Charles University in Prague
Faculty of Mathematics and Physics

HABILITATION THESIS

Libor Barto

Universal Algebra and the Constraint Satisfaction Problem

Department of Algebra

Prague 2014

Contents

Preface

This thesis presents selected results on the complexity of the constraint satisfaction problem and related topics in universal algebra:

[16] L. Barto, M. Kozik, T. Niven, The CSP dichotomy holds for digraphs with no sources and no sinks (a positive answer to a conjecture of Bang-Jensen and Hell), SIAM Journal on Computing 38/5 (2009), 1782-1802.

[13] L. Barto, M. Kozik, Constraint satisfaction problems solvable by local consistency methods, Journal of the ACM 61/1 (2014), 3:1-3:19.

[11] L. Barto, M. Kozik, Absorbing subalgebras, cyclic terms and the constraint satisfaction problem, Logical Methods in Computer Science 8/1:07 (2012), 1-26.

[6] L. Barto, Finitely related algebras in congruence distributive varieties have near unanimity terms, Canadian Journal of Mathematics 65/1 (2013), 3-21.

[5] L. Barto, The dichotomy for conservative constraint satisfaction problems revisited, Proceedings of the 26th IEEE Symposium on Logic in Computer Science, LICS'11, 301-310.

[17] L. Barto, M. Kozik, R. Willard, Near unanimity constraints have bounded pathwidth duality, Proceedings of the 27th ACM/IEEE Symposium on Logic in Computer Science, LICS'12, 125-134.

[12] L. Barto, M. Kozik, Robust satisfiability of constraint satisfaction problems, Proceedings of the 44th symposium on Theory of Computing, STOC'12 (2012), 931-940.

The thesis is structured as follows. The first chapter is an introduction to the topic intended for a non-specialist. A survey based on this chapter is being prepared for the *Bulletin of Symbolic Logic*. A shorter version will appear in *SIGLOG Newsletter*. The second chapter briefly summarizes the main results of the papers listed above. Appendices contain reprints of these papers.

Introduction

The Constraint Satisfaction Problem (CSP) provides a common framework for expressing a wide range of both theoretical and real-life combinatorial problems [61]. One solves an instance of CSP by assigning values to the variables so that the constraints are satisfied.

The topic of this thesis is a very active theoretical subfield which studies the computational complexity of the CSP over a fixed constraint language. This restricted framework is still broad enough to include many decision problems in the class NP, yet it is narrow enough to potentially allow a complete classification of all such CSP problems.

One particularly important achievement is the understanding of what makes a problem in this class computationally easy or hard. It is not surprising that hardness comes from lack of symmetry. However, usual objects capturing symmetry, automorphisms (or endomorphisms) and their groups (or semigroups), are not sufficient in this context. It turned out that the complexity of CSP is determined by more general symmetries: polymorphisms and their clones.

My aim in this chapter is to introduce the basics of this exciting area and highlight selected deeper results, in a way that is understandable to readers with a basic knowledge of computational complexity (see [58, 1]). The presentation of the material is based on my talk "Universal algebra and the constraint satisfaction problem" delivered at the Association of Symbolic Logic North American Annual Meeting held in Boulder, Colorado, in 2014.

1 CSP over a fixed constraint language

A constraint – such as $R(x_3, x_1, x_4)$ – restricts the allowed values for a tuple of variables – in this case (x_3, x_1, x_4) – to be an element of a particular relation on the domain – in this case $R \subseteq D^3$.[1] By an n-ary *relation* R on a domain D we mean a subset of the n-th cartesian power D^n. It is sometimes convenient to work with the corresponding *predicate* which is a mapping from D^n to $\{\text{true}, \text{false}\}$ specifying which tuples are in R. We will use both formalism, so e.g. $(a, b, c) \in R$ and $R(a, b, c)$ both mean that the triple $(a, b, c) \in D^3$ is from the relation R.

An instance of CSP is a list of constraints, e.g.,

$$R(x),\ S(y, y, z),\ T(y, w),$$

where R, S, T are relations of appropriate arity on a common domain D and x, y, z, w are variables. A mapping f assigning values from the domain to variables is a *solution* if it satisfies all the constraints, that is, in our example,

$$R(f(x)) \text{ and } S(f(y), f(y), f(z)) \text{ and } T(f(y), f(w)) \ .$$

[1]There are also different types of constraints considered in the literature, see e.g. Chapter 7 in [61].

A standard formal definition of an instance of the CSP over a finite domain goes as follows.

Definition 1.1. *An* instance of the CSP *is a triple* $P = (V, D, \mathcal{C})$ *with*

- V *a finite set of* variables*,*
- D *a finite* domain*,*
- $\mathcal{C}$ *a finite list of constraints, where each constraint is a pair* $C = (\mathbf{x}, R)$ *with*
  - $\mathbf{x}$ *a tuple of variables of length* n*, called the* scope *of* C*, and*
  - R *an* n*-ary relation on* D*, called the* constraint relation *of* C*.*

An assignment*, that is, a mapping* $f : V \to D$*,* satisfies *a constraint* $C = (\mathbf{x}, R)$ *if* $f(\mathbf{x}) \in R$*, where* f *is applied component-wise. An assignment* f *is a* solution *if it satisfies all constraints.*

Three basic computational problems associated with an instance are the following:[2]

- **Satisfiability.** Does the given instance have a solution? (A related problem, the *search problem*, is to find some solution if at least one solution exists.)
- **Optimization.** Even if the instance has no solution, find an optimal assignment, i.e., one that satisfies the maximum possible number of constraints. (*Approximation algorithms* are extensively studied, where the aim is, for example, to find an assignment that satisfies at least 80% of the number of constraints satisfied by an optimal assignment.)
- **Counting.** How many solutions does the given instance have? (This problem also has an approximation version: *approximate counting.*)

1.1 Satisfiability over a fixed constraint language

Even the easiest of the problems, satisfiability, is computationally hard: It contains many NP-complete problems including, e.g., 3-SAT (see Example 1.3). However, certain natural restrictions to CSP satisfiability ensure tractability. The main types of restrictions that have been studied are *structural restrictions*, which limit how constraints interact, and *language restrictions*, which limit the choice of constraint relations.

In this chapter, we focus just on satisfiability problems with language restrictions. Please see [66] for optimization problems and a generalization to valued CSPs, [43] for approximation, [32] for counting, and [20] for a generalization to infinite domains.

Definition 1.2. *A* constraint language*,* $\mathcal{D}$*, is a set of relations on a common finite domain,* D*. We use* CSP($\mathcal{D}$) *to denote the set of CSP satisfiability problems whose relations are drawn from* $\mathcal{D}$*.*

[2] To study the computational complexity of these problems we need to specify a representation of instances. We will assume that the constraint relation in every constraint is given by a list of all its members. Note, however, that for most of the problems considered in this column any reasonable representation can be taken.

1.2 Examples

Example 1.3. *An instance of the standard NP-complete problem,* 3-SAT*, is a Boolean formula in conjunctive normal form with exactly three literals per clause. For example, the formula,*

$$\varphi \quad = \quad (x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_4 \vee x_5 \vee \neg x_1) \wedge (\neg x_1 \vee \neg x_4 \vee \neg x_3)$$

is a satisfiable instance of 3-SAT*. (Any assignment making x_1 and x_2 false, satisfies φ.)* 3-SAT *is equivalent to* $\mathrm{CSP}(\mathcal{D}_{3SAT})$*, where* $D_{3SAT} = \{0,1\}$ *and*

$$\mathcal{D}_{3SAT} = \{S_{ijk} : i,j,k \in \{0,1\}\}, \text{ where } S_{ijk} = \{0,1\}^3 \setminus \{(i,j,k)\} \ .$$

For example, the above formula φ corresponds to the following instance of $\mathrm{CSP}(\mathcal{D}_{3SAT})$

$$S_{010}(x_1,x_2,x_3),\ S_{101}(x_4,x_5,x_1),\ S_{111}(x_1,x_4,x_3) \ .$$

*More generally, for a natural number k, k-*SAT *denotes a similar problem where each clause is a disjunction of k literals.*

Since 3-SAT *is NP-complete, it follows that k-*SAT *is NP-complete for each $k \geq 3$. On the other hand,* 2-SAT *is solvable in polynomial time, and is in fact complete for the complexity class NL (non-deterministic logarithmic space).*

Example 1.4. HORN-3-SAT *is a restricted version of* 3-SAT*, where each clause may have at most one positive literal. This problem is equivalent to* $\mathrm{CSP}(\mathcal{D}_{HornSAT})$ *for* $\mathcal{D}_{HornSAT} = \{S_{011}, S_{101}, S_{110}, S_{111}\}$ *(or just* $\mathcal{D}_{HornSAT} = \{S_{011}, S_{111}\}$*).* HORN-3-SAT *is solvable in polynomial time, in fact, it is a P-complete problem.*

Example 1.5. *For a fixed natural number k, the k-*COLORING *problem is to decide whether it is possible to assign colors $\{0,1,\ldots,k-1\}$ to vertices of an input graph in such a way that adjacent vertices receive different colors. This problem is equivalent to* $\mathrm{CSP}(\mathcal{D}_{kCOLOR})$*, where* $D_k = \{1,2,\ldots,k\}$ *and* $\mathcal{D}_{kCOLOR} = \{\neq_k\}$ *consists of a single relation – the binary inequality relation* $\neq_k = \{(a,b) \in D_k^2 : a \neq b\}$*.*

Indeed, given an instance of $\mathrm{CSP}(\mathcal{D})$*, we can form a graph whose vertices are the variables and edges correspond to the binary constraints (that is, x has an edge to y iff the instance contains the constraint $x \neq_k y$). It is easily seen that the original instance has a solution if and only if the obtained graph is k-colorable. The translation in the other direction is similar.*

*The k-*COLORING *problem is NP-complete for $k \geq 3$.* 2-COLORING *is equivalent to deciding whether an input graph is bipartite. It is solvable in polynomial time, in fact, it is an L-complete problem (where L stands for logarithmic space) by a celebrated result of Reingold [60].*

Example 1.6. *Let p be a prime number. An input of* $3\text{-LIN}(p)$ *is a system of linear equations over the p-element field* $\mathrm{GF}(p)$*, where each equation contains* 3 *variables, and the question is whether the system has a solution. This problem is equivalent to* $\mathrm{CSP}(\mathcal{D})$*, where* $D_{3LINp} = \mathrm{GF}(p)$ *and* $\mathcal{D}_{3LINp}$ *consists of all affine subspaces of* $\mathrm{GF}(p)^3$ *of dimension* 2 *or* 3*:*

$$\mathcal{D}_{3LINp} = \{R_{abcd} : a,b,c,d \in \mathrm{GF}(p)\},\, , \text{ where } R_{abcd} = \{(x,y,z) \in \mathrm{GF}(p)^3 : ax+by+cz = d\} \ .$$

This problem is solvable in polynomial time, e.g. by Gaussian elimination.[3] *It is complete for a somewhat less familiar class* $\mathrm{Mod}_p\mathrm{L}$.

[3] The problem of solving general systems of linear equations over $\mathrm{GF}(p)$ without the restriction on number of variables cannot be faithfully phrased as $\mathrm{CSP}(\mathcal{D})$ with $\mathcal{D}$ consisting of all affine subspaces, since the input representation of the latter problem can be substantially larger. However, a system of linear equation can be easily rewritten to an instance of 3-LIN(p) by introducing new variables.

Example 1.7. *An instance of the s, t-connectivity problem,* STCON, *is a directed graph and two vertices s, t. The question is whether there exists a directed path from s to t.*

A closely related (but not identical) problem is $\mathrm{CSP}(\mathcal{D}_{STCON})$*, where* $D_{STCON} = \{0, 1\}$ *and* $\mathcal{D}_{STCON} = \{C_0, C_1, \leq\}$*,* $C_0 = \{0\}$*,* $C_1 = \{1\}$*,* $\leq = \{(0, 0), (0, 1), (1, 1)\}$*. Indeed, given an instance of* $\mathrm{CSP}(\mathcal{D}_{STCON})$ *we form a directed graph much as we did in Example 1.5 and label some vertices* 0 *or* 1 *according to the unary constraints. Then the original instance has a solution if and only if there is no directed path from a vertex labeled* 1 *to a vertex labeled* 0*. Thus* $\mathrm{CSP}(\mathcal{D}_{STCON})$ *can be solved by invoking the complement of* STCON*, the s, t-non-connectivity problem, several times.*

Both STCON *and* $\mathrm{CSP}(\mathcal{D}_{STCON})$ *can clearly be solved in polynomial time. By the theorem of Immerman and Szelepcsényi [47, 64] both problems are NL-complete.*

In the same way, the s, t-connectivity problem for undirected graphs is closely related to $\mathrm{CSP}(\mathcal{D}_{USTCON})$*, where* $D_{USTCON} = \{0, 1\}$ *and* $\mathcal{D}_{USTCON} = \{C_0, C_1, =\}$*. These problems are L-complete by [60].*

1.3 The dichotomy conjecture

The most fundamental problem in the area was formulated in the landmark paper by Feder and Vardi [40].

Conjecture 1.8 (The dichotomy conjecture)**.** *For every finite*[4] *constraint language $\mathcal{D}$, the problem* $\mathrm{CSP}(\mathcal{D})$ *is in P or is NP-complete.*

Recall that if P $\neq$ NP, then there are problems of intermediate complexity [53]. Feder and Vardi argued that CSPs over a fixed constraint language is a good candidate for a largest natural class of problems with P versus NP-complete dichotomy.

At that time the conjecture was supported by two major cases: the dichotomy theorem for all languages over the two-element domain by Schaefer [62] and the dichotomy theorem for languages consisting of a single binary symmetric relation by Hell and Nešetřil [44].

Feder and Vardi have identified two sources of polynomial-time solvability and made several important contributions toward understanding these sources. In particular, they observed that the known polynomial cases were tied to algebraic closure properties and asked whether polynomial solvability for CSP can always be explained in such a way. Subsequent papers have shown that this is indeed the case and this connection to algebra brought the area to another level.

The algebraic approach is outlined in section 2 and some fruits of the theory discussed in section 3.

1.4 Alternative views

Note that a constraint language $\mathcal{D}$ with domain D can be viewed as a relational structure $(D; R_1, R_2, \dots)$, or equivalently relational database, with universe D.

Recall that a *conjunctive query* over the database $\mathcal{D}$ is an existential sentence whose quantifier-free part is a conjunction of atoms. $\mathrm{CSP}(\mathcal{D})$ is exactly the problem of deciding

[4] It is conjectured in [24] that the dichotomy remains true without the finiteness assumption. Namely, the local-global conjecture states that $\mathrm{CSP}(\mathcal{D})$ is in P (NP-complete) whenever $\mathrm{CSP}(\mathcal{D}')$ is in P (NP-complete) for every (some) finite $\mathcal{D}' \subseteq \mathcal{D}$.

whether $\mathcal{D}$ satisfies a given conjunctive query. For example, the instance

$$R(x),\ S(y,y,z),\ T(y,w)$$

has a solution if and only if the sentence

$$(\exists x,y,z,w \in D)\ R(x) \wedge S(y,y,z) \wedge T(y,w)$$

is true in $\mathcal{D}$.

From this perspective, it is natural to ask what happens if we allow some other combination of logical connectives $\{\exists, \forall, \wedge, \vee, \neg, =, \neq\}$. It turns out that out of the 2^7 cases only 3 are interesting (the other cases either reduce to these, or are almost always easy or hard by known results): $\{\exists, \wedge\}$ which is CSP, $\{\exists, \forall, \wedge\}$ which is so called *quantified CSP*, and $\{\exists, \forall, \wedge, \vee\}$. The complexity of quantified CSP is also an active research area [34] with possible trichotomy P, NP-complete or Pspace-complete. Recently, a tetrachotomy was obtained for the last choice [56] – for every $\mathcal{D}$, the corresponding problem is either in P, NP-complete, co-NP-complete, or Pspace-complete.

The CSP over a fixed language can also be formulated as the homomorphism problem between relational structures with a fixed target structure [40]. The idea of the translation is shown in Examples 1.5, 1.7.

2 Universal algebra in CSP

If a computational problem $\mathcal{A}$ can simulate (in some sense) another problem $\mathcal{B}$, then $\mathcal{A}$ is at least as hard as $\mathcal{B}$. This simple idea is widely used in computational complexity; for instance, NP-completeness is often shown by a gadget reduction of a known NP-complete problem to the given one. A crucial fact for the algebraic theory of CSP is that so called primitive positive (pp-, for short) interpretation between constraint languages gives such a reduction between corresponding CSPs (more precisely, if $\mathcal{D}$ pp-interprets $\mathcal{E}$, then $\mathrm{CSP}(\mathcal{E})$ is reducible to $\mathrm{CSP}(\mathcal{D})$). Pp-interpretations have been, indirectly, the main subject of universal algebra for the last 80 years!

The algebraic theory of CSPs was developed in a number of papers including [49, 48, 24, 54]. The viewpoint taken here is close to [20]. All results in this section come from these sources unless stated otherwise.

To simplify formulations, all structures (relational or algebraic) are assumed to have finite domains, all constraint languages are assumed to contain *finitely many* relations, all of them *nonempty*. By a *reduction* we mean a logarithmic space reduction (although first-order reductions are often possible under additional weak assumptions).

2.1 Primitive positive interpretations

An important special case of pp-interpretability is pp-definability.

Definition 2.1. *Let $\mathcal{D}, \mathcal{E}$ be constraint languages on the same domain $D = E$. We say that $\mathcal{D}$* pp-defines *$\mathcal{E}$ (or $\mathcal{E}$ is pp-definable from $\mathcal{D}$) if each relation in $\mathcal{E}$ can be defined by a first order formula which only uses relations in $\mathcal{D}$, the equality relation, conjunction and existential quantification.*

Theorem 2.2. *If $\mathcal{D}$ pp-defines $\mathcal{E}$, then* $\mathrm{CSP}(\mathcal{E})$ *is reducible to* $\mathrm{CSP}(\mathcal{D})$.

Proof by example. Let R be an arbitrary ternary relation on a domain D. Consider the relations on D defined by

$$S(x,y) \text{ iff } (\exists z)R(x,y,z) \wedge R(y,y,x), \quad T(x,y) \text{ iff } R(x,x,x) \wedge (x=y) \ ,$$

where the existential quantification is understood over D. The relations S and T are defined by pp-formulae, therefore the constraint language $\mathcal{D} = \{R\}$ pp-defines the constraint language $\mathcal{E} = \{S, T\}$.

We sketch the reduction of CSP($\mathcal{E}$) to CSP($\mathcal{D}$) using the instance

$$S(x_3, x_2),\ T(x_1, x_4),\ S(x_2, x_4) \ .$$

We first replace S and T with their pp-definitions by introducing a new variable for each quantified variable:

$$R(x_3,x_2,y_1),\ R(x_2,x_2,x_3), \quad R(x_1,x_1,x_1),\ x_1 = x_4, \quad R(x_2,x_4,y_2),\ R(x_4,x_4,x_2)$$

and then we get rid of the equality constraint $x_1 = x_4$ by identifying these variables. This way we obtain an instance of CSP($\mathcal{D}$):

$$R(x_3,x_2,y_1),\ R(x_2,x_2,x_3),\ R(x_1,x_1,x_1),\ R(x_2,x_1,y_2),\ R(x_1,x_1,x_2) \ .$$

Clearly, the new instance of CSP($\mathcal{D}$) has a solution if and only if the original instance does. □

This simple theorem provides a quite powerful tool for comparing CSPs over different languages *on the same domain*. A more powerful tool, which can also be used to compare languages with different domains, is pp-interpretability. Informally, a constraint language $\mathcal{D}$ pp-interprets $\mathcal{E}$, if the domain of $\mathcal{E}$ is a pp-definable relation (from $\mathcal{D}$) modulo a pp-definable equivalence, and the relations of $\mathcal{E}$ (viewed, in a natural way, as relations on D) are also pp-definable from $\mathcal{D}$.[5] Formally:

Definition 2.3. *Let $\mathcal{D}, \mathcal{E}$ be constraint languages. We say that $\mathcal{D}$* pp-interprets *$\mathcal{E}$ if there exists a natural number n, $F \subseteq D^n$, and an onto mapping $f : F \to E$ such that $\mathcal{D}$ pp-defines*

- *the relation F,*
- *the f-preimage of the equality relation on E, and*
- *the f-preimage of every relation in $\mathcal{E}$,*

where by the f-preimage of a k-ary relation S on E we mean the nk-ary relation $f^{-1}(S)$ on D defined by

$$f^{-1}(S)(x_{11},\dots,x_{1k},x_{21},\dots,x_{2k},\dots,x_{n1},\dots,x_{nk}) \textit{ iff } S(f(x_{11},\dots,x_{n1}),\dots,f(x_{1k},\dots,x_{nk}))$$

Theorem 2.4. *If $\mathcal{D}$ pp-interprets $\mathcal{E}$, then* CSP($\mathcal{E}$) *is reducible to* CSP($\mathcal{D}$).

Proof sketch. The properties of the mapping f from Definition 2.3 allow us to rewrite an instance of CSP($\mathcal{E}$) to an instance of the CSP over a constraint language which is pp-definable from $\mathcal{D}$. Then we apply Theorem 2.2. □

[5] This is the classical notion of interpretation from model theory restricted to pp-formulas.

Pp-interpretability is a reflexive and transitive relation on the class of constraint languages. By identifying equivalent languages, i.e. languages which mutually pp-interpret each other, we get a partially ordered set, the *pp-interpretability poset*. Theorem 2.4 then says that the "higher" we are in the poset the "easier" CSP we deal with. 3-SAT is terribly hard – we will see later that its constraint language is the least element of this poset. Surprisingly, this is "almost" the case for all known NP-complete CSPs! For precise formulation we need a reduction described in the following subsection.

2.2 Cores and singleton expansions

Let $\mathcal{D}$ be a constraint language on a finite set D. A mapping $f : D \to D$ is called an *endomorphism* if it preserves every relation $\mathcal{D}$, that is, $f(R) := \{f(\mathbf{a}) : \mathbf{a} \in R\} \subseteq R$ for every $R \in \mathcal{D}$.

Theorem 2.5. *Let $\mathcal{D}$ be a constraint language and f an endomorphism of $\mathcal{D}$. Then* $\mathrm{CSP}(\mathcal{D})$ *is reducible to* $\mathrm{CSP}(f(\mathcal{D}))$ *and vice versa, where $f(\mathcal{D})$ is a constraint language with domain $f(D)$ defined by $f(\mathcal{D}) = \{f(R) : R \in \mathcal{D}\}$.*

Proof sketch. An instance of the $\mathrm{CSP}(\mathcal{D})$ has a solution if and only if the corresponding instance of $\mathrm{CSP}(f(\mathcal{D}))$, obtained by replacing each $R \in \mathcal{D}$ with $f(R)$, has a solution. □

A language $\mathcal{D}$ is a *core* if every endomorphism of $\mathcal{D}$ is a bijection. It is not hard to show that if f is an endomorphism of a constraint language $\mathcal{D}$ with minimal range, then $f(\mathcal{D})$ is a core. Moreover, this core is unique up to isomorphism, therefore we speak about *the core* of $\mathcal{D}$.

An important fact is that we can add all singleton unary relations to a core constraint language without increasing the complexity of its CSP:

Theorem 2.6. *Let $\mathcal{D}$ be a core constraint language and $\mathcal{E} = \mathcal{D} \cup \bigcup_{a \in D} C_a$, where C_a denotes the unary relation $C_a = \{a\}$. Then* $\mathrm{CSP}(\mathcal{E})$ *is reducible to* $\mathrm{CSP}(\mathcal{D})$.

Proof idea. The crucial step is to observe that the set of endomorphisms of $\mathcal{D}$, viewed as a $|D|$-ary relation, is pp-definable from $\mathcal{D}$. More precisely, the relation

$$S = \{(f(a_1), \dots, f(a_n)) : f \text{ is an endomorphism of } \mathcal{D}\} \ ,$$

where $a_1, \dots, a_n$ is a list of all elements of D, is pp-definable from $\mathcal{D}$ (even without existential quantification). Indeed, f is, by definition, an endomorphism of $\mathcal{D}$ if for every $R \in \mathcal{D}$ of arity $\mathrm{ar}(R)$ and every $(b_1, \dots, b_{\mathrm{ar}(R)}) \in R$ we have $(f(b_1), \dots, f(b_{\mathrm{ar}(R)})) \in R$. This directly leads to a pp-definition of S:

$$S(x_{a_1}, \dots, x_{a_n}) \text{ iff } \bigwedge_{R \in \mathcal{D}} \bigwedge_{(b_1, \dots, b_{\mathrm{ar}R}) \in R} R(x_{b_1}, \dots, x_{b_{\mathrm{ar}(R)}}) \ .$$

Given an instance of $\mathrm{CSP}(\mathcal{E})$ we introduce new variables $x_{a_1}, \dots, x_{a_n}$, replace every constraint of the form $C_a(x)$ by $x = x_a$, and add the constraint $S(x_{a_1}, \dots, x_{a_n})$. In this way we obtain an instance of $\mathrm{CSP}(\mathcal{D} \cup \{=\})$. Clearly, if the original instance has a solution, then the new instance has a solution as well. In the other direction, if g is a solution to the new instance, then its values on $x_{a_1}, \dots, x_{a_n}$ determine an endomorphism f of $\mathcal{D}$. As $\mathcal{D}$ is a core, f is a bijection, thus f^{-1} is an endomorphism as well, and $f^{-1} \circ g$ restricted to the original variables is a solution of the original instance. □

We will call constraint languages containing all singletons *idempotent*. Note that an idempotent constraint language is automatically a core as the only endomorphism is the identity. By Theorems 2.5, 2.6, CSP over $\mathcal{D}$ is reducible to CSP over the singleton expansion of the core of $\mathcal{D}$ and vice versa. It is therefore enough to study CSPs over idempotent constraint languages.

An interesting consequence of these reductions is that the search problem for CSP($\mathcal{D}$) is solvable in polynomial time whenever CSP($\mathcal{D}$) is. The idea is to gradually guess values for variables using the unary singleton constraints.

2.3 Example

Example 2.7. *We show that* 3-SAT *is reducible to* 3-COLORING.

Recall the constraint language $\mathcal{D}_{3COLOR} = \{\neq_{\{0,1,2\}}\}$ *of* 3-COLORING *from Example 1.5 and the constraint language* $\mathcal{D}_{3SAT} = \{S_{000}, \ldots, S_{111}\}$ *of* 3-SAT *from Example 1.3.*

Since $\mathcal{D}_{3COLOR}$ *is a core,* CSP($\mathcal{D}'_{3COLOR}$)*, where* $\mathcal{D}'_{3COLOR} = \{\neq, C_0, C_1, C_2\}$*, is reducible to* CSP($\mathcal{D}_{3COLOR}$) *by Theorem 2.6. By Theorem 2.4, it is now enough to show that* $\mathcal{D}'_{3COLOR}$ *pp-interprets* $\mathcal{D}_{3SAT}$*. We give a pp-interpretation with* $n = 1$*,* $F = \{0, 1\}$*, and* f *the identity map (see Definition 2.3). The set (=unary relation)* $\{0, 1\}$ *can be pp-defined by*

$$E(x) \quad \textit{iff} \quad (\exists y)\ C_2(y) \wedge x \neq y \quad (\textit{iff } x \neq 2)\ .$$

The preimage of the equality relation is the equality relation on $\{0, 1\}$ *which is clearly pp-definable. The relation* S_{000} *can be defined by*

$$S_{000}(x_1, x_2, x_3) \quad \textit{iff} \quad (\exists y_1, y_2, y_3, z)\ C_2(z) \wedge y_1 \neq y_2 \wedge y_2 \neq y_3 \wedge y_1 \neq y_3 \\ \wedge \bigwedge_{i=1,2,3} z \neq x_i \wedge T(x_i, y_i)\ ,$$

where T *is the binary relation*

$$T(x, y) \quad \textit{iff} \quad (\exists u, v)\ C_1(u) \wedge u \neq v \wedge x \neq v \wedge y \neq v$$

The other relations S_{ijk} *are defined similarly.*

While it is easy to verify that the presented pp-definitions work, it is not so easy to find them without any tools. The proof of Theorem 2.9 gives an algorithm to produce pp-definitions whenever they exist (although the obtained definitions will be usually very long).

2.4 Tractability conjecture

Now we return to the pp-interpretability poset. Recall that "higher" in the poset means "easier" CSP and that 3-SAT corresponds to the least (the hardest) element. When we restrict to idempotent constraint languages (which we can do by the previous discussion), all known NP-complete CSPs are at the bottom of the poset. Bulatov, Jeavons and Krokhin conjectured that this is not a coincidence.[6]

Conjecture 2.8 (Tractability conjecture)**.** *If an idempotent constraint language* $\mathcal{D}$ *does not pp-interpret the language of* 3-SAT*, then* CSP($\mathcal{D}$) *is solvable in polynomial time.*

This conjecture is also known as the *algebraic dichotomy conjecture* because many equivalent formulations, including the original one, are algebraic.

[6] Similar hardness results and conjectures are formulated for other computational/descriptive complexity classes.

2.5 Algebraic counterpart of pp-definability

The link between relations and operations is provided by a natural notion of compatibility. An n-ary operation f on a finite set D (that is, a mapping $f : D^n \to D$) is *compatible* with a k-ary relation $R \subseteq D^k$ if f applied component-wise to any n-tuple of elements of R gives an element of R. In more detail, whenever (a_{ij}) is an $n \times k$ matrix such that every row is in R, then f applied to the columns gives a k-tuple which is in R as well.

We say that an operation f on D is a *polymorphism* of a constraint language $\mathcal{D}$ if f is compatible with every relation in $\mathcal{D}$. Note that unary polymorphism is the same as endomorphism. Endomorphisms can be thought of as symmetries, polymorphisms are then symmetries of higher arity.

The set of all polymorphisms of $\mathcal{D}$ will be denoted by $\mathbf{D}$. This algebraic object has the following two properties.

- $\mathbf{D}$ *contains all projections*, that is, for every natural number n and $i \leq n$ the n-ary projection onto i-th coordinate defined by

$$\pi_i^n(a_1, \dots, a_n) = a_i$$

  is in $\mathbf{D}$.

- $\mathbf{D}$ *is closed under composition*, that is, for any n-ary $g \in \mathbf{D}$ and k-ary $f_1, \dots, f_n \in \mathbf{D}$ their (k-ary) composition $g(f_1, \dots, f_n)$ defined by

$$g(f_1, \dots, f_n)(a_1, \dots, a_k) = g(f_1(a_1, \dots, a_k), \dots, f_n(a_1, \dots, a_k))$$

  is in $\mathbf{D}$.

Sets of operations with these properties are called *concrete clones* (or simply *clones*), therefore we refer to $\mathbf{D}$ as the *clone of polymorphisms* of $\mathcal{D}$.

The clone of polymorphisms controls pp-definability in the sense of the following old result [41, 21].

Theorem 2.9. *Let $\mathcal{D}$, $\mathcal{E}$ be constraint languages with $D = E$. Then $\mathcal{D}$ pp-defines $\mathcal{E}$ if and only if $\mathbf{D} \subseteq \mathbf{E}$.*[7]

Proof sketch. The implication "$\Rightarrow$" is quite easy. For the other implication it is enough to prove that whenever R is a relation compatible with every polymorphism of $\mathcal{D}$, then R is pp-definable from $\mathcal{D}$. A crucial step is a more general version of the observation made in the proof of Theorem 2.6: For any k, the set of k-ary polymorphisms of $\mathcal{D}$ can be viewed as a $|D|^k$-ary relation S on D, and this relation is pp-definable from $\mathcal{D}$. Now R can be defined from such a relation S (where k is the number of tuples in R) by existential quantification over suitable coordinates as in Example 2.11. $\square$

In view of this result, Theorem 2.2 says that the complexity of CSP($\mathcal{D}$) only depends on the clone $\mathbf{D}$. More precisely, if $\mathbf{D} \subseteq \mathbf{E}$, then CSP($\mathcal{E}$) is reducible to CSP($\mathcal{D}$). Moreover, the proof of Theorem 2.9 gives a generic pp-definition of $\mathcal{E}$ from $\mathcal{D}$, which gives us a generic reduction of CSP($\mathcal{E}$) to CSP($\mathcal{D}$).

[7] Moreover, every concrete clone is the clone of polymorphisms of some (possibly infinite) constraint language.

Example 2.10. *It is a nice exercise to show that the language $\mathcal{D}_{3SAT}$ of* 3-SAT *has no polymorphisms but projections. This means that $\mathcal{D}_{3SAT}$ pp-defines every constraint language with domain $\{0,1\}$. It follows (see also Theorem 2.14) that $\mathcal{D}_{3SAT}$ pp-interprets every constraint language, so it is the least element of the pp-interpretability poset, as claimed before.*

Example 2.11. *Another nice exercise is to show that the language $\mathcal{D}'_{3COLOR} = \{\neq, C_0, C_1, C_2\}$ on the domain $\{0,1,2\}$ (see Example 2.7) also does not have any polymorphisms except for projections.*

We show how the proof of Theorem 2.9 produces a pp-definition of the relation

$$R = \{(0,1),(0,2),(1,1),(2,2)\} \ .$$

Since R contains 4 pairs, we pp-define the 3^4-ary relation

$$S = \{(f(0,0,0,0), f(0,0,0,1), \dots, f(2,2,2,2)) : \ f \text{ is a 4-ary polymorphism of } \mathcal{D}'_{3COLOR}\}$$

which corresponds to the set of all 4-ary polymorphisms of $\mathcal{D}'_{3COLOR}$:

$$S(x_{0000}, \dots, x_{2222}) \quad \text{iff} \quad \bigwedge_i x_{iiii} = i \wedge \bigwedge_{i_1 \neq i_2, j_1 \neq j_2, k_1 \neq k_2, l_1 \neq l_2} x_{i_1 j_1 k_1 l_1} \neq x_{i_2 j_2 k_2 l_2} \ .$$

Now we existentially quantify over all variables but x_{0012} and x_{1212} – the exceptions are those variables which correspond to the i-th coordinates of pairs in R, $i \in \{1,2\}$. The obtained binary relation $R'(x_{0012}, x_{1212})$ contains R since S contains the projections, and is contained in R since R is compatible with every polymorphism of $\mathcal{D}'_{3COLOR}$.

Note that the definition of S_{000} from Example 2.7 obtained in this way contains 3^7 variables. This is the price we need to pay for genericity.

2.6 Algebraic counterpart of pp-interpretability

For the algebraic description of pp-interpretability we introduce three constructions which are clone versions of standard constructions for groups, rings, etc.

Let $\mathbf{D}$ be a (concrete) clone.

The domain D of $\mathbf{D}$ is also called the *universe* of $\mathbf{D}$. We say that $E \subseteq D$ is a *subuniverse* of $\mathbf{D}$ if it is closed under all operations of $\mathbf{D}$. In this situation, we can form a clone $\mathbf{E}$ by restricting all operations of $\mathbf{D}$ to the set E. The clone $\mathbf{E}$ is called a *subalgebra* of $\mathbf{D}$ (the word *subclone* is reserved for set theoretic inclusion).

For a natural number n we can form the *n-th power* $\mathbf{D}^n$ of $\mathbf{D}$ with domain D^n and operations from $\mathbf{D}$ acting coordinate-wise. (More generally, we can form the X-th power $\mathbf{D}^X$ of $\mathbf{D}$ for any set X.) A *subpower* is a subuniverse (or a subalgebra, depending on the context) of a power. Note that if $\mathbf{D}$ is the clone of polymorphisms of a constraint language $\mathcal{D}$, then R is a subpower of $\mathbf{D}$ if and only if R is pp-definable from $\mathcal{D}$ (by Theorem 2.9).

Finally, let $\phi : D \to E$ be an onto mapping such that for any operation $f \in \mathbf{D}$ (say of arity n), the formula

$$f_\phi(\phi(a_1), \dots, \phi(a_n)) = \phi(f(a_1, \dots, a_n)) \quad \forall a_1, \dots, a_n \in D$$

correctly defines an operation f_ϕ on E. Then $\mathbf{E} = \{f_\phi : f \in \mathbf{D}\}$ is a clone with domain E called a *concrete homomorphic image* of $\mathbf{D}$ and ϕ is called a *concrete homomorphism.*

The definition of pp-interpretability can be translated into algebraic terms as follows.

Theorem 2.12. *Let $\mathcal{D}$, $\mathcal{E}$ be constraint languages. Then $\mathcal{D}$ pp-interprets $\mathcal{E}$ if and only if* $\mathbf{E}$ *contains a concrete homomorphic image of a subpower of* $\mathbf{D}$.

2.7 Identities and Mal'tsev conditions

An alternative algebraic characterization of pp-interpretability, which is missing on the relational side, follows from the foundation stone of universal algebra, the Birkhoff HSP theorem [19]: pp-interpretability depends on the identities (i.e. universally quantified equations) satisfied by polymorphisms.

We first present a formulation using abstract clone homomorphisms and then explain the connection to identities.

Definition 2.13. *A mapping H from a clone $\mathbf{D}$ to a clone $\mathbf{E}$ is called a* clone homomorphism *if*

- *it preserves the arities of operations,*
- *it maps projections to projections (that is, $H(\pi_i^n) = \pi_i^n$, where the projection on the left hand side works on the set D, while on the right hand side on the set E), and*
- *it preserves the composition (that is, $H(g(f_1, \dots, f_n)) = H(g)(H(f_1), \dots, H(f_n))$ if $g, f_1, \dots, f_n$ are from $\mathbf{D}$ and have appropriate arities).*

Theorem 2.14. *Let $\mathcal{D}$, $\mathcal{E}$ be constraint languages. Then $\mathcal{D}$ pp-interprets $\mathcal{E}$ if and only if there exists an abstract clone homomorphism from $\mathbf{D}$ to $\mathbf{E}$.*

Proof sketch. There are natural abstract clone homomorphisms associated to the three constructions on clones (taking sublagebras, powers and concrete homomorphic images). The implication $\Rightarrow$ follows from this observation and Theorem 2.12.

