A|$ we have

$$|Z| = \sum_{q=1}^{k-1} (-1)^{q+1} \binom{m}{q} \cdot \zeta_q.$$

To conclude the proof that $|Z|$ is divisible by $|A|$ observe that for $q = 1, \dots, k-1$ all the binomial coefficients $\binom{m}{q}$ are divisible by $|A|$, as m was set to be $(k-1)! \cdot |A|$. \square

10. SIMULTANEOUS SATISFIABILITY OF MANY CIRCUITS

This section is devoted to the problem MCSAT. All we have to do is to prove Corollary 2.10.

Corollary 2.10. *Let \mathbf{A} be a finite algebra from a congruence modular variety.*

- (1) *If \mathbf{A} has no quotient \mathbf{A}' with $\text{MCSAT}(\mathbf{A}')$ being NP-complete then \mathbf{A} is isomorphic to a direct product $\mathbf{M} \times \mathbf{D}$, where \mathbf{M} is an affine algebra and \mathbf{D} is a subdirect product of 2-element algebras each of which is polynomially equivalent to the 2-element lattice.*
- (2) *If \mathbf{A} decomposes into a direct product $\mathbf{M} \times \mathbf{D}$, where \mathbf{M} is an affine algebra and \mathbf{D} is a subdirect product of 2-element algebras each of which is polynomially equivalent to the 2-element lattice, then for every quotient \mathbf{A}' of \mathbf{A} the problem $\text{CSAT}(\mathbf{A}')$ is solvable in polynomial time.*

Proof. First note that every instance of $\text{CSAT}(\mathbf{A})$ is also a instance of $\text{MCSAT}(\mathbf{A})$, so that $\text{MCSAT}(\mathbf{A})$ is NP-complete whenever $\text{CSAT}(\mathbf{A})$ is NP-complete. Moreover, $\text{MCSAT}(\mathbf{A})$ can be treated as a problem of satisfiability of systems of equations of the form

$$\mathbf{g}_1(x_1, \dots, x_n) = \mathbf{g}_2(x_1, \dots, x_n) = \dots = \mathbf{g}_k(x_1, \dots, x_n).$$

Thus $\text{MCSAT}(\mathbf{A})$ is in fact a subproblem of $\text{SCSAT}(\mathbf{A})$, so that Gaussian elimination type algorithms for affine \mathbf{A} 's show that in this case $\text{MCSAT}(\mathbf{A}) \in$

P (see also Theorem 2.6). Therefore, by Theorem 2.9 we are left with the following two classes of algebras:

- nilpotent non-affine algebras,
- subdirect products of algebras each of which is polynomially equivalent to the 2-element lattice.

In case \mathbf{A} is nilpotent we can easily interpret $\text{SCSAT}(\mathbf{A})$ into $\text{MCSAT}(\mathbf{A})$. Indeed, nilpotent \mathbf{A} has a Malcev polynomial \mathbf{d} , such that for all $a, b \in A$ the function $x \mapsto \mathbf{d}(x, a, b)$ is a permutation. Thus, after arbitrarily fixing $a \in A$, the system of equations over \mathbf{A}

$$\begin{aligned} \mathbf{g}_1(x_1, \dots, x_n) &= \mathbf{h}_1(x_1, \dots, x_n) \\ &\vdots \\ \mathbf{g}_k(x_1, \dots, x_n) &= \mathbf{h}_k(x_1, \dots, x_n), \end{aligned}$$

can be equivalently rewritten to the following instance of $\text{MCSAT}(\mathbf{A})$

$$\mathbf{d}(\mathbf{g}_1(x_1, \dots, x_n), \mathbf{h}_1(x_1, \dots, x_n), a) = \dots = \mathbf{d}(\mathbf{g}_k(x_1, \dots, x_n), \mathbf{h}_k(x_1, \dots, x_n), a) = a.$$

This interpretation, together with Theorem 2.6, makes $\text{MCSAT}(\mathbf{A})$ NP-complete whenever \mathbf{A} is nilpotent but not affine.