Now assume that $H : \mathbf{D} \to \mathbf{E}$ is a clone homomorphism. For simplicity, let $E = \{1, 2, \dots, n\}$. It is easy to check that the set F of all n-ary operations in $\mathbf{D}$ is a subuniverse of $\mathbf{D}^{D^n}$. Let $\mathbf{F}$ be the corresponding subalgebra of $\mathbf{D}^{D^n}$. (This important object, the *n-generated free algebra for* $\mathbf{D}$, already appeared in the proof of Theorem 2.9. Indeed, if $\mathbf{D}$ is the clone of polymorphisms of a constraint language $\mathcal{D}$, then F is the set of all n-ary polymorphisms of $\mathcal{D}$.) A simple calculation shows that the mapping $\phi : F \to E$, defined by $\phi(f) = (H(f))(1, 2, \dots, n)$, is a concrete clone homomorphism from $\mathbf{F}$ onto $H(\mathbf{D}) \subseteq \mathbf{E}$ and thus $\mathcal{D}$ pp-interprets $\mathcal{E}$ by Theorem 2.12. $\square$

Observe that the existence of an abstract clone homomorphism $H : \mathbf{D} \to \mathbf{E}$ does not depend on concrete operations in $\mathbf{D}$ and $\mathbf{E}$ – it only depends on the way how operations compose and which operations are projections. The torso of a concrete clone which only remembers projections and composition is called an *abstract clone*.[8]

We now explain the promised link to identities, first on an example. A binary operation f on D is a *semilattice* operation if satisfies the identities

$$f(f(x, y), z) \approx f(x, f(y, z)),\ f(x, y) \approx f(y, x),\ \text{and}\ f(x, x) \approx x,$$

meaning that $f(f(a, b), c) = f(a, f(b, c))$, $f(a, b) = f(b, a)$, and $f(a, a) = a$ hold for any $a, b, c \in D$. This can be expressed in terms of composition and projections: f is a semilattice operation if and only if

$$f(f(\pi_1^3, \pi_2^3), \pi_3^3) = f(\pi_1^3, f(\pi_2^3, \pi_3^3)),\ f(\pi_1^2, \pi_2^2) = f(\pi_2^2, \pi_1^2),\ \text{and}\ f(\pi_1^1, \pi_1^1) = \pi_1^1\ .$$

[8] The relation between abstract clones and concrete clones is similar to the relation between groups and permutation groups, or between monoids and transformation monoids.

It follows that if $H : \mathbf{D} \to \mathbf{E}$ is an abstract clone homomorphism and $\mathbf{D}$ contains a semilattice operation f, then $\mathbf{E}$ contains a semilattice operation as well, namely $H(f)$.

More generally, if there exists an abstract clone homomorphism from $\mathbf{D}$ to $\mathbf{E}$, then $\mathbf{E}$ satisfies all properties of the form "there exist operations ... satisfying identities ... " which are satisfied by $\mathbf{D}$. Such properties are called *Mal'tsev conditions.*[9] It is not hard to see that the converse is also true: if no abstract clone homomorphism $\mathbf{D} \to \mathbf{E}$ exists, then there is some Mal'tsev condition which is satisfied by $\mathbf{D}$ while not satisfied by $\mathbf{E}$. In short:

> The complexity of CSP($\mathcal{D}$) only depends on Mal'tsev conditions satisfied by the clone of polymorphisms of $\mathcal{D}$.

To illustrate this, we state one of increasingly many (e.g., [65, 45, 57, 52, 63]) characterizations of the conjectured borderline between P and NP-complete CSPs by means of cyclic operations [11].

Theorem 2.15. *Let $\mathcal{D}$ be an idempotent constraint language and $p > |D|$ a prime. Then the following are equivalent.*

- *$\mathcal{D}$ does not interpret the language of* 3-SAT.
- *$\mathbf{D}$ contains an operation t (equivalently, $\mathcal{D}$ has a polymorphism t) of arity p such that*

$$t(x_1, \ldots, x_p) \approx t(x_2, \ldots, x_p, x_1) \ .$$

Even if the tractability conjecture or the dichotomy conjecture (or finer classification conjectures) turns out to be incorrect, we *know* that classes of CSPs in P, L, NL, ... *can be* characterized by Malt'sev conditions on polymorphisms.

Example 2.16. *We show how to apply cyclic operations to prove the dichotomy theorem for undirected graphs [44].*

Let R be a symmetric binary relation viewed as an undirected graph and $\mathcal{D} = \{R\}$. Let $\mathcal{D}' = \{R', ...\}$ be the singleton expansion of the core of $\mathcal{D}$. If R contains a loop then CSP($\mathcal{D}$) *is trivially tractable. If R is bipartite, then the core of R is an edge and* CSP($\mathcal{D}$) *is essentially* 2-COLORING, *which is tractable.*

Finally, if R is not bipartite and does not contain a loop, then R' does not contain a loop and does contain a closed walk $a_1, a_2, \ldots, a_p, a_1$ for some prime $p > |D'|$. Assume that $\mathbf{D}'$ contains a cyclic operation t of arity p. Since t is a polymorphism, the pair

$$t((a_1, a_2), \ldots, (a_{p-1}, a_p), (a_p, a_1)) = (t(a_1, \ldots, a_p), t(a_2, \ldots, a_p, a_1))$$

is in R', but it is a loop since t is cyclic. This contradiction shows that $\mathbf{D}'$ does not contain a cyclic operation of arity p, therefore CSP($\mathcal{D}'$) *(and thus* CSP($\mathcal{D}$)*) is NP-complete.*

3 Results

Universal algebra serves the investigation in two ways: as a toolbox containing heavy hammers (such as the Tame Congruence Theory by Hobby and McKenzie [45]) and as a guideline for identifying interesting intermediate cases, which are hard to spot from the purely relational perspective. Major results include the following.

[9] The number of operations and identities appearing in a Mal'tsev condition can be infinite. If it is finite, then we speak about a *strong Mal'tse condtion.*

- The dichotomy theorem of Schaefer for CSPs over a two-element domain was generalized to a three-element domain by Bulatov [27]. A simplification of this result and a generalization to four-element domains was announced by Marković et al.

- The dichotomy theorem of Hell and Nešetřil for CSPs over undirected graphs was generalized to digraphs with no sources or sinks [16].

- The dichotomy conjecture was proved for all constraint languages containing all unary relations by Bulatov [29] (a simpler proof is in [5]).

Notably, all known tractable cases are solvable by a combination of two basic algorithms, or rather algorithmic principles – local consistency, and the "few subpowers" algorithm. It is another significant success of the algebraic approach that the applicability of these principles is now understood.

3.1 Local consistency

The CSP over some constraint languages can be decided in polynomial time by constraint propagation algorithms, or, in other words, by enforcing local consistency. Such CSPs are said to have *bounded width*.

This notion comes in various versions and equivalent forms. We refer to [40] for formalizations using Datalog programs and games, to [31] for description using dualities, and to [29, 7] for a notion suitable for infinite languages.

We informally sketch one possible definition. Let $k \leq l$ be positive integers. The (k,l)-*algorithm* derives the strongest possible constraints on k variables by considering l variables at a time. If a contradiction is found, the algorithm answers "no (solution)", otherwise it answers "yes". These algorithms work in polynomial time (for fixed k,l) and "no" answers are always correct. A constraint language $\mathcal{D}$ (or CSP($\mathcal{D}$)) has *width* (k,l), if "yes" answers are correct for every instance of CSP($\mathcal{D}$). If $\mathcal{D}$ has width (k,l) for some k,l, we say that $\mathcal{D}$ has *bounded width*.

As an example, we consider the constraint language $\mathcal{D}_{2\mathrm{COLOR}}$ and the instance

$$x_1 \neq x_2,\ x_2 \neq x_3,\ x_3 \neq x_4,\ x_4 \neq x_5,\ x_5 \neq x_1\ \ .$$

The $(2,3)$-algorithm can certify that this instance has no solution as follows:

- We consider the variables x_1, x_2, x_3. Using $x_1 \neq x_2$, $x_2 \neq x_3$ we derive $x_1 = x_3$.

- We consider x_1, x_3, x_4. Using $x_3 \neq x_4$ and the already derived constraint $x_1 = x_3$ we derive $x_1 \neq x_4$.

- We consider x_1, x_4, x_5 and using $x_1 \neq x_4$, $x_4 \neq x_5$ and $x_5 \neq x_1$ we derive a contradiction.

In fact, 2-COLORING has width $(2,3)$, that is, such reasoning finds a contradiction for every unsatisfiable instance. Other examples of bounded width problems include HORN-3-SAT and 2-SAT.

Feder and Vardi [40] proved that problems 3-LIN(p) (and more generally, similar problems 3-LIN($\mathbf{M}$) over finite modules) do not have bounded width and conjectured that linear equations are essentially the only obstacle for having bounded width. An algebraic formulation was given by Larose and Zádori [55]. They proved that analogues of results in section 2 hold for bounded width, therefore no problem which pp-interprets the language of 3-LIN($\mathbf{M}$) has bounded width, and conjectured that the converse is also true. After a sequence of partial results [51, 33, 8, 28], the conjecture was eventually confirmed in [13][10] and independently in [22].

Theorem 3.1. *An idempotent constraint language $\mathcal{D}$ has bounded width if and only if $\mathcal{D}$ does not interpret the language of* 3-LIN($\mathbf{M}$) for a finite module $\mathbf{M}$.[11]

3.2 Few subpowers

Gaussian elimination not only solves 3-LIN(p), it also describes all the solutions in the sense that the algorithm can output a small (polynomially large) set of points in $\mathrm{GF}(p)^n$ so that the affine hull of these points is equal to the solution set of the original instance. A sequence of papers [40, 25, 23, 37] culminating in [46, 18] pushed this idea, in a way, to its limit.

We need some terminology to state the result. Let $\mathcal{D}$ be a constraint language and $\mathbf{D}$ its clone of polymorphism. Recall that a relation on D is a *subpower* of $\mathbf{D}$ if and only if it is pp-definable from $\mathcal{D}$. Note that the set of solutions of any instance of CSP($\mathcal{D}$) can be viewed as a subpower of $\mathbf{D}$. Now $\mathbf{D}$ has *few subpowers* if each subpower can be obtained as a closure under polymorphisms of a small set (polynomially large with respect to the arity).[12]

Theorem 3.2. *Let $\mathcal{D}$ be an idempotent constraint language. If $\mathbf{D}$ has few subpowers, then* CSP($\mathcal{D}$) *can be solved in polynomial time.*

4 Conclusion

We have seen that the complexity of the satisfiability problem for CSP over a fixed constraint language depends on "higher arity symmetries" – polymorphisms of the language. (We have only discussed languages with finite domains. The algebraic theory extends to interesting subclasses of infinite domain CSP [20]). Significant progress has been achieved using this insight, but the main problem, the dichotomy conjecture, is still open.

A similar approach can be applied to other variants of CSP over a fixed constraint language. In two of them, the main goal has been reached: the dichotomy for the counting problem was proved in [30] (substantially simplified in [39]) and for the robust satisfiability problem in [12]. A generalization of the theory for the optimization problem and valued CSPs was given in [35], and some links to universal algebra are emerging from research in the area of approximation algorithms (such as [59]).

Is this approach only applicable to CSPs over fixed languages? Or are we merely seeing a piece of a bigger theory?

[10] A modification required to handle infinite languages was given in [7].

[11] Moreover, if $\mathcal{D}$ has bounded width, then it has width $(2,3)$ with an appropriate notion of width. Also, the property of having bounded width can be checked in polynomial time given an idempotent $\mathcal{D}$ on input.

[12] The name comes from an equivalent property that $\mathbf{D}$ has only exponentially many subpowers.

Comments on the included articles

Smooth digraphs [16]

A *digraph* R with vertex set D is a binary relation on the set D. It is called *smooth* if it has no sources or sinks, that is, every vertex $a \in D$ has an incoming edge $(b, a) \in R$ and an outgoing edge $(a, c) \in R$. Hell and Nešetřil [44] proved the dichotomy for the special case of symmetric digraphs:[13] CSP($\{R\}$) is in P if the core of R has at most 2 vertices, and is NP-complete otherwise.

Bang-Jensen and Hell [3] conjectured a generalization for smooth digraphs: CSP($\{R\}$) is in P if the core of R is a disjoint union of cycles, and is NP-complete otherwise. We confirmed their conjecture. This result was presented at the conference STOC'08 [15] and published in the SIAM Journal on Computing [16]. It follows from [4] that we also get a characterization of hereditarily hard digraphs, that is, digraphs whose CSP is NP-complete and remains NP-complete after adding arbitrary vertices and non-loop edges.

Since CSP over a disjoint union of cycles is easily seen to be polynomially solvable, the interesting part is to prove NP-completeness of the remaining cases. Our main tool was a characterization of the conjectured borderline between P and NP-complete CSPs in terms of *weak near unanimity* operations by Maróti and McKenzie [57]. It turned out that our result can be used to obtain alternative characterizations of the borderline – by means of a 4-ary operation satisfying a single identity [50] (this result was inspired by an earlier surprising characterization of Siggers [63] by means of a 6-ary operation) and by means of cyclic operations [11].

Bounded width [13]

This result was discussed in subsection 3.1: we confirmed the conjectures from [40, 27, 55] which characterize finite constraint languages whose CSP is solvable by local consistency methods (see Theorem 3.1). This result was presented at the conference FOCS'09 [9] and an improved result published in [13].

Techniques and concepts, such as *Prague instances* and *absorption* (already implicit in [16]), which we discovered while working on this project have found further significant applications.

[13] Note that a symmetric digraph is smooth unless it has an isolated vertex. Isolated vertices can be safely ignored.

Absorption and cyclic terms [11]

This work further developed the technique of absorption. A subuniverse B of an idempotent clone $\mathbf{A}$ is absorbing if $\mathbf{A}$ contains an operation t such that $t(a_1, \ldots, a_n) \in B$ whenever all but at most one a_i is in B. This simple concept is useful, because (1) a problem about $\mathbf{A}$ can often be reduced to a problem about an absorbing subuniverse of $\mathbf{A}$, and (2) "many" clones have nontrivial absorbing subuniverse. The second claim is witnessed by the Absorption Theorem proved in this paper.

As an application we gave a different proof of the dichotomy theorem for smooth digraphs which is shorter and does not require the result of Maróti and McKenzie [57] on weak near unanimity operations. In fact, our second application strengthened their characterization to cyclic operations (see Theorem 2.15). This new characterization is currently the syntactically strongest Mal'tsev condition for the conjectured borderline between P and NP-complete CSPs.

This work was presented at LICS'10 [10]. It was selected, as one of seven papers from this conference, to a special issue of the journal Logical Methods in Computer Science [11].

Congruence distributive finitely related algebras [6]

This paper applies CSP techniques from [8][14] to settle an algebraic problem – the Zádori conjecture.

I explain only one consequence of the main result. An operation t on a set A is a *near unanimity* operation if $t(a, \ldots, a, b, a, \ldots, a) = a$ for any $a, b \in A$ and any position of b. Clones containing near unanimity operations are important both in universal algebra (as they generalize lattices and have some pleasant properties [2]) and in CSP, where they characterize problems of *bounded strict width* [40] – these are, roughly, such CSPs whose solution can be obtained by a greedy algorithm after performing the (k, l)-algorithm described in subsection 3.1. A consequence of the main result of this paper is that it is decidable whether a finite constraint language has a near unanimity polymorphism, therefore it is decidable whether a finite constraint language defines a CSP of bounded strict width.

Conservative CSPs [5]

One of the main achievements toward the Feder–Várdi conjecture is the dichotomy theorem of Bulatov [26, 29] for conservative CPSs, that is, CSPs over languages that contain all unary relations. Bulatov's proof is very long and technical. My paper [5] presented at LICS'11 gives a new, quite short proof using (a refinement of) techniques from the bounded width paper [13].

Near unanimity in NL [17]

This paper contributes to finer (computational and descriptive) complexity classification of CSPs, namely, to classification of CSPs in NL.

The "obvious" obstructions for CSP($\mathcal{D}$) to be in the class NL is that $\mathcal{D}$ pp-interprets the language of 3-LIN($\mathbf{M}$) for a finite module $\mathbf{M}$ (since it is unlikely that these problems are in

[14] This paper contains a partial result on bounded width problems and it was the first paper which explicitly uses Prague instances and absorption.

NL) or that $\mathcal{D}$ pp-interprets the language of HORN-3-SAT (since HORN-3-SAT is P-complete, it is unlikely in NL). It is conjectured that these are the only obstacles – if $\mathcal{D}$ interprets neither 3-LIN($\mathbf{M}$) nor HORN-3-SAT, then CSP($\mathcal{D}$) is in NL, moreover it can be solved by a linear Datalog program, which is, roughly, a non-branching local consistency checking algorithm.

This conjecture remains open, but we confirmed it for what seems to be (in light of [6]) the last reasonable intermediate case – for languages with a near unanimity polymorphism.[15] The proof again uses the absorption technique in an essential way.

The result was presented at LICS'12 [17] and we were invited to submit the result to the Journal of the ACM (work on the journal version is in progress).

Robust satisfiability [12]

Robust satisfiability is one of the few variants of the CSP where the main goal – the classification of constraint languages over which the problem is polynomially solvable – has been achieved.

The task is a robust version of the search problem: we not only want to output a solution if a solution exists, we also want to output an "almost solution" of an "almost satisfiable" instance. More formally, we say that CSP($\mathcal{D}$) is robustly solvable, if there exists a function $f : [0,1] \to [0,1]$ with $\lim_{\varepsilon \to 0} f(\varepsilon) = 0$ and a polynomial algorithm which finds an assignment satisfying at least $(1 - f(\varepsilon))$-fraction of the constraints given a $(1 - \varepsilon)$-satisfiable instance of CSP($\mathcal{D}$).

This problem was introduced by Zwick [67] who also found such algorithms for 2-SAT and HORN-3-SAT based on linear programming and semidefinite programming relaxations of the CSP. From a deep result of Håstad [43] it follows that the problems 3-LIN($\mathbf{M}$) are not robustly solvable (unless P=NP). Guruswami and Zhou [42] recognized the similarity with the characterization of bounded width problems and conjectured that CSP($\mathcal{D}$) is robustly solvable if and only if CSP($\mathcal{D}$) has bounded width (assuming P $\neq$ NP). One direction was settled by Dalmau and Krokhin [38] who proved analogues of the results in section 2, thus CSP($\mathcal{D}$) can be robustly solvable only if CSP($\mathcal{D}$) has bounded width. We confirmed the Guruswami-Zhou conjecture by designing a robust polynomial algorithm for bounded width CSPs.

This result was presented at STOC'12 [12] and we were invited to submit it for publication in a special issue of the SIAM Journal of Computing for the best papers from the conference [14].

[15] The previous important step was done by Dalmau and Krokhin [36] who proved the conjecture for languages with a ternary near unanimity polymorphism.

Bibliography

[1] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.

[2] Kirby A. Baker and Alden F. Pixley. Polynomial interpolation and the Chinese remainder theorem for algebraic systems. *Mathematische Zeitschrift*, 143:165–174, 1975.

[3] Jørgen Bang-Jensen and Pavol Hell. The effect of two cycles on the complexity of colourings by directed graphs. *Discrete Appl. Math.*, 26(1):1–23, 1990.

[4] Jørgen Bang-Jensen, Pavol Hell, and Gary MacGillivray. Hereditarily hard H-colouring problems. *Discrete Math.*, 138(1-3):75–92, 1995. 14th British Combinatorial Conference (Keele, 1993).

[5] Libor Barto. The dichotomy for conservative constraint satisfaction problems revisited. In *26th Annual IEEE Symposium on Logic in Computer Science—LICS 2011*, pages 301–310. IEEE Computer Soc., Los Alamitos, CA, 2011.

[6] Libor Barto. Finitely related algebras in congruence distributive varieties have near unanimity terms. *Canad. J. Math.*, 65(1):3–21, 2013.

[7] Libor Barto. The collapse of the bounded width hierarchy. submitted, 2014.

[8] Libor Barto and Marcin Kozik. Congruence distributivity implies bounded width. *SIAM Journal on Computing*, 39(4):1531–1542, 2009.

[9] Libor Barto and Marcin Kozik. Constraint satisfaction problems of bounded width. In *FOCS'09: Proceedings of the 50th Symposium on Foundations of Computer Science*, pages 595–603, 2009.

[10] Libor Barto and Marcin Kozik. New conditions for Taylor varieties and CSP. In *Proceedings of the 2010 25th Annual IEEE Symposium on Logic in Computer Science*, LICS '10, pages 100–109, Washington, DC, USA, 2010. IEEE Computer Society.

[11] Libor Barto and Marcin Kozik. Absorbing subalgebras, cyclic terms, and the constraint satisfaction problem. *Logical Methods in Computer Science*, 8(1), 2012.

[12] Libor Barto and Marcin Kozik. Robust satisfiability of constraint satisfaction problems. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 931–940, New York, NY, USA, 2012. ACM.

[13] Libor Barto and Marcin Kozik. Constraint satisfaction problems solvable by local consistency methods. *J. ACM*, 61(1):3:1–3:19, January 2014.

[14] Libor Barto and Marcin Kozik. Robustly solvable constraint satisfaction problems. submitted, 2014.

[15] Libor Barto, Marcin Kozik, and Todd Niven. Graphs, polymorphisms and the complexity of homomorphism problems. In *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 789–796, New York, NY, USA, 2008. ACM.

[16] Libor Barto, Marcin Kozik, and Todd Niven. The CSP dichotomy holds for digraphs with no sources and no sinks (a positive answer to a conjecture of Bang-Jensen and Hell). *SIAM J. Comput.*, 38(5):1782–1802, 2008/09.

[17] Libor Barto, Marcin Kozik, and Ross Willard. Near unanimity constraints have bounded pathwidth duality. In *Proceedings of the 2012 27th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 125–134. IEEE Computer Soc., Los Alamitos, CA, 2012.

[18] Joel Berman, Pawel Idziak, Petar Marković, Ralph McKenzie, Matthew Valeriote, and Ross Willard. Varieties with few subalgebras of powers. *Transactions of The American Mathematical Society*, 362:1445–1473, 2009.

[19] Garrett Birkhoff. On the structure of abstract algebras. *Proc. Camb. Philos. Soc.*, 31:433–454, 1935.

[20] Manuel Bodirsky. Constraint satisfaction problems with infinite templates. In Nadia Creignou, Phokion G. Kolaitis, and Heribert Vollmer, editors, *Complexity of Constraints*, volume 5250 of *Lecture Notes in Computer Science*, pages 196–228. Springer, 2008.

[21] V. G. Bodnarčuk, L. A. Kalužnin, V. N. Kotov, and B. A. Romov. Galois theory for Post algebras. I, II. *Kibernetika (Kiev)*, (3):1–10; ibid. 1969, no. 5, 1–9, 1969.

[22] Andrei Bulatov. Bounded relational width. manuscript, 2009.

[23] Andrei Bulatov and Víctor Dalmau. A simple algorithm for Mal′tsev constraints. *SIAM J. Comput.*, 36(1):16–27 (electronic), 2006.

[24] Andrei Bulatov, Peter Jeavons, and Andrei Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM J. Comput.*, 34:720–742, March 2005.

[25] Andrei A. Bulatov. Mal'tsev constraints are tractable. *Electronic Colloquium on Computational Complexity (ECCC)*, (034), 2002.

[26] Andrei A. Bulatov. Tractable conservative constraint satisfaction problems. In *Proceedings of the 18th Annual IEEE Symposium on Logic in Computer Science*, pages 321–, Washington, DC, USA, 2003. IEEE Computer Society.

[27] Andrei A. Bulatov. A graph of a relational structure and constraint satisfaction problems. In *Proceedings of the Nineteenth Annual IEEE Symposium on Logic in Computer Science (LICS 2004)*, pages 448–457. IEEE Computer Society Press, July 2004.

[28] Andrei A. Bulatov. Combinatorial problems raised from 2-semilattices. *J. Algebra*, 298(2):321–339, 2006.

[29] Andrei A. Bulatov. Complexity of conservative constraint satisfaction problems. *ACM Trans. Comput. Logic*, 12(4):24:1–24:66, July 2011.

[30] Andrei A. Bulatov. The complexity of the counting constraint satisfaction problem. *J. ACM*, 60(5):34:1–34:41, October 2013.

[31] Andrei A. Bulatov, Andrei Krokhin, and Benoit Larose. Complexity of constraints. chapter Dualities for Constraint Satisfaction Problems, pages 93–124. Springer-Verlag, Berlin, Heidelberg, 2008.

[32] Jin-Yi Cai and Xi Chen. Complexity of counting csp with complex weights. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 909–920, New York, NY, USA, 2012. ACM.

[33] Catarina Carvalho, Víctor Dalmau, Petar Marković, and Miklós Maróti. CD(4) has bounded width. *Algebra Universalis*, 60(3):293–307, 2009.

[34] Hubie Chen. Meditations on quantified constraint satisfaction. In RobertL. Constable and Alexandra Silva, editors, *Logic and Program Semantics*, volume 7230 of *Lecture Notes in Computer Science*, pages 35–49. Springer Berlin Heidelberg, 2012.

[35] David A. Cohen, Martin C. Cooper, Páidí Creed, Peter Jeavons, and Stanislav Živný. An algebraic theory of complexity for discrete optimisation. *SIAM Journal on Computing*, 42(5):1915–1939, 2013.

[36] Vctor Dalmau and Andrei Krokhin. Majority constraints have bounded pathwidth duality. *European Journal of Combinatorics*, 29(4):821 – 837, 2008. Homomorphisms: Structure and Highlights Homomorphisms: Structure and Highlights.

[37] Víctor Dalmau. Generalized majority-minority operations are tractable. *Log. Methods Comput. Sci.*, 2(4):4:1, 14, 2006.

[38] Víctor Dalmau and Andrei Krokhin. Robust satisfiability for csps: Hardness and algorithmic results. *ACM Trans. Comput. Theory*, 5(4):15:1–15:25, November 2013.

[39] Martin E. Dyer and David Richerby. An effective dichotomy for the counting constraint satisfaction problem. *SIAM J. Comput.*, 42(3):1245–1274, 2013.

[40] Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic snp and constraint satisfaction: A study through datalog and group theory. *SIAM Journal on Computing*, 28(1):57–104, 1998.

[41] David Geiger. Closed systems of functions and predicates. *Pacific J. Math.*, 27:95–100, 1968.

[42] Venkatesan Guruswami and Yuan Zhou. Tight bounds on the approximability of almost-satisfiable Horn SAT and exact hitting set. In Dana Randall, editor, *SODA*, pages 1574–1589. SIAM, 2011.

[43] Johan Håstad. On the efficient approximability of constraint satisfaction problems. In *Surveys in combinatorics 2007*, volume 346 of *London Math. Soc. Lecture Note Ser.*, pages 201–221. Cambridge Univ. Press, Cambridge, 2007.

[44] Pavol Hell and Jaroslav Nešetřil. On the complexity of H-coloring. *J. Combin. Theory Ser. B*, 48(1):92–110, 1990.

[45] David Hobby and Ralph McKenzie. *The structure of finite algebras*, volume 76 of *Contemporary Mathematics*. American Mathematical Society, Providence, RI, 1988.

[46] Pawel Idziak, Petar Marković, Ralph McKenzie, Matthew Valeriote, and Ross Willard. Tractability and learnability arising from algebras with few subpowers. In *Proceedings of the Twenty-Second Annual IEEE Symposium on Logic in Computer Science (LICS 2007)*, pages 213–222. IEEE Computer Society Press, July 2007.

[47] Neil Immerman. Nondeterministic space is closed under complementation. *SIAM J. Comput.*, 17(5):935–938, October 1988.

[48] Peter Jeavons. On the algebraic structure of combinatorial problems. *Theoretical Computer Science*, 200(12):185 – 204, 1998.

[49] Peter Jeavons, David Cohen, and Marc Gyssens. Closure properties of constraints. *J. ACM*, 44(4):527–548, 1997.

[50] Keith Kearnes, Petar Markovi, and Ralph McKenzie. Optimal strong malcev conditions for omitting type 1 in locally finite varieties. *Algebra universalis*, 72(1):91–100, 2014.

[51] Emil Kiss and Matthew Valeriote. On tractability and congruence distributivity. *Log. Methods Comput. Sci.*, 3(2):2:6, 20 pp. (electronic), 2007.

[52] Gábor Kun and Mario Szegedy. A new line of attack on the dichotomy conjecture. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 725–734, New York, NY, USA, 2009. ACM.

[53] Richard E. Ladner. On the structure of polynomial time reducibility. *J. Assoc. Comput. Mach.*, 22:155–171, 1975.

[54] Benoît Larose and Pascal Tesson. Universal algebra and hardness results for constraint satisfaction problems. *Theor. Comput. Sci.*, 410:1629–1647, April 2009.

[55] Benoit Larose and László Zádori. Bounded width problems and algebras. *Algebra Universalis*, 56(3-4):439–466, 2007.

[56] F. Madelaine and B. Martin. A tetrachotomy for positive first-order logic without equality. In *Logic in Computer Science (LICS), 2011 26th Annual IEEE Symposium on*, pages 311–320, June 2011.

[57] Miklós Maróti and Ralph McKenzie. Existence theorems for weakly symmetric operations. *Algebra Universalis*, 59(3-4):463–489, 2008.

[58] Christos H. Papadimitriou. *Computational complexity.* Addison-Wesley Publishing Company, Reading, MA, 1994.

[59] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *STOC'08*, pages 245–254, 2008.

[60] Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4):17:1–17:24, September 2008.

[61] Francesca Rossi, Peter van Beek, and Toby Walsh. *Handbook of Constraint Programming (Foundations of Artificial Intelligence)*. Elsevier Science Inc., New York, NY, USA, 2006.

[62] Thomas J. Schaefer. The complexity of satisfiability problems. In *Conference Record of the Tenth Annual ACM Symposium on Theory of Computing (San Diego, Calif., 1978)*, pages 216–226. ACM, New York, 1978.

[63] MarkH. Siggers. A strong malcev condition for locally finite varieties omitting the unary type. *Algebra universalis*, 64(1-2):15–20, 2010.

[64] R. Szelepcsényi. The method of forced enumeration for nondeterministic automata. *Acta Inf.*, 26(3):279–284, November 1988.

[65] Walter Taylor. Varieties obeying homotopy laws. *Canad. J. Math.*, 29(3):498–527, 1977.

[66] Stanislav Živný. *The complexity of valued constraint satisfaction problems*. Cognitive Technologies. Springer, Heidelberg, 2012.

[67] Uri Zwick. Finding almost-satisfying assignments. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, STOC '98, pages 551–560, New York, NY, USA, 1998. ACM.

Appendix A – Smooth digraphs

SIAM J. Comput.
Vol. 38, No. 5, pp. 1782–1802

THE CSP DICHOTOMY HOLDS FOR DIGRAPHS WITH NO SOURCES AND NO SINKS (A POSITIVE ANSWER TO A CONJECTURE OF BANG-JENSEN AND HELL)*

LIBOR BARTO†, MARCIN KOZIK‡, AND TODD NIVEN†

Abstract. Bang-Jensen and Hell conjectured in 1990 (using the language of graph homomorphisms) a constraint satisfaction problem (CSP) dichotomy for digraphs with no sources or sinks. The conjecture states that the CSP for such a digraph is tractable if each component of its core is a cycle and is NP-complete otherwise. In this paper we prove this conjecture and, as a consequence, a conjecture of Bang-Jensen, Hell, and MacGillivray from 1995 classifying hereditarily hard digraphs. Further, we show that the CSP dichotomy for digraphs with no sources or sinks agrees with the algebraic characterization conjectured by Bulatov, Jeavons, and Krokhin in 2005.

Key words. constraint satisfaction problem, graph homomorphism, smooth digraphs

AMS subject classifications. 68R10, 08A70

DOI. 10.1137/070708093

1. Introduction. The history of the constraint satisfaction problem (CSP) goes back more than thirty years and begins with the work of Montanari [Mon74] and Mackworth [Mac77]. Since that time many combinatorial problems in artificial intelligence and other areas of computer science have been formulated in the language of CSPs. The study of such problems, under this common framework, has applications in database theory [Var00], machine vision recognition [Mon74], temporal and spatial reasoning [SV98], truth maintenance [DD96], technical design [NL], scheduling [LALW98], natural language comprehension [All94], and programming language comprehension [Nad]. Numerous attempts to understand the structure of different CSPs has been undertaken, and a wide variety of tools ranging from statistical physics (e.g., [ANP05, KMRT$^+$07]) to universal algebra (e.g., [JCG97]) has been employed. Methods and results developed in seemingly disconnected branches of mathematics transformed the area. The conjecture proved in this paper resisted the approaches based in combinatorics and theoretical computer science for nearly twenty years. Only recent developments in the structural theory of finite algebras provided tools strong enough to solve this problem.

For the last ten years the study of CSPs has also been a driving force in theoretical computer science. The dichotomy conjecture of Feder and Vardi, published in [FV99], has origins going back to 1993. The conjecture states that a CSP, for any fixed language, is solvable in polynomial time or NP-complete. Therefore the class of CSPs would be a subclass of NP avoiding problems of intermediate difficulty. The logical

*Received by the editors November 14, 2007; accepted for publication (in revised form) August 1, 2008; published electronically January 9, 2009. A part of this article appeared, in a preliminary form, in the Proceedings of the 40th ACM Symposium on Theory of Computing, STOC'08. This work was partly supported by the Eduard Cech Center grant LC505.

http://www.siam.org/journals/sicomp/38-5/70809.html

†Eduard Čech Center, Charles University, 186 75 Praha, Czech Republic (barto@karlin.mff.cuni.cz, niven@karlin.mff.cuni.cz). The first author was supported by the Grant Agency of the Czech Republic under grant 201/06/0664 and by the project of Ministry of Education under grant MSM 0021620839.

‡Eduard Čech Center, Charles University, 186 75 Praha, Czech Republic, and Algorithmics Research Group, Jagiellonian University, 30-072 Kraków, Poland (marcin.kozik@tcs.uj.edu.pl).

characterization of the class of CSPs (see [FV99] and [Kun]) provides arguments in support of the dichotomy; nevertheless the conjecture remains open.

One of the results of [FV99] shows that the CSP dichotomy conjecture is equivalent to the CSP dichotomy conjecture restricted to digraphs. Therefore the CSPs can be defined in terms of the (di)graph homomorphisms studied in graph theory for over forty years (cf. [Sab61, HP64, Lev73]). It adds a new dimension to a well-established problem and shows the importance of solving CSPs for digraphs. The classification of the complexity of the **H**-coloring problems for undirected graphs, discovered by Hell and Nešetřil [HN90], is an important step and provides a starting point towards proving, or refuting, the CSP dichotomy conjecture. There have since appeared many papers on the complexity of digraph coloring problems (see, e.g., [BJH90, BJHM95, Fed01, GWW92, HNZ96a, HNZ96b, HNZ96c, HZZ93, Mac91, Zhu95]), but as yet, no plausible conjecture on a graph theoretical classification has been proposed. Bang-Jensen and Hell [BJH90] did, however, conjecture a classification (implying the dichotomy) for the class of digraphs with no sources or sinks. Their conjecture significantly generalizes the result of Hell and Nešetřil.

In 1995, Bang-Jensen, Hell, and MacGillivray (in [BJHM95]) introduced the notion of hereditarily hard digraphs and conjectured their classification. Surprisingly, they were able to show that this conjecture and the one given in [BJH90] are equivalent. In this paper we prove the conjecture of Bang-Jensen and Hell and, as a consequence, the conjecture of Bang-Jensen, Hell, and MacGillivray.

Our paper relies on the interconnection between the CSP and algebra as first discovered by Jeavons, Cohen, and Gyssens in [JCG97] and refined by Bulatov, Jeavons, and Krokhin in [BJK05]. Using this connection, Bulatov, Jeavons, and Krokhin conjectured a full classification of the NP-complete CSPs. For a small taste of results in the direction of proving this classification, see [BIM$^+$06, Bul06, Dal05, Dal06, KV07]. A particularly interesting example, demonstrating the potency of the algebraic approach, is Bulatov's proof of the result of Hell and Nešetřil (see [Bul05]). A recent, purely algebraic result of Maróti and McKenzie [MM07] is one of the key ingredients in the proof of the conjecture of Bang-Jensen and Hell. This provides further evidence supporting the extremely strong bond between the CSP and universal algebra.

2. Preliminaries. We assume that the reader possesses a basic knowledge of universal algebra and graph theory. For an easy introduction to the notions of universal algebra that are not defined in this paper, we invite the reader to consult the monographs [BS81] and [MMT87]. Further information concerning the structural theory of finite algebras (called tame congruence theory) can be found in [HM88]. For an explanation of the basic terms in graph theory and graph homomorphisms, we recommend [HN04]. Finally, for an introduction to the connections between universal algebra and the CSP, we recommend [BJK05].

Throughout the paper we deviate from the standard definition of the CSP, with respect to a fixed language (found in, e.g., [BKJ00]), in favor of an equivalent definition from [FV99, LZ06]. The definitions of a *relational structure*, a *homomorphism*, or a *polymorphism* are presented further in this section in their full generality as well as in restriction to directed graphs.

A *directed graph* (or *digraph*) is a pair $\mathbf{G} = (V, E)$, where V is a set of vertices and $E \subseteq V \times V$ is a set of edges. More generally *a relational structure* $\mathcal{T} = (T, \mathcal{R})$ is an ordered pair, where T is a finite nonempty set and $\mathcal{R}$ is a finite set of finitary relations on T indexed by a set J. Let d_j denote the arity of the relation $R_j \in \mathcal{R}$. The indexed set of all the d_j constitutes the *signature* of $\mathcal{T}$.

A vertex of a digraph is called a *source* (resp., a *sink*) if it has no incoming (resp., outgoing) edges. An *oriented walk* is a sequence of vertices $(v_0, \dots, v_{n-1})$ such that $(v_i, v_{i+1}) \in E$ or $(v_{i+1}, v_i) \in E$ for any $i < n-1$ and the *length* of such a walk is $n-1$. A *walk* is an oriented walk such that $(v_i, v_{i+1}) \in E$ for any $i < n-1$. A *closed walk* is a walk such that $v_0 = v_{n-1}$. Given a digraph $\mathbf{G}$, we sometimes denote the set of vertices of $\mathbf{G}$ by $V(\mathbf{G})$ and similarly the edges of $\mathbf{G}$ by $E(\mathbf{G})$. A graph with n vertices is a *cycle* if its vertices can be ordered (i.e., $V = \{v_0, \dots, v_{n-1}\}$) in such a way that $E = \{(v_i, v_j) | \, j = i+1 \mod n\}$.

A *graph homomorphism* is a function between sets of vertices of two graphs mapping edges to edges. A graph is 3-colorable if and only if it maps homomorphically to the complete graph on three vertices (without loops). The notion of *colorability* is generalized using graph homomorphisms: a digraph, say $\mathbf{G}$, is $\mathbf{H}$-colorable if there exists a homomorphism mapping $\mathbf{G}$ to $\mathbf{H}$. For two relational structures of the same signature, say $\mathcal{T} = (T, \mathcal{R})$ and $\mathcal{U} = (U, \mathcal{S})$, a map $h : T \to U$ is a *homomorphism* if $h(T_j) \subseteq R_j$ for all $j \in J$ (where $h(T_j)$ is computed pointwise).

A *digraph polymorphism* is a homomorphism from a finite Cartesian power of a graph to the graph itself. Precisely, for a digraph $\mathbf{G} = (V, E)$ a function $h : V^n \to V$ is a *polymorphism* of $\mathbf{G}$ if, for any vertices $a_0, \dots, a_{n-1}, b_0, \dots, b_{n-1} \in V$,

$$\text{if } (a_i, b_i) \in E \text{ for all } i < n, \text{ then } (h(a_0, \dots, a_{n-1}), h(b_0, \dots, b_{n-1})) \in E.$$

The notion of a polymorphism is defined for relational structures as well. A *polymorphism* h of $\mathcal{T}$ is an operation $h : T^n \to T$ such that, for all relations $R \in \mathcal{R}$ of arity m, if

$$(a_{i,0}, a_{i,1}, \dots, a_{i,m-1}) \in R \qquad \text{for all } i < n,$$

then

$$(h(a_{0,0}, a_{1,0}, \dots, a_{n-1,0}), \dots, h(a_{0,m-1}, a_{1,m-1}, \dots, a_{n-1,m-1})) \in R.$$

A digraph $\mathbf{G} = (V, E)$ *retracts* to an induced subgraph $\mathbf{H} = (W, F)$ if there is an endomorphism $h : V \to V$ such that $h(V) = W$ and $h(a) = a$ for all $a \in W$. Such a map h is called a *retraction*. A *core* of a digraph is a minimal induced subgraph to which the digraph retracts. The definition of retraction and core clearly generalize to relational structures. It is a trivial fact that, for any digraph $\mathbf{H}$, and for a core of $\mathbf{H}$, say $\mathbf{H}'$, the set of $\mathbf{H}$-colorable digraphs coincides with the set of $\mathbf{H}'$-colorable digraphs.

An *algebra* is a tuple $\mathbf{A} = (A, f_0, \dots)$ consisting of a nonempty set A (called a *universe* of $\mathbf{A}$) and *operations* on A. An *operation* f_i is an n_i-ary function $f_i : A^{n_i} \to A$. With each operation we associate an *operation symbol* and, by an abuse of notation, denote it also by f_i. A set $B \subseteq A$ is a *subuniverse* of an algebra $\mathbf{A}$ if, for any number i, the operation f_i restricted to B^{n_i} has all the results in B. For a nonempty subuniverse B of an algebra $\mathbf{A}$ the algebra $\mathbf{B} = (B, f'_0, \dots)$ (where f'_i is a restriction of f_i to B^{n_i}) is a *subalgebra* of $\mathbf{A}$. A *power* of an algebra $\mathbf{A}$ has a universe A^k and the operations f''_i derived from the operations of $\mathbf{A}$ by computing coordinatewise. A subalgebra of a power of an algebra is often called a *subpower*. A *term function* of an algebra is any function that can be obtained by a composition using the operations of the algebra together with all the projections. A *term* is a formal way of denoting such a composition; i.e., a term function is attached to an algebra, but a term can be computed in a subalgebra or a power as well as in the original algebra. A set $C \subseteq A$

generates a subuniverse B in an algebra $\mathbf{A}$ if B is the smallest subuniverse containing C—such a subuniverse always exists and can be obtained by applying all the term functions of the algebra $\mathbf{A}$ to all the choices of arguments coming from C.

In this paper all relational structures, digraphs, and algebras are assumed to be finite.

3. The main result. For a relational structure $\mathcal{T} = (T, R)$ we define the language CSP($\mathcal{T}$), of relational structures with the same signature as $\mathcal{T}$, to be

$$\mathrm{CSP}(\mathcal{T}) = \{\mathcal{U} \mid \text{there is a homomorphism from } \mathcal{U} \text{ to } \mathcal{T}\}.$$

Alternatively we can view CSP($\mathcal{T}$) as a decision problem:

INPUT: a relational structure $\mathcal{U}$ with the same signature as $\mathcal{T}$
QUESTION: does there exist a homomorphism from $\mathcal{U}$ to $\mathcal{T}$?

In either approach we are concerned with the computational complexity (of membership of the language, or of the decision problem, respectively) for a given relational structure. The CSP dichotomy conjecture proposed in [FV99] can be stated as follows.

The CSP dichotomy conjecture. *For a relational structure $\mathcal{T}$ the problem* CSP($\mathcal{T}$) *is solvable in polynomial time or NP-complete.*

The (di)graph coloring problems can be viewed as special cases of the CSP. Although a digraph $\mathbf{H} = (W, F)$ is technically different from a relational structure, the set of $\mathbf{H}$-colorable digraphs is obviously polynomially equivalent to the CSP for an appropriate relational structure, and therefore we denote the class of all $\mathbf{H}$-colorable digraphs by CSP($\mathbf{H}$). Due to the reduction presented in [FV99], every CSP is polynomially equivalent to a digraph homomorphism problem. Thus we can restate the CSP dichotomy conjecture in the following way.

The CSP dichotomy conjecture. *For a fixed digraph $\mathbf{H}$, deciding whether a given digraph is $\mathbf{H}$-colorable is either NP-complete or solvable in polynomial time.*

This brings us to the main problem of the paper, a conjecture nearly ten years older than the CSP dichotomy conjecture, and a special case of it. It deals with digraphs with no sources or sinks and was first formulated by Bang-Jensen and Hell in [BJH90].

The conjecture of Bang-Jensen and Hell. *Let $\mathbf{H}$ be a digraph without sources or sinks. If each component of the core of $\mathbf{H}$ is a cycle, then* CSP($\mathbf{H}$) *is polynomially decidable. Otherwise* CSP($\mathbf{H}$) *is NP-complete.*

Note that the above conjecture is a substantial generalization of the $\mathbf{H}$-coloring result of Hell and Nešetřil [HN90].

The notion of hereditarily hard digraphs was introduced by Bang-Jensen, Hell, and MacGillivray in [BJHM95]. A digraph $\mathbf{H}$ is said to be *hereditarily hard* if the $\mathbf{H}'$-coloring problem is NP-complete for all loopless digraphs $\mathbf{H}'$ that contain $\mathbf{H}$ as a subgraph (not necessarily induced). The following conjecture was posed and shown to be equivalent to the Bang-Jensen and Hell conjecture in [BJHM95].

The conjecture of Bang-Jensen, Hell, and MacGillivray. *Let $\mathbf{H}$ be a digraph. If the digraph $R(\mathbf{H})$ (which is obtained by iteratively removing the sources and sinks from $\mathbf{H}$ until none remain) does not admit a homomorphism to a cycle of length greater than one, then $\mathbf{H}$ is hereditarily hard. Otherwise there exists a loopless digraph $\mathbf{H}'$ containing $\mathbf{H}$ (as a not necessarily induced subgraph) such that $\mathbf{H}'$-coloring is solvable in a polynomial time.*

In this section we prove the Bang-Jensen and Hell conjecture and therefore the conjecture of Bang-Jensen, Hell, and MacGillivray. In this proof we assume Theorem 3.1, which will be proved in the subsequent sections of the paper. The reasoning uses weak near unanimity operations[1] and Taylor operations (used only to connect Theorems 3.2 and 3.3, and therefore not defined here [HM88, Tay77, LZ06]).

THEOREM 3.1. *If a digraph without sources or sinks admits a weak near unanimity polymorphism, then it retracts to the disjoint union of cycles.*

It is easy to see that the colorability by digraphs retracting to a disjoint union of cycles is tractable (see, e.g., [BJH90]). It remains to prove the NP-completeness of the digraphs not retracting to such a union. Before we do so, we recall two fundamental results.

It follows from [HM88, Lemma 9.4 and Theorem 9.6] that a part of the result of Maróti and McKenzie [MM07, Theorem 1.1] can be stated as follows.

THEOREM 3.2 (see [MM07]). *A finite relational structure $\mathcal{T}$ admits a Taylor polymorphism if and only if it admits a weak near unanimity polymorphism.*

The following result was originally proved in [BKJ00] and [LZ03] and, as stated below, can be found in [LZ06, Theorem 2.3]. It relies on a connection between relational structures and varieties generated by algebras of their polymorphisms. A lack of a Taylor polymorphism in such an algebra implies an existence of a "trivial" algebra in a variety and NP-completeness of the associated CSP.

THEOREM 3.3 (see [LZ06]). *Let $\mathcal{T}$ be a relational structure which is a core. If $\mathcal{T}$ does not admit a Taylor polymorphism, then* $\mathrm{CSP}(\mathcal{T})$ *is NP-complete.*

If a digraph $\mathbf{H}$ without sources or sinks does not retract to a disjoint union of cycles, then its core $\mathbf{H}'$ also does not. Thus, by Theorem 3.1, it follows that $\mathbf{H}'$ does not admit a weak near unanimity polymorphism, and by Theorems 3.2 and 3.3 it follows that $\mathrm{CSP}(\mathbf{H}')$ is NP-complete, completing the proof of the conjecture of Bang-Jensen and Hell.

The conjecture (posed in [BKJ00]), classifying the CSPs from the algebraic point of view, can be stated as follows (see, e.g., [LZ06]).

THE ALGEBRAIC CSP DICHOTOMY CONJECTURE. *Let $\mathcal{T}$ be a relational structure that is a core. If $\mathcal{T}$ admits a Taylor polymorphism, then* $\mathrm{CSP}(\mathcal{T})$ *is polynomial time solvable. Otherwise* $\mathrm{CSP}(\mathcal{T})$ *is NP-complete.*

Note that the proof of the conjecture of Bang-Jensen and Hell immediately implies that the structure of the NP-complete digraph coloring problems agrees with the algebraic CSP dichotomy conjecture. The remainder of the paper is dedicated to the proof of the Theorem 3.1.

4. Notation.

In this section we introduce the notation required throughout the remainder of the paper.

4.1. Neighborhoods in graphs.

For a fixed digraph $\mathbf{G} = (V, E)$ we denote $(a, b) \in E$ by $a \to b$, and we use $a \xrightarrow{k} b$ to say that there is a directed walk from a to b of length precisely k. More generally we call a digraph $\mathbf{H}$ a *pattern* if $V(\mathbf{H}) = \{0, \ldots, n-1\}$ and $(u, v) \in E$ if and only if $|u - v| = 1$ and $(v, u) \notin E$. We denote patterns by lowercase Greek letters and, for a pattern α, we write $a \xrightarrow{\alpha} b$ if there exists a homomorphism ϕ from α into $\mathbf{G}$ such that $\phi(0) = a$ and $\phi(n-1) = b$. In such a case we say that a and b can be connected via the pattern α. The oriented walk connecting vertices a and b and consisting of the images of elements of α under ϕ is a *realization* of the pattern. For any $W \subseteq V$ we define

$$W^{+n} = \{v \in V \mid (\exists w \in W)\ w \xrightarrow{n} v\}$$

[1] A weak near unanimity operation is a function such that, for any choice of arguments a, b, $w(b, a, \ldots, a) = w(a, b, \ldots, a) = \cdots = w(a, a, \ldots, b)$ and $w(a, \ldots, a) = a$. These operations are described in more detail in section 4.

and similarly

$$W^{-n} = \{v \in V \mid (\exists w \in W)\ v \xrightarrow{n} w\}.$$

We define $W^0 = W$, and write a^{+n} (resp., a^{-n}, a^0) instead of $\{a\}^{+n}$ (resp., $\{a\}^{-n}, \{a\}^0$) for any $a \in V$. More generally, for a pattern α, we write

$$W^{\alpha} = \{v \in V \mid (\exists w \in W)\ w \xrightarrow{\alpha} v\}.$$

As before, we use a^{α} for $\{a\}^{\alpha}$. Sometimes, for ease of presentation, we write $a \xrightarrow{k,n} b$ to denote $a \xrightarrow{k} b$ and $a \xrightarrow{n} b$.

4.2. Digraph path powers. Let $\mathbf{G} = (V, E)$ be a digraph and α be a pattern. We define a path power of the digraph $\mathbf{G}$, which we denote by $\mathbf{G}^{\alpha}$, in the following way: the vertices of the power are the vertices of the digraph $\mathbf{G}$, and a pair $(c, d) \in V^2$ is an edge in $\mathbf{G}^{\alpha}$ if and only if $c \xrightarrow{\alpha} d$ in $\mathbf{G}$. Moreover, we set $\mathbf{G}^{+n} = \mathbf{G}^{\alpha}$ for the pattern α consisting of n arrows pointing forward. Note that if $f : V^m \to V$ is a polymorphism of $\mathbf{G}$, then it is also a polymorphism of any path power of this digraph. Path powers are special cases of primitive positive definitions (used in, e.g., [Bul05]) or indicator constructions introduced in [HN90] in order to deal with the colorability problem for undirected graphs.

4.3. Components. A *connected* digraph is a digraph such that there exists an oriented walk, consisting of at least one edge, between every choice of two vertices. A *strongly connected* digraph is a digraph such that, for every choice of two vertices, there is a walk connecting them. By a *component* (resp., *strong component*) of a digraph $\mathbf{G}$, we mean a maximal (under inclusion) induced subgraph that is connected (resp., strongly connected). Note that, according to this definition, a single vertex with the empty set of edges is not connected, and thus not every digraph decomposes into a union of components (or strong components). Given a digraph $\mathbf{G}$ with no sources or sinks, we say that a strong component $\mathbf{H}$ of $\mathbf{G}$ is a *top component* if $V(\mathbf{H})^{+1} = V(\mathbf{H})$. Similarly, we say that a strong component $\mathbf{H}$ of $\mathbf{G}$ is a *bottom component* if $V(\mathbf{H})^{-1} = V(\mathbf{H})$.

4.4. Algebraic length. The following definition is taken from [HNZ96b]. For a pattern α we define the *algebraic length* $al(\alpha)$ to be

$$al(\alpha) = |\{\text{edges going forward in } \alpha\}| - |\{\text{edges going backward in } \alpha\}|.$$

An *algebraic length of an oriented walk* is a shorthand expression for an algebraic length of a pattern which can be realized as such an oriented walk—the pattern is always clear from the context. For a digraph $\mathbf{G} = (V, E)$ we set

$$al(\mathbf{G}) = \min\{i > 0 \mid (\exists v \in V)\ (\exists \text{ a pattern } \alpha)\ v \xrightarrow{\alpha} v \text{ and } al(\alpha) = i\}$$

whenever the set on the right-hand side is nonempty and ∞ otherwise. In case of strongly connected digraphs in section 7 the algebraic length can be equivalently defined (cf. Corollary 5.7) as the greatest common divisor of the lengths of closed walks in a digraph. We note that for digraphs with no sources or sinks (or with a closed walk) the algebraic length of a nonempty digraph is always a natural number. It is folklore (cf. [HN04, Proposition 5.19]) that a connected digraph $\mathbf{G}$ retracts to a cycle if and only if it contains a closed walk of length $al(\mathbf{G})$.

4.5. Algebraic notation. By $\overline{a}$ we denote the tuple $(a, a, \dots, a)$ (the arity will always be clear by the context), and by $\overrightarrow{a}$ we denote the tuple $(a_0, a_1, \dots, a_n)$. Further, we extend the notation $\overline{a}$ to the sets in the following way. For a set W let $\overline{W}$ be an appropriate Cartesian power of W. Thus, for example, given a vertex a of a digraph $\mathbf{G}$, the set $\overline{a^{+n}}$ is the collection of all tuples whose coordinates are vertices reachable by a walk of length n from a.

An idempotent operation on a set A is an operation, say $f : A^n \to A$, such that $f(\overline{a}) = a$ for all $a \in A$. In accordance with [MM07], by a weak near unanimity operation we understand an idempotent operation $w(x_0, \dots, x_{n-1})$ that satisfies

$$w(y, x, \dots, x) = w(x, y, \dots, x) = \cdots = w(x, x, \dots, y),$$

for any choice of x and y in the underlying set. Moreover, for a term t of arity n, we define

$$t^{(i)}(x_0, x_1, \dots, x_{n-1}) = t(x_{n-i}, x_{n-i+1}, \dots, x_0, x_1, \dots, x_{n-i-1}),$$

for each $0 \le i < n$, where addition on the indices is performed modulo n.

5. Preliminary results on digraphs. We start with a number of basic results describing the connection between digraphs and their path powers. The following lemma reveals the behavior of the algebraic lengths of oriented walks in powers of a digraph.

LEMMA 5.1. *Let* $\mathbf{G}$ *be a digraph without sources or sinks. Let* α *be a pattern of algebraic length* k*, and let* $a \xrightarrow{\alpha} b$ *in* $\mathbf{G}$*. Then* $a \xrightarrow{\beta} b$ *in* $\mathbf{G}^{+k}$ *for some pattern* β *of algebraic length one.*

Proof. For a fixed, large enough number j, consider all oriented walks in $\mathbf{G}$ of the form $a \xrightarrow{l_1} a_1 \xleftarrow{l_2} a_2 \xrightarrow{l_3} \cdots \leftrightarrows a_{l_j} = b$, where $l_1 - l_2 + \cdots \pm l_j = k$. We will show that at least one of these walks has all the l_i's divisible by k. Let us choose an oriented walk in which k divides all the l_i in a maximal initial segment of the i, and let l_{i_0} be the last element of this segment. If $i_0 + 1 < j$, then (assuming without loss of generality that i_0 is odd) the walk

$$a \xrightarrow{l_1} \cdots \xrightarrow{l_{i_0}} a_{i_0} \xleftarrow{l_{i_0+1}} a_{i_0+1} \xrightarrow{l_{i_0+2}} a_{i_0+2} \cdots$$

can be altered, using the fact that a_{i_0+1} (and possibly other vertices) is not a source, to obtain

$$a \xrightarrow{l_1} \cdots \xrightarrow{l_{i_0}} a_{i_0} \xleftarrow{l'_{i_0+1}} a'_{i_0+1} \xrightarrow{l'_{i_0+2}} a_{i_0+2} \cdots ,$$

where l'_{i_0+1} is greater than l_{i_0+1} and is divisible by k. This contradicts the choice of i_0.

If, on the other hand, $i_0 + 1 = j$, the number k divides $l_1 - l_2 + \cdots \pm l_{i_0}$ and, using the fact that $l_1 - l_2 + \cdots \pm l_{i_0} \mp l_{i_0+1} = k$, we infer that k divides l_{i_0+1}, again contradicting the choice of i_0. Thus $i_0 = j$ and we can find an oriented walk $a \xrightarrow{l_1} a_1 \xleftarrow{l_2} a_2 \xrightarrow{l_3} \cdots a_{l_j} = b$ with $l_1 - l_2 + \cdots \pm l_j = k$, where each l_i is divisible by k. This shows that a is connected to b via a pattern of algebraic length one in $\mathbf{G}^{+k}$. $\square$

As a consequence we obtain the following fact.

COROLLARY 5.2. *Let* $\mathbf{G}$ *be a digraph, without sources or sinks, such that* $al(\mathbf{G}) = 1$*. Then* $al(\mathbf{G}^{+k}) = 1$ *for any natural number* k*.*

Proof. Let $a \xrightarrow{\alpha} a$, where α is a pattern of algebraic length one. Then, by following a realization of α k-many times, we obtain $a \xrightarrow{\beta} a$ in $\mathbf{G}$ for a pattern β of algebraic length k. Now the statement follows from the previous lemma. $\square$

Theorem 3.1 is proved in section 7 for strongly connected digraphs first, and therefore we need some preliminary results on such digraphs. The following very simple lemma is needed to prove some of the further corollaries in this section.

LEMMA 5.3. *Let c be a vertex in a strongly connected digraph. Then the greatest common divisor* (*GCD*) *of the lengths of the closed walks in this digraph is equal to the GCD of the lengths of the closed walks containing c.*

Proof. Suppose, for contradiction, that the GCD, say n', of the lengths of the closed walks containing c is bigger than the GCD of the lengths of the closed walks for the entire digraph. Then there exists a walk $d \xrightarrow{l} d$ of length l such that n' does not divide l. On the other hand, since the digraph is strongly connected, $c \xrightarrow{l'} d$ and $d \xrightarrow{l''} c$ for some numbers l', l''. The number n', by definition, divides $l' + l''$ and $l' + l + l''$ and thus divides l, a contradiction. $\square$

Moreover, the following easy proposition holds.

PROPOSITION 5.4. *Let $\mathbf{G}$ be a connected digraph $\mathbf{G}$ and α be a pattern. If $a \xrightarrow{\alpha} a$ for a vertex a in $\mathbf{G}$, then the number $al(\mathbf{G})$ divides $al(\alpha)$.*

Proof. Let $\mathbf{G}$ be a connected digraph and, for some vertex a, $a \xrightarrow{\alpha} a$ via a pattern α. Let b be a vertex in $\mathbf{G}$ such that $b \xrightarrow{\beta} b$ for a pattern β satisfying $al(\beta) = al(\mathbf{G})$. Since $\mathbf{G}$ is connected there is a pattern γ such that $b \xrightarrow{\gamma} a$ and thus $b \xrightarrow{\gamma} a \xrightarrow{\alpha} a \xrightarrow{\gamma'} b$ with $al(\gamma') = -al(\gamma)$. Following appropriate walks, we can obtain an oriented walk, from b to b, of algebraic length $al(\alpha) - k \cdot al(\mathbf{G})$, for any number k. The minimality of $al(\mathbf{G})$ implies that $al(\mathbf{G})$ divides $al(\alpha)$. $\square$

The following lemma is heavily used in the proof of Theorem 3.1 for strongly connected digraphs in section 7.

LEMMA 5.5. *If, for a strongly connected digraph $\mathbf{G} = (V, E)$, the GCD of the lengths of the closed walks in $\mathbf{G}$ is equal to one, then*

$$(\exists m)\ (\forall a, b \in V)\ (\forall n)\ \text{if } n \geq m, \text{ then } a \xrightarrow{n} b.$$

Proof. Fix an arbitrary element $c \in V$. By Lemma 5.3 we find some closed walks containing c such that their lengths $k_1, \dots, k_i$ satisfy $GCD(k_1, \dots, k_i) = 1$. Thus c is contained in a closed walk of length l whenever l is a linear combination of $k_1, \dots, k_i$ with nonnegative integer coefficients. It is easy to see that there is a natural number m' such that, for every $n' \geq m'$, n' can be expressed as such a linear combination; hence c is in a closed walk of length n' for each such n'. Now it suffices to set $m = m' + 2|V|$ since, for arbitrary vertices $a, b \in V$, there are walks of length at most $|V|$ from a to c and from c to b. $\square$

The following easy corollary follows.

COROLLARY 5.6. *For a strongly connected digraph $\mathbf{G}$ with GCD of the lengths of the closed walks equal to one, and for any number n, the digraph $\mathbf{G}^{+n}$ is strongly connected.*

For strongly connected digraphs, the GCD of the lengths of the closed walks and the algebraic length of the digraph coincide.

COROLLARY 5.7. *For a strongly connected digraph, the GCD of the lengths of the closed walks is equal to the algebraic length of the digraph.*

Proof. Let us fix a digraph $\mathbf{G} = (V, E)$ and denote by n the GCD of the lengths of the closed walks in $\mathbf{G}$. Since, by Proposition 5.4, the algebraic length of $\mathbf{G}$ divides the length of every closed walk in $\mathbf{G}$, $al(\mathbf{G})$ divides n.

Conversely, let $a = a_0 \xrightarrow{l_0} b_0 \xleftarrow{k_0} a_1 \xrightarrow{l_1} \cdots \xleftarrow{k_{m-1}} a_m = a$ be a realization of a pattern of algebraic length $al(\mathbf{G})$. Let k_i' be such that $b_i \xleftarrow{k_i} a_{i+1} \xleftarrow{k_i'} b_i$ for all i. Note that n divides $k_i + k_i'$ and $\sum_{i<m} l_i + \sum_{i<m} k_i'$. Thus n divides $\sum_{i<m} l_i - \sum_{i<m} k_i = al(\mathbf{G})$, which shows that $n \leq al(\mathbf{G})$, and the lemma is proved. ∎

Finally, we remark that if α is a pattern of algebraic length one and $\mathbf{G}$ has no sources and no sinks, then $E(\mathbf{G}^\alpha) \supseteq E(\mathbf{G})$. In particular, if $al(\mathbf{G}) = 1$, then $al(\mathbf{G}^\alpha) = 1$.

6. A connection between graphs and algebra. In this section we present basic definitions and results concerning the connection between digraphs and algebras. Let $\mathbf{G} = (V, E)$ be a digraph admitting a weak near unanimity polymorphism $w(x_0, x_1, \ldots, x_{h-1})$. We associate with $\mathbf{G}$ an algebra $\mathbf{A} = (V, w)$ and note that E is a subuniverse of $\mathbf{A}^2$. Note that for any subuniverse of $\mathbf{A}$, say W, we can define the digraph $\mathbf{G}_{|W} = (W, E \cap W \times W)$ (or $(W, E_{|W})$) which admits the weak near unanimity polymorphism $w|_{W^h}$, and the algebra $(W, w|_{W^h})$ is a subalgebra of $\mathbf{A}$. For the remainder of this section we assume that $\mathbf{G}$ and $\mathbf{A}$ are as above.

The first lemma describes the influence of the structure of the digraph on the subuniverses of the algebra.

LEMMA 6.1. *For any subuniverse W of $\mathbf{A}$ the sets W^{+1} and W^{-1} are subuniverses of $\mathbf{A}$.*

Proof. Take any elements $a_0, \ldots, a_{h-1}$ from W^{+1} and choose $b_0, \ldots, b_{h-1} \in W$ such that $b_i \rightarrow a_i$ for all i. Then $w(b_0, \ldots, b_{h-1}) \rightarrow w(a_0, \ldots, a_{h-1})$ showing that $w(a_0, \ldots, a_{h-1}) \in W^{+1}$, and the claim is proved. The proof for W^{-1} is similar. ∎

Since the weak near unanimity operation is idempotent, all the one element subsets of V are subuniverses of $\mathbf{A}$. Using the previous lemma, the following result follows trivially.

COROLLARY 6.2. *For any $a \in V$, any pattern α, and any number n, the sets a^{+n}, a^{-n}, and a^α are subuniverses of $\mathbf{A}$.*

Subuniverses of $\mathbf{A}$ can also be obtained in another way.

LEMMA 6.3. *Let $\mathbf{H}$ be a strong component of $\mathbf{G}$. Assume that the GCD of the lengths of the cycles in $\mathbf{H}$ is equal to one. Then $V(\mathbf{H})$ is a subuniverse of $\mathbf{A}$.*

Proof. Using Lemma 5.5, we find a number m such that there is a walk $b \xrightarrow{m} c$ in $\mathbf{H}$ for all $b, c \in V(\mathbf{H})$. Fix a vertex $a \in V(\mathbf{H})$. There is a walk $a \xrightarrow{m} b$ for all $b \in V(\mathbf{H})$ and a walk $c \xrightarrow{m} a$ for all $c \in V(\mathbf{H})$. Thus, $V(\mathbf{H}) = a^{+m} \cap a^{-m}$ is a subuniverse. ∎

We present a second construction leading to a subuniverse of the algebra.

LEMMA 6.4. *If $\mathbf{H} = (W, F)$ is the largest induced subgraph of $\mathbf{G}$ without sources or sinks, then W is a subuniverse of $\mathbf{A}$.*

Proof. Clearly, the vertices of $\mathbf{H}$ can be described as those having arbitrarily long walks to and from them. Since $\mathbf{G}$ is finite, there exists a natural number k such that

$$W = \{w \mid (\exists v, v' \in V)\ v \xrightarrow{k} w \text{ and } w \xrightarrow{k} v'\}.$$

Thus $W = V^{+k} \cap V^{-k}$, and we are done, since both sets on the right-hand side are subuniverses. ∎

7. Strongly connected digraphs. In this section we present a proof Theorem 3.1 in the case of strongly connected digraphs.

THEOREM 7.1. *If a strongly connected digraph of algebraic length k admits a weak near unanimity polymorphism, then it contains a closed walk of length k* (*and thus retracts to a cycle of length k*).

Using Corollary 5.7, the result can be restated in terms of the GCD of the lengths of closed walks in $\mathbf{G}$, and we will freely use this duality. Theorem 7.1 is a consequence of the following result.

THEOREM 7.2. *If a strongly connected digraph $\mathbf{G}$ of algebraic length one admits a weak near unanimity polymorphism, then it contains a loop.*

We present a proof of Theorem 7.1, assuming Theorem 7.2, and devote the remainder of this section to proving Theorem 7.2.

Proof of Theorem 7.1. Fix an arbitrary vertex c in a strongly connected digraph of algebraic length k. Using Lemma 5.3 and Corollary 5.7, we obtain closed walks containing c with the GCD of their lengths equal to k. Thus, in the path power $\mathbf{G}^{+k}$, the GCD of lengths of closed walks containing c is equal to one. Let $\mathbf{H}$ be the strong component of $\mathbf{G}^{+k}$ containing c. Using Lemma 6.3, we infer that $V(\mathbf{H})$ is a subuniverse of the algebra $(V(\mathbf{G}^{+k}), w)$, and thus $\mathbf{H}$ admits a weak near unanimity polymorphism. The algebraic length of $\mathbf{H}$ (again by Corollary 5.7) is one, and therefore by Theorem 7.2 it follows that there is a loop in $\mathbf{G}^{+k}$. This trivially implies a closed walk of length k in $\mathbf{G}$, and the theorem is proved using the folklore proposition from section 4.4. $\square$

The remaining part of this section is devoted to the proof of Theorem 7.2. We start by choosing a digraph $\mathbf{G} = (V, E)$ to be a minimal (with respect to the number of vertices) counterexample to Theorem 7.2. We fix a weak near unanimity polymorphism $w(x_0, \dots, x_{h-1})$ of this digraph and associate with it the algebra $\mathbf{A} = (V, w)$. The proof will proceed by a number of claims.

CLAIM 7.3. *The digraph $\mathbf{G}$ can be chosen to contain a closed walk of length* 2.

Proof. Using Lemma 5.5, we find a minimal k such that a closed walk of length 2^k is contained in $\mathbf{G}$. Consider the path power $\mathbf{G}^{+2^{k-1}}$. It contains a closed walk of length 2 and admits a weak near unanimity polymorphism. Moreover, since k was chosen to be minimal and $\mathbf{G}$ did not contain a loop, the path power $\mathbf{G}^{+2^{k-1}}$ does not contain a loop either. By Corollary 5.6 the path power is strongly connected, and by Corollary 5.2 it has algebraic length equal to one. Thus, the digraph $\mathbf{G}^{+2^{k-1}}$ is also a counterexample to Theorem 7.2 (with the same number of vertices as $\mathbf{G}$), and therefore we can use it as a substitute for $\mathbf{G}$. $\square$

From this point on we assume that $\mathbf{G}$ contains a closed walk of length 2 (an undirected edge). The next claim allows us to choose and fix an undirected edge with special properties.

CLAIM 7.4. *There are vertices $a, b \in V$ forming an undirected edge in $\mathbf{G}$ and a binary term t of $\mathbf{A}$ such that $a = t(w(\overline{a}, b), w(\overline{b}, a))$.*

Proof. Let $M \subseteq V$ be a minimal (under inclusion) subuniverse of $\mathbf{A}$ containing an undirected edge, and let $a, b \in M$ be vertices in such an edge. Since vertices $w(\overline{a}, b)$, $w(\overline{b}, a) \in M$ form an undirected edge in $\mathbf{G}$, the set $\{w(\overline{a}, b), w(\overline{b}, a)\}$ generates, in the algebraic sense, the set M (by the minimality of M). Since every vertex in a subuniverse is a result of an application of some term function to the generators of the subuniverse, there exists a term t such that $t(w(\overline{a}, b), w(\overline{b}, a)) = a$. $\square$

In the following claims we fix vertices a, b and a term $t(x, y)$ such that $a \to b \to a$ and $a = t(w(\overline{a}, b), w(\overline{b}, a))$ (provided by the previous claim). Note that, by the definition of the operation $w(x_0, \dots, x_{h-1})$, for any numbers $i, j < h$, we obtain $a = t(w^{(i)}(\overline{a}, b), w^{(j)}(\overline{b}, a))$.

Using Lemma 5.5, we find and fix a minimal number n such that $a^{+(n+1)} = V$. We put $W = a^{+n}$ and $F = (W \times W) \cap E$ so that $\mathbf{H} = (W, F)$ is an induced subgraph of the digraph $\mathbf{G}$. Using Corollary 6.2, we infer that W is a subuniverse of $\mathbf{A}$ and thus $\mathbf{H}$ admits a weak near unanimity polymorphism. In the following claims we will show that the algebraic length of some strong component of $\mathbf{H}$ is one, which will contradict the minimality of $\mathbf{G}$.

Claim 7.5. *For any vertex in W there exists a closed walk in $\mathbf{H}$ and a walk (also in $\mathbf{H}$) connecting the closed walk to this vertex.*

Proof. Let d_0 denote an arbitrary vertex of W. Since $a^{+(n+1)} = W^{+1} = V$ there is $d_1 \in W$ such that $d_1 \to d_0$. Similarly, there exists $d_2 \in W$ such that $d_2 \to d_1$. By repeating this procedure, we get both statements of the claim. ∎

The next claim will allow us to fix some more vertices necessary for further construction.

Claim 7.6. *There exist vertices $c, c' \in W$ and a number k such that*

1. $c' \to a$,
2. $c \xrightarrow{k} c$ *in* $\mathbf{H}$, *and*
3. $c \xrightarrow{k-n-1} c'$ *in* $\mathbf{H}$.

Proof. Since $W^{+1} = V$ there exists $c' \in W$ such that $c' \to a$. Let l be the length of a closed walk provided by Claim 7.5 for $c' \in W$. For a sufficiently large multiple k of l there is a walk in $\mathbf{H}$ of length $k - n - 1$ from some vertex of the closed walk to c'; we call this vertex c. This finishes the proof. ∎

From this point on we fix vertices c and c' in W and a number k to satisfy the conditions of the last claim. The following claims focus on uncovering the structure of the strong component containing c in $\mathbf{H}$.

Claim 7.7. *For any $m \leq n$ either $a^{+m} \subseteq a^{+n}$ or $a^{+m} \subseteq b^{+n}$.*

Proof. Since a is in a closed walk of length 2, we obviously have $a^{+n} \supseteq a^{+(n-2)} \supseteq a^{+(n-4)} \cdots$, which proves the claim for even m's. If, on the other hand, m is odd, we have $b^{+n} \supseteq a^{+(n-1)} \supseteq a^{+(n-3)} \cdots$, completing the proof. ∎

The next two claims are of major importance for the proof of Theorem 7.2. They are used to show that the algebraic length of the strong component of $\mathbf{H}$ containing c is one.

Claim 7.8. *For any $m \leq n$ and for any $0 \leq i, j < h$ the following inclusion holds:*

$$t(w^{(i)}(\overline{a^{+n}}, a^{+m}), w^{(j)}(\overline{a^{+m}}, a^{+n})) \subseteq a^{+n}.$$

Proof. Note that $a = t(w^{(i)}(\overline{a}, b), w^{(j)}(\overline{b}, a))$ and therefore, for any choice of arguments of the term reachable by walks of length n from corresponding arguments of $t(w^{(i)}(\overline{a}, b), w^{(j)}(\overline{b}, a))$, the result is reachable by a walk of the same length from a, i.e.,

$$a^{+n} \supseteq t(w^{(i)}(\overline{a^{+n}}, b^{+n}), w^{(j)}(\overline{b^{+n}}, a^{+n})).$$

By the same token, using $a = t(w^{(i)}(\overline{a}, a), w^{(j)}(\overline{a}, a))$ provided by the idempotency of the terms, we obtain

$$a^{+n} \supseteq t(w^{(i)}(\overline{a^{+n}}, a^{+n}), w^{(j)}(\overline{a^{+n}}, a^{+n})).$$

Now the claim follows directly from Claim 7.7. ∎

The following technical claim will allow us to find walks in the strong component of $\mathbf{H}$ containing c.