To see that subdirect products of algebras each of which is polynomially equivalent to the 2-element lattice stay on the polynomial side recall, from the proof of Theorem 9.3, that such algebras have Uniform Solution Property. Thus checking if

$$\mathbf{g}_1(x_1, \dots, x_n) = \mathbf{g}_2(x_1, \dots, x_n) = \dots = \mathbf{g}_k(x_1, \dots, x_n)$$

has a solution in \mathbf{A} , reduces to finding $a \in A$ with

$$\mathbf{g}_1(a, \dots, a) = \mathbf{g}_2(a, \dots, a) = \dots = \mathbf{g}_k(a, \dots, a).$$

□

11. EQUIVALENCE OF CIRCUITS

This section considers the equivalence of circuits as defined in the problem CEQV. Our results in this direction are covered by Theorem 2.11.

Theorem 2.11. *Let \mathbf{A} be a finite algebra from a congruence modular variety. If \mathbf{A} has no quotient \mathbf{A}' with $\text{CEQV}(\mathbf{A}')$ being co-NP-complete then \mathbf{A} is nilpotent.*

Proof. First note that if $\mathbf{3}$ or $\mathbf{4}$ is in $\text{typ}\{\mathbf{A}\}$ then \mathbf{A} has a minimal set $U = \{0, 1\}$ such that $\mathbf{A}|_U$ is polynomially equivalent to either 2-element Boolean algebra or 2-element lattice. But CEQV is co-NP-complete for both of these small algebras (see Example 2.4). Arguing like in the proof of Theorem 5.1 this intractability can be carried up to $\text{CEQV}(\mathbf{A})$.

Thus we are left with solvable algebras, i.e. with $\text{typ}\{\mathbf{A}\} \subseteq \{\mathbf{2}\}$. To force such algebra \mathbf{A} to be nilpotent we can argue to the contrary like in the proof of Corollary 7.2 to produce its subdirectly irreducible quotient monolith of

which does not centralize 1. This allows us to mimic the proof of Lemma 7.1. Indeed the polynomials $\mathbf{t}_G(\bar{x})$ and $\mathbf{t}_\Phi(\bar{x})$ produced there (to encode graph colorability or Boolean satisfiability, respectively) take only two values: a and e . Now, instead of considering the satisfiability of the equation $\mathbf{t}(\bar{x}) = a$ we check whether $\mathbf{t}(\bar{x})$ always return the value e . \square

REFERENCES

- [1] E. Aichinger and N. Mudrinski, Some applications of higher commutators in Malcev algebras, *Algebra Universalis*, **63**(2010), 367–403.
- [2] L. Barto and M. Kozik, Constraint satisfaction problems solvable by local consistency methods, *Journal of the ACM*, **61**(2014), 3:1-3:19.
- [3] J. Berman and P.M. Idziak, *Generative complexity in algebra*, Memoirs of the American Mathematical Society, **175**(2005), no. 828, viii+159.
- [4] P. Broniek, *Computational Complexity of Solving Equation Systems*, Springer, 2015.
- [5] A. Bulatov, On the number of finite Maltsev algebras, *Contributions to general algebra*, **13**(2000) 41–54.
- [6] S. Burris and J. Lawrence, Results on the equivalence problem for finite groups, *Algebra Universalis*, **52**(2005), 495–500.
- [7] S. Burris and J. Lawrence, The equivalence problem for finite rings, *Journal of Symbolic Computation*, **15**(1993) 67–71.
- [8] T. Feder and M. Y. Vardi, The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory, *SIAM Journal on Computing*, **28**(1998), 57–104.
- [9] T. Feder, F. Madelaine, and I. A. Stewart. Dichotomies for classes of homomorphism problems involving unary functions, *Theoretical Computer Science*, **314**(2004), 1–43.
- [10] R. Freese and R. McKenzie, *Commutator Theory for Congruence Modular Varieties*, London Math. Soc. Lecture Notes, No. 125, Cambridge U. Press, Cambridge, 1987.
- [11] M. Goldmann and A. Russell, The complexity of solving equations over finite groups, Proceedings of the Fourteenth Annual IEEE Conference on Computational Complexity, 1999, pp. 80–86.
- [12] T. Gorzad and J. Krzaczkowski, Term equation satisfiability over finite algebras, *International Journal of Algebra and Computation*, **20**(2010),1001–1020.
- [13] T. Gorzad and J. Krzaczkowski, The complexity of problems connected with two-element algebras, *Reports on Mathematical Logic*, **46**(2011), 91–108.
- [14] R. L. Graham, B. L. Rothschild and J. H. Spencer, *Ramsey Theory*, 2nd ed., John Wiley & Sons, 1990.
- [15] H. P. Gumm, *Geometrical methods in congruence modular varieties*, Memoirs of the American Mathematical Society, **289**(1983).
- [16] J. Hagemann and C. Herrmann, A concrete ideal multiplication for algebraic systems and its relation to congruence distributivity, *Archiv der Mathematik*, **32**(1979), 234–245.
- [17] D. Hobby and R. McKenzie, *The Structure of Finite Algebras*, Contemporary Mathematics vol. 76, Amer. Math. Soc., Providence, RI, 1988.
- [18] G. Horváth, The complexity of the equivalence and equation solvability problems over nilpotent rings and groups, *Algebra Universalis*, **66**(2011), 391–403.
- [19] G. Horváth, The complexity of the equivalence and equation solvability problems over meta-Abelian groups, *Journal of Algebra*, **433**(2015), 208–230.
- [20] G. Horváth, J. Lawrence, L. Mérai and Cs. Szabó, The complexity of the equivalence problem for nonsolvable groups, *Bulletin of the London Mathematical Society*, **39**(2007), 433–438.