CLAIM 7.9. *The following implication holds in* $\mathbf{H}$ (*i.e., all the walks and vertices lie inside* $\mathbf{H}$). *For any numbers* $0 \le i, j < h$ *and all* $e, e', f \in W$ *and* $\overrightarrow{d}, \overrightarrow{d'}, \overrightarrow{g} \in \overline{W}$,

$$\text{if} \quad \begin{array}{c} e \\ \downarrow^{k} \\ e' \end{array} \quad \text{and} \quad \begin{array}{c} d_l \\ \downarrow^{k} \\ d'_l \end{array} \quad \text{for all } l, \quad \text{then} \quad \begin{array}{c} t(w^{(i)}(\overrightarrow{d}, c), w^{(j)}(\overline{c}, e)) \\ \downarrow^{k} \\ t(w^{(i)}(\overrightarrow{d'}, f), w^{(j)}(\overrightarrow{g}, e')). \end{array}$$

Proof. Note that, by looking at the tuples of vertices pointwise, we can find the following walks in $\mathbf{G}$:

$$\begin{array}{cccc} \overrightarrow{d} & c & \overline{c} & e \\ \Big| & \downarrow^{k-n-1} & \downarrow^{k-n-1} & \Big| \\ \Big| & c' & \overline{c'} & \Big| \\ \Big|^{k} & \downarrow & \downarrow & \Big|^{k} \\ \Big| & a & \overline{a} & \Big| \\ \downarrow & \downarrow^{n} & \downarrow^{n} & \downarrow \\ \overrightarrow{d'} & f & \overrightarrow{g} & e' \end{array}$$

where the walks from c to c' are provided by Claim 7.6 and lie entirely in $\mathbf{H}$. Applying the appropriate term to the consecutive vertices of the walks (rows in the diagram above), we obtain a walk of length k connecting $t(w^{(i)}(\overrightarrow{d}, c), w^{(j)}(\overline{c}, e))$ to $t(w^{(i)}(\overrightarrow{d'}, f), w^{(j)}(\overrightarrow{g}, e'))$. It remains to prove that all the vertices of this walk are in W. The first $k-n-1$ vertices of the walks are in W, since W is a subuniverse and they are results of an application of a term to vertices of the subuniverse. For $m \ge 0$, the $(k-n+m)$th vertex of the walk is a member of $t(w^{(i)}(\overline{a^{+n}}, a^{+m}), w^{(j)}(\overline{a^{+m}}, a^{+n}))$ and thus in W by Claim 7.8. ∎

We now construct a closed walk in $\mathbf{H}$, that contains c, of length coprime to k.

CLAIM 7.10. *There exists a closed walk* $c \xrightarrow{(h+1)k-1} c$ *in digraph* $\mathbf{H}$.

Proof. In the proof of this claim we use only vertices and walks that lie inside $\mathbf{H}$. Fix $d \in W$ (provided by Claim 7.6) such that $c \to d \xrightarrow{k-1} c$ in $\mathbf{H}$. By repeatedly applying Claim 7.9 we obtain

$$\begin{array}{ccc} t(w(c,\dots,c,c,c), w(c,c\dots,c)) & & \\ \downarrow^{k} & & \\ t(w(c,\dots,c,c,d), w(d,c,\dots,c)) & = & t(w^{(1)}(c,\dots,c,d,c), w^{(1)}(c,\dots,c,d)) \\ & & \downarrow^{k} \\ t(w^{(2)}(c,\dots,d,d,c), w^{(1)}(c,\dots,c,d)) & = & t(w^{(1)}(c,\dots,c,d,d), w^{(1)}(c,\dots,c,d)) \\ \vdots & & \\ & = & t(w^{(h-1)}(d,\dots,d,d,c), w^{(1)}(c,\dots,c,d)) \\ & & \downarrow^{k} \\ & & t(w^{(h-1)}(d,\dots,d,d,d), w^{(1)}(d,\dots,d,d)) \end{array}$$

and since the algebra is idempotent, the starting point of this walk is c and the ending point is d. Thus $c \xrightarrow{hk} d$ (for h the arity of the operation $w(x_0,\dots,x_{h-1})$), which immediately gives us the claim. ∎

By Claims 7.6 and 7.10, the strong component of $\mathbf{H}$ containing c has GCD of the lengths of its closed walks equal to one, and thus, by Lemma 6.3, its vertex set forms a subuniverse of the algebra $\mathbf{A}$. As a digraph it admits a weak near unanimity polymorphism. By Corollary 5.7 it has algebraic length one, and (as an induced subgraph of $\mathbf{G}$) it has no loops. Since $\mathbf{H}$ was chosen to be strictly smaller than $\mathbf{G}$ we obtain a contradiction with the minimality of $\mathbf{G}$, and the proof of Theorem 7.2 is complete.

8. The general case. In this section we prove Theorem 3.1 in its full generality. Nevertheless the majority of this section is devoted to the proof of the following result.

THEOREM 8.1. *If a digraph with no sources or sinks has algebraic length one and admits a weak near unanimity polymorphism, then it contains a loop.*

Using the above result, we prove the core theorem of the paper, Theorem 3.1.

Proof of Theorem 3.1. Let $\mathbf{G}$ be a digraph with no sources or sinks which admits a weak near unanimity polymorphism. Let n be the algebraic length of some component of $\mathbf{G}$. The path power $\mathbf{G}^{+n}$ admits a weak near unanimity polymorphism, has no sources or sinks, and, by Lemma 5.1, has algebraic length equal to one. Thus, Theorem 8.1 applied to $\mathbf{G}^{+n}$ provides a loop in the path power and therefore a closed walk of length n in $\mathbf{G}$.

Let n be minimal, under divisibility, in the set of algebraic lengths of components of $\mathbf{G}$. Since the algebraic length of a component divides (by Proposition 5.4) the length of any closed walk in it, every closed walk of length n (for such a minimal n) forms a subgraph which is a cycle. Moreover, by the same reasoning, cycles obtained for two different minimal n's cannot belong to the same component. Thus each component of $\mathbf{G}$ maps homomorphically to an n-cycle (for any minimal n dividing the algebraic length of this component), and it is not difficult to see that these homomorphisms can be chosen so that their union is a retraction. This proves the theorem. ∎

Therefore the only missing piece of the proof to the conjecture of Bang-Jensen and Hell is Theorem 8.1. We prove this result by way of contradiction. Suppose that $\mathbf{G} = (V, E)$ is a minimal (with respect to the number of vertices) counterexample to Theorem 8.1, and let $\mathbf{A} = (V, w(x_0, \dots, x_{h-1}))$ be the algebra associated with $\mathbf{G}$, in the sense of section 6, for some weak near unanimity polymorphism $w(x_0, \dots, x_{h-1})$.

The first part of the proof is dedicated to finding a particular counterexample satisfying more restrictive conditions than $\mathbf{G}$. To do so we need to define a special family of digraphs called *tambourines*. The n-tambourine is the digraph $(\{d_0, \dots, d_{n-1}, u_0, \dots, u_{n-1}\}, F_n)$ such that

$$F_n = \bigcup_i \{(d_i, d_{i+1}), (d_i, u_i), (d_i, u_{i+1}), (u_i, u_{i+1})\},$$

where the addition on the indices is computed modulo n. The 12-tambourine can be found in Figure 1. We begin the proof of the theorem with the following claim.

CLAIM 8.2. *We can choose a digraph* $\mathbf{G}$ *and a number* n *such that*

1. *the* n*-tambourine maps homomorphically to* $\mathbf{G}$*,*
2. *every vertex of* $\mathbf{G}$ *is in a closed walk of length* n*, and*
3. $\mathbf{G}^{+(mn+1)} = \mathbf{G}$ *for any number* m*.*

To prove this claim, we begin with an easy subclaim and work towards replacing $\mathbf{G}$ with a particular path power of $\mathbf{G}$ which satisfies the additional conditions. Note that, for any pattern α, the path power G^{α} admits $w(x_0, \dots, x_{h-1})$ as a polymorphism and has no sources or sinks. If such a path power has algebraic length one and does not contain a loop, then it can be taken as a substitute for $\mathbf{G}$.

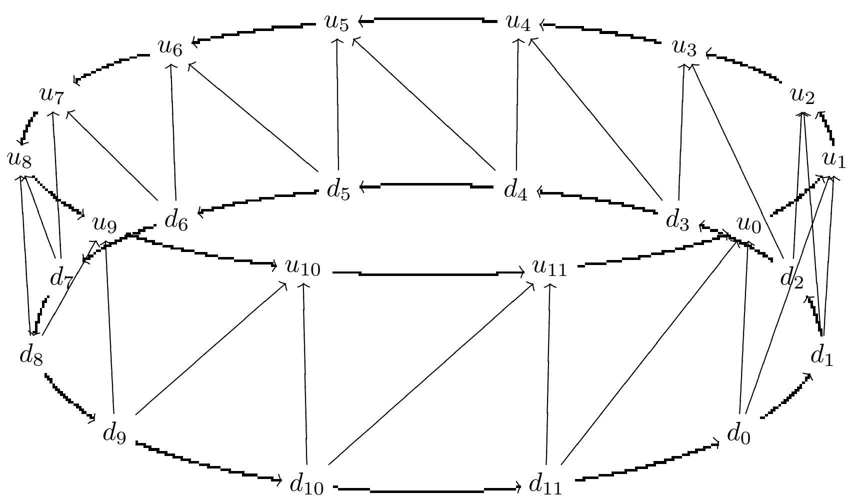


FIG. 1. *The 12-tambourine.*

SUBCLAIM 8.2.1. *The digraph* $\mathbf{G}$ *contains vertices* d *and* u *such that* $d \xrightarrow{|V|,|V|+1} u$.

Proof. Let α be the pattern

$$\underbrace{\to \ \cdots \to}_{|V|+1} \underbrace{\leftarrow \ \cdots \leftarrow}_{|V|}.$$

Using the fact that $al(\alpha) = 1$ and that $\mathbf{G}$ has no sources or sinks, it follows that $E(\mathbf{G}) \subseteq E(\mathbf{G}^\alpha)$. Moreover, let a, b be vertices in $\mathbf{G}$ such that b is contained in a closed walk and $a \xrightarrow{k} b$ for some k. Then $a \xrightarrow{k'} b$ for some $k' \leq |V|$, and choosing b' (from the closed walk containing b) such that $b' \xrightarrow{k'+1} b$, we obtain

$$b' \xrightarrow{k'+1} b \xrightarrow{(|V|+1)-(k'+1)} c \xleftarrow{(|V|+1)-(k'+1)} b \xleftarrow{k'} a \quad \text{ for some } c.$$

Thus $b' \xrightarrow{\alpha} a$, and this implies that every component of $\mathbf{G}$ becomes a strong component of $\mathbf{G}^\alpha$.

Let $\mathbf{H} = (W, F)$ be a component of $\mathbf{G}$ with a closed walk realizing a pattern of algebraic length one. Then, for an appropriate F', containing F, the digraph $\mathbf{H}' = (W, F')$ is a strong component of $\mathbf{G}^\alpha$. The digraph $\mathbf{H}'$ contains $\mathbf{H}$ as a subgraph, and therefore its algebraic length is one. The path power $\mathbf{G}^\alpha$ admits $w(x_0, \ldots, x_{h-1})$ as a polymorphism, and thus, by Lemma 6.3, the digraph $\mathbf{H}'$ admits an appropriate restriction of $w(x_0, \ldots, x_{h-1})$. Theorem 7.2 provides a loop in $\mathbf{H}'$, which in turn implies the existence of vertices $d, u \in W$ such that $d \xrightarrow{|V|,|V|+1} u$ in $\mathbf{G}$. ∎

Proof of Claim 8.2. We fix $n = |V|!$ and argue that, for some k, the path power $\mathbf{G}_k = \mathbf{G}^{+(kn+1)}$ satisfies the assertions of the claim and therefore can be taken as a substitute for $\mathbf{G}$. Note that, for any number k, the digraph $\mathbf{G}_k$ admits $w(x_0, \ldots, x_{h-1})$ as a polymorphism, has no sources or sinks, and, by Corollary 5.2, has algebraic length one.

We first prove that, for all k, the digraph $\mathbf{G}_k$ does not contain a loop. If $\mathbf{G}_k$ does contain a loop, then there exists a closed walk of length $kn+1$ in some strong component of $\mathbf{G}$. In the same strong component in $\mathbf{G}$ there exists a closed walk of length smaller than n and thus coprime to $kn+1$; therefore the GCD of the lengths of closed walks in this strong component is one, and, using Corollary 5.7, Lemma 6.3, and Theorem 7.2, we obtain a loop in this strong component and therefore also in $\mathbf{G}$, a contradiction. Thus, to prove the claim, it remains to verify the additional required properties.

We now show that, for the fixed number n, the n-tambourine maps homomorphically to $\mathbf{G}_k$ for $k \geq 4$. Let d, u be vertices of $\mathbf{G}$ provided by Subclaim 8.2.1. Since $\mathbf{G}$ has no sources or sinks, we can find vertices d', u', each contained in a closed walk, such that d' is connected by a walk to d and u is connected by a walk to u'. By following the closed walks containing d' and u' multiple times, we get d'_0, u'_0, each contained in a closed walk, such that $d'_0 \xrightarrow{3n,3n+1} u'_0$. Moreover, again following the closed walks multiple times, we obtain

$$d'_0 \xrightarrow{n} d'_0 \xrightarrow{3n,3n+1} u'_0 \xrightarrow{n} u'_0.$$

Let d'_i denote the ith vertex of the closed walk $d'_0 \xrightarrow{n} d'_0$ and, similarly, u'_i the ith vertex of the closed walk $u'_0 \xrightarrow{n} u'_0$. Then, for any number $k \geq 4$ and any $i < n$, we have $d'_i \xrightarrow{kn+1} u'_i$ and $d'_i \xrightarrow{kn+1} u'_{(i+1) \bmod n}$. On the other hand, $d'_i \xrightarrow{kn+1} d'_{(i+1) \bmod n}$ and $u'_i \xrightarrow{kn+1} u'_{(i+1) \bmod n}$. Thus, for any $k \geq 4$, the map $d_i \mapsto d'_i$, $u_i \mapsto u'_i$ is a homomorphism from the n-tambourine in the path power $\mathbf{G}_k$.

To prove the second assertion of the claim we need to show that if $k \geq 4$, then any vertex of $\mathbf{G}_k$ is in a closed walk of length n. We fix such a number k and let $W \subset V$ be the subuniverse of $\mathbf{A}$ generated by $\{d'_0, \ldots, d'_{n-1}, u'_0, \ldots, u'_{n-1}\}$. Let $\mathbf{G}'_k$ be the subgraph induced by $\mathbf{G}_k$ on W. The digraph $\mathbf{G}'_k$ obviously admits a restriction of $w(x_0, \ldots, x_{h-1})$ and (since the n-tambourine maps homomorphically to it) has algebraic length one. Choose an arbitrary $a \in W$. Then, by the definition of W, we have a term $t(x_0, \ldots, x_{n-1}, y_0, \ldots, y_{n-1})$ such that $a = t(d'_0, \ldots, d'_{n-1}, u'_0, \ldots, u'_{n-1})$. Therefore,

$$\left.\begin{array}{c} t(d'_0, \ldots, d'_{n-2}, d'_{n-1}, u'_0, \ldots, u'_{n-2}, u'_{n-1}) \\ \downarrow \\ t(d'_1, \ldots, d'_{n-1}, d'_0, u'_1, \ldots, u'_{n-1}, u'_0) \\ \vdots \\ t(d'_{n-1}, \ldots, d'_{n-3}, d'_{n-2}, u'_{n-1}, \ldots, u'_{n-3}, u'_{n-2}) \\ \downarrow \\ t(d'_0, \ldots, d'_{n-2}, d'_{n-1}, u'_0, \ldots, u'_{n-2}, u'_{n-1}) \end{array}\right\} n,$$

and thus a is in a closed walk of length n. This proves that $\mathbf{G}'_k$ has no sources and no sinks, and since it cannot be a counterexample smaller than $\mathbf{G}$, we infer that $W = V$. Therefore the second assertion holds for all the digraphs $\mathbf{G}_k$ with $k \geq 4$.

In the digraph $\mathbf{G}_4$ every vertex is in a closed walk of length n, and therefore $E({\mathbf{G}_4}^{+(nm+1)}) \subseteq E({\mathbf{G}_4}^{+(n(m+1)+1)})$ for any number m. Thus, there is a number l such that for any $m \geq l$ we have ${\mathbf{G}_4}^{+(nm+1)} = {\mathbf{G}_4}^{+(nl+1)}$. Take

$$\mathbf{G}' = {\mathbf{G}_4}^{+(nl+1)} = \mathbf{G}^{+(4n+1)(nl+1)} = \mathbf{G}_{(4nl+l+4)n+1}$$

and note that, according to the previous paragraphs of this proof, such a digraph satisfies all but the last assertion of the claim. Let m be arbitrary. Then $(\mathbf{G}')^{+(mn+1)} = \mathbf{G}_4{}^{+((mnl+l+m)n+1)} = \mathbf{G}_4{}^{+(nl+1)} = \mathbf{G}'$, and thus $\mathbf{G}'$ can be taken to substitute for $\mathbf{G}$ and the claim is proved. $\square$

From this point on we substitute $\mathbf{G}$ with a digraph provided by the previous claim and fix it together with the number n. For ease of notation we denote the number modulo n using brackets (e.g., $[n+1] = 1$). We already know that the n-tambourine maps homomorphically to $\mathbf{G}$, but we must choose such a homomorphism carefully.

Claim 8.3. *The n-tambourine can be mapped homomorphically to $\mathbf{G}$ in such a way that, for some term $t(x_0, \ldots, x_{n-1})$ of algebra $\mathbf{A}$,*

$$d'_i = t^{(i)}(w(\overline{d'_0}, d'_1), w(\overline{d'_1}, d'_2), \ldots, w(\overline{d'_{n-1}}, d'_0)) \quad \text{for all } i < n,$$

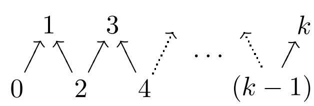
where d'_i is the image of d_i.

Proof. Let $d_i \mapsto d'_i, u_i \mapsto u'_i$ be a homomorphism from the n-tambourine to $\mathbf{G}$. Then, for any i, we have

$$\begin{tikzcd}
w(\overline{u'_i}, u'_{[i+1]}) \arrow[r] & w(\overline{u'_{[i+1]}}, u'_{[i+2]}) \arrow[r] & \cdots \\
w(\overline{d'_i}, d'_{[i+1]}) \arrow[r] \arrow[u] \arrow[ur] & w(\overline{d'_{[i+1]}}, d'_{[i+2]}) \arrow[r] \arrow[u] \arrow[ur, phantom, "\cdots"] & \cdots
\end{tikzcd}$$

and thus $d_i \mapsto w(\overline{d'_i}, d'_{[i+1]}), u_i \mapsto w(\overline{u'_i}, u'_{[i+1]})$ is also a homomorphism from the n-tambourine to $\mathbf{G}$. By repeating this procedure, we obtain an infinite sequence of homomorphisms from the n-tambourine to $\mathbf{G}$, and thus some homomorphism has to appear twice in this sequence. This homomorphism satisfies the claim, since the term $t(x_0, \ldots, x_{n-1})$ can be easily obtained as a composition of the polymorphism $w(x_0, \ldots, x_{h-1})$ used in the construction of the sequence. $\square$

In the remaining part of the proof we fix vertices $d'_0, \ldots, d'_{n-1}, u'_0, \ldots, u'_{n-1}$ provided by the previous claim and a term $t(x_0, \ldots, x_{n-1})$ associated with them. Let φ_k be the pattern $0 \xrightarrow{\varphi_k} k$



with exactly k edges. (The last edge of the pattern is pointing forward for odd k, as in the above picture, and backward for even k.)

Claim 8.4. *The neighborhood $(d'_0)^{\varphi_n}$ contains all vertices of $\mathbf{G}$.*

Proof. Note that, in the n-tambourine, we have

$$(d_0)^{\varphi_n} = \{d_0, \ldots, d_{n-1}, u_0, \ldots, u_{n-1}\},$$

and thus in the digraph $\mathbf{G}$ we have

$$(d'_0)^{\varphi_n} \supseteq \{d'_0, \ldots, d'_{n-1}, u'_0, \ldots, u'_{n-1}\}.$$

Let $\mathbf{G}'$ denote the subgraph of $\mathbf{G}$ induced on the set $(d'_0)^{\varphi_n}$. Then, by Corollary 6.2, $\mathbf{G}'$ admits a restriction of $w(x_0, \ldots, x_{h-1})$ as a polymorphism and has algebraic length one. Further restricting the digraph $\mathbf{G}'$, denote the largest induced subgraph of $\mathbf{G}'$ without sources or sinks by $\mathbf{G}''$. By Lemma 6.4 $\mathbf{G}''$ admits a weak near unanimity polymorphism. Moreover, the vertices $\{d'_0, \dots, d'_{n-1}, u'_0, \dots, u'_{n-1}\}$ are among the vertices of $\mathbf{G}''$. Thus $\mathbf{G}''$ is a counterexample to Theorem 8.1 and therefore has to be equal to $\mathbf{G}$. This proves the claim. $\square$

We choose (and fix) k to be a minimal number such that $(d'_0)^{\varphi_{k+1}} = V$. Define $W_i = (d'_i)^{\varphi_k}$, for each $i < n$. We set

$$W = \bigcap_{i<n} W_i,$$

and since W is an intersection of subuniverses of $\mathbf{A}$, by Corollary 6.2, it is itself a subuniverse of $\mathbf{A}$. We denote by $\mathbf{H}$ the subgraph of $\mathbf{G}$ induced by W and prove that $\mathbf{H}$ is a counterexample to Theorem 8.1, contradicting the minimality of $\mathbf{G}$.

The most involved part of the proof deals with constructing a closed realization of a pattern with the algebraic length one in $\mathbf{H}$. Two following claims introduce tools for "projecting" certain walks from $\mathbf{G}$ to $\mathbf{H}$.

CLAIM 8.5. *There exists a term $s(x_0, \dots, x_{p-1})$ of algebra $\mathbf{A}$ such that for every coordinate $q < p$ there exists i such that*

$$s^{(q)}(W_l, W, \dots, W) \subseteq W_{[i-l]} \cap W_{[i-l+1]} \quad \text{for any } l < n.$$

Proof. Let $p = hn$ and let $s(x_0, \dots, x_{p-1})$ be defined by

$$t\left(w(x_0, \dots, x_{h-1}), w(x_h, \dots, x_{2h-1}), \dots, w(x_{(n-1)h}, \dots, x_{hn-1})\right).$$

For all $q < p$, let i be maximal such that $q = ih + q''$ for some nonnegative q''. Then, for all $l < n$

$$\begin{aligned}
s^{(q)}(W_l, \overline{W}) &\subseteq t^{(i)}\left(w^{(q'')}(W_l, \overline{W}), w(\overline{W}), \dots, w(\overline{W})\right) \\
&\subseteq t^{(i)}\left(w^{(q'')}(\overline{W_l}, W_{[l+1]}), w(\overline{W_{[l+1]}}, W_{[l+2]}), \dots, w(\overline{W_{[l+n-1]}}, W_l)\right) \\
&= t^{([i-l])}\left(w(\overline{W_0}, W_1), \dots, w^{(q'')}(\overline{W_l}, W_{[l+1]}), \dots, w(\overline{W_{n-1}}, W_0)\right) \\
&\subseteq W_{[i-l]},
\end{aligned}$$

where the last inclusion follows from Claim 8.3 and the fact that

$$\begin{aligned}
d'_{[i-l]} &= t^{([i-l])}(w(\overline{d'_0}, d'_1), \dots, w(\overline{d'_l}, d'_{[l+1]}), \dots, w(\overline{d'_{n-1}}, d'_0)) \\
&= t^{([i-l])}(w(\overline{d'_0}, d'_1), \dots, w^{(q'')}(\overline{d'_l}, d'_{[l+1]}), \dots, w(\overline{d'_{n-1}}, d'_0)).
\end{aligned}$$

Similar reasoning shows that

$$\begin{aligned}
s^{(q)}(W_l, \overline{W}) &\subseteq t^{(i)}\left(w^{(q'')}(W_l, \overline{W}), w(\overline{W}), \dots, w(\overline{W})\right) \\
&\subseteq t^{(i)}\left(w^{(q'')}(W_l, \overline{W_{[l-1]}}), w(W_{[l+1]}, \overline{W_l}), \dots, w(W_{[l+n-1]}, \overline{W_{[l+n-2]}})\right) \\
&= t^{[i-l+1]}\left(w(W_1, \overline{W_0}), \dots, w^{(q'')}(W_l, \overline{W_{[l-1]}}), \dots, w(W_0, \overline{W_{n-1}})\right) \\
&\subseteq W_{[i-l+1]},
\end{aligned}$$

and the proof is finished. $\square$

Further, using the term constructed in the last claim, we can construct a term satisfying stronger conditions.

Claim 8.6. *There exists a term $r(x_0, \ldots, x_{m-1})$ of algebra $\mathbf{A}$ such that for every coordinate $q < m$*

$$r^{(q)}\left(\bigcup_{l<n} W_l, W, \ldots, W\right) \subseteq W.$$

Proof. Let $s(x_0, \ldots, x_{p-1})$ be the p-ary term provided by the previous claim. Note that the term

$$s_2(x_0, x_1, \ldots, x_{p^2-1}) = s(s(x_0, \ldots, x_{p-1}), \ldots, s(x_{p^2-p}, \ldots, x_{p^2-1}))$$

has the property that for every coordinate $q < p^2 - 1$ there exists an i such that

$$s_2^{(q)}(W_l, \overline{W}) \subseteq W_{[i-l]} \cap W_{[i-l+1]} \cap W_{[i-l+2]}.$$

To prove a more general statement we recursively define a sequence of terms

- $s_1(x_0, \ldots, x_{p-1}) = s(x_0, \ldots, x_{p-1})$ and
- $s_{j+1}(x_0, \ldots, x_{p^j-1}) = s(s_j(x_0, \ldots, x_{p^{j-1}-1}), \ldots, s_j(x_{(p-1)p^{j-1}}, \ldots, x_{p^j-1}))$.

We claim that for any j, any $q < p^j$, and any $l < n$ there is an i such that

$$s_j^{(q)}\big(W_l, W, \ldots, W\big) \subseteq W_{[i-l]} \cap \cdots \cap W_{[i-l+j]}.$$

We prove this fact by induction on j. The first step of the induction holds via Claim 8.5. Assume that the fact holds for j; then for any l (setting q' to be the result of integer division of q by p^{j-1}, and q'' the remainder of this division) there exist i and i' such that

$$\begin{aligned} s_{j+1}^{(q)}(W_l, \overline{W}) &\subseteq s^{(q')}\big(s_j^{(q'')}(W_l, \overline{W}), s_j(\overline{W}), \ldots, s_j(\overline{W})\big) \\ &\subseteq s^{(q')}\big(W_{[i-l]} \cap \cdots \cap W_{[i-l+j]}, \overline{W}\big) \\ &\subseteq W_{[i'+i-l]} \cap \cdots \cap W_{[i'+i-l+(j+1)]}, \end{aligned}$$

where the second inclusion follows from the induction step and the last one from Claim 8.5. Setting $r(x_0, \ldots, x_{m-1})$ equal to $s_{n-1}(x_0, \ldots, x_{p^n-1})$ proves the claim. ∎

From this point on we fix a term $r(x_0, \ldots, x_{m-1})$ (of arity m) provided by the previous claim. To prove additional properties of the set W (e.g., the fact that it is not empty) we require the following easy claim.

Claim 8.7. *Let α be a pattern, and let $a_0 \to a_1$ and $b_0 \to b_1$ be edges that belong to closed walks. If $a_0 \xrightarrow{\alpha} b_0$, then $a_1 \xrightarrow{\alpha} b_1$.*

Proof. We prove the claim by induction with respect to the number of edges in α. Let the vertices a_0, a_1, b_0, b_1 be as in the statement of the claim. Assume that $a_0 \to b_0$. If i is the length of the closed walk containing the edge $a_0 \to a_1$, then, following this walk almost n times, $a_1 \xrightarrow{in-1} a_0 \to b_0 \to b_1$ and, by Claim 8.2, $a_1 \to b_1$. The same reasoning can be applied to the case of $a_0 \leftarrow b_0$, and the first step of the induction is proved.

For a pattern α consisting of more than one edge we can assume, without loss of generality, that the last edge is going forward. Then $a_0 \xrightarrow{\alpha'} a_0' \to b_0$ for some vertex a_0' (where α' is the pattern obtained by removing the last edge of α). By Claim 8.2, it follows that a_0' is in a closed walk of length n, and therefore $a_0' \to a_1' \xrightarrow{n-1} a_0'$ for some a_1'. By the induction hypothesis, $a_1 \xrightarrow{\alpha'} a_1'$ and, by the first step of the induction, $a_1' \to b_1$, which proves the claim. ∎

We recall the definition of the top and bottom components of the graph from subsection 4.3 and prove some basic properties of W.

CLAIM 8.8. *The digraph* $\mathbf{H}$ *has no sources and no sinks and*

1. *if* k *is even, then every bottom component is contained in* W*, and*
2. *if* k *is odd, then every top component is contained in* W*.*

Proof. First we show that, for any vertices a, b such that $a \xrightarrow{i} b \xrightarrow{j} a$ in $\mathbf{G}$ for some i, j,

$$\text{if } a \in W_l, \quad \text{then } b \in W_{[l+i]}.$$

To see this note that if $d'_l \xrightarrow{\varphi_k} a$ and $a \to b \xrightarrow{j} a$, then, using Claim 8.7 and the edge $d'_l \to d'_{[l+1]}$, we infer that $d'_{[l+1]} \xrightarrow{\varphi_k} b$. The same procedure repeated i-many times provides the result for arbitrary i.

Let $a \in W$ be arbitrary and b be such that $a \xrightarrow{i} b \xrightarrow{j} a$ for some numbers i, j. Since $a \in W$ it follows, using the note above, that $b \in \bigcap_{l<n} W_{[l+i]} = W$, and this implies that W is a union of strong components. Since, by Claim 8.2, every vertex in $\mathbf{G}$ belongs to a closed walk of length n, the digraph $\mathbf{H}$ has no sources or sinks.

Let k be even and let a be a member of a bottom component. Since every vertex of the graph, by Claim 8.2, belongs to a closed walk, there exists b in the bottom component containing a such that $a \to b$. Since $(d'_0)^{\varphi_{k+1}} = V$, we have $d'_0 \xrightarrow{\varphi_{k-1}} c \leftarrow a' \to b$ for some a' and c. The vertex a is in a bottom component, and therefore a' must be a member of the same bottom component. This implies that $a' \to b \xrightarrow{i} a'$, for some i, and following the closed walk containing b almost n times, $a \to b \xrightarrow{n(i+1)-1} a' \to c$. Thus, by Claim 8.2, we have $a \to c$ and $a \in W_0$. Therefore every bottom component is contained in W_0. To see that every a from a bottom component is contained in an arbitrary W_l we find a b satisfying $a \xrightarrow{l} b \xrightarrow{i} a$ for some i and apply the note from the beginning of the proof of the claim. The claim is proved for even k's, and the same reasoning provides a proof for odd k and top components. ∎

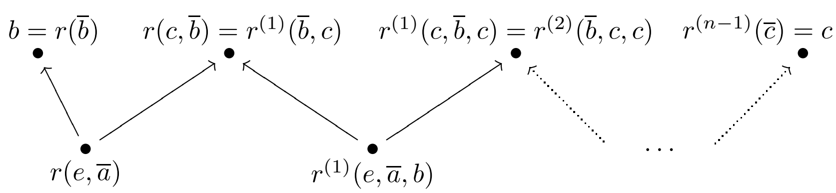
Now we are ready to prove the final claim of this section.

CLAIM 8.9. *The algebraic length of* $\mathbf{H}$ *is one.*

Proof. In the case where k is odd, we want to find $a, b, c \in W$ and $e \in W_0$ such that

$$\begin{tikzcd} a \arrow[r] & b \arrow[r] & c \\ & e \arrow[u] \arrow[ur] & \end{tikzcd}$$

To find such vertices we set $e = d'_1$ and find, using Claim 8.8, $b \in W$ from a top component such that $u'_{[2]} \xrightarrow{in-1} b$ for some i. There exist a and c in the same component (and thus in W by Claim 8.8) such that $a \to b \to c$. Since $d'_1 \xrightarrow{1,2} u'_{[2]}$, we have $e \xrightarrow{in+1} b$ and $e \xrightarrow{in+1} c$, and therefore, by Claim 8.2, the vertices a, b, c, and e satisfy the required properties. Then, using the term $r(x_0, \dots, x_{m-1})$, we produce the following oriented walk:



By Claim 8.6, all the vertices of this walk belong to W. Thus we have constructed an oriented walk in $\mathbf{H}$ realizing a pattern of algebraic length zero connecting b to c. Since $b \to c$ we immediately obtain that the algebraic length of $\mathbf{H}$ is one.

In the case where k is even, we similarly find $a, b, c \in W$ and $e \in W_0$ (using u_1' for e) such that

$$\begin{tikzcd} & & e & \\ a & b \arrow[l] \arrow[ur] & & c \arrow[ll] \arrow[ul] \end{tikzcd}$$

The construction of a closed oriented walk realizing a pattern of algebraic length one is the same as it is for odd k, with the exception that the direction of the edges is reversed. $\square$

Thus $\mathbf{H}$ is a digraph without sources or sinks (by Claim 8.8), admitting a weak near unanimity polymorphism and, by the last claim, having algebraic length equal to one. Since, by the definition of W, the number of vertices in $\mathbf{H}$ is strictly smaller than the number of vertices in $\mathbf{G}$, we obtain a contradiction with the minimality of $\mathbf{G}$, and Theorem 8.1 is proved.

REFERENCES

[All94] J. Allen, *Natural Language Understanding*, 2nd ed., Benjamin Cummings, San Francisco, CA, 1994.

[ANP05] D. Achlioptas, A. Naor, and Y. Peres, *Rigorous location of phase transitions in hard optimization problems*, Nature, 435 (2005), pp. 759–764.

[BIM+06] J. Berman, P. Idziak, P. Marković, R. McKenzie, M. Valeriote, and R. Willard, *Varieties with few subalgebras of powers*, Trans. Amer. Math. Soc., to appear.

[BJH90] J. Bang-Jensen and P. Hell, *The effect of two cycles on the complexity of colourings by directed graphs*, Discrete Appl. Math., 26 (1990), pp. 1–23.

[BJHM95] J. Bang-Jensen, P. Hell, and G. MacGillivray, *Hereditarily hard H-colouring problems*, Discrete Math., 138 (1995), pp. 75–92.

[BJK05] A. Bulatov, P. Jeavons, and A. Krokhin, *Classifying the complexity of constraints using finite algebras*, SIAM J. Comput., 34 (2005), pp. 720–742.

[BKJ00] A. A. Bulatov, A. A. Krokhin, and P. Jeavons, *Constraint satisfaction problems and finite algebras*, in Automata, Languages and Programming (Geneva, 2000), Lecture Notes in Comput. Sci. 1853, Springer, Berlin, 2000, pp. 272–282.

[BS81] S. Burris and H. P. Sankappanavar, *A Course in Universal Algebra*, Graduate Texts in Math. 78, Springer-Verlag, New York, 1981.

[Bul05] A. A. Bulatov, *H-coloring dichotomy revisited*, Theoret. Comput. Sci., 349 (2005), pp. 31–39.

[Bul06] A. A. Bulatov, *A dichotomy theorem for constraint satisfaction problems on a* 3-*element set*, J. ACM, 53 (2006), pp. 66–120.

[Dal05] V. Dalmau, *A new tractable class of constraint satisfaction problems*, Ann. Math. Artif. Intell., 44 (2005), pp. 61–85.

[Dal06] V. Dalmau, *Generalized majority-minority operations are tractable*, Log. Methods Comput. Sci., 2 (2006), pp. 1–15.

[DD96] R. Dechter and A. Dechter, *Structure-driven algorithms for truth maintenance*, Artificial Intelligence, 82 (1996), pp. 1–20.

[Fed01] T. Feder, *Classification of homomorphisms to oriented cycles and of k-partite satisfiability*, SIAM J. Discrete Math., 14 (2001), pp. 471–480.

[FV99] T. Feder and M. Y. Vardi, *The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory*, SIAM J. Comput., 28 (1998), pp. 57–104.

[GWW92] W. Gutjahr, E. Welzl, and G. Woeginger, *Polynomial graph-colorings*, Discrete Appl. Math., 35 (1992), pp. 29–45.

[HM88] D. Hobby and R. McKenzie, *The Structure of Finite Algebras*, Contemp. Math. 76, American Mathematical Society, Providence, RI, 1988.

[HN90] P. Hell and J. Nešetřil, *On the complexity of H-coloring*, J. Combin. Theory Ser. B, 48 (1990), pp. 92–110.

[HN04] P. Hell and J. Nešetřil, *Graphs and Homomorphisms*, Oxford Lecture Series in Math. Appl. 28, Oxford University Press, Oxford, 2004.

[HNZ96a] P. Hell, J. Nešetřil, and X. Zhu, *Complexity of tree homomorphisms*, Discrete Appl. Math., 70 (1996), pp. 23–36.

[HNZ96b] P. Hell, J. Nešetřil, and X. Zhu, *Duality and polynomial testing of tree homomorphisms*, Trans. Amer. Math. Soc., 348 (1996), pp. 1281–1297.

[HNZ96c] P. Hell, J. Nešetřil, and X. Zhu, *Duality of graph homomorphisms*, in Combinatorics, Paul Erdős is Eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud. 2, János Bolyai Mathematical Society, Budapest, 1996, pp. 271–282.

[HP64] Z. Hedrlín and A. Pultr, *Relations (graphs) with given finitely generated semigroups*, Monatsh. Math., 68 (1964), pp. 213–217.

[HZZ93] P. Hell, H. S. Zhou, and X. Zhu, *Homomorphisms to oriented cycles*, Combinatorica, 13 (1993), pp. 421–433.

[JCG97] P. Jeavons, D. Cohen, and M. Gyssens, *Closure properties of constraints*, J. ACM, 44 (1997), pp. 527–548.

[KMRT+07] F. Krząkała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, *Gibbs states and the set of solutions of random constraint satisfaction problems*, Proc. Natl. Acad. Sci. USA, 104 (2007), pp. 10318–10323.

[Kun] G. Kun, *Constraints, MMSNP and Expander Relational Structures*, manuscript, 2007; http://arxiv.org/abs/0706.1701.

[KV07] E. Kiss and M. Valeriote, *On tractability and congruence distributivity*, Log. Methods Comput. Sci., 3 (2007), 2:6 (electronic).

[LALW98] D. Lesaint, N. Azarmi, R. Laithwaite, and P. Walker, *Engineering dynamic scheduler for Work Manager*, BT Technology J., 16 (1998), pp. 16–29.

[Lev73] L. A. Levin, *Universal enumeration problems*, Problemy Peredači Informacii, 9 (1973), pp. 115–116.

[LZ03] B. Larose and L. Zádori, *The complexity of the extendability problem for finite posets*, SIAM J. Discrete Math., 17 (2003), pp. 114–121.

[LZ06] B. Larose and L. Zádori, *Taylor terms, constraint satisfaction and the complexity of polynomial equations over finite algebras*, Internat. J. Algebra Comput., 16 (2006), pp. 563–581.

[Mac77] A. Mackworth, *Consistency in networks of relations*, Artificial Intelligence, 8 (1977), pp. 99–118.

[Mac91] G. MacGillivray, *On the complexity of colouring by vertex-transitive and arc-transitive digraphs*, SIAM J. Discrete Math., 4 (1991), pp. 397–408.

[MM07] M. Maróti and R. McKenzie, *Existence theorems for weakly symmetric operations*, Algebra Universalis, to appear.

[MMT87] R. N. McKenzie, G. F. McNulty, and W. F. Taylor, *Algebras, Lattices, Varieties, Vol.* I, The Wadsworth & Brooks/Cole Mathematics Series, Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA, 1987.

[Mon74] U. Montanari, *Networks of constraints: Fundamental properties and applications to picture processing*, Information Sci., 7 (1974), pp. 95–132.

[Nad] B. A. Nadel, *Constraint satisfaction in Prolog: Complexity and theory-based heuristics*, Inform. Sci., 83 (1995), pp. 113–131.

[NL] B. A. Nadel and J. Lin, *Automobile transmission design as a constraint satisfaction problem: Modeling the kinematic level*, Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AI EDAM), 53 (1991), pp. 137–171.

[Sab61] G. Sabidussi, *Graph derivatives*, Math. Z., 76 (1961), pp. 385–401.

[SV98] E. Schwalb and L. Vila, *Temporal constraints: A survey*, Constraints, 3 (1998), pp. 129–149.

[Tay77] W. Taylor, *Varieties obeying homotopy laws*, Canad. J. Math., 29 (1977), pp. 498–527.

[Var00] M. Vardi, *Constraint satisfaction and database theory: A tutorial*, in Proceedings of 19th ACM Symposium on Principles of Database Systems (PODS'00), ACM, New York, 2000, pp. 76–85.

[Zhu95] X. Zhu, *A polynomial algorithm for homomorphisms to oriented cycles*, J. Algorithms, 19 (1995), pp. 333–345.

Appendix B – Bounded width

Constraint Satisfaction Problems Solvable by Local Consistency Methods

LIBOR BARTO, McMaster University and Charles University in Prague
MARCIN KOZIK, Jagiellonian University

We prove that constraint satisfaction problems without the ability to count are solvable by the local consistency checking algorithm. This settles three (equivalent) conjectures: Feder–Vardi [SICOMP'98], Bulatov [LICS'04] and Larose–Zádori [AU'07].

Categories and Subject Descriptors: F.1.3 [**Computation by Abstract Devices**]: Complexity Measures and Classes—*Relations among complexity classes*; F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*Computations on discrete structures*; F.4.1 [**Mathematical Logic and Formal Languages**]: Mathematical Logic—*Logic and constraint programming*

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Constraint satisfaction problem, local consistency checking

ACM Reference Format:
Barto, L. and Kozik, M. 2014. Constraint Satisfaction Problems solvable by local consistency methods. *J. ACM* 61, 1, Article 3 (January 2014), 19 pages.
DOI:http://dx.doi.org/10.1145/2556646

1. INTRODUCTION

The Constraint Satisfaction Problem (CSP) asks if, given variables and constraints, there is an assignment such that all the constraints are satisfied. Each constraint is presented by a list of admissible evaluations for a few variables. Deciding if such an assignment exists is NP-complete.

In a seminal paper [Feder and Vardi 1998], a concept of nonuniform CSPs was introduced. A nonuniform CSPs restricts the admissible instances by limiting the allowed constraints – the language of the CSP. Thus, each constraint language defines a computational problem of some complexity. The most important question in this area is the Dichotomy Conjecture of Feder and Vardi [1998] postulating that each constraint language defines a problem which is NP-complete or solvable in polynomial time.

Two major algorithmic principles solve CSPs for wide classes of constraint languages. One of them, generalizations of Gaussian elimination [Berman et al. 2009; Bulatov 2006a; Bulatov and Dalmau 2006; Idziak et al. 2007], is beyond the scope of

A preliminary version of this article appeared in *Proceedings of the 50th Symposium on Foundations of Computer Science (FOCS 09)* [Barto and Kozik 2009b].
L. Barto was supported by the Grant Agency of the Czech Republic, grant 201/09/P223 and by the Ministry of Education of the Czech Republic, grant MSM 0021620839. M. Kozik was partially supported by National Science Center grant DEC-2011/01/B/ST6/01006.
Authors' addresses: L. Barto, Department of Algebra, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, 186 75 Praha 8, Czech Republic; M. Kozik (corresponding author), Theoretical Computer Science Department, Faculty of Mathematics and Computer Science, Jagiellonian University, ul. Prof. St. Łojasiewicza 6, 30-348 Kraków, Poland; email: marcin.kozik@uj.edu.pl.

DOI:http://dx.doi.org/10.1145/2556646

this article. The other algorithm, called the local consistency checking algorithm, determines whether an instance has a consistent set of local solutions. For a wide variety of constraint languages the existence of such a set implies a global solution. These constraint languages are said to have bounded width [Feder and Vardi 1998].

The CSPs of bounded width appear naturally in different approaches to the subject. The class can be equivalently described as the CSPs having complements recognizable by a Datalog program with a goal predicate, using pebble games, or as problems with a bounded tree-width duality.

An obstruction to having bounded width was recognized in Feder and Vardi [1998] and called the ability to count. A constraint language has this property if it can, in some sense, simulate relations encoding linear equations over a finite field. Feder and Vardi [1998] proved that a bounded width constraint language cannot have the ability to count and conjectured that lack of this ability implies bounded width. A different conjecture characterizing bounded width was stated by Bulatov [2004b] by means of his very successful [Bulatov 2003, 2004a, 2006c] technique based on studying the local structure of the constraint language. Finally, Larose and Zádori [2007] used Tame Congruence Theory [Hobby and McKenzie 1988] to propose a characterization using the types in the variety associated with the constraint language. These three conjectures were shown to be equivalent in Larose et al. [2009] and Bulatov and Valeriote [2008].

The progress toward proving the bounded width conjecture(s) [Barto and Kozik 2009a; Bulatov 2006b; Carvalho et al. 2009; Kiss and Valeriote 2007] culminated in its confirmation announced by the authors [Barto and Kozik 2009b] and independently by Bulatov [2009]. This article contains a full proof of the result announced in Barto and Kozik [2009b] presented in a self-contained way (modulo basic algebraic and computational notions). The result is stated in a slightly more general way than in Barto and Kozik [2009b]. This change is motivated by recent developments in robust approximability of CSPs [Barto and Kozik 2012].[1] The presented proof is also more elementary than Barto and Kozik [2009b] in the sense that we do not require the algebraic results from Maróti and McKenzie [2008]. A generalization of our result in a different direction appears in Barto [2013].

The material presented in this article is divided in the following way. The next section contains preliminary information on CSPs, universal algebra and the connection between constraint languages and algebras. In Section 3, we introduce the local consistency checking algorithm and use it to define CSPs of bounded width. Section 4 states the three conjectures. Sections 5 and 6 contain reductions of the conjecture of Larose and Zádori to the case of special instances with binary constraints only. In Section 7, we introduce tools and concepts of universal algebra necessary for the proof. In Section 8, we use these tools to prove the conjecture.

2. PRELIMINARIES

The preliminaries are split into three parts.

2.1. Constraint Satisfaction Problems

The following definition is standard for nonuniform Constraint Satisfaction Problems.

Definition 2.1. An *instance of the* CSP is a triple $\mathcal{I} = (V, D, \mathcal{C})$ with V a finite set of *variables*, D a finite *domain*, and $\mathcal{C}$ a finite list of *constraints*, where each constraint is

[1] The characterization of CSPs admitting robust approximation, given in Barto and Kozik [2012], requires a result stronger than the one provided in Barto and Kozik [2009b].

a pair $C = (S, R)$ with S a tuple of variables of length k, called the *scope* of C, and R a k-ary relation on D (i.e., a subset of D^k), called the *constraint relation* of C.

An *assignment* for $\mathcal{I}$ is a mapping $F : V \to D$. We say that F *satisfies* a constraint $C = (S, R)$ if $F(S) \in R$ (where F is applied component-wise). An assignment which satisfies all the constraints of the instance is a *solution*.

A finite set of relations $\mathbb{D}$ over a common domain D is called a *constraint language* (the arities of the relations form the *signature* of the language). An *instance of* CSP($\mathbb{D}$) is an instance of the CSP such that all the constraint relations are from $\mathbb{D}$. The *decision problem* for CSP($\mathbb{D}$) asks whether an input instance $\mathcal{I}$ of CSP($\mathbb{D}$) has a solution.

The definition of the CSP does not specify how the constraint relations are given on the input. Note, however, that for a nonuniform problem CSP($\mathbb{D}$), where $\mathbb{D}$ is finite, standard representations lead to log-space equivalent problems. We can, for instance, represent relations by listing all their tuples in every constraint (this is how relations are usually represented when the constraint language is not finite), or we can just use names of the relations.

2.2. Universal Algebra

The following paragraphs cover most of the algebraic concepts which are used in this paper, for a more exhaustive introduction refer to Burris and Sankappanavar [1981] and Bergman [2011].

An *algebra* denoted in boldface, for example $\mathbf{A}$, consists of a set A (the *universe* of $\mathbf{A}$) and operations (sometimes called the *basic operations*) – functions from finite powers of A to A. The symbols and arities of the operations of an algebra form the *signature* of this algebra, for example, algebras $\mathbf{A}$ and $\mathbf{B}$ in the same signature containing an n-ary symbol t have basic operations $t^{\mathbf{A}} : A^n \to A$ and $t^{\mathbf{B}} : B^n \to B$ respectively.[2]

A *subuniverse* of $\mathbf{A}$ is a nonempty set B, contained in A, and such that every operation of $\mathbf{A}$ evaluated on arguments from B produces a result in B. An algebra $\mathbf{B}$ is a *subalgebra* of $\mathbf{A}$ (denoted by $\mathbf{B} \leq \mathbf{A}$) if B is a subuniverse of $\mathbf{A}$ and the operations of $\mathbf{B}$ are the operations of $\mathbf{A}$ restricted to B.

The *product* of algebras $\mathbf{A}$ and $\mathbf{B}$ (of the same signature), denoted by $\mathbf{A} \times \mathbf{B}$, is the algebra with universe $A \times B$ and operations computed coordinatewise. A product of more than two, or an infinite number of algebras is defined analogically. A *power* of $\mathbf{A}$ is a product of copies of $\mathbf{A}$ and a *subpower* of $\mathbf{A}$ is a subalgebra of a power of $\mathbf{A}$. A subalgebra $\mathbf{B}$ of a product $\mathbf{A}_1 \times \cdots \times \mathbf{A}_n$ is called *subdirect* if the projection of B on each coordinate is full (i.e., the projection on the ith coordinate is equal to A_i), in such a case we write $\mathbf{B} \leq_{sd} \mathbf{A}_1 \times \cdots \times \mathbf{A}_n$.

A function f from A to B is a *homomorphism* from $\mathbf{A}$ to $\mathbf{B}$ (the signatures of $\mathbf{A}$ and $\mathbf{B}$ are identical) if, for every operation t with arity n, we have $f(t^{\mathbf{A}}(a_1, \ldots, a_n)) = t^{\mathbf{B}}(f(a_1), \ldots, f(a_n))$, for every choice of $a_1, \ldots, a_n \in A$.

An equivalence relation α is a *congruence* of $\mathbf{A}$ if every operations of $\mathbf{A}$, computed on two coordinatewise α-related tuples, produces a pair in α. The congruences of an algebra $\mathbf{A}$ form a lattice: $\alpha \wedge \beta$ is the largest congruence contained in both α and β, and $\alpha \vee \beta$ is the smallest congruence containing both. An algebra with just two (or one) congruences (i.e., the identity congruence and the full congruence) is called *simple*. If α is a congruence of $\mathbf{A}$ one can form the *quotient* algebra $\mathbf{A}/\alpha$: the universe of $\mathbf{A}/\alpha$ is A/α and the operations are derived from the operations on $\mathbf{A}$ by taking representatives of the congruence classes (since α is a congruence the operations are well defined). If $\mathbf{B} \leq \mathbf{A}_1 \times \cdots \times \mathbf{A}_n$, then the kernel of projection to $\mathbf{A}_i$ is denoted by π_i and is a

[2]We will skip the superscript, writing t instead $t^{\mathbf{A}}$, whenever the algebra is clear from the context.

congruences of $\mathbf{B}$. If, on the other hand, α is a congruence of such $\mathbf{B}$, then the projection of α to ith coordinate is the smallest congruence of $\mathbf{A}_i$ containing all (a_i, b_i) such that $((a_1, \ldots, a_n), (b_1, \ldots, b_n)) \in \alpha$.

A *term* t is a syntactical description of composition of operations in a given signature. For a given algebra $\mathbf{A}$ (in this signature) a *term operation* $t^{\mathbf{A}}$ is an operation obtained by composition of basic operation according to t. A class of algebras is a *variety* if it is closed under taking products, subalgebras and homomorphic images of algebras in the class (quotients of algebras in a variety are also inside the variety). The smallest variety containing an algebra $\mathbf{A}$ is called the variety generated by $\mathbf{A}$ and denoted by $\mathcal{V}(\mathbf{A})$. By Birkhoff's theorem [Birkhoff 1935], each variety $\mathcal{V}$ is determined by a set of pairs of terms $s \approx t$ (called *identities*) in the following way: an algebra $\mathbf{A}$ is in $\mathcal{V}$ if and only if $t^{\mathbf{A}} = s^{\mathbf{A}}$ for each pair of terms from the set.

2.3. Polymorphisms

Much of the recent progress on the complexity of the decision problem for CSP was achieved by the algebraic approach. The notion linking relations and operations is at the center of these developments.

Definition 2.2. An l-ary operation f on D is a *polymorphism* of a k-ary relation R (or R is *compatible* with f), if

$$(f(a_1^1, \ldots, a_1^l), f(a_2^1, \ldots, a_2^l), \ldots, f(a_k^1, \ldots, a_k^l)) \in R$$

whenever $(a_1^1, \ldots, a_k^1)$, $(a_1^2, \ldots, a_k^2)$, $\ldots$, $(a_1^l, \ldots, a_k^l) \in R$.

An operation on D is a *polymorphism* of a constraint language $\mathbb{D}$ (with domain D) if it is a polymorphism of every relation in $\mathbb{D}$.

The set of all polymorphisms of a constraint language $\mathbb{D}$ will be denoted by $\mathrm{Pol}(\mathbb{D})$. Among the earliest results of the algebraic approach to the CSP is a theorem [Jeavons et al. 1997] stating that adding to $\mathbb{D}$ relations compatible with all operations of $\mathrm{Pol}(\mathbb{D})$ produces a CSP computationally equivalent to $\mathrm{CSP}(\mathbb{D})$. Earlier results [Bodnarčuk et al. 1969; Geiger 1968] show that these are exactly the relations definable from $\mathbb{D}$ by pp-formulas – formulas using existential quantification, conjunction and equality. Moreover, adding such relations does not alter the set of polymorphisms of the constraint language.

With each constraint language $\mathbb{D}$, an algebra $\mathbf{D}$ is associated. The algebra $\mathbf{D}$ has universe D and the set of operations $\mathrm{Pol}(\mathbb{D})$. The relations compatible with all the operations of $\mathrm{Pol}(\mathbb{D})$ (from the previous paragraph) are, using algebraic terms, the subpowers of $\mathbf{D}$ (i.e., subalgebras of powers of $\mathbf{D}$).

Further developments in the algebraic approach [Bulatov et al. 2000, 2005; Larose and Tesson 2009] showed that, given $\mathbb{D}$, for any algebra $\mathbf{E}$ in the variety generated by $\mathbf{D}$, any constraint language with domain E and with subpowers of $\mathbf{E}$ as relations defines a CSP reducible (in log-space) to $\mathrm{CSP}(\mathbb{D})$. Therefore, the complexity of the decision problem for $\mathrm{CSP}(\mathbb{D})$ depends only on the variety generated by $\mathbf{D}$, that is, by the result of Birkhoff, on the identities that hold in $\mathbf{D}$.

A constraint language $\mathbb{D}$ is a *core*, if all its unary polymorphisms are bijections. It is clear that, for any constraint language $\mathbb{D}$, there is a core constraint language $\mathbb{D}'$ (the *core of* $\mathbb{D}$) such that $\mathrm{CSP}(\mathbb{D}) = \mathrm{CSP}(\mathbb{D}')$. Further, if $\mathbb{D}'$ is a core one can construct $\mathbb{D}''$ by adding, for every $d \in D'$, the unary constraint relation $\{d\}$ (this relation allows to fix an evaluation of a variable to d). The constraint languages $\mathbb{D}'$ and $\mathbb{D}''$ are log-space equivalent [Bulatov et al. 2005]. Any $t \in \mathrm{Pol}(\mathbb{D}'')$ has the property that $t(d, \ldots, d) = d$ for all $d \in D$ which means that $\mathbf{D}$ satisfies the identity $t(x, \ldots, x) \approx x$ for every operation t. Such algebras are called *idempotent*. Summarizing, every constraint language

ALGORITHM 1: The (k,l)-consistency checking algorithm

Input: An instance $\mathcal{I} = (V, D, \mathcal{C})$
1 $\mathcal{F}$ = all functions from at most l-elements subsets of V into D;
2 **for** $f \in \mathcal{F}$ **do**
3   **for** $((x_1,\dots,x_n),R) \in \mathcal{C}$ **do**
4   **if** $x_1,\dots,x_n \in \operatorname{dom} f$ *and* $(f(x_1),\dots,f(x_n)) \notin R$ **then**
5             $\mathcal{F} = \mathcal{F} \setminus \{f\}$;
6   break;
7 **repeat**
8   **for** $f \in \mathcal{F}$ **do**
9   **foreach** W *at most* l*-element subset of* V **do**
10   **if**
11             $(|\operatorname{dom} f| \leq k$, $\operatorname{dom} f \subseteq W$ *and there is no* $g \in \mathcal{F}$ *with* $\operatorname{dom} g = W$ *and* $g_{|\operatorname{dom} f} = f)$ *or*
12             $(W \subseteq \operatorname{dom} f$ *and* $f_{|W} \notin \mathcal{F})$
13   **then**
14                 $\mathcal{F} = \mathcal{F} \setminus \{f\}$;
15   break; // proceed to the next $f \in \mathcal{F}$
16 **until** $\mathcal{F}$ *was not altered*;
17 **if** $\mathcal{F} = \emptyset$ **then return** *NO* **else return** *YES*

has a computationally equivalent constraint language associated with an idempotent algebra. This fact is essential to the algebraic classification of CSPs.

3. PROBLEMS OF BOUNDED WIDTH

In the seminal paper [Feder and Vardi 1998], the class of CSPs of bounded width was introduced. A constraint language $\mathbb{D}$ has, according to Feder and Vardi [1998], bounded width if the complement of CSP($\mathbb{D}$) can be recognized by a Datalog program. The class can be also described as the CSPs recognizable by certain pebble games or having bounded tree-width duality. In this article, the class of CSPs of bounded width is introduced as the CSPs solvable by the local consistency checking algorithm. A more exhaustive overview of this and other related classes of problems is provided in Bulatov et al. [2008].

The local consistency algorithm, presented as Algorithm 1, is parametrized by two natural numbers $k \leq l$. The algorithm starts with the set $\mathcal{F}$ of all partial assignments of variables (of at most l variables) into the domain D. In the first loop (lines 2–6), the assignments which falsify constraints are removed. In the second loop (lines 7–15), an assignment $f \in \mathcal{F}$ is removed if it falsifies one of two conditions:

(1) if $|\operatorname{dom} f| \leq k$, for any set consisting of at most l variables and containing $\operatorname{dom} f$, there is an assignment in $\mathcal{F}$ which extends f to this set (line 11), and
(2) every restriction of f to a subset of its domain is in $\mathcal{F}$ (line 12).

Finally, the algorithm answers *NO* if $\mathcal{F}$ is empty and *YES* otherwise (note that, at this stage, if $\mathcal{F}$ contains no functions with domain W, and $|W| \leq l$, then $\mathcal{F}$ is empty).

Intuitively, the algorithm constructs a set of partial assignments (restricted to at most l variables) which are consistent on small (at most k element) sets of variables. If the instance on the input of the algorithm has a global solution, then none of its restrictions can be removed from $\mathcal{F}$. Thus, if the local consistency checking algorithm outputs *NO* there is no solution. Constraint languages of bounded width are those, for which the *YES* answer of the algorithm is always correct.

Definition 3.1. A constraint language $\mathbb{D}$ has *width* (k,l) if (k,l)-consistency checking correctly decides CSP($\mathbb{D}$). A constraint language $\mathbb{D}$ (or the problem CSP($\mathbb{D}$)) has *bounded width* if it has width (k,l) for some k,l.

Let $\mathbb{D}$ be a constraint language and let $\mathcal{I}$ be an instance of CSP in this language. Let us assume that the (k,l)-consistency checking algorithm answered *YES* and $\mathcal{F}$ is the set of partial assignments of variables at the end of the run. Note that, for a fixed set of variables $\{x_1,\dots,x_n\}$, all the functions in $\mathcal{F}$ with domain $\{x_1,\dots,x_n\}$ can be viewed as a set, say E, of tuples from D^n (by putting $(f(x_1),\dots,f(x_n))$ into E for each such f). It is crucial for the proof, and a part of folklore, that such obtained E is a subpower of $\mathbf{D}$, that is, E is a relation preserved by all[3] the polymorphisms of $\mathbb{D}$.

Finally, if a constraint language $\mathbb{D}$ has bounded width, then so do the languages defined (in the same way as in the fourth paragraph of Section 2.3) on algebras in the variety generated by $\mathbf{D}$ [Larose and Zádori 2007]. Therefore, again by Birkhoff's result, the property of having bounded width is determined by identities in the algebra of polymorphisms.

4. THE THREE CONJECTURES

The conjectures in this section are mostly concerned with constraint languages which are cores. In the last paragraph of Section 2.3 we argued that this restriction does not decrease the generality of the statements.

Feder and Vardi [1998] introduced the notion of *ability to count*. A constraint language has the ability to count if it can simulate the set $\{1,\dots,p\}$ with two relations similar to the relations "sum of x,y and z is equal to 1" and "x is 0" in a cyclic group. The precise definition can be found in Larose et al. [2009]. Constraint languages with the ability to count are on the opposite end of the spectra from those of bounded width (Feder and Vardi [1998] called them languages "in Datalog"). The conjecture of Feder and Vardi in the orginal form states the following.

CONJECTURE 4.1 (CONJECTURE 1 IN FEDER AND VARDI [1998]). *A constraint-satisfaction problem is not in Datalog if and only if the associated core $\mathbb{T}$ can simulate a core $\mathbb{T}'$ consisting of two relations C,Z that give the ability to count. This is equivalent to simulating either $\mathbb{Z}_p$ or one-in-three SAT.*

It is shown in Feder and Vardi [1998] that if a constraint language can simulate one-in-three SAT or $\mathbb{Z}_p$ then the associated CSP does not have bounded width. The conjecture postulates that the reverse implication holds as well.

Given a constraint language $\mathbb{D}$, Bulatov [2004b] constructs a graph $\mathcal{G}(\mathbb{D})$ with colored edges and a vertex set D. Two elements of the domain $a,b \in D$ are connected with a blue edge if the smallest subalgebra of $\mathbf{D}$ containing a and b, say $\mathbf{E}$, has a congruence α and there is an affine operation of $\mathbf{E}/\alpha$ (and no congruence on this algebra produces a quotient with a semilattice or majority operation).

CONJECTURE 4.2 (CONJECTURE 2 IN BULATOV [2004B]). *Let $\mathbb{D}$ be a constraint language which is a core, and let $\mathbb{D}'$ consist of all the relations from $\mathbb{D}$ plus all the single element relations $\{a\}$ for $a \in D$. Then,* CSP($\mathbb{D}$) *has bounded width if and only if*

(*1*) *for every algebra $\mathbf{A} \le \mathbf{D}'$ and congruence α of $\mathbf{A}$ the algebra $((\mathrm{Pol}(\mathbb{D}'))_{|A})/\alpha$ contains an operation which is not essentially unary, and*

[3]The reason is that, when we apply all polymorphisms of $\mathbb{D}$ to a set $\mathcal{F}$ satisfying (1) and (2), we get a set satisfying the same conditions. On the other hand, the (k,l)-consistency checking algorithm produces the largest set $\mathcal{F}$ of partial assignments satisfying the constraints and conditions (1) and (2). Therefore, the set $\mathcal{F}$ is closed under all polymorphisms.

(2) *the graph* $\mathcal{G}(\mathbb{D})$ *has no blue edges,*

Similarly, as in the case of the conjecture of Feder and Vardi, the implication from left to right is proved [Bulatov 2004b].

Larose and Zádori [2007] used Tame Congruence Theory (TCT) to conjecture the classification of problems of bounded width. TCT, developed in Hobby and McKenzie [1988], introduces five types of local behaviors of algebras. TCT then classifies varieties according to types of the behavior present in its finite members. The TCT-types **1** and **2** correspond to essentially unary and affine behavior, respectively.

CONJECTURE 4.3 (CONJECTURE IN LAROSE AND ZÁDORI [2007]). *For a core* $\mathbb{D}$*, the* CSP($\mathbb{D}$) *has bounded width if and only if* **D** *generates a variety omitting TCT-types* **1** *and* **2**.

Also in this case, the authors show that if type **1** or **2** appears in the variety then the CSP in question cannot have bounded width. The reverse implication is the question asked.

The equivalence of the conjectures above has been shown [Bulatov and Valeriote 2008; Larose et al. 2009] which allows us to focus on proving the conjecture of Larose and Zádori. The knowledge of types of Tame Congruence Theory is not required for the proof presented in this article. Our proof uses a condition from Hobby and McKenzie [1988] stated, in Section 7, as Theorem 7.1.

5. REDUCTION TO BINARY RELATIONS

In this section we prove a reduction of Conjecture 4.3 to the case of languages with binary constraints only. The precise form of the binary instances we produce will simplify the remainder of the proof. We say that an instance of CSP is *syntactically simple* if

— every constraint is binary, that is, the scope of each constraint is a pair of *distinct* variables,
— for every pair (x, y) of distinct variables there is at most one constraint with scope (x, y). The corresponding constraint relation is denoted by $R_{x,y}$,[4] and
— if (x, y) is the scope of some constraint, then so is (y, x) and $R_{y,x} = \{(b, a) : (a, b) \in R_{x,y}\}$.

Let $\mathbb{D}$ be a constraint language with the maximum arity of a constraint relation equal to n and such that the algebra **D** generates a variety omitting types **1** and **2**. Let $\mathcal{I}$ be an instance in the language $\mathbb{D}$. We run the $(2\lceil\frac{n}{2}\rceil, 3\lceil\frac{n}{2}\rceil)$-consistency checking algorithm on $\mathcal{I}$. If the answer is *NO*, there is no solution to $\mathcal{I}$. If the answer is *YES*, we produce a new instance $\mathcal{I}'$ in a language $\mathbb{D}'$ such that:

(1) the instance $\mathcal{I}'$ is syntactically simple (with one constraint $((x, y), R'_{x,y})$ for every pair (x, y) of distinct variables),
(2) the $(2, 3)$-consistency checking algorithm on $\mathcal{I}'$ answers *YES* and stops with a set $\mathcal{F}'$ such that $R'_{x,y} = \{(f'(x), f'(y)) : f' \in \mathcal{F}' \text{ and } \operatorname{dom} f' = \{x, y\}\}$,
(3) the algebra $\mathbf{D}'$ generates a variety omitting types **1** and **2**, and
(4) $\mathcal{I}$ has a solution if and only if $\mathcal{I}'$ does.

The construction of $\mathcal{I}'$ was presented in Section 5 of Barto and Kozik [2009a] (although with a different notation). For the sake of completeness, we now give a sketch.

[4]The reduction of Conjecture 4.3 to Theorem 6.6 presented in this section produces an instance with constraints $((x, y), R_{x,y})$ for every pair (x, y) of distinct variables. The statement of Theorem 6.6 in full generality, that is, with some of the constraints missing, is referred, and required, in Barto and Kozik [2012].

Assume that the $(2\lceil \frac{n}{2} \rceil, 3\lceil \frac{n}{2} \rceil)$-consistency checking algorithm on $\mathcal{I}$ returns *YES* and stops with the set $\mathcal{F}$ of partial assignments of variables. The constraint language $\mathbb{D}'$ has domain $D^{\lceil \frac{n}{2} \rceil}$ and the relations in the new, syntactically simple instance $\mathcal{I}'$ are defined as follows:

(1) for every $\lceil \frac{n}{2} \rceil$-tuple of variables in $\mathcal{I}$ we introduce a variable in $\mathcal{I}'$, and
(2) if x is a variable for $(x_1, \ldots, x_{\lceil \frac{n}{2} \rceil})$ and y for $(y_1, \ldots, y_{\lceil \frac{n}{2} \rceil})$, $x \neq y$, we introduce a constraint $((x, y), R'_{x,y})$ where

$$R'_{x,y} = \{((a_1, \ldots, a_{\lceil \frac{n}{2} \rceil}), (b_1, \ldots, b_{\lceil \frac{n}{2} \rceil})) : \exists f \in \mathcal{F}\ f(x_i) = a_i \text{ and } f(y_i) = b_i\}.$$

Let $\mathcal{F}'$ be the set of assignments obtained by the $(2, 3)$-consistency checking for $\mathcal{I}'$. It is clear that, after the first loop (lines 2–6 of Algorithm 1) f' is in $\mathcal{F}'$ if and only if there is $f \in \mathcal{F}$ such that for every variable $x \in \operatorname{dom} f'$ with corresponding tuple $(x_1, \ldots, x_{\lceil \frac{n}{2} \rceil})$ we have $f'(x)_i = f(x_i)$. That is, condition (2) imposed on $\mathcal{F}'$ and $R'_{x,y}$ in $\mathcal{I}'$ holds after the first loop of the algorithm. The second loop of Algorithm 1 does not remove any functions from $\mathcal{F}'$. Therefore, the $(2, 3)$-consistency checking algorithm for $\mathcal{I}'$ answers *YES* and condition (2) required for $\mathcal{I}'$ holds.

The domain $D^{\lceil \frac{n}{2} \rceil}$ is the universe of the algebra $\mathbf{D}^{\lceil \frac{n}{2} \rceil}$ (a power of $\mathbf{D}$) which lies in the variety generated by $\mathbf{D}$. Each relation

$$\{(a_1, \ldots, a_{\lceil \frac{n}{2} \rceil}, b_1, \ldots, b_{\lceil \frac{n}{2} \rceil}) : \exists f \in \mathcal{F}\ f(x_i) = a_i \text{ and } f(y_i) = b_i\}$$

is a subpower of $\mathbf{D}$ (by the discussion in the fourth paragraph of Section 3), and thus all $R'_{x,y}$'s are subpowers of $\mathbf{D}^{\lceil \frac{n}{2} \rceil}$. The variety generated by $\mathbf{D}^{\lceil \frac{n}{2} \rceil}$ is contained in the variety generated by $\mathbf{D}$ and therefore omits types **1** and **2**. The algebra of polymorphisms of $\mathbb{D}'$ contains all the operations of $\mathbf{D}^{\lceil \frac{n}{2} \rceil}$ and thus omits types **1** and **2** as well [Hobby and McKenzie 1988].

6. REDUCTION TO WEAK PRAGUE INSTANCES

From this point on, all instances, without mentioning it, are assumed to be syntactically simple. By considerations of Section 5, this assumption does not decrease the generality of the result.

We define two consistency notions very important for the proof. The first notion is the weaker one.

Definition 6.1. An instance $\mathcal{I} = (V, D, \mathcal{C})$ is called *1-minimal*, if there are sets P_x, $x \in V$, such that the projection of each $R_{x,y}$ on the first coordinate is P_x and on the second P_y.

In order to define a stronger consistency notion, a weak Prague instance, we define patterns and realizations.

Definition 6.2 (*Pattern and Step*). A *step* in an instance $\mathcal{I}$ is a pair of variables which is the scope of a constraint in $\mathcal{I}$. A *pattern from x to y* (in $\mathcal{I}$) is a sequence of variables $p = (x = x_1, x_2, \ldots, x_k = y)$ such that every pair (x_i, x_{i+1}), $1 \leq i \leq k - 1$, is a step.

For a pattern $p = (x_1, \ldots, x_k)$, we define $-p = (x_k, \ldots, x_1)$. If $p = (x_1, \ldots, x_k)$, $q = (y_1, \ldots, y_l)$, $x_k = y_1$, then the sum of p and q is the pattern $p + q = (x_1, x_2, \ldots, x_k = y_1, y_2, \ldots, y_k)$. For a pattern p from x to x and a natural number k, kp denotes the k-fold sum of p with itself.

Note that the addition of patterns is allowed only if the terminal point of the first pattern coincides with the initial point of the second one. Observe also that from the definition of a syntactically simple instance it follows that $-p$ is a pattern whenever p is.

Definition 6.3 (*Realization, Addition*). Let $p = (x = x_1, x_2, \ldots, x_k = y)$ be a pattern from x to y in an instance $\mathcal{I}$. A *realization of* p is a sequence $(a_1, \ldots, a_k) \in D^k$ such that $(a_i, a_{i+1}) \in R_{x_i, x_{i+1}}$ for every $1 \le i \le k-1$.

For a subset $A \subseteq D$, we define $A+p$ as the set of the last elements of those realizations of p whose first element is in A, that is,

$$A + p = \{b \in D : (\exists\, a_1, \ldots, a_{k-1} \in D)\ a_1 \in A \text{ and } (a_1, \ldots, a_{k-1}, b) \text{ is a realization of } p\}.$$

Finally, we define $A - p = A + (-p)$.

The addition of patterns is associative, that is, $(A + p) + q = A + (p + q)$. Also note that, in a 1-minimal instance, we have $A \subseteq A + p - p$ for any $A \subseteq P_x$ and any pattern p from x, in particular, $A + p$ is nonempty when $A \neq \emptyset$. A weak Prague instance is a 1-minimal instance with additional requirements concerning addition of patterns.

Definition 6.4 (*Weak Prague Instance*). An instance $\mathcal{I}$ is a *weak Prague instance* if

(P1) $\mathcal{I}$ is 1-minimal (with sets P_x from Definition 6.1),
(P2) for every $A \subseteq P_x$ and every pattern p from x to x, if $A + p = A$, then $A - p = A$, and
(P3) for any patterns p, q from x to x and every $A \subseteq P_x$, if $A + p + q = A$, then $A + p = A$.

The instance $\mathcal{I}$ is *nontrivial*, if $P_x \neq \emptyset$ for every $x \in V$.

To clarify the definition, let us consider the following digraph $\mathbb{G}$ for a 1-minimal instance: vertices of $\mathbb{G}$ are all the pairs (A, x) with $x \in V$ and $A \subseteq P_x$, and $((A, x), (B, y))$ forms a directed edge iff $A + (x, y) = B$. Condition (P3) means that no strong component of $\mathbb{G}$ contains (A, x) and (A', x) with $A \neq A'$. Condition (P2) is equivalent (albeit this fact requires a reasoning) to the fact that every strong component of $\mathbb{G}$ contains only "undirected" edges (that is, if $((A, x), (B, y))$ is an edge then so is $((B, y), (A, x))$).

LEMMA 6.5. *The instance $\mathcal{I}'$ constructed in Section 5 is a weak Prague instance.*

PROOF. Let $\mathcal{F}'$ be the set of partial assignments of variables after the $(2,3)$-consistency checking algorithm answered *YES* on $\mathcal{I}'$. We put $P'_x = \{f(x) : f \in \mathcal{F}' \text{ and } \operatorname{dom} f = \{x\}\}$.

To see that condition (P1) is satisfied, let x and y be arbitrary variables. If $a \in P'_x$, then there is $f \in \mathcal{F}'$ with domain $\{x\}$ and $f(x) = a$. Then some function f' with $\operatorname{dom} f' = \{x, y\}$ and $f'(x) = a$ belongs to $\mathcal{F}'$ (as otherwise f would be removed by condition on line 11 of Algorithm 1). But this implies that a is in the first projection of $R'_{x,y}$. On the other hand, if a is in the first projection of $R'_{x,y}$ then there is $f \in \mathcal{F}'$ with $f(x) = a$ and thus $f_{|\{x\}} \in \mathcal{F}'$ provides $a \in P_x$ ($f_{|\{x\}} \in \mathcal{F}'$ as otherwise f would be dropped from $\mathcal{F}'$ by condition on line 12 of Algorithm 1). The proof for the second projection is analogical.

To show (P2) and (P3), we first prove that for any (a, b) in $R'_{x,y}$ and for any pattern p from x to y, we have $b \in \{a\} + p$. The proof is by induction on the length of p. If p is a step, then $p = (x, y)$ and the claim is obvious. For the induction step from n to $n+1$, take any pattern $p = (x, \ldots, w, v, y)$ of length $n+1$. By induction hypothesis used for the pattern $(x, \ldots, w, y)$, there is $c \in \{a\} + (x, \ldots, w)$ such that $(c, b) \in R'_{w,y}$. Therefore, there exists $f \in \mathcal{F}'$ such that $\operatorname{dom} f = \{w, y\}$, $f(w) = c$ and $f(y) = b$. By condition on line 11 of Algorithm 1, we can find $f' \in \mathcal{F}'$ which extends f to $\{w, v, y\}$. Then, the element $d = f'(v)$ satisfies $(c, d) \in R'_{w,v}, (d, b) \in R'_{v,y}$ (by condition on line 12) which implies $b \in c + (w, v, y)$, and $b \in \{a\} + p$.