- [21] G. Horváth and Cs. Szabó, The Complexity of Checking Identities over Finite Groups, *International Journal of Algebra and Computation*, **16**(2006), 931–940.
- [22] G. Horváth and Cs. Szabó, The extended equivalence and equation solvability problems for groups, *Discrete Mathematics & Theoretical Computer Science*, **13**(2011), 23–32.
- [23] G. Horváth and Cs. Szabó, Equivalence and equation solvability problems for the alternating group A_4 , *Journal of Pure and Applied Algebra*, **216**(2012), 2170–2176.
- [24] H. B. Hunt and R. E. Stearns, The complexity of equivalence for commutative rings, *Journal of Symbolic Computation*, **10**(1990), 411–436.
- [25] P. Idziak, P. Marković, R. McKenzie, M. Valeriote, and R. Willard, Tractability and learnability arising from algebras with few subpowers. *SIAM Journal on Computing*, **39**(2010), 3023–3037.
- [26] A. Kazda, M. Kozik, R. McKenzie and M. Moore, Absorption and directed Jónsson terms, preprint, arXiv:1502.01072.
- [27] K. Kearnes, Congruence modular varieties with small free spectra, *Algebra Universalis*, **42**(1999), 165–181.
- [28] O. Klíma, Complexity issues of checking identities in finite monoids, *Semigroup Forum*, **79**(2009), 435–444
- [29] O. Klíma, P. Tesson and Denis Thérien, Dichotomies in the Complexity of Solving Systems of Equations over Finite Semigroups, *Theory of Computing Systems*, **40**(2007), 263–297.
- [30] L. Zádori, Taylor operations on finite reflexive structures, *International Journal of Mathematics and Computer Science*, **1**(2006), 1–21.
- [31] B. Larose and L. Zádori, Taylor terms, constraint satisfaction and the complexity of polynomial equations over finite algebras, *International Journal of Algebra and Computation*, **16**(2006), 563–581.
- [32] Yu. V. Matiyasevich, Enumerable Sets are Diophantine, *Soviet Mathematics Doklady*, **11**(1970), 354–357
- [33] R. McKenzie, G. McNulty, W. Taylor, *Algebras, Lattices, Varieties*, Wadsworth/Brooks Cole, Monterrey, CA, 1987.
- [34] R. McKenzie and M. Valeriote, *The Structure of Decidable Locally Finite Varieties*, Birkhäuser, Boston, 1989.
- [35] T.J. Schaefer, The complexity of satisfiability problems, *Proceedings of the 10th Annual ACM Symposium on Theory of Computing*, 1978, pp. 216–226.
- [36] B. Schwarz, The complexity of satisfiability problems over finite lattices, *Annual Symposium on Theoretical Aspects of Computer Science*, Springer 2004, pp. 31–43.
- [37] J.D.H. Smith, *Malcev Varieties*, Lecture Notes in Mathematics, vol. 554, Springer 1976.
- [38] M. Valeriote, On Decidable Locally Finite Varieties, PhD Dissertation, U.C.Berkeley, 1986.
- [39] L. Zádori, On solvability of systems of polynomial equations, *Algebra Universalis*, **65**(2011), 277–283.