To see (P3), we note that for any $A \subseteq P'_x$ and any pattern p from x to x we have $A \subseteq A + p$. Indeed, take $a \in A$, and let $p = (x, \ldots, y, x)$. As our instance is 1-minimal, there is some $b \in P'_y$ with $(a,b) \in R'_{x,y}$. By the previous paragraph, $b \in \{a\} + (x,\dots,y)$ and therefore $a \in \{a\} + p$ and $A \subseteq A + p$. Now condition (P3) follows: if $A + p + q = A$, then $A \subseteq A + p \subseteq A + p + q = A$.

Finally, for (P2), let $A \subseteq P'_x$ and $A + (x,\dots,y,x) = A$. It suffices to show that $A + (x,\dots,y) = A - (y,x)$ as we can apply the same argument to the sets $A' = A + p'$ for initial segments p' of p (the condition $A' + p'' = A'$ will be satisfied for a cyclic shift p'' of p). The inclusion $A + (x,\dots,y) \subseteq A - (y,x)$ follows from 1-minimality as $A + (x,\dots,y) \subseteq A + (x,\dots,y) + (y,x) - (y,x)$ and $A + (x,\dots,y) + (y,x) = A$. For the reverse inclusion, we take an arbitrary $b \in A - (y,x)$. Then, $(a,b) \in R'_{x,y}$ for some $a \in A$, and, by a paragraph above, $b \in \{a\} + (x,\dots,y)$ as required. □

In order to confirm Conjecture 4.3, it remains to prove the following theorem.

THEOREM 6.6. *Every nontrivial weak Prague instance in a constraint language $\mathbb{D}$, with $\mathbf{D}$ in a variety omitting types $\mathbf{1}$ and $\mathbf{2}$, has a solution.*

7. ALGEBRAIC TOOLS

Our proof relies heavily on the algebraic properties implied by omitting types **1** and **2**. We use the following charaterization [Hobby and McKenzie 1988]:

THEOREM 7.1. *Let $\mathcal{V}$ be a variety generated by a finite algebra. The following are equivalent:*

(1) *$\mathcal{V}$ omits types $\mathbf{1}$ and $\mathbf{2}$;*
(2) *for every $\mathbf{A} \in \mathcal{V}$ and three congruences α, β, γ on $\mathbf{A}$, if $\alpha \wedge \beta = \alpha \wedge \gamma$, then $\alpha \wedge \beta = \alpha \wedge (\beta \vee \gamma)$.*

In this section, we tacitly assume that all the algebras are idempotent, that is, every operation of every algebra satisfies the identity $t(x,\dots,x) \approx x$.

7.1. Absorption

One of the main algebraic concepts behind the proof is the notion of an absorbing subuniverse.

Definition 7.2. We say that B is an *absorbing subuniverse* of an algebra $\mathbf{A}$ (denoted by $B \lhd \mathbf{A}$) if $B \leq \mathbf{A}$ and there exists a term t of $\mathbf{A}$ such that $t(B,B,\dots,B,A,B,B,\dots,B) \subseteq B$ for any position of A.

In varieties omiting type **1**, lack of absorption has interesting consequences for a particular type of subdirect subalgebras.

Definition 7.3. A subdirect subalgebra $\mathbf{R}$ of $\mathbf{A} \times \mathbf{B}$ is called *linked*, if $\pi_1 \vee \pi_2 = 1_{\mathbf{R}}$.

The set $R \subseteq A \times B$ can be viewed as a bipartite graph with partite sets A and B. Then, $\mathbf{R}$ is linked if and only if this graph is connected.

One of the main tools of Barto and Kozik [2010] is extensively used in the proof.

THEOREM 7.4 (ABSORPTION THEOREM). *If $\mathbf{A}$ and $\mathbf{B}$ are algebras in a variety omitting type $\mathbf{1}$, $R \leq_{sd} \mathbf{A} \times \mathbf{B}$ is linked and $R \neq A \times B$, then $\mathbf{A}$ or $\mathbf{B}$ has a proper absorbing subuniverse.*

7.2. Pointed Terms

The second algebraic tool – pointed terms – is studied in more detail in Barto et al. [2013]. Here we provide only the facts necessary for the proof of the main result of the paper.

Definition 7.5. Let $\mathbf{A}$ be an algebra and let $a \in A$. An n-ary term t of $\mathbf{A}$ *points to* a if there exist $a_1, \dots, a_n \in A$ such that

$$t(b_1, \dots, b_n) = a \text{ whenever } b_i \in A \text{ and } |\{i : a_i \neq b_i\}| \leq 1.$$

In the remaining part of this section, we will be working towards a proof of the following lemma.

LEMMA 7.6. *Let $\mathbf{A}$ be a simple algebra with no proper absorbing subuniverse, generating a variety omitting types $\mathbf{1}$ and $\mathbf{2}$. Then, for every $a \in A$, there exists a term of $\mathbf{A}$ which points to a.*

We begin with a basic property of (absorbing) subuniverses of algebras.

LEMMA 7.7. *Let $R \leq_{sd} \mathbf{A} \times \mathbf{B}$ and let C be an (absorbing) subuniverse of $\mathbf{A}$. Then the set $\{d \in B : \exists c \in C\, (c, d) \in R\}$ is an (absorbing) subuniverse of $\mathbf{B}$. The absorption is realized by the same term.*

PROOF. Let C be a subuniverse of $\mathbf{A}$. Put $D = \{d \in B : \exists c \in C\, (c, d) \in R\}$ and let t be a n-ary term. For any $d_1, \dots, d_n \in D$, we can find $c_1, \dots, c_n \in C$ such that $(c_i, d_i) \in R$ for all i. Therefore, $t((c_1, d_1), \dots, (c_n, d_n)) = (t(c_1, \dots, c_n), t(d_1, \dots, d_n)) \in R$ and, since C is a subuniverse of $\mathbf{A}$, we get $t(d_1, \dots, d_n) \in D$ as required – this proves that D is a subuniverse of $\mathbf{B}$.

If C were absorbing $\mathbf{A}$ with an m-ary term r, then for, $d_1, \dots, d_{k-1}, b, d_{k+1}, \dots, d_m$ with $d_i \in D$, we find $c_1, \dots, c_{k-1}, c_{k+1}, \dots, c_m$ as previously stated and use subdirectness of R to get $a \in A$ with $(a, b) \in R$. We proceed as in the previous case and use absorption to conclude that $r(c_1, \dots, c_{k-1}, a, c_{k+1}, \dots, c_m) \in C$. □

The following lemma states that a product of algebras with no proper absorbing subuniverses has no proper absorbing subuniverse.

LEMMA 7.8. *Let $\mathbf{A}_1, \dots, \mathbf{A}_n$ be algebras and let $\mathbf{R} \lneqq \prod_{i=1}^{n} \mathbf{A}_i$. If $\mathbf{R} \lhd \prod_{i=1}^{n} \mathbf{A}_i$, then some $\mathbf{A}_i$ contains a proper absorbing subalgebra.*

PROOF. Suppose, for a contradiction, that the lemma holds for $n-1$ and fails for $\mathbf{R} \lneqq \prod_{i=1}^{n} \mathbf{A}_i$. The projection of $\mathbf{R}$ on the first coordinate is an absorbing subuniverse of $\mathbf{A}_1$, so it has to be equal to $\mathbf{A}_1$. Therefore, there exists $a \in A_1$ such that

$$\emptyset \neq R' = \{(b_2, \dots, b_n) \mid (a, b_2, \dots, b_n) \in R\} \neq \prod_{i=2}^{n} \mathbf{A}_i.$$

Since $\mathbf{A}_1$ is idempotent, R' is a subuniverse of $\prod_{i=2}^{n} \mathbf{A}_i$. Moreover, since $\mathbf{R}$ absorbs the full product with an idempotent term, it is easily seen that $\mathbf{R}'$ absorbs $\prod_{i=2}^{n} \mathbf{A}_i$ with the same term. Using the induction hypothesis for $\mathbf{R}'$, we get a proper absorbing subuniverse of one of the $\mathbf{A}_i$'s for some $i > 1$ which is a contradiction. □

Next, we argue that certain products of algebras (in varieties omitting types $\mathbf{1}$ and $\mathbf{2}$) have no nontrivial subdirect subuniverses.

LEMMA 7.9. *Let $\mathbf{A}_1, \dots, \mathbf{A}_n$ be simple algebras with no proper absorbing subuniverses and in a variety omitting types $\mathbf{1}$ and $\mathbf{2}$. If $\mathbf{R} \leq_{sd} \prod_{i=1}^{n} \mathbf{A}_i$ and $\pi_i \vee \pi_j = 1_{\mathbf{R}}$ for every $i \neq j$, then $\mathbf{R} = \prod_{i=1}^{n} \mathbf{A}_i$.*

PROOF. First, we prove that there exists j such that $\bigwedge_{i \neq j} \pi_i \neq 0_{\mathbf{R}}$, unless $0_{\mathbf{R}} = 1_{\mathbf{R}}$ (which implies that $\mathbf{R}$ has one element and the lemma is trivially true). Suppose otherwise, that is, for every j, we have $\bigwedge_{i \neq j} \pi_i = 0_{\mathbf{R}}$. We prove by induction on $n - |I|$ that $\bigwedge_{i \in I} \pi_i = 0_{\mathbf{R}}$ for every $I \subseteq \{1, \dots, n\}$. By the assumption, it is true for $|I| = n - 1$;

suppose it holds for all I with $|I| = k$ and let J be such that $|J| = k - 1$. Let $l \notin J, m \notin J, l \neq m$. Theorem 7.1 (with $\alpha = \bigwedge_{i \in J} \pi_i, \beta = \pi_l, \gamma = \pi_m$) provides

$$0_{\mathbf{R}} = \bigwedge_{i \in J \cup \{l\}} \pi_i = \bigwedge_{i \in J} \pi_i \wedge (\pi_l \vee \pi_m) = \bigwedge_{i \in J} \pi_i$$

and the induction step is proved. Finally, for $|I| = 1$, we get $\pi_1 = \pi_2 = 0_{\mathbf{R}}$ which contradicts $\pi_1 \vee \pi_2 = 1_{\mathbf{R}}$.

Without loss of generality, we assume that $\bigwedge_{i \neq 1} \pi_i \neq 0_{\mathbf{R}}$. The projection of $\pi_1 \vee \bigwedge_{i \neq 1} \pi_i$ to the first coordinate cannot be $0_{\mathbf{A}_1}$ (since for some $a \neq b$ we have $(a, b) \in \bigwedge_{i \neq 1} \pi_i$, we immediately get that the projections of a and b on the first coordinate are different and related by the projection of $\pi_1 \vee \bigwedge_{i \neq 1} \pi_i$). Since $\mathbf{A}_1$ is simple, $\pi_1 \vee \bigwedge_{i \neq 1} \pi_i = 1_{\mathbf{R}}$.

Suppose, for a contradiction, that the lemma holds for $n - 1$ and fails for $\mathbf{R} \leq_{sd} \prod_{i=1}^{n} \mathbf{A}_i$. By this assumption, the projection of $\mathbf{R}$ to the coordinates $2, \ldots, n$ is equal to $\prod_{i=2}^{n} \mathbf{A}_i$ and $\mathbf{R}$ is a subdirect product of $\mathbf{A}_1$ and $\prod_{i=2}^{n} \mathbf{A}_i$. By the previous paragraph, R is linked and therefore, by the Absorption Theorem, R is either equal to $A_1 \times \prod_{i=1}^{n} A_i$, or $\mathbf{A}_1$ contains a proper absorbing subuniverse – which contradicts the hypotheses, or $\prod_{i=2}^{n} \mathbf{A}_i$ contains a proper absorbing subuniverse – which contradicts Lemma 7.8. □

Finally, we argue that if a simple algebra with no proper absorbing subuniverses generates a variety omitting types **1** and **2**, then it has many term operations. In the statement, we use the notation $Sg_{\mathbf{A}}(B)$ for the subalgebra of $\mathbf{A}$ generated by B, that is, the smallest subalgebra of $\mathbf{A}$ containing B.

LEMMA 7.10. *Let $\mathbf{A}$ be a simple algebra in a variety omitting types* **1** *and* **2***, with no proper absorbing subuniverse. Let n be an arbitrary positive integer, and let $\mathcal{Z} \subseteq A^n$ be such that*

$$\forall a \neq b \in \mathcal{Z}\ \exists i, j\ (a_i = a_j \wedge b_i \neq b_j) \vee (a_i \neq a_j \wedge b_i = b_j)$$

and

$$\forall a \in \mathcal{Z}\ Sg_{\mathbf{A}}(\{a_1, \ldots, a_n\}) = \mathbf{A}.$$

Then every function from $\mathcal{Z}$ to A is a restriction of some n-ary term operation of $\mathbf{A}$.

PROOF. Let $\mathbf{F} \leq \mathbf{A}^{A^n}$ be the free algebra on n-generators (the universe consists of the n-ary term operations of $\mathbf{A}$).

Choose a, b to be two arbitrary tuples from $\mathcal{Z}$ and let $\mathbf{R}$ be the projection of $\mathbf{F}$ to the coordinates a, b. Since $\mathbf{F}$ is generated by projections, we have $(a_i, b_i) \in \mathbf{R}$ for every $i \leq n$. By the second condition on $\mathcal{Z}$, the algebra $\mathbf{R}$ is subdirect in $\mathbf{A} \times \mathbf{A}$. The first condition provides i, j with, say, $a_i = a_j$ and $b_i \neq b_j$ (the other case is similar). Thus, the projection of $\pi_1 \vee \pi_2$ (computed in $\mathbf{R}$) to the second coordinate is not the equality congruence, and, by simplicity of $\mathbf{A}$, it is the full congruence. That means that $\mathbf{R}$ is linked, so $\pi_a \vee \pi_b = 1_{\mathbf{F}}$.

By the previous paragraph, the restriction of $\mathbf{F}$ to $\mathcal{Z}$ satisfies the hypotheses of Lemma 7.9, and therefore is the full relation. In other words, every function from $\mathcal{Z}$ to $\mathbf{A}$ extends to a term operation of $\mathbf{A}$. □

To finish the proof of Lemma 7.6, we take a list $a_1, \ldots, a_n$ of all elements in $\mathbf{A}$. We put $b = (a_1, a_1, a_2, a_2, \ldots, a_n, a_n) \in \mathbf{A}^{2n}$ and

$$\mathcal{Z} = \{c \in \mathbf{A}^{2n} : |\{i : c_i \neq b_i\}| \leq 1\}.$$

It is straightforward to verify that the set $\mathcal{Z}$ satisfies the hypotheses of Lemma 7.10 and therefore, for every $a \in A$, the constant function mapping $\mathcal{Z}$ to a extends to a term operation. The term associated with this operation points to a.

7.3. Additional Properties of Absorbing Subuniverses

We require the following observation.

LEMMA 7.11. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be algebras in a variety omitting types* $\mathbf{1}$ *and* $\mathbf{2}$ *and such that neither* $\mathbf{A}$ *nor* $\mathbf{B}$ *has a proper absorbing subuniverse. Moreover, let* $R \leq_{sd} \mathbf{A} \times \mathbf{B}$ *and let* α *be a maximal congruence of* $\mathbf{A}$*. Then*

(1) *either* $(a,b),(a',b) \in R$ *implies that* $(a,a') \in \alpha$ *for all* $a,a' \in A$*, or*
(2) *for every* $a \in A$ *and* $b \in B$ *there exists* $a' \in A$ *such that* $(a,a') \in \alpha$ *and* $(a',b) \in R$*.*

PROOF. Consider the subdirect product $R' = \{(a/\alpha, b) : (a,b) \in R\} \leq_{sd} \mathbf{A}/\alpha \times \mathbf{B}$. The algebra $\mathbf{A}/\alpha$ has no absorbing subuniverse (since it is easily seen that the preimage of an absorbing subuniverse of $\mathbf{A}/\alpha$ is an absorbing subuniverse of $\mathbf{A}$) and is simple (as α is maximal).

If (1) is not satisfied, then the projection of $\pi_1 \vee \pi_2$ to the first coordinate (i.e., $\mathbf{A}/\alpha$) is not the equality congruence, therefore, since $\mathbf{A}/\alpha$ is simple, $\pi_1 \vee \pi_2 = \mathbf{1}_{\mathbf{R}'}$, so R' is linked. By the Absorption Theorem, $R' = A/\alpha \times B$ which is a restatement of (2). $\square$

8. A PROOF OF THEOREM 6.6

Theorem 6.6 is a generalization of a result in Barto and Kozik [2009b]. The more general version is required for the approximability results which appear in Barto and Kozik [2012].

To prove Theorem 6.6, we work with a weak Prague instance $\mathcal{I}$ in a constraint language $\mathbb{D}$ such that the associated algebra $\mathbf{D}$ generates a variety omitting types $\mathbf{1}$ and $\mathbf{2}$. Since $\mathcal{I}$ is syntactically simple, the only constraints are of the form $((x,y), R_{x,y})$ and each $R_{x,y}$ is a subuniverse of $\mathbf{D}^2$. Similarly, each P_x from the definition of 1-minimality is a subuniverse of $\mathbf{D}$.

Our last preliminary step is to drop, from $\mathbf{D}$, all the nonidempotent operations. The algebra $\mathbf{D}'$ obtained in this way is idempotent and, still, generates a variety omitting types $\mathbf{1}$ and $\mathbf{2}$ (it follows from part (2) of Theorem 9.10 in Hobby and McKenzie [1988]). All $R_{x,y}$'s and P_x's are subuniverses of $(\mathbf{D}')^2$ and $\mathbf{D}'$ respectively. This allows us to substitute, in the remainder of the proof, $\mathbf{D}$ with $\mathbf{D}'$ and work within a variety generated by an idempotent algebra. In particular, all the results of Section 7 can be applied in this variety.

8.1. Pointed Decomposition

Given a weak Prague instance, we proceed by reducing it to a smaller induced subinstance.

Definition 8.1. Let $\mathcal{J}$ be an instance. An instance $\mathcal{J}'$ is an *induced subinstance of* $\mathcal{J}$ *with potatoes* $P'_x \leq \mathbf{D}$*,* $x \in V$, if $R'_{x,y} = R_{x,y} \cap (P'_x \times P'_y)$.

In order to succeed with such a reduction, we first decompose the instance.

Definition 8.2. A *decomposition* of a 1-minimal instance $\mathcal{J}$ consists of induced subinstances $\mathcal{J}^1, \dots, \mathcal{J}^l$ of $\mathcal{J}$ with potatoes $P^i_x \leq \mathbf{P}_x$, relations $R^i_{x,y} = R_{x,y} \cap (P^i_x \times P^i_y)$ $x,y \in V$, $i \leq l$ and a subset X of V such that

(1) if $x \notin X$, then $P^i_x = P_x$ for all $i \leq l$,
(2) if $x \in X$, then

(a) $P_x^i \cap P_x^j = \emptyset$ for all $i \neq j$, and
(b) for any step (x,y), either $P_x^i + (x,y) = P_y^i$ for all $i \leq l$, or $P_x^i + (x,y) = P_y$ for all $i \leq l$.

Note that any decomposition of a 1-minimal $\mathcal{J}$ consists of instances which are 1-minimal as well.

Definition 8.3. We say that the decomposition is *pointed*, if there exists a term t of $\mathbf{D}$ of arity m and indices $k_1, \ldots, k_m$ such that

$$\forall i \leq m \ \forall x \in V \ t(P_x^{k_1}, \ldots, P_x^{k_{i-1}}, P_x, P_x^{k_{i+1}}, \ldots, P_x^{k_m}) \subseteq P_x^1.$$

The decomposition is *proper* if $\mathcal{J}^1$ is a nontrivial instance and $P_x^1 \subsetneq P_x$ for some $x \in V$.

In the following two sections, we prove:

THEOREM 8.4. *Every nontrivial weak Prague instance $\mathcal{J}$ with $|P_x| > 1$ for some $x \in V$ has a proper pointed decomposition.*

In the last section, we show that $\mathcal{J}^1$ is a weak Prague instance. As P_x^1's are assumed to be subuniverses of $\mathbf{D}$, then $R_{x,y}^1$'s are subuniverses of $\mathbf{D}^2$. That means that throwing all the $R_{x,y}^1$'s into $\mathbb{D}$ does not affect polymorphisms of $\mathbb{D}$ and that we reduced $\mathcal{I}$ to a strictly smaller instance still satisfying hypotheses of Theorem 6.6. Continuing this process, we obtain a minimal induced subinstance of $\mathcal{I}$ and it must have $|P_x| = 1$ for all $x \in V$. Any such instance, trivially, has a solution which sends x to the unique element of P_x. This finishes the proof of Theorem 6.6.

8.2. Pointed Decomposition when Absorption Is Present

We prove Theorem 8.4 when some algebra $\mathbf{P}_x$ has a proper absorbing subuniverse A. In this case, we find a decomposition into a single subinstance $\mathcal{J}^1$.

We define a preorder $\sqsubseteq$ on the set of all pairs (C,y), $C \subsetneq P_y$, by $(C,y) \sqsubseteq (D,z)$ if $(C,y) = (D,z)$ or there exists a pattern p from y to z such that $C + p = D$. Among the equivalence classes of this preorder which are greater or equal to the equivalence class containing (A,x), we choose a maximal one and denote it by $\mathcal{M}$. Let X denote the set of all $y \in V$ for which there exists some C such that $(C,y) \in \mathcal{M}$.

CLAIM 1. *The set C is uniquely determined by y.*

PROOF. If $(C,y), (D,y) \in \mathcal{M}$, then by the fact that (C,y) and (D,y) are in the same equivalence class of $\sqsubseteq$ we get $C = D$, or $C + p = D$ and $D + q = C$ for some patterns p and q from y to y, and, by (P3), $C = D$ again. □

For $y \in X$ we define P_y^1 as the unique set with $(P_y^1, y) \in \mathcal{M}$. For $y \notin X$, we put $P_y^1 = P_y$. Note that, as $\mathcal{J}$ is 1-minimal, we have $P_y^1 \neq \emptyset$ for all $y \in V$. Conditions (1) and (2a) from Definition 8.2 hold by the construction; it remains to verify condition (2b) and find a term witnessing that the decomposition is pointed.

To show condition (2b) of Definition 8.2, let $y \in X$ be arbitrary and let (y,z) be any step. If $(P_y^1 + (y,z), z) \in \mathcal{M}$, then $P_y^1 + (y,z) = P_z^1$ as required. On the other hand, if $(P_y^1 + (y,z), z) \notin \mathcal{M}$, then this pair is outside of the preorder; therefore, $P_y^1 + (y,z) = P_z$.

It remains to show that the induced subinstance $\mathcal{J}^1$ of $\mathcal{J}$ with potatoes P_y^1 for $y \in V$ is a pointed decomposition of $\mathcal{J}$.

Note that Lemma 7.7 implies that, for an absorbing subuniverse B of P_y and a step (y,v), the set $B+(y,v)$ is an absorbing subuniverse of P_v. By induction, we get that P_y^1 (which is equal to $A+p$ for some p from x to y) is an absorbing subuniverse of P_y and the absorbing term is the same as for A and $\mathbf{P}_x$. This term satisfies the condition in Definition 8.2 with $k_i=1$ for all i.

8.3. Pointed Decomposition when Absorption Is Missing

Now we prove Theorem 8.4 in the case that none of the algebras $\mathbf{P}_x, x\in V$ has a proper absorbing subuniverse.

LEMMA 8.5. *Let α be a maximal congruence on some $\mathbf{P}_x$. Then, for every step (x,y)*

(case1) either $\big(a/\alpha+(x,y)\big)\cap\big(b/\alpha+(x,y)\big)\neq\emptyset$ implies $(a,b)\in\alpha$ for all $a,b\in P_x$,
(case2) or $a/\alpha+(x,y)=P_y$ for all $a\in P_x$.

PROOF. The lemma is a restatement of Lemma 7.11. □

If $\mathbf{P}_x$ and α are as in the previous lemma, then we can define an equivalence relation $\alpha+(x,y)$. The equivalence is given by the following partition of P_y

$$\{a/\alpha+(x,y) : a\in P_x\}.$$

This defines a trivial partition $\{P_y\}$ in (case2) of Lemma 8.5 or a proper partition in (case1). The equivalence $\alpha+(x,y)$ is easily seen to be a congruence of $\mathbf{P}_y$. Moreover, in (case1), the function

$$\big\{\big(a/\alpha, b/(\alpha+(x,y))\big) : (a,b)\in R_{x,y}\big\}$$

provides an isomorphism between $\mathbf{P}_x/\alpha$ and $\mathbf{P}_y/(\alpha+(x,y))$. Therefore, in this case, $\alpha+(x,y)$ is a maximal congruence of $\mathbf{P}_y$ and $(\alpha+(x,y))+(y,x)=\alpha$.

LEMMA 8.6. *Let α be a maximal congruence on $\mathbf{P}_x$, $a\in P_x$, and let p be a pattern from x to y. If $a/\alpha+p\subsetneq P_y$, then for any $a'\in P_x$ we have $a'/\alpha+p-p=a'/\alpha$.*

PROOF. The proof is by induction on the length of p. Let $p=(x,z)+p'$. Then, $a/\alpha+(x,z)\subsetneq P_z$ and (case1) of Lemma 8.5 (used for the step (x,z)) applies. Therefore $\alpha+(x,z)$ is a maximal congruence on $\mathbf{P}_z$, both $a/\alpha+(x,z)$ and $a'/\alpha+(x,z)$ are its congruence classes and $a'/\alpha+(x,z)+(z,x)=a'/\alpha$. By induction hypothesis, $(a'/\alpha+(x,z))+p'-p'=(a'/\alpha+(x,z))$ and the lemma is proved. □

If $\mathbf{P}_x$ and α are as in the previous lemma, $a/\alpha+p\subsetneq P_y$, and q is another pattern from x to y such that $a/\alpha+q\subsetneq P_y$, then, by the previous lemma, $(a/\alpha+p)-p+q=a/\alpha+q$ and $(a/\alpha+q)-q+p=a/\alpha+p$, and using (P3), we get $a/\alpha+p=a/\alpha+q$, that is, the result of addition is independent of the pattern.

Now we are ready to define the decomposition. We assume that P_x has at least two elements and put α to be a maximal congruence on $\mathbf{P}_x$. We denote by $P_x^1,\dots,P_x^l$ the equivalence classes of α and will decompose $\mathcal{J}$ into $\mathcal{J}^1,\dots,\mathcal{J}^l$. We include y into X if there is a pattern p from x to y such that $P_x^1+p\subsetneq P_y$ and, in this case, we set $P_y^i=P_x^i+p$ for all $i\leq l$ (by the discussion after Lemma 8.6, the definition is independent on the choice of p as long as $P_x^i+p\subsetneq P_y^i$). If $y\notin X$, we put $P_y^i=P_y$. Condition (1) of Definition 8.2 holds by the construction.

Let y be an arbitrary variable in X. From Lemma 8.6, it follows that $P_y^i\cap P_y^{i'}=\emptyset$ for any $i\neq i'$, that is, condition (2a) of Definition 8.2 holds. To prove condition (2b), take any step (y,z), and suppose that $P_y^i+(y,z)\neq P_z^i$ for some $i\leq l$. By Lemma 8.5, the sets $(P_y^j+(y,z))$ (for different j's) are either pairwise disjoint, or all equal to P_z. If they were disjoint, $P^i_y + (y,z)$ would be a candidate for P^i_z (given by a different path). This is impossible (by the discussion after Lemma 8.6) and therefore $P^j_y + (y,z) = P_z$ for all $j \leq l$ – this proves condition (2b) of Definition 8.2. Therefore, the induced subinstances $\mathcal{J}^1, \dots, \mathcal{J}^l$ with potatoes $P^i_x, x \in V$ form a decomposition.

It remains to show that the decomposition is pointed. Clearly, for any i, P^i_x is a subuniverse of $\mathbf{P}_x$ and therefore, by Lemma 7.7, P^i_y is a subuniverse of $\mathbf{P}_y$. In order to find a pointed term for this decomposition consider $\mathbf{P}_x/\alpha$. This algebra is simple, has no proper absorbing subuniverses (since $\mathbf{P}_x$ does not) and lies in a variety omitting types **1** and **2**. Therefore, by Lemma 7.6, there exists a term $t(x_1, \dots, x_m)$ in $\mathbf{P}_x/\alpha$ pointing to P^1_x/α.

This term clearly satisfies the required condition for the variable x and for all the variables outside the set X. On the other hand, if $y \in X$, then $\mathbf{P}_y/(\alpha + p)$ is isomorphic to $\mathbf{P}_x/\alpha$ via an isomorphism sending $\mathbf{P}^1_x$ to $\mathbf{P}^1_y$ and therefore $t(x_1, \dots, x_m)$ satisfies the condition for $\mathbf{P}_y$ as well. This finishes the case when there is no absorption in the weak Prague instance.

8.4. The Reduction to $\mathcal{J}^1$

The following theorem finishes the proof of the main result.

THEOREM 8.7. *Let $\mathcal{J}$ be a weak Prague instance with a pointed decomposition $\mathcal{J}^1, \dots, \mathcal{J}^l$ with potatoes P^i_x, $x \in V$, $i \leq l$. Then, $\mathcal{J}^1$ is a weak Prague instance.*

In this section, we need to consider realizations and addition in the instance $\mathcal{J}$ as well as in the subinstance $\mathcal{J}^1$. To distinguish them, we write $A + p$ for addition in $\mathcal{J}$ and $A +^1 p$ for addition in $\mathcal{J}^1$ (and similarly for subtraction).

When we keep adding a pattern p from x to x to a set $A \subseteq P_x$, the process will stabilize on a set containing A.

LEMMA 8.8. *Let $x \in V$, $A \subseteq P_x$ and let p be a pattern from x to x. There exists a natural number i such that $A + ip + jp = A + ip$ for every integer*[5] *j and, moreover, $A \subseteq A + ip$.*

PROOF. Since the domain is finite, there exist $i, i' > 0$ such that $A + ip + i'p = A + ip$. Put $B = A + ip$. Then, $B + p + (i' - 1)p = B$ and, by (P3), $B + p = B$. From (P2) we get that $B - p = B$ and now clearly $B + jp = B$ for every integer j. Finally, we have $A \subseteq A + ip - ip = B - ip = B$. □

For every $A \subseteq P_x$ and pattern p from x to x, the set $A + ip$ given by Lemma 8.8 is denoted by $[A]_p$. We have $A \subseteq [A]_p$ and $[A]_p + jp = [A]_p$ for every integer j.

Theorem 8.7 is a consequence of the following lemma.

LEMMA 8.9. *Let $x \in V$, $A \subseteq P^1_x$ and let p be a pattern from x to x. If $A +^1 p = A$, then $A = [A]_p \cap P^1_x$.*

Using this lemma, we prove Theorem 8.7.

PROOF OF THEOREM 8.7. It follows from the definition of decomposition that $\mathcal{J}^1$ (as well as $\mathcal{J}^2, \dots, \mathcal{J}^l$) is 1-minimal.

(P2). If $A +^1 p = A$, where $A \subseteq P_x$ and p is a pattern from x to x, then $A \subseteq A +^1 p -^1 p = A -^1 p$. To prove the reverse inclusion, we use the properties of $[A]_p$ stated after Lemma 8.8 and Lemma 8.9. Since $A \subseteq [A]_p = [A]_p - p$, we have $A -^1 p \subseteq A - p \subseteq [A]_p - p = [A]_p$, and then $A -^1 p \subseteq [A]_p \cap P^1_x = A$. Thus, $A = A -^1 p$ as required.

[5] Compare Definition 6.2 for pattern multiplicities.

(P3). Let $A \subseteq P_x^1$ and let p,q be patterns from x to x such that $A +^1 p +^1 q = A$. Using (P3) for the instance $\mathcal{J}$ and $[A]_{p+q} = [A]_{p+q} + p + q$, we get $[A]_{p+q} + p = [A]_{p+q}$. Now $A +^1 p \subseteq A + p \subseteq [A]_{p+q} + p = [A]_{p+q}$ and, by Lemma 8.9, we obtain $A +^1 p \subseteq [A]_{p+q} \cap P_x^1 = A$.

Similarly, the set $A = (A +^1 p) +^1 q$ is a subset of $A +^1 p$, and then $A = A +^1 p$ as required. □

The remaining part of this section is devoted to the proof of Lemma 8.9. Let x, A, p be as in the statement of the lemma.

For any $i > 0$, we have $A +^1 ip = A +^1 p$ and $[A]_p = [A]_{ip}$, therefore we can, without loss of generality, replace p with the pattern ip. We do it for large enough i so that $B + p = [B]_p$ for every $B \subseteq P_x$.

As $A \subseteq [A]_p$, the inclusion $A \subseteq [A]_p \cap P_x^1$ is satisfied and we proceed to prove the reverse containment. Let $b \in [A]_p \cap P_x^1$ be arbitrary. We need to show that $b \in A$. As $[A]_p = A + p$, there exists $a \in A$ such that $b \in \{a\} + p$ and we fix such an element a.

We split the pattern p into parts. Let $p = (x = x_1, x_2, \ldots, x_n = x)$ and define r, s:

— let r be the largest index such that $P_{x_z}^1 + (x_z, x_{z+1}) = P_{x_{z+1}}^1$ for all $z < r$,
— let s be the smallest index such that $P_{x_z}^1 + (x_z, x_{z-1}) = P_{x_{z-1}}^1$ for all $z > s$.

If it is not true that $r + 1 \leq s - 1$, then $b \in \{a\} + p$ implies $b \in \{a\} +^1 p \in A$ since no realization of p can leave $\mathcal{J}^1$ and return to it, thus in this case the proof is concluded.

Let t be an m-ary term and $k_1, \ldots, k_m$ be indices from the definition of pointed decomposition. We find a matrix of domain elements with m rows such that its ith row

$$(a = a_{11}^i, a_{12}^i, \ldots, a_{1n}^i = a_{21}^i, \ldots, a_{2n}^i = a_{31}^i, \ldots, a_{(m-1)n}^i = a_{m1}^i, \ldots, a_{mn}^i = b)$$

is a realization of the pattern mp in the instance $\mathcal{J}$ satisfying the following conditions for every $1 \leq j \leq m$.

(1) The elements $a_{j1}^i, \ldots, a_{jr}^i, a_{js}^i, \ldots, a_{jn}^i$ lie in the instance $\mathcal{J}^1$ (i.e., more precisely, $a_{j1}^i \in P_{x_1}^1, \ldots, a_{jr}^i \in P_{x_r}^1$, etc.), and
(2) if $j \neq i$, then $a_{j(r+1)}^i, \ldots, a_{j(s-1)}^i$ lie in the instance $\mathcal{J}^{k_i}$ (i.e., for all z such that $r < z < s$, we have $a_{jz}^i \in P_{x_z}^{k_i}$).

First, we find the initial part $(a_{11}^i, \ldots, a_{i1}^i)$.

The first segment $(a_{11}^i, \ldots, a_{1n}^i)$ is found as follows. We start with $a_{11}^i = a$. By 1-minimality of $\mathcal{J}^1$, we can choose elements $a_{12}^i \in P_{x_2}^1, \ldots, a_{1r}^i \in P_{x_r}^1$ one by one so that $(a_{1z}^i, a_{1(z+1)}^i)$ is a realization of (x_z, x_{z+1}) in the instance $\mathcal{J}^1$ for all $z < r$. By the choice of r, we have $P_{x_r}^1 + (x_r, x_{r+1}) \neq P_{x_{r+1}}^1$ and, by the definition of decomposition, $P_{x_r}^1 + (x_r, x_{r+1}) = P_{x_{r+1}}$. In particular, $P_{x_{r+1}}^{k_i} + (x_{r+1}, x_r)$ intersects $P_{x_r}^1$. If $k_i \neq 1$, then it follows from the definition of decomposition that $P_{x_{r+1}}^{k_i} + (x_{r+1}, x_r) = P_{x_r}$, and we can therefore choose $a_{1(r+1)}^i \in P_{x_{r+1}}^{k_i}$ such that $(a_{1r}^i, a_{1(r+1)}^i)$ is a realization of (x_r, x_{r+1}). If $k_i = 1$, we can find $a_{1(r+1)}^i$ from 1-minimality of $\mathcal{J}^1$. Using 1-minimality of $\mathcal{J}^{k_i}$, we find $a_{1(r+2)}^i, \ldots, a_{1(s-1)}^i$ such that $(a_{1(r+1)}^i, \ldots, a_{1(s-1)}^i)$ is a realization of $(x_{r+1}, \ldots, x_{s-1})$ in the instance $\mathcal{J}^{k_i}$. By the choice of s, we have $P_{x_s}^1 + (x_s, x_{s-1}) \neq P_{x_{s-1}}^1$ and, by the definition of decomposition again, we get $P_{x_s}^1 + (x_s, x_{s-1}) = P_{x_{s-1}}$; therefore, there exists

$a^i_{1s} \in P^1_{x_s}$ such that $(a^i_{1(s-1)}, a^i_{1s})$ is a realization of (x_{s-1}, x_s). Finally, by 1-minimality of $\mathcal{J}^1$, we find $a^i_{1(s+1)}, \ldots, a^i_{1n}$ such that $(a^i_{1s}, \ldots, a^i_{1n})$ is a realization of $(x_s, \ldots, x_n)$ in $\mathcal{J}^1$.

By the same argument, we construct the remaining $(i-2)$ segments of $(a^i_{11}, \ldots, a^i_{(i-1)n} = a^i_{i1})$. These elements satisfy, by construction, both (1) and (2) for every $j < i$. Using an analogical reasoning starting from $b = a^i_{mn}$ (and following the pattern in reverse direction), we can find a realization $(a^i_{in} = a^i_{(i+1)1}, \ldots, a^i_{mn})$ of the pattern $(m-i)p$ satisfying (1) and (2) for every $j > i$.

It remains to fill in the middle part $(a^i_{i1}, \ldots, a^i_{in})$ of the ith row of the matrix. Let $a' = a^i_{i1}$ and $b' = a^i_{in}$. By construction, $a', b' \in P^1_x$, $a' \in \{a\} + (i-1)p$ and $b' \in \{b\} - (m-i)p$. We observe that $[\{a'\}]_p$ contains b'. Indeed, since $a \in \{a'\} - (i-1)p$, we have $a \in [\{a'\}]_p - (i-1)p = [\{a'\}]_p$, then, using $b \in \{a\} + p$, we get $b \in [\{a'\}]_p + p = [\{a'\}]_p$, and, by $b' \in \{b\} - (m-i)p$, we obtain $b' \in [\{a'\}]_p - (m-i)p = [\{a'\}]_p$. Recall that $[B]_p = B + p$ for every $B \subseteq P_x$, therefore $b' \in \{a'\} + p$ and hence we can find a realization $(a' = a^i_{i1}, a^i_{i2}, \ldots, a^i_{in} = b')$ of the pattern p in the instance $\mathcal{J}$. By the choice of r, we have $P^1_{x_z} + (x_z, x_{z+1}) = P^1_{x_{z+1}}$ for all $z < r$, which guarantees $a^i_{iz} \in P^1_{x_z}$ for all $z \leq r$. Similarly, by the choice of s, we have $a^i_{iz} \in P^1_{x_z}$ for $z \geq s$ and the construction of the matrix is concluded.

Finally, we apply the operation t to the columns of the matrix and obtain a tuple $(b_{11}, \ldots, b_{mn})$. Note that $b_{11} = a$ and $b_{mn} = b$ from idempotency of t. Since every row of the matrix is a realization of the pattern mp and t is a polymorphism of the relations $R_{x_1,x_2}, \ldots R_{x_{n-1},x_n}$, the tuple $(b_{11}, \ldots, b_{mn})$ is a realization of mp. Every element b_{jz} lies in $P^1_{x_z}$ since either t was applied to elements of $P^1_{x_z}$ (in the case that $z \leq r$ or $z \geq s$), or t was applied to $(a^1_{jz}, \ldots, a^m_{jz})$ where $a^i_{jz} \in P^{k_i}_{x_z}$ for all $i \neq j$ (in the case that $r < z < s$) and we can then use the property of t from the definition of pointed decomposition. Thus $(a = b_{11}, \ldots, b_{mn} = b)$ is a realization of mp in the instance $\mathcal{J}^1$; therefore, $b \in \{a\} +^1 mp \subseteq A +^1 mp = A$ as required.

REFERENCES

Libor Barto. 2013. The collapse of the bounded width hierarchy. In preparation.

Libor Barto and Marcin Kozik. 2009a. Congruence distributivity implies bounded width. *SIAM J. Comput. 39*, 4, 1531–1542. DOI:http://dx.doi.org/10.1137/080743238.

Libor Barto and Marcin Kozik. 2009b. Constraint satisfaction problems of bounded width. In *Proceedings of the 50th Symposium on Foundations of Computer Science (FOCS'09)*. 595–603.

Libor Barto and Marcin Kozik. 2010. New conditions for Taylor varieties and CSP. In *Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science (LICS'10)*. IEEE Computer Society, Los Alamitos, CA, 100–109. DOI:http://dx.doi.org/10.1109/LICS.2010.34.

Libor Barto and Marcin Kozik. 2012. Robust satisfiability of constraint satisfaction problems. In *Proceedings of the 44th Symposium on Theory of Computing (STOC'12)*. ACM, New York, 931–940. DOI:http://dx.doi.org/10.1145/2213977.2214061.

Libor Barto, Marcin Kozik, and David Stanovský. 2013. Mal'tsev conditions, lack of absorption, and solvability. Submitted.

C. Bergman. 2011. *Universal Algebra: Fundamentals and Selected Topics*. Taylor and Francis. http://books.google.ca/books?id=QXi3BZWoMRwC.

Joel Berman, Pawel Idziak, Petar Marković, Ralph McKenzie, Matthew Valeriote, and Ross Willard. 2009. Varieties with few subalgebras of powers. *Trans. Amer. Math. Soc. 362*, 1445–1473. DOI:http://dx.doi.org/10.1090/S0002-9947-09-04874-0.

Garrett Birkhoff. 1935. On the structure of abstract algebras. *Proc. Cambridge Philos. Soc. 31*, 433–454.

V. G. Bodnarčuk, L. A. Kalužnin, V. N. Kotov, and B. A. Romov. 1969. Galois theory for Post algebras. I, II. *Kibernetika (Kiev) 3*, 1–10; ibid. 1969, no. 5, 1–9.

Andrei Bulatov. 2006a. Complexity of Maltsev constraints. *Algebra i Logika 26*, 1, (In Russian).

Andrei Bulatov. 2009. Bounded relational width. Manuscript.
http://www.cs.sfu.ca/~abulatov/papers/relwidth.pdf.

Andrei Bulatov and Víctor Dalmau. 2006. A simple algorithm for Mal′tsev constraints. *SIAM J. Comput. 36*, 1, 16–27 (electronic).

Andrei Bulatov, Peter Jeavons, and Andrei Krokhin. 2005. Classifying the complexity of constraints using finite algebras. *SIAM J. Comput. 34*, 720–742. Issue 3.
DOI:http://dx.doi.org/10.1137/S0097539700376676.

Andrei A. Bulatov. 2003. Tractable conservative constraint satisfaction problems. In *Proceedings of the 18th Annual IEEE Symposium on Logic in Computer Science*. IEEE Computer Society, Los Alamitos, CA, 321–. DOI:http://dl.acm.org/citation.cfm?id=788023.789067.

A. A. Bulatov. 2004a. Complexity of the conservative generalized satisfiability problem. *Dokl. Akad. Nauk 397*, 5, 583–585.

Andrei A. Bulatov. 2004b. A graph of a relational structure and constraint satisfaction problems. In *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science (LICS'04)*. IEEE Computer Society Press, Los Alamitos, CA, 448–457.

Andrei A. Bulatov. 2006b. Combinatorial problems raised from 2-semilattices. *J. Algebra 298*, 2, 321–339.

Andrei A. Bulatov. 2006c. A dichotomy theorem for constraint satisfaction problems on a 3-element set. *J. ACM 53*, 1, 66–120.

Andrei A. Bulatov and Matthew Valeriote. 2008. Recent results on the algebraic approach to the CSP. In *Complexity of Constraints*, Nadia Creignou, Phokion G. Kolaitis, and Heribert Vollmer Eds., Lecture Notes in Computer Science, vol. 5250, Springer, 68–92.
http://dblp.uni-trier.de/db/conf/dagstuhl/coc2008.html#BulatovV08.

Andrei A. Bulatov, Andrei A. Krokhin, and Peter Jeavons. 2000. Constraint satisfaction problems and finite algebras. In *Automata, Languages and Programming*. Lecture Notes in Computer Science, vol. 1853, Springer, Berlin, 272–282.

Andrei A. Bulatov, Andrei Krokhin, and Benoit Larose. 2008. Dualities for constraint satisfaction problems. In *Complexity of Constraints*, Springer-Verlag, Berlin, 93–124.
DOI:http://dx.doi.org/10.1007/978-3-540-92800-3_5.

Stanley N. Burris and H. P. Sankappanavar. 1981. *A Course in Universal Algebra*. Graduate Texts in Mathematics, Vol. 78, Springer-Verlag, New York.

Catarina Carvalho, Víctor Dalmau, Petar Marković, and Miklós Maróti. 2009. CD(4) has bounded width. *Algebra Universalis 60*, 3, 293–307.

Tomás Feder and Moshe Y. Vardi. 1998. The computational structure of monotone monadic SNP and constraint satisfaction: A study through datalog and group theory. *SIAM J. Comput. 28*, 1, 57–104.
DOI:http://dx.doi.org/10.1137/S0097539794266766.

David Geiger. 1968. Closed systems of functions and predicates. *Pacific J. Math. 27*, 95–100.

David Hobby and Ralph McKenzie. 1988. *The Structure of Finite Algebras*. Contemporary Mathematics, Vol. 76, American Mathematical Society, Providence, RI.

Pawel Idziak, Petar Marković, Ralph McKenzie, Matthew Valeriote, and Ross Willard. 2007. Tractability and learnability arising from algebras with few subpowers. In *Proceedings of the 22nd Annual IEEE Symposium on Logic in Computer Science (LICS'07)*. IEEE Computer Society Press, Los Alamitos, CA, 213–222.

Peter Jeavons, David Cohen, and Marc Gyssens. 1997. Closure properties of constraints. *J. ACM 44*, 4, 527–548.

Emil Kiss and Matthew Valeriote. 2007. On tractability and congruence distributivity. *Log. Meth. Comput. Sci. 3*, 2, 2:6, (electronic).

Benoît Larose and Pascal Tesson. 2009. Universal algebra and hardness results for constraint satisfaction problems. *Theoret. Comput. Sci. 410*, 18, 1629–1647.
DOI:http://dx.doi.org/10.1016/j.tcs.2008.12.048.

Benoît Larose and László Zádori. 2007. Bounded width problems and algebras. *Algebra Universalis 56*, 3–4, 439–466.

Benoît Larose, Matt Valeriote, and László Zádori. 2009. Omitting types, bounded width and the ability to count. *Internat. J. Algebra Comput. 19*, 5, 647–668.
DOI:http://dx.doi.org/10.1142/S021819670900524X.

Miklós Maróti and Ralph McKenzie. 2008. Existence theorems for weakly symmetric operations. *Algebra Universalis 59*, 3–4, 463–489.

Received March 2012; revised February 2013; accepted November 2013

Appendix C – Absorption and cyclic terms

Logical Methods in Computer Science
Vol. 8 (1:07) 2012, pp. 1–26
www.lmcs-online.org

Submitted Dec. 30, 2010
Published Feb. 20, 2012

ABSORBING SUBALGEBRAS, CYCLIC TERMS, AND THE CONSTRAINT SATISFACTION PROBLEM *

LIBOR BARTO [a] AND MARCIN KOZIK [b]

[a] Department of Algebra, Charles University, Sokolovská 83, 186 75 Prague 8, Czech Republic and Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
e-mail address: libor.barto@gmail.com

[b] Theoretical Computer Science Department, Faculty of Mathematics and Computer Science, Jagiellonian University, ul. Prof. St. Łojasiewicza 6, 30-348 Kraków, Poland
e-mail address: marcin.kozik@tcs.uj.edu.pl

Abstract. The Algebraic Dichotomy Conjecture states that the Constraint Satisfaction Problem over a fixed template is solvable in polynomial time if the algebra of polymorphisms associated to the template lies in a Taylor variety, and is NP-complete otherwise. This paper provides two new characterizations of finitely generated Taylor varieties. The first characterization is using absorbing subalgebras and the second one cyclic terms. These new conditions allow us to reprove the conjecture of Bang-Jensen and Hell (proved by the authors) and the characterization of locally finite Taylor varieties using weak near-unanimity terms (proved by McKenzie and Maróti) in an elementary and self-contained way.

Introduction

The Constraint Satisfaction Problem (CSP) is a generic problem in computer science. An instance consists of a number of variables and constraints imposed on them and the objective is to determine whether variables can be assigned values in such a way that all the constraints are met. As CSP provides a common framework for many theoretical problems as well as for many real-life applications, it has been studied by computer scientists for over forty years.

1998 ACM Subject Classification: F.2.2, F.4.1.

2000 Mathematics Subject Classification: 08A70, 68Q17.

Key words and phrases: Constraint Satisfaction Problem, Taylor variety, cyclic term, absorbing subalgebra.

* A part of this work has appeared in our paper *New conditions for Taylor varieties and CSP*, Proceedings of the 25th IEEE Symposium on Logic in Computer Science, LICS'10, 100-109.

[a] Research supported by the Grant Agency of the Czech Republic under the grant No. 201/09/P223 and by the Ministry of Education of the Czech Republic under the grant No. MSM 0021620839.

[b] Research supported by the Foundation for Polish Science under the grant No. HOM/2008/7 (supported by MF EOG), and Ministry of Science and Higher Education of Poland under the grant No. N206 357036.

DOI:10.2168/LMCS-8 (1:07) 2012

The results contained in this paper follow a long line of research devoted to verifying the Constraint Satisfaction Problem Dichotomy Conjecture of Feder and Vardi [FV99]. It deals with so called *non-uniform* CSP — the same decision problem as the ordinary CSP, but in this case the set of allowed constraint relations is finite and fixed. The conjecture states that, for every finite, fixed set of constraint relations (a fixed *template*), the CSP defined by it is NP-complete or solvable in polynomial time, i.e. the class of CSPs exhibits a dichotomy.

The conjecture of Feder and Vardi dates back to 1993. At that time it was supported by two major results, Schaefer's dichotomy theorem for two–element templates [Sch78], and the dichotomy theorem for undirected graphs by Hell and Nešetřil [HN90]. The first breakthrough in the research appeared in 1997 in the work of Jeavons, Cohen and Gyssens [JCG97], refined later by Bulatov, Jeavons and Krokhin [BKJ00, BJK05]. At heart of the new approach lies a proof that the complexity of CSP, for a fixed template, depends only on a set of certain operations — polymorphisms of the template. Thus the study of templates gives rise to the study of algebras associated to them.

The algebraic approach has lead to a better understanding of the known results and brought a number of new results which were out of reach for pre-algebraic methods. The theorem of Schaefer [Sch78] has been extended by Bulatov [Bul06] to three–element domains. Another major result of Bulatov [Bul03, Bul11] establishes the dichotomy for templates containing all unary relations. The conjecture of Bang-Jensen and Hell [BJH90], generalizing Hell's and Nešetřil's dichotomy theorem [HN90], was confirmed [BKN08, BKN09]. New algorithms were devised [BD06, Dal06, IMM$^+$07] and pre-algebraic algorithms were characterized in algebraic terms [BK09a, BK09b].

The hardness parts in the dichotomy results mentioned above were obtained using a theorem of Bulatov, Jeavons and Krokhin [BKJ00, BJK05] stating that whenever an algebra associated with a core template does not lie in a Taylor variety then the CSP defined by the template is NP-complete. In the same paper the authors conjecture that in all the other cases the associated CSP is solvable in polynomial time. All the known partial results agree with this proposed classification, which is now commonly referred to as the Algebraic Dichotomy Conjecture.

In order to prove the Algebraic Dichotomy Conjecture one has to devise an algorithm that works for any relational structure with the corresponding algebra in a Taylor variety. As the characterization originally provided by Taylor [Tay77] is difficult to work with, a search for equivalent conditions is ongoing. A technical, but useful condition was obtained by Bulatov who used it to prove his dichotomy theorems [Bul03, Bul06]. Another powerful tool is the characterization of (locally finite) Taylor varieties in terms of weak near-unanimity operations due to Maróti and McKenzie [MM08]. Unfortunately, their proof uses a deep algebraic theory of Hobby and McKenzie [HM88], therefore is not easily accessible for a nonspecialist. The proof of the conjecture of Bang-Jensen and Hell hinges on this characterization; also the algebraic characterization of problems of bounded width [BK09b] relies on a similar characterization of congruence meet semi-distributive varieties provided in the same paper [MM08]. Recently, a surprisingly simple condition for Taylor varieties was found by Siggers [Sig10], and an analytical characterization was given by Kun and Szegedy [KS09].

In this paper we provide two new conditions for (finitely generated) Taylor varieties. These new characterizations already proved to be useful. Not only they provide new tools for attacking the algebraic dichotomy conjecture, but they also allow us to present easy and elementary proofs for some of the results mentioned above. Moreover, their proofs are self-contained and do not require heavy algebraic machinery.

The first, structural characterization (the Absorption Theorem) is expressed in terms of *absorbing subalgebras* developed and successfully applied by the authors in [BKN08, BKN09, BK09a, BK09b]. We use it to present an elementary proof of the conjecture of Bang-Jensen and Hell. Recently, the Absorption Theorem was applied to give a short proof of Bulatov's dichotomy theorem for conservative CSPs [Bar11]. The second, equational characterization involves *cyclic terms* and is a stronger version of the weak near-unanimity condition. We use it to restate the Algebraic Dichotomy Conjecture in simple combinatorial terms and to provide a very short proof of the theorem of Hell and Nešetřil.

The results of this paper also show that the tools developed for the CSP can be successfully applied to algebraic questions which indicates a deep connection between the CSP and universal algebra.

Organization of the paper. In section 1 we introduce the necessary notions concerning algebras and the CSP. In section 2 we define absorbing subalgebras and present the Absorption Theorem and its corollaries. In section 3 we use the absorbing subalgebra characterization to provide an elementary proof of the conjecture of Bang-Jensen and Hell in a slightly stronger version which is needed in section 4. Finally, in section 4 we prove the characterization using cyclic terms and its corollaries: the theorem of Hell and Nešetřil [HN90] and the weak near-unanimity characterization of locally finite Taylor varieties of Maróti and McKenzie [MM08].

1. Preliminaries

1.1. **Notation for sets.** For a set A and a natural number n, elements of A^n are the n-tuples of elements of A. We index its coordinates starting from zero, for example $(a_0, a_1, \dots, a_{n-1}) \in A^n$.

Let R be a subset of a Cartesian product $A_1 \times A_2 \times \cdots \times A_n$. R is called *subdirect* ($R \subseteq_S A_1 \times \cdots \times A_n$) if, for every $i = 1, 2, \dots, n$, the projection of R to the i-th coordinate is the whole set A_i.

Given $R \subseteq A \times B$ and $S \subseteq B \times C$, by $S \circ R$ we mean the following subset of $A \times C$:

$$S \circ R = \{(a, c) : \exists\, b \in B \quad (a, b) \in R, (b, c) \in S\}.$$

If $R \subseteq A \times A$ and n is a natural number greater than zero, then we define

$$R^{\circ n} = \underbrace{R \circ R \circ \cdots \circ R}_{n}.$$

1.2. **Algebras and varieties.** An *algebraic signature* is a finite set of function symbols with a natural number (the *arity*) associated to each of them. An *algebra* of a signature Σ is a pair $\mathbf{A} = (A, (t^{\mathbf{A}})_{t\in\Sigma})$, where A is a set, called the *universe* of $\mathbf{A}$, and $t^{\mathbf{A}}$ is an operation on A of arity $\mathrm{ar}(t)$, that is, a mapping $A^{\mathrm{ar}(t)} \to A$. We always use a boldface letter to denote an algebra and the same letter in a plain type to denote its universe. We often omit the superscripts of operations when the algebra is clear from the context.

A *term* in a signature Σ is a formal expression using variables and compositions of symbols in Σ. In this paper we introduce a special notation for a particular case of composition of terms: given a k-ary term t_1 and an l-ary term t_2 we define a kl-ary term $t_1 * t_2$ by

$$t_1 * t_2(x_0, x_1, \ldots, x_{kl-1}) = t_1(t_2(x_0, \ldots, x_{l-1}), t_2(x_l, \ldots, x_{2l-1}), \ldots, t_2(x_{(k-1)l} \ldots, x_{kl-1})).$$

For an algebra $\mathbf{A}$ and a term h in the same signature Σ, $h^{\mathbf{A}}$ has the natural meaning in $\mathbf{A}$ and is called a *term operation* of $\mathbf{A}$. Again, we usually omit the superscripts of term operations when the algebra is clear from the context. The set of all term operations of $\mathbf{A}$ is called the *clone of term operations* of $\mathbf{A}$ and it is denoted $\mathrm{Clo}(\mathbf{A})$.

For a pair of terms s, t over a signature Σ, we say that an algebra $\mathbf{A}$ in the signature Σ *satisfies the identity* $s \approx t$ if the term operations $s^{\mathbf{A}}$ and $t^{\mathbf{A}}$ are the same.

There are three fundamental operations on algebras of a fixed signature Σ: forming subalgebras, factoralgebras and products. A subset B of the universe of an algebra $\mathbf{A}$ is called a *subuniverse*, if it is closed under all operations (equivalently term operations) of $\mathbf{A}$. Given a subuniverse B of $\mathbf{A}$ we can form the algebra $\mathbf{B}$ by restricting all the operations of $\mathbf{A}$ to the set B. In this situation we write $B \leq \mathbf{A}$ or $\mathbf{B} \leq \mathbf{A}$. We call the subuniverse B (or the subalgebra $\mathbf{B}$) *proper* if $\emptyset \neq B \neq A$. The smallest subalgebra of $\mathbf{A}$ containing a set $B \subseteq A$ is called the subalgebra *generated by* B and will be denoted by $\mathrm{Sg}_{\mathbf{A}}(B)$. It can be equivalently described as the set of elements which can be obtained by applying term operations of $\mathbf{A}$ to elements of B.

Given a family of algebras $\mathbf{A}_i, i \in I$ we define its product $\prod_{i\in I} \mathbf{A}_i$ to be the algebra with the universe equal to the cartesian product of the A_i's and with operations computed coordinatewise. The product of algebras $\mathbf{A}_1, \ldots, \mathbf{A}_n$ will be denoted by $\mathbf{A}_1 \times \cdots \times \mathbf{A}_n$ and the product of n copies of an algebra $\mathbf{A}$ by $\mathbf{A}^n$. $\mathbf{R}$ is a *subdirect subalgebra* of $\mathbf{A}_1 \times \mathbf{A}_2 \times \cdots \times \mathbf{A}_n$ if R is subdirect in $A_1 \times A_2 \times \cdots \times A_n$ and, in such a case, we write $\mathbf{R} \leq_S \mathbf{A}_1 \times \cdots \times \mathbf{A}_n$.

An equivalence relation $\sim$ on the universe of an algebra $\mathbf{A}$ is a *congruence*, if it is a subalgebra of $\mathbf{A}^2$. The corresponding *factor algebra* $\mathbf{A}/\sim$ has, as the universe, the set of $\sim$-blocks and the operations are defined using (arbitrarily chosen) representatives. A congruence is *nontrivial*, if it is not equal to the diagonal or to the full relation $A \times A$.

A *variety* is a class of algebras of the same signature closed under forming isomorphic copies, subalgebras, factoralgebras and products. For a pair of terms s, t over a signature Σ, we say that a class of algebras $\mathcal{V}$ in the signature Σ *satisfies the identity* $s \approx t$ if every algebra in the class does. By Birkhoff's theorem, a class of algebras is a variety if and only if there exists a set of identities E such that the members of $\mathcal{V}$ are precisely those algebras which satisfy all the identities from E.

A variety $\mathcal{V}$ is called *locally finite*, if every finitely generated algebra (that is, an algebra generated by a finite subset) contained in $\mathcal{V}$ is finite. $\mathcal{V}$ is called *finitely generated*, if there exists a finite set $\mathcal{K}$ of finite algebras such that $\mathcal{V}$ is the smallest variety containing $\mathcal{K}$. In such a case $\mathcal{V}$ is actually generated by a single, finite algebra, the product of members of $\mathcal{K}$. Every finitely generated variety is locally finite, and if a variety is generated by a single algebra then the identities satisfied in this algebra are exactly the identities satisfied in the variety.

For a more in depth introduction to universal algebra and proofs of the above mentioned results we recommend [BS81].

1.3. **Taylor varieties.** A term s is *idempotent* in a variety (or an algebra), if it satisfies the identity

$$s(x, x, \dots, x) \approx x.$$

An algebra (a variety) is idempotent if all its terms are.

A term t of arity at least 2 is called a *weak near-unanimity* term of a variety (or an algebra), if t is idempotent and satisfies

$$t(y, x, x, \dots, x) \approx t(x, y, x, x, \dots, x) \approx \dots\cdots \approx t(x, x, \dots, y, x) \approx t(x, x, \dots, x, y).$$

A term t of arity at least 2 is called a *cyclic* term of a variety (or an algebra), if t is idempotent and satisfies

$$t(x_0, x_1, \dots, x_{k-1}) \approx t(x_1, x_2, \dots, x_{k-1}, x_0).$$

Finally, a term t of arity k is called a *Taylor term* of a variety (or an algebra), if t is idempotent and for every $j < k$ it satisfies an identity of the form

$$t(\square_0, \square_1, \dots, \square_{k-1}) \approx t(\triangle_0, \triangle_1, \dots, \triangle_{k-1}),$$

where all $\square_i$'s and $\triangle_i$'s are substituted with either x or y, but $\square_j$ is x while $\triangle_j$ is y.

Definition 1.1. An idempotent variety $\mathcal{V}$ is called *Taylor* if it has a Taylor term.

Study of Taylor varieties has been a recurring subject in universal algebra for many years. One of the first characterizations is due to Taylor [Tay77]

Theorem 1.2 (Taylor [Tay77])**.** *Let $\mathcal{V}$ be an idempotent variety. The following are equivalent.*

- *$\mathcal{V}$ is a Taylor variety.*
- *$\mathcal{V}$ does not contain a two-element algebra whose every (term) operation is a projection.*

Further research led to discovery of other equivalent conditions [HM88, MM08, Sig10, KS09]. One of the most important ones is the result of Maróti and McKenzie [MM08].

Theorem 1.3 (Maróti and McKenzie [MM08])**.** *Let $\mathcal{V}$ be an idempotent, locally finite variety. The following are equivalent.*

- *$\mathcal{V}$ is a Taylor variety.*
- *$\mathcal{V}$ has a weak near-unanimity term.*

This result, together with a similar characterization provided in the same paper for congruence meet semi-distributive varieties, found deep applications in CSP [BKN08, BKN09, BK09b].

1.4. **Relational structures and CSP.** A convenient formalization of non-uniform CSP is via homomorphisms between relational structures [FV99].
A *relational signature* is a finite set of relation symbols with arities associated to them. A *relational structure* of the signature Σ is a pair $\mathbb{A} = (A, (R^{\mathbb{A}})_{R\in\Sigma})$, where A is a set, called the *universe* of $\mathbb{A}$, and $R^{\mathbb{A}}$ is a relation on A of arity $\mathrm{ar}(R)$, that is, a subset of $A^{\mathrm{ar}(R)}$.

Let $\mathbb{A}, \mathbb{B}$ be relational structures of the same signature. A mapping $f : A \to B$ is a *homomorphism* from $\mathbb{A}$ to $\mathbb{B}$, if it preserves all $R \in \Sigma$, that is, $(f(a_0), f(a_1), \dots, f(a_{\mathrm{ar}(R)-1})) \in R^{\mathbb{B}}$ for any $(a_0, \dots, a_{\mathrm{ar}(R)-1}) \in R^{\mathbb{A}}$. A finite relational structure $\mathbb{A}$ is a *core*, if every homomorphism from $\mathbb{A}$ to itself is bijective.

For a fixed relational structure $\mathbb{A}$ of a signature Σ, $\mathrm{CSP}(\mathbb{A})$ is the following decision problem:

INPUT: A relational structure $\mathbb{X}$ of the signature Σ.
QUESTION: Does $\mathbb{X}$ map homomorphically to $\mathbb{A}$?

It is easy to see that if $\mathbb{A}'$ is a core of $\mathbb{A}$ (i.e. a core which is contained in $\mathbb{A}$ and such that $\mathbb{A}$ can be mapped homomorphically into it) then $\mathrm{CSP}(\mathbb{A})$ and $\mathrm{CSP}(\mathbb{A}')$ are identical.

The celebrated conjecture of Feder and Vardi [FV99] states that the class of CSPs exhibits a dichotomy:

The dichotomy conjecture of Feder and Vardi. *For any relational structure $\mathbb{A}$, the problem $\mathrm{CSP}(\mathbb{A})$ is solvable in polynomial time, or NP-complete.*

1.5. **Algebraic approach to CSP.** A mapping $f : A^n \to A$ is *compatible* with an m-ary relation R on A if the tuple

$$\left(f(a_0^0, a_0^1, \dots, a_0^{n-1}), \dots, f(a_{m-1}^0, a_{m-1}^1, \dots, a_{m-1}^{n-1})\right)$$

belongs to R whenever $(a_0^i, \dots, a_{m-1}^i) \in R$ for all $i < n$. A mapping compatible with all the relations in a relational structure $\mathbb{A}$ is a *polymorphism* of this structure.

For a given relational structure $\mathbb{A} = (A, (R^{\mathbb{A}})_{R\in\Sigma})$ we define an algebra $\mathrm{IdPol}(A)$ (often denoted by just $\mathbf{A}$). This algebra $\mathbf{A}$ has its universe equal to A and the operations of $\mathbf{A}$ are the idempotent polymorphisms of $\mathbb{A}$ (we formally define a signature of $\mathbf{A}$ to be identical with the set of its operations).

It follows from an old result [BKKR69, Gei68] that a relation R of arity k is a subuniverse of $\mathrm{IdPol}(\mathbb{A})^k$ if and only if R can be positively primitively defined from relations in $\mathbb{A}$ and singleton unary relations identifying every element of A. That is, R can be defined by a first-order formula which uses relations in $\mathbb{A}$, singleton unary relations on A, the equality relation on A, conjunction and existential quantification.

Already the first results on the algebraic approach to CSP [JCG97, BKJ00, BJK05] show that whenever a relational structure $\mathbb{A}$ is a core then $\mathrm{IdPol}(\mathbb{A})$ fully determines the computational complexity of $\mathrm{CSP}(\mathbb{A})$. Moreover, Bulatov, Jeavons and Krokhin showed [BKJ00, BJK05]:

Theorem 1.4 (Bulatov, Jeavons and Krokhin [BKJ00, BJK05])**.** *Let $\mathbb{A}$ be a finite relational structure which is a core. If* $\mathrm{IdPol}(\mathbb{A})$ *does not lie in a Taylor variety, then* $\mathrm{CSP}(\mathbb{A})$ *is NP-complete.*

In the same paper they conjectured that these are the only cases of finite cores which give rise to NP-complete CSPs.

The Algebraic Dichotomy Conjecture. *Let $\mathbb{A}$ be a finite relational structure which is a core. If* $\mathrm{IdPol}(\mathbb{A})$ *does not lie in a Taylor variety, then* $\mathrm{CSP}(\mathbb{A})$ *is NP-complete. Otherwise is it solvable in polynomial time.*

This conjecture is supported by many partial results on the complexity of CSPs [Bul03, Bul06, BKN08, BKN09, BK09b, IMM+07] and it renewed interest in properties of finitely generated Taylor varieties.

2. Absorbing subalgebras and absorption theorem

In this section we introduce the concept of an absorbing subalgebra and prove the Absorption Theorem and its corollaries. The proof is self-contained and elementary. In section 3 we use Theorem 2.3 to reprove a stronger version of the "Smooth Theorem" [BKN08, BKN09] which, in turn, will be used to prove the second main result of this article, Theorem 4.1. This approach simplifies significantly the known proof of the Smooth Theorem, and does not rely on the involved algebraic results results from [MM08]. It has also lead to a simple proof [Bar11] of the dichotomy theorem for conservative CSPs [Bul03].

2.1. **Absorption.** A subalgebra B of an algebra $\mathbf{A}$ is an absorbing subalgebra, if there exists a term operation of $\mathbf{A}$ which outputs an element of B whenever all but at most one of its arguments are from B. More precisely

Definition 2.1. Let $\mathbf{A}$ be an algebra and $t \in \mathrm{Clo}(\mathbf{A})$. We say that a subalgebra $\mathbf{B}$ of $\mathbf{A}$ is an *absorbing subalgebra of* $\mathbf{A}$ *with respect to* t if, for any $k < \mathrm{ar}(t)$ and any choice of $a_i \in A$ such that $a_i \in B$ for all $i \neq k$, we have $t(a_0, \dots, a_{\mathrm{ar}(t)-1}) \in B$.

We say that $\mathbf{B}$ is an *absorbing subalgebra of* $\mathbf{A}$, or that $\mathbf{B}$ *absorbs* $\mathbf{A}$ (and write $\mathbf{B} \triangleleft \mathbf{A}$), if there exists $t \in \mathrm{Clo}(\mathbf{A})$ such that $\mathbf{B}$ is an absorbing subalgebra of $\mathbf{A}$ with respect to t.

We also speak about *absorbing subuniverses*, i.e. universes of absorbing subalgebras. Recall that an (absorbing) subalgebra $\mathbf{B}$ of $\mathbf{A}$ is *proper*, if $\emptyset \neq B \subsetneq A$.

The Absorption Theorem says that the existence of a certain kind of subuniverse R of a product of two Taylor algebras $\mathbf{A}$ and $\mathbf{B}$ forces a proper absorbing subuniverse in one of these algebras. It is helpful to draw R as a bipartite undirected graph in the following sense: the vertex set is the disjoint union of A (draw it on the left) and B (on the right) and two elements $a \in A$ from the left side and $b \in B$ from the right side are adjacent if $(a, b) \in R$. We say that two vertices are linked if they are connected in this graph, and we call R linked if the graph is connected after deleting the isolated vertices. Note that $R \leq_S \mathbf{A} \times \mathbf{B}$ if and only if there are no isolated vertices.

Definition 2.2. Let $R \subseteq A \times B$ and let $a, a' \in A$. We say that $a, a' \in A$ are *linked in* R, or R*-linked*, via $c_0, \dots, c_{2n}$, if $a = c_0, c_{2n} = a'$ and $(c_{2i}, c_{2i+1}) \in R$ and $(c_{2i+2}, c_{2i+1}) \in R$ for all $i = 0, 1, \dots, n-1$.

In a similar way we define when $a \in A, a' \in B$ (or $a \in B, a' \in A$, or $a \in B, a' \in B$) are R-linked.

We say that R is *linked*, if a, a' are R-linked for any elements a, a' of the projection of R to the first coordinate.

These definitions allow us to state the Absorption Theorem which is the first main result of the paper.

Theorem 2.3. *Let $\mathcal{V}$ be an idempotent, locally finite variety, then the following are equivalent.*

- *$\mathcal{V}$ is a Taylor variety;*
- *for any finite $\mathbf{A}, \mathbf{B} \in \mathcal{V}$ and any linked $\mathbf{R} \leq_S \mathbf{A} \times \mathbf{B}$:*
  - *$\mathbf{R} = \mathbf{A} \times \mathbf{B}$ or*
  - *$\mathbf{A}$ has a proper absorbing subuniverse or*
  - *$\mathbf{B}$ has a proper absorbing subuniverse.*

2.2. **Proof of Absorption Theorem.** We start with a couple of useful observations. The first one says that absorbing subalgebras are closed under taking intersection, and that $\triangleleft$ is a transitive relation:

Proposition 2.4. *Let $\mathbf{A}$ be an algebra.*

- *If $\mathbf{C} \triangleleft \mathbf{B} \triangleleft \mathbf{A}$, then $\mathbf{C} \triangleleft \mathbf{A}$.*
- *If $\mathbf{B} \triangleleft \mathbf{A}$ and $\mathbf{C} \triangleleft \mathbf{A}$, then $B \cap C \triangleleft \mathbf{A}$.*

Proof. We start with a proof of the first item. Assume that $\mathbf{B}$ absorbs $\mathbf{A}$ with respect to t (of arity m) and that $\mathbf{C}$ absorbs $\mathbf{B}$ with respect to s (of arity n). We will show that $\mathbf{C}$ is an absorbing subalgebra of $\mathbf{A}$ with respect to $s * t$. Indeed, take any tuple $(a_0, \ldots, a_{mn-1}) \in A^{mn}$ such that $a_i \in C$ for all but one index, say j, and consider the evaluation of $s * t(a_0, \ldots, a_{mn-1})$. Every evaluation of the term t appearing in $s * t$ is of the form

$$t(a_{im}, \ldots, a_{im+m-1})$$

and therefore whenever j does not fall into the interval $[im, im+m-1]$ the result of it falls in C (as C is a subuniverse of $\mathbf{A}$). In the case when j is in that interval we have a term t evaluated on the elements of C (and therefore elements of B) in all except one coordinate. The result of such an evaluation falls in B (as $\mathbf{B}$ absorbs $\mathbf{A}$ with respect to t). Thus s is applied to a tuple consisting of elements of C on all but one position, and on this position the argument comes from B. Since $\mathbf{C}$ absorbs $\mathbf{B}$ with respect to s the results falls in C and the first part of the proposition is proved.

For the second part we consider $\mathbf{B} \triangleleft \mathbf{A}$ and $\mathbf{C} \triangleleft \mathbf{A}$; it follows easily that $B \cap C \triangleleft \mathbf{C}$ with respect to the same term as $\mathbf{B} \triangleleft \mathbf{A}$. Now it is enough to apply the first part. $\square$

Let R be a subuniverse of $\mathbf{A} \times \mathbf{B}$. We use the following notation for the neighborhoods of $X \subseteq A$ or $Y \subseteq B$:

$$\begin{aligned} X^{+R} &= \{b \in B : \exists\, a \in X \quad (a, b) \in R\} \\ Y^{-R} &= \{a \in A : \exists\, b \in Y \quad (a, b) \in R\} \end{aligned}$$

When R is clear from the context we write just X^+ and Y^-. The next lemma shows that these operations preserve (absorbing) subalgebras.

Lemma 2.5. *Let $R \leq \mathbf{A} \times \mathbf{B}$, where $\mathbf{A}, \mathbf{B}$ are algebras of the same signature. If $X \leq \mathbf{A}$ and $Y \leq \mathbf{B}$, then $X^+ \leq \mathbf{B}$ and $Y^- \leq \mathbf{A}$. Moreover, if $R \leq_S \mathbf{A} \times \mathbf{B}$ and $X \triangleleft \mathbf{A}$ and $Y \triangleleft \mathbf{B}$, then $X^+ \triangleleft \mathbf{B}$ and $Y^- \triangleleft \mathbf{A}$.*

Proof. Suppose $X \leq \mathbf{A}$ and take any term t, say of arity j, in the given signature. Let $b_0, \ldots, b_{j-1} \in X^+$ be arbitrary. From the definition of X^+ we can find $a_0, \ldots, a_{j-1} \in X$ such that $(a_i, b_i) \in R$ for all $0 \leq i < j$. Since R is a subuniverse of $\mathbf{A} \times \mathbf{B}$, the pair

$(t(a_0,\dots,a_{j-1}),t(b_0,\dots,b_{j-1}))$ is in R. But $t(a_0,\dots,a_{j-1})\in X$ as X is a subuniverse of $\mathbf{A}$. Therefore $t(b_0,\dots,b_{j-1})\in X^+$ and we have shown that X^+ is closed under all term operations of $\mathbf{B}$, i.e. $X^+\le\mathbf{B}$.

Suppose X absorbs $\mathbf{A}$ with respect to a term t of arity j. Let $0\le k<j$ be arbitrary and let $b_0,\dots,b_j\in B$ be elements such that $b_i\in X^+$ for all $i\neq k$. Then, for every $i, i\neq k$, we can find $a_i\in X$ such that $(a_i,b_i)\in R$. Also, since the projection of R to the second coordinate is B, we can find $a_k\in A$ such that $(a_k,b_k)\in R$. We again have $(t(a_0,\dots,a_{j-1}),t(b_0,\dots,b_{j-1}))\in R$ and $t(a_0,\dots,a_{j-1})\in X$ (as X absorbs $\mathbf{A}$ with respect to t). It follows that $t(b_0,\dots,b_{j-1})\in X^+$ and that $X^+\triangleleft\mathbf{B}$ with respect to t.

The remaining two statements are proved in an identical way. □

The subalgebra of $\mathbf{A}$ generated by B can be obtained by applying term operations of $\mathbf{A}$ to elements of B. The following auxiliary lemma provides a single term for all subsets B.

Lemma 2.6. *Let $\mathbf{A}$ be a finite idempotent algebra. Then there exists an operation $s\in\mathrm{Clo}(\mathbf{A})$ such that for any $B\subseteq A$ and any $b\in\mathrm{Sg}_{\mathbf{A}}(B)$ there exists $a_0,\dots,a_{\mathrm{ar}(s)-1}\in B$ such that $s(a_0,\dots,a_{\mathrm{ar}(s)-1})=b$.*

Proof. From the definition of $\mathrm{Sg}_{\mathbf{A}}(B)$ it follows that for every $B\subseteq A$ and every $b\in\mathrm{Sg}_{\mathbf{A}}(B)$ there exists an operation $s_{(B,b)}\in\mathrm{Clo}(\mathbf{A})$ of arity n and elements $a_0,\dots,a_{n-1}\in B$ such that $s_{(B,b)}(a_0,\dots,a_{n-1})=b$. This operation is idempotent, as $\mathbf{A}$ is.

For any two idempotent operations t_1,t_2 on $\mathbf{A}$ (of arities n_1, n_2) and any $a_0,\dots,a_{n_1-1}$, $b_0,\dots,b_{n_2-1}\in A$ we have

$$t_1*t_2(\underbrace{a_0,\dots,a_0}_{n_2},\underbrace{a_1,\dots,a_1}_{n_2},\dots,\underbrace{a_{n_1-1},\dots,a_{n_1-1}}_{n_2})$$

equal to $t_1(a_0,\dots,a_{n_1-1})$ and

$$t_1*t_2(b_0,b_1,\dots,b_{n_2-1},\dots,b_0,b_1,\dots,b_{n_2-1})$$

equal to $t_2(b_0,\dots,b_{n_2-1})$. Therefore the term operation

$$s=s_{(B_1,b_1)}*s_{(B_2,b_2)}*\dots*s_{(B_l,b_l)},$$

where $(B_1,b_1),(B_2,b_2),\dots,(B_l,b_l)$ is a complete list of pairs such that $b_i\in\mathrm{Sg}_{\mathbf{A}}(B_i)$, satisfies the conclusion of the lemma. □

The following proposition is the only place in this article, where we use a Taylor term. Although the proof is quite easy, we believe that this proposition is of an independent interest.

Proposition 2.7. *Let $\mathbf{A}$ be a finite algebra in a Taylor variety and suppose that $\mathbf{A}$ has no proper absorbing subalgebra. Then there exists an operation $v\in\mathrm{Clo}(\mathbf{A})$ such that for any $b,c\in A$ and any coordinate $i<\mathrm{ar}(v)$ there exist $a_0,\dots,a_{\mathrm{ar}(v)-1}\in A$ such that $a_i=b$ and $v(a_0,\dots,a_{\mathrm{ar}(v)-1})=c$.*

Proof. For a term operation $t\in\mathrm{Clo}(\mathbf{A})$ of arity k, an element $b\in A$, and a coordinate $i<\mathrm{ar}(t)$ we set

$$W(t,b,i)=\{t(a_0,\dots,a_{k-1}):a_i=b\text{ and }\ a_j\in A\ \forall j\}.$$

Our aim is to find a term v such that $W(v,b,i)=A$ for any $b\in A$ and any coordinate i. We will achieve this goal by gradually enlarging the sets $W(t,b,i)$.

Let $n < |A|$ and assume we already have an operation $v^{(n)} \in \mathrm{Clo}(\mathbf{A})$ such that each $W(v^{(n)}, b, i)$ contains a subuniverse of $\mathbf{A}$ with at least n elements. From idempotency it follows that all the one-element subsets of A are subuniverses of $\mathbf{A}$, thus any operation in $\mathrm{Clo}(\mathbf{A})$ can be taken as $v^{(1)}$.

For an induction step we first find an operation $w^{(n+1)} \in \mathrm{Clo}(\mathbf{A})$ such that each $W(w^{(n+1)}, b, i)$ has at least $(n+1)$-elements:

Claim 2.8. Let $t \in \mathrm{Clo}(\mathbf{A})$ be a Taylor term operation and put $w^{(n+1)} = t * v^{(n)}$. Then $|W(w^{(n+1)}, b, i)| > n$ for all $b \in A$ and all coordinates $i < \mathrm{ar}(w^{(n+1)})$.

Proof. Let $j = i$ div $\mathrm{ar}(t)$, $k = i$ mod $\mathrm{ar}(t)$ and let $B \subseteq W(v^{(n)}, b, k)$ be a subuniverse of $\mathbf{A}$ with $|B| \geq n$.

First we observe that $B \subseteq W(w^{(n+1)}, b, i)$. Indeed, take an arbitrary element $c \in B$, and find a tuple $a_0, \ldots, a_{\mathrm{ar}(v^{(n)})-1} \in A$ such that $a_k = b$ and that $v^{(n)}(a_0, \ldots, a_{\mathrm{ar}(v^{(n)})-1}) = c$. The application of $t * v^{(n)}$ to a concatenation of $\mathrm{ar}(t)$-many copies of $(a_0, \ldots, a_{\mathrm{ar}(v^{(n)})-1})$ produces $t(c, c, \ldots, c) = c$. Since on the i-th coordinate of this catenation we have b, we showed that $c \in W(w^{(n+1)}, b, i)$. Therefore if $B = A$ the claim holds and we can assume $B \subsetneq A$.

As t is a Taylor operation, it satisfies an identity of the form

$$t(\square_0, \square_1, \ldots, \square_{m-1}) \approx t(\triangle_0, \triangle_1, \ldots, \triangle_{m-1}),$$

where all $\square_l$'s and $\triangle_l$'s are substituted with either x or y, but $\square_j$ is x while $\triangle_j$ is y.

Let $r(x, y) = t(\square_0, \square_1, \ldots, \square_{m-1})$. Clearly $r \in \mathrm{Clo}(\mathbf{A})$. Since $\mathbf{A}$ has no proper absorbing subuniverses, the subuniverse B is not an absorbing subuniverse of $\mathbf{A}$ with respect to the operation r. Therefore there exist $c \in B$ and $d \in A$ such that either $r(c, d) \notin B$ or $r(d, c) \notin B$. We will show that $r(c, d), r(d, c) \in W(w^{(n+1)}, b, i)$.

For each $e \in \{r(c, d), r(d, c)\}$ we can find a tuple $f_0, \ldots, f_{\mathrm{ar}(t)-1} \in \{c, d\}$ such that $f_j = c$ and that $t(f_0, \ldots, f_{\mathrm{ar}(t)-1}) = e$. To obtain this we put

- $f_l = c$ if $\square_l = x$, and $f_l = d$ if $\square_l = y$ in the case that $e = r(c, d)$ and
- $f_l = c$ if $\triangle_l = x$, and $f_l = d$ if $\triangle_l = y$ in the case that $e = r(d, c)$.

Further, since $c \in B \subseteq W(v^{(n)}, b, k)$, we can find elements $a_0, \ldots, a_{\mathrm{ar}(v^{(n)})-1} \in A$ such that $a_k = b$ and $v^{(n)}(a_0, \ldots, a_{\mathrm{ar}(v^{(n)})-1}) = c$. To construct the argument for $t * v^{(n)}$ we expand each element of the tuple $(f_0, \ldots, f_{\mathrm{ar}(t)-1})$ into $\mathrm{ar}(v^{(n)})$-many identical copies of itself except f_j which is substituted by $(a_0, \ldots, a_{\mathrm{ar}(v^{(n)})-1})$. It is easy to verify that $t * v^{(n)}$ applied to such an argument produces e.

We have proved that $B \cup \{r(c, d), r(d, c)\} \subseteq W(w^{(n+1)}, b, i)$. As $|B| \geq n$ and $r(c, d) \notin B$ or $r(d, c) \notin B$, we are done $\square$

Now we are ready to define an operation $v^{(n+1)}$ such that each $W(v^{(n+1)}, b, i)$ contains a subuniverse with at least $(n+1)$ elements:

Claim 2.9. Let s be the operation from Lemma 2.6 and let $v^{(n+1)} = s * w^{(n+1)}$. Then, for all $b \in A$ and all coordinates $i < \mathrm{ar}(v^{(n+1)})$, $W(v^{(n+1)}, b, i)$ contains a subuniverse with more than n elements.

Proof. Let $j = i$ div $\mathrm{ar}(t)$, $k = i$ mod $\mathrm{ar}(t)$ and let $B = W(w^{(n+1)}, b, k)$. We will show that $\mathrm{Sg}_{\mathbf{A}}(B) \subseteq W(v^{(n+1)}, b, i)$.

Choose an arbitrary $c \in \mathrm{Sg}_{\mathbf{A}}(B)$. By Lemma 2.6, there exist $f_0, \dots, f_{\mathrm{ar}(s)-1} \in B$ such that $s(f_0, \dots, f_{\mathrm{ar}(s)-1}) = c$. As before we prepare the tuple of arguments for $s * w^{(n)}$ by expanding the tuple $(f_0, \dots, f_{\mathrm{ar}(s)-1})$. Each f_i gets expanded into $\mathrm{ar}(w^{(n+1)})$-many identical copies of itself, except f_j which gets expanded into a tuple $(a_0, \dots, a_{\mathrm{ar}(w^{(n+1)})-1}) \in A$ with $a_k = b$ and such that $w^{(n+1)}(a_0, \dots, a_{\mathrm{ar}(w^{(n+1)})-1}) = f_j$ (such a tuple exists as $f_j \in B$). It is clear that $s * w^{(n+1)}$ applied to such a tuple produces c and the claim is proved. □

To finish the proof of Proposition 2.7, it is enough to set $v = v^{(|A|)}$. □

It is an easy corollary that for two (or any finite number of) algebras in a Taylor variety we can find a common term satisfying the conclusion of Proposition 2.7.

Corollary 2.10. *Let $\mathbf{A}, \mathbf{B}$ be finite algebras in a Taylor variety without proper absorbing subalgebras. Then there exists a term v such that for any $b, c \in A$ (resp. $b, c \in B$) and any coordinate $j < \mathrm{ar}(v)$ there exist $a_0, \dots, a_{\mathrm{ar}(v)} \in A$ (resp. $a_0, \dots, a_{\mathrm{ar}(v)} \in B$) such that $a_j = b$ and $v(a_0, \dots, a_{\mathrm{ar}(v)}) = c$.*

Proof. If v_1 (resp. v_2) is the term obtained from Proposition 2.7 for the algebra $\mathbf{A}$ (resp. $\mathbf{B}$), then we can put $v = v_1 * v_2$. □

We are now ready to prove Theorem 2.3. One direction of the proof is straightforward: if an idempotent variety $\mathcal{V}$ is not a Taylor variety, then, by Theorem 1.2, it contains a two-element algebra whose every operation is a projection. Such an algebra has no absorbing subuniverses and any three-element subset of its square is a linked subdirect subalgebra which falsifies the second condition of Theorem 2.3. Therefore it remains to prove the following.

Theorem 2.11. *Let $\mathbf{A}, \mathbf{B}$ be finite algebras in a Taylor variety and let R be a proper, subdirect and linked subalgebra of $\mathbf{A} \times \mathbf{B}$. Then $\mathbf{A}$ or $\mathbf{B}$ has a proper absorbing subalgebra.*

Proof. For contradiction, assume that $R, \mathbf{A}, \mathbf{B}$ form a counterexample to the theorem. Thus neither $\mathbf{A}$ nor $\mathbf{B}$ has a proper absorbing subalgebra and $R \leq_S \mathbf{A} \times \mathbf{B}$ is a linked, proper subset of $A \times B$.

First we find another counterexample satisfying $R^{-1} \circ R = A \times A$. As R is linked, there exists a natural number k such that $(R^{-1} \circ R)^{\circ k} = A^2$. Take the smallest such k. If $k = 1$, then $R^{-1} \circ R = A \times A$ and we need not to do anything. Otherwise we replace $\mathbf{B}$ by $\mathbf{A}$ and R by $(R^{-1} \circ R)^{\circ(k-1)}$. Our new choice of $R, \mathbf{A}, \mathbf{B}$ is clearly a counterexample to the theorem satisfying $R^{-1} \circ R = A \times A$.

From now on we assume that our counterexample satisfies $R^{-1} \circ R = A \times A$. In other words, for any $a, c \in A$, there exists $b \in B$ such that $(a, b), (c, b) \in R$.

For a $X \subseteq A$ we set

$$N(X) \quad = \quad \{b \in B : \forall\, a \in X \quad (a, b) \in R\} = \bigcap_{a \in X} \{a\}^+$$

Claim 2.12. $N(X) = N(\mathrm{Sg}_{\mathbf{A}}(X))$.

Proof. If t is a k-ary term, $a_0, \dots, a_{k-1}$ are elements of X and $b \in N(X)$, then $(a_i, b) \in R$ for any $i = 0, 1, \dots, k-1$. Therefore $(t(a_0, \dots, a_{k-1}), b) \in R$. This shows that $b \in \{t(a_0, \dots, a_{k-1})\}^+$. □

Claim 2.13. $N(A) \neq \emptyset$.

Proof. We call a subset $X \subseteq A$ *good*, if $(N(X))^- = A$. Since $R^{-1} \circ R = A \times A$, every one-element subset of A is good. We prove the claim by showing that A is good.

Let X be a maximal, with respect to inclusion, good subset of A. We know that $\emptyset \neq X$, since each one-element subset is good, and also $X \neq A$, otherwise the claim is proved. As $N(X) = N(\mathrm{Sg}_{\mathbf{A}}(X))$ due to the Claim 2.12, X is a subuniverse of $\mathbf{A}$. Let $v \in \mathrm{Clo}(\mathbf{A})$ be the operation from Proposition 2.7. Due to our assumption that $\mathbf{A}$ has no proper absorbing subuniverses, X is not an absorbing subuniverse of $\mathbf{A}$ with respect to the operation v. It follows that there exists a coordinate $j < \mathrm{ar}(v)$ and elements $a_0, \dots, a_{\mathrm{ar}(v)-1} \in A$ such that $a_i \in X$ for all $i \neq j$, and $b := v(a_0, \dots, a_{\mathrm{ar}(v)-1}) \notin X$.

We will prove that the set $X \cup \{b\}$ is good, which will contradict the maximality of X. Let $c \in A$ be arbitrary. From Proposition 2.7 we obtain $d_0, \dots, d_{\mathrm{ar}(v)-1} \in A$ such that $d_j = a_j$ and $v(d_0, \dots, d_{\mathrm{ar}(v)-1}) = c$. Since $(N(X))^- = A$, we can find $e_0, \dots, e_{\mathrm{ar}(v)-1} \in N(X)$ such that $(d_i, e_i) \in R$ for all i. Put $f = v(e_0, \dots, e_{\mathrm{ar}(v)-1})$. As R is a subuniverse of $\mathbf{A} \times \mathbf{B}$ and $(d_i, e_i) \in R$ for all i, it follows that $(v(d_0, \dots, d_{\mathrm{ar}(v)-1}), v(e_0, \dots, e_{\mathrm{ar}(v)-1})) = (c, f) \in R$. The set $N(X)$ is a subuniverse of $\mathbf{B}$ thus we have $f \in N(X)$. For all $i \neq j$, we have $a_j \in X$ and $e_j \in N(X)$, hence $(a_j, e_j) \in R$. But also $(a_i = d_i, e_i) \in R$ and, again, R is a subuniverse of $\mathbf{A} \times \mathbf{B}$, therefore $(v(a_0, \dots, a_{\mathrm{ar}(v)-1}), v(e_0, \dots, e_{\mathrm{ar}(v)-1})) = (b, f) \in R$. We have proved that, for any $c \in A$, there exists $f \in N(X) \cap \{b\}^+ = N(X \cup \{b\})$ such that $(c, f) \in R$. Therefore $X \cup \{b\}$ is good, a contradiction. This contradiction shows that $N(A)$ is nonempty. □

Since R is a proper subset of $A \times B$, $N(A)$ is a proper subset of B. This set is an intersection of subuniverses of $\mathbf{B}$, thus $N(A)$ a subuniverse of $\mathbf{B}$. Since $N(A)$ is not an absorbing subuniverse of $\mathbf{B}$ with respect to v, there exists a coordinate $j < \mathrm{ar}(v)$ and a tuple $b_0, \dots, b_{\mathrm{ar}(v)-1} \in B$ such that $b_i \in N(A)$ for all $i \neq j$, and $c := v(b_0, \dots, b_{\mathrm{ar}(v)-1}) \notin N(A)$.

We will prove that $(d, c) \in R$ for all $d \in A$, which will contradict the definition of $N(A)$. Let $a \in A$ be any element of A such that $(a, b_j) \in R$ (we use subdirectness of R here) and let $a_i \in A$ be obtained from Proposition 2.7 in such a way that $a_j = a$ and $v(a_0, \dots, a_{\mathrm{ar}(v)}) = d$. For all $i \neq j$, we have $(a_i, b_i) \in R$ as $b_i \in N(A)$, and also $(a_j = a, b_j) \in R$. Thus $(v(a_0, \dots, a_{\mathrm{ar}(v)-1}), v(b_0, \dots, b_{\mathrm{ar}(v)-1})) = (d, c) \in R$. □

2.3. **Minimal absorbing subalgebras.**

We present a number of properties of absorbing subuniverses required in the proof of Theorem 4.1. Most of them are corollaries of the Absorption Theorem and they give us some information about minimal absorbing subalgebras:

Definition 2.14. If $\mathbf{B} \triangleleft \mathbf{A}$ and no proper subalgebra of $\mathbf{B}$ absorbs $\mathbf{A}$, we call $\mathbf{B}$ a *minimal absorbing subalgebra* of $\mathbf{A}$ (and write $\mathbf{B} \triangleleft\!\triangleleft\, \mathbf{A}$).

Alternatively, we can say that $\mathbf{B}$ is a minimal absorbing subalgebra of $\mathbf{A}$, if $\mathbf{B} \triangleleft \mathbf{A}$ and $\mathbf{B}$ has no proper absorbing subalgebras. Equivalence of these definitions follows from transitivity of $\triangleleft$ (proved in Proposition 2.4). Observe also that two minimal absorbing subuniverses of $\mathbf{A}$ are either disjoint or coincide, but the union of all minimal absorbing subuniverses need not be the whole set A.

Proposition 2.15. *Let $\mathcal{V}$ be a Taylor variety, let $\mathbf{A}$ and $\mathbf{B}$ be finite algebras in $\mathcal{V}$ and let $\mathbf{R} \leq_S \mathbf{A} \times \mathbf{B}$.*

(i) *If R is linked and $\mathbf{E} \triangleleft \mathbf{R}$, then E is linked.*

(ii) *If* $\mathbf{C} \lhd\lhd \mathbf{A}$*,* $\mathbf{D} \lhd\lhd \mathbf{B}$*, and* $(C \times D) \cap R \neq \emptyset$*, then* $(\mathbf{C} \times \mathbf{D}) \cap R \leq_S \mathbf{C} \times \mathbf{D}$*.*
(iii) *If* R *is linked,* $\mathbf{C} \lhd\lhd \mathbf{A}$*,* $\mathbf{D} \lhd\lhd \mathbf{B}$*, and* $(C \times D) \cap R \neq \emptyset$*, then* $\mathbf{C} \times \mathbf{D} \lhd\lhd \mathbf{R}$*.*
(iv) *If* R *is linked, and* $\mathbf{C} \lhd\lhd \mathbf{A}$*, then there exists* $\mathbf{D} \lhd\lhd \mathbf{B}$ *such that* $C \times D \subseteq R$*.*
(v) *If* R *is linked,* $\mathbf{C} \lhd\lhd \mathbf{A}$ *or* $\mathbf{C} \lhd\lhd \mathbf{B}$*,* $\mathbf{D} \lhd\lhd \mathbf{A}$ *or* $\mathbf{D} \lhd\lhd \mathbf{B}$*,* $c \in C$*, and* $d \in D$*, then* c *and* d *can be linked via* $c_0, \dots, c_j$ *where each* c_i *is a member of some minimal absorbing subalgebra of* $\mathbf{A}$ *or* $\mathbf{B}$*.*

To avoid ambiguity in the statement of item (v), assume that the algebras $\mathbf{A}, \mathbf{B}$ are disjoint. When we apply the corollary this need not be the case, but the assumptions (and therefore conclusions) of the corollary will be satisfied when we substitute the algebras $\mathbf{A}, \mathbf{B}$ with their isomorphic, disjoint copies.

Proof.

(i) Suppose that $\mathbf{E}$ absorbs $\mathbf{R}$ with respect to an operation t. Let $(a, b), (a', b')$ be arbitrary elements of E. As R is linked, there exist $c_0, c_1, \dots, c_{2n} \in A \cup B$ such that $c_0 = a$, $c_{2n} = a'$, $(c_{2i}, c_{2i+1}) \in R$ and $(c_{2i+2}, c_{2i+1}) \in R$ for all $i = 0, 1, \dots, n-1$. The pair
$$t((c_{2i}, c_{2i+1}), (a, b), (a, b), \dots, (a, b)),$$
which is, by definition of the product of two algebras, equal to
$$(t(c_{2i}, a, a, \dots, a), t(c_{2i+1}, b, b, \dots, b))$$
is in E for all i, since E absorbs R with respect to t. Similarly,
$$(t(c_{2i+2}, a, a, \dots, a), t(c_{2i+1}, b, b, \dots, b)) \in E.$$
Therefore the elements $a = t(a, a, \dots, a)$ and $t(a', a, a, \dots a)$ are linked in E via $t(c_0, a, \dots, a)$, $t(c_1, b, \dots, b)$, $\dots$, $t(c_{2n}, a, \dots, a)$.

Using the same reasoning, the pairs
$$(t(a', c_{2i}, a, \dots, a), t(b', c_{2i+1}, b, \dots, b))$$
and
$$(t(a', c_{2i+2}, a, \dots, a), t(b', c_{2i+1}, b, \dots, b))$$
are in E and it follows that $t(a', a, a, \dots, a)$ and $t(a', a', a, a, \dots, a)$ are linked in E. By continuing similarly we get that $a = t(a, a, \dots, a)$ and $a' = t(a', a', \dots, a')$ are linked in E as required.

(ii) By Lemma 2.5 $\mathbf{D}^- \lhd \mathbf{A}$, therefore $\emptyset \neq (\mathbf{D}^- \cap \mathbf{C}) \lhd \mathbf{A}$ (by Proposition 2.4) and, as $\mathbf{C} \lhd\lhd \mathbf{A}$, we get $\mathbf{D}^- \supseteq \mathbf{C}$. A symmetric reasoning shows that $\mathbf{C}^+ \supseteq \mathbf{D}$ and the item is proved.

(iii) Let $E = (C \times D) \cap R$ and let $\mathbf{E}$ be the subalgebra of $\mathbf{A} \times \mathbf{B}$ with universe E. From (ii) it follows that $E \leq_S \mathbf{C} \times \mathbf{D}$. Clearly $E \lhd \mathbf{R}$, therefore E is linked by (i). Theorem 2.11 together with the minimality of $\mathbf{C}$ and $\mathbf{D}$ now gives $E = C \times D$.

Let $\emptyset \neq F \lhd \mathbf{E}$. The projection of F to the first (resp. the second) coordinate is clearly an absorbing subuniverse of $\mathbf{C}$ (resp. $\mathbf{D}$). Therefore $F \leq_S \mathbf{C} \times \mathbf{D}$. Using (i) and Theorem 2.11 as above we conclude that $F = C \times D$.

(iv) Let $D' = C^+$. According to Lemma 2.5, D' is an absorbing subuniverse of $\mathbf{B}$. Let $\mathbf{D}'$ be the subalgebra of $\mathbf{A}$ with universe D' and let $\mathbf{D}$ be a minimal absorbing subalgebra of $\mathbf{D}'$. The claim now follows from (iii).

(v) We prove this fact by induction on the length of the path connecting c and d. If the length is 2, then we have $c, d \in A$ (thus $\{c\}^+ \cap \{d\}^+ \neq \emptyset$), or $c, d \in B$ (thus $\{c\}^- \cap \{d\}^- \neq \emptyset$). Without loss of generality we assume the first case and, conclude using Lemma 2.5 and Proposition 2.4, that $\emptyset \neq (\mathbf{C}^+ \cap \mathbf{D}^+) \triangleleft \mathbf{B}$. Let E be any subuniverse such that $E \lll (\mathbf{C}^+ \cap \mathbf{D}^+)$. Then, as $(C \times E) \cap R \neq \emptyset$ and $(D \times E) \cap R \neq \emptyset$, by (iii), we obtain $C \times E \subseteq R$ and $D \times E \subseteq R$ and the first case is proved.

For the induction step, we assume, without loss of generality, that $\mathbf{C} \lll \mathbf{A}$ and define $C_0 = C, C_1 = C_0^+, C_2 = C_1^-, C_3 = C_2^+, \ldots$ with $d \in C_n$. Suppose, for simplicity of the presentation, that d appears on the right side (i.e. $d \in B$) and consider $(C_{n-1} \cap \mathbf{D}^-) \triangleleft \mathbf{A}$. Let $\mathbf{E} \triangleleft\triangleleft (C_{n-1} \cap \mathbf{D}^-)$. By (iii) we have $E \times D \subseteq R$ and, by inductive assumption we have an element of E, say e, linked inside minimal absorbing subuniverses to some element of $\mathbf{C}$ say c'. Therefore d is linked (through e) inside minimal sets to some $c' \in \mathbf{C}$. By (iv) we link, inside minimal absorbing subuniverses, c' to c and the item is proved.

$\square$

3. New proof of the Smooth Theorem

The Smooth Theorem classifies the computational complexity of CSPs generated by smooth digraphs (digraphs, where every vertex has at least one incoming and at least one outgoing edge). This classification was conjectured by Bang-Jensen and Hell [BJH90] and confirmed by the authors in [BKN08, BKN09]. The proof presented in those papers heavily relied on the results of McKenzie and Maroti [MM08] which characterized the locally finite Taylor varieties in terms of weak near-unanimity operations. We present an alternative proof which depends only on Theorem 2.3. The Smooth Theorem states:

Theorem 3.1. *Let $\mathbb{H}$ be a smooth digraph. If each component of the core of $\mathbb{H}$ is a circle, then* CSP($\mathbb{H}$) *is polynomially decidable. Otherwise* CSP($\mathbb{H}$) *is NP-complete.*

3.1. **Basic digraph notions.** A *digraph* is a pair $\mathbb{G} = (V, E)$, where V is a finite set of vertices and $E \subseteq V \times V$ is a set of edges. If the digraph is fixed we write $a \to b$ instead of $(a, b) \in E$. The induced subgraph of $\mathbb{G}$ with vertex set $W \subseteq V$ is denoted by $\mathbb{G}_{|W}$, that is, $\mathbb{G}_{|W} = (W, E \cap (W \times W))$. A *loop* is an edge of the form (a, a). $\mathbb{G}$ is said to be *smooth* if every vertex has an incoming and an outgoing edge, in other words, $\mathbb{G}$ is smooth, if E is a subdirect product of V and V. The *smooth part* of $\mathbb{G}$ is the largest subset W of V such that $\mathbb{G}_{|W}$ is smooth (it can be empty).

An *oriented path* is a digraph $\mathbb{P}$ with vertex set $P = \{p_0, \ldots, p_k\}$ and edge set consisting of k edges — for all $i < k$ either (p_i, p_{i+1}), or (p_{i+1}, p_i) is an edge of $\mathbb{P}$. An *initial segment* of such a path is any path induced by $\mathbb{P}$ on vertices $\{p_0, \ldots, p_i\}$ for some $i < k$. We denote the oriented path consisting of k edges pointing forward by $\cdot \xrightarrow{k} \cdot$ and, similarly the oriented path consisting of k edges pointing backwards by $\cdot \xleftarrow{k} \cdot$. The concatenation of paths is performed in the natural way. A (k, n)-*fence* (denoted by $\mathbb{F}[k, n]$) is the oriented path consisting of $2kn$ edges, k forward edges followed by k backward edges, n times i.e.:

$$\underbrace{\cdot \xrightarrow{k} \cdot \xleftarrow{k} \cdot \cdots \cdot \xrightarrow{k} \cdot \xleftarrow{k} \cdot}_{n}$$

The *algebraic length* of an oriented path is the number of forward edges minus the number of backward edges (and thus all the fences have algebraic length zero). Let $\mathbb{G}$ be a digraph, let $\mathbb{P}$ be an oriented path with vertex set $P = \{p_0, \dots, p_k\}$, and let a, b be vertices of $\mathbb{G}$. We say that a is *connected to* b via $\mathbb{P}$, if there exists a homomorphism $f : \mathbb{P} \to \mathbb{G}$ such that $f(p_0) = a$ and $f(p_k) = b$. We sometimes write $a \xrightarrow{k} b$ when a is connected to b via $\cdot \xrightarrow{k} \cdot$. If $a \xrightarrow{k} a$ (for some k) then a is *in a cycle* and any image of the path $\cdot \xrightarrow{k} \cdot$ with the same initial and final vertex is *a cycle*. *A circle* is a cycle which has no repeating vertices and no chords.

The relation "a is connected to b (via some path)" is an equivalence, its blocks (or sometimes the corresponding induced subdigraphs) are called the *weak components* of $\mathbb{G}$. The vertices a and b are in the same *strong component* if $a \xrightarrow{k} b \xrightarrow{k'} a$ for some k, k'. For a subset B of A and an oriented path $\mathbb{P}$ we set

$$B^{\mathbb{P}} = \{c : \exists b \in B \quad b \text{ is connected to } c \text{ via } \mathbb{P} \ \}.$$

Note that $B^{\cdot \xrightarrow{k} \cdot}$ is formally equal to $B^{+E^{\circ k}}$ but we prefer the first notation.

Finally, $\mathbb{G}$ has *algebraic length* k, if there exists a vertex a of $\mathbb{G}$ such that a is connected to a via a path of algebraic length k and k is the minimal positive number with this property. The following proposition summarizes easy results concerning reachability via paths:

Proposition 3.2. *Let $\mathbb{G}$ be a smooth digraph, then:*

- *for any vertices a, b in $\mathbb{G}$ if a is connected to b via $\cdot \xrightarrow{k} \cdot$ then a is connected to b via every path of algebraic length k;*
- *for any vertex a and any path $\mathbb{P}$ there exists a vertex b and a path $\mathbb{Q}$ which is an initial segment of some fence such that $\{a\}^{\mathbb{P}} \subseteq \{b\}^{\mathbb{Q}}$;*
- *if $H \subseteq G$ is such that $H^{\cdot \to \cdot} \supseteq H$ or $H^{\cdot \leftarrow \cdot} \supseteq H$ then the digraph $\mathbb{G}_{|H}$ contains a cycle (i.e.* the smooth part *of $\mathbb{G}_{|H}$ is non-empty)*

Proof. The first item of the proposition follows directly form the definition of a smooth digraph.

We prove the second item by induction on the length of $\mathbb{P}$. If the length is zero there is nothing to prove. Therefore we take an arbitrary path $\mathbb{P}$ of length n and arbitrary $a \in A$. The proof splits into two cases depending on the direction of the last edge in $\mathbb{P}$. We consider the case when the last edge of $\mathbb{P}$ points forward first and set $\mathbb{P}'$ to be $\mathbb{P}$ take away the last edge. The inductive assumption for a and $\mathbb{P}'$ provides a vertex b and a path $\mathbb{Q}'$ (an initial fragment of a fence $\mathbb{F}[k, l]$). If the algebraic length of $\mathbb{Q}'$ is strictly smaller than k, we put $\mathbb{Q}'''$ to be a path such that the concatenation of $\mathbb{Q}'$ and $\mathbb{Q}'''$ is an initial fragment of the fence $\mathbb{F}[k, l+1]$ and such that the algebraic length of $\mathbb{Q}'''$ is one; then the concatenation of $\mathbb{Q}'$ and $\mathbb{Q}'''$ proves the second item of the proposition (as, by the first item of the proposition, every element reachable from $\{b\}^{\mathbb{Q}'}$ by $\cdot \to \cdot$ is also reachable by $\mathbb{Q}'''$). If the algebraic length of $\mathbb{Q}'$ equals k we consider a path $\mathbb{Q}''$ obtained from $\mathbb{Q}'$ by substituting each subpath of the shape $\cdot \to \cdot \leftarrow \cdot$ with $\cdot \xrightarrow{2} \cdot \xleftarrow{2} \cdot$. The path $\mathbb{Q}''$ is an initial fragment of $\mathbb{F}[k+1, l]$ and we have $\{b\}^{\mathbb{Q}'} \subseteq \{b\}^{\mathbb{Q}''}$ (as the digraph is smooth). Now we can find $\mathbb{Q}'''$ as in the previous case.

If the last edge of $\mathbb{P}$ points backwards, we proceed with dual reasoning. If the algebraic length of $\mathbb{Q}'$ is greater than zero we obtain $\mathbb{Q}'''$ of algebraic length -1 as before and the proposition is proved. If the algebraic length of $\mathbb{Q}'$ is zero we substitute b with any vertex

b' such that $b' \to b$ and alter $\mathbb{Q}'$ by substituting each $\cdot \leftarrow \cdot \to \cdot$ with $\cdot \overset{2}{\leftarrow} \cdot \overset{2}{\to} \cdot$. The new path is an initial fragment of $\mathbb{F}[k+1, l]$ and we can proceed as in previous case.

For the third item of the proposition. Without loss of generality we can assume the first possibility and choose an arbitrary $b_0 \in H$. As $H \subseteq H^{\cdot \to \cdot}$ there is an element $b_1 \in H$ such that $b_1 \to b_0$. Repeating the same reasoning for $b_1, b_2, \ldots$ we obtain a sequence of vertices in H such that $b_{i+1} \to b_i$. As H is finite, we obtain a cycle in H and the last item of the proposition is proved. $\square$

The following lemma shows that the smooth part of an induced subdigraph of a smooth digraph shares some algebraic properties with the induced subdigraph.

Lemma 3.3. *Let $\mathbf{A}$ be a finite algebra and let $\mathbb{G} = (A, E)$ be a smooth digraph such that E is a subuniverse of $\mathbf{A}^2$. If B is a subuniverse of $\mathbf{A}$ (an absorbing subuniverse of $\mathbf{A}$) then the smooth part of $\mathbb{G}_{|B}$ forms a subuniverse of $\mathbf{A}$ (an absorbing subuniverse of $\mathbf{A}$ respectively).*

Proof. Note that if the smooth part of $\mathbb{G}_{|B}$ is empty then the lemma holds. Assume it is non-empty and let $\mathbf{A}$, $\mathbb{G}$, B be as in the statement of the lemma. We put $B_1 \subseteq B$ to be the set of all the vertices in B with at least one outgoing and at least one incoming edge in $\mathbb{G}_{|B}$ (i.e. an outgoing edge and an incoming edge to elements of B). As $B_1 = B \cap B^{+E} \cap B^{-E}$ Lemma 2.5 implies that B_1 is a subuniverse (absorbing subuniverse resp.) of $\mathbf{A}$. We put $B_2 = B_1 \cap B_1^{+E} \cap B_1^{-E}$ and continue the reasoning. Since $\mathbf{A}$ is finite we obtain some k such that $B_k = B_{k+1}$. Since $\mathbb{G}_{|B_k}$ has no sources and no sinks the lemma is proved. $\square$

3.2. **Reduction of the problem.** The first part of Theorem 3.1 is easy: if a digraph $\mathbb{H}$ has a core which is a disjoint union of circles then CSP($\mathbb{H}$) is solvable in polynomial time (see [BJH90]). On the other hand, using Theorem 1.4 and the fact that CSPs of a relational structure and its core are the same, it suffices to prove that:

Theorem 3.4. *If a smooth digraph admits a Taylor polymorphism then it retracts onto the disjoint union of circles.*

Finally, Theorem 3.4 reduces to the theorem below. An elementary proof of this reduction can be found in [BKN08, BKN09].

Theorem 3.5. *If a smooth digraph has algebraic length one and admits a Taylor polymorphism then it contains a loop.*

In fact, in the remainder of this section, we prove a stronger version of Theorem 3.5:

Theorem 3.6. *Let $\mathbf{A}$ be a finite algebra in a Taylor variety and let $\mathbb{G} = (A, E)$ be a smooth digraph of algebraic length one such that E is a subuniverse of $\mathbf{A}^2$. Then $\mathbb{G}$ contains a loop. Moreover, if there exists an absorbing subuniverse I of $\mathbf{A}$ which is contained in a weak component of $\mathbb{G}$ of algebraic length 1, then the loop can be found in some J such that $J \lhd\!\lhd\ \mathbf{A}$.*

3.3. **The proof.** Our proof of Theorem 3.6 proceeds by induction on the size of the vertex set of $\mathbb{G} = (A, E)$. If $|A| = 1$ there is nothing to prove (as the only smooth digraph on such a set contains a loop); for the induction step we assume that Theorem 3.6 holds for all smaller digraphs.

Claim 3.7. Let H be a weak component of $\mathbb{G}$ of algebraic length one, then there exists $a \in H$ and a path $\mathbb{P}$ such that $\{a\}^{\mathbb{P}}$ contains a cycle.

Proof. We choose $a \in H$ to be the element of the component H such that there is a path $\mathbb{Q}$ of algebraic length one connecting a to a. We define the sequence of sets $B_0 = \{a\}$ and $B_i = B_{i-1}^{\mathbb{Q}}$ recursively. As a is connected to a via $\mathbb{Q}$ we have $B_0 \subseteq B_1$ and therefore $B_i \subseteq B_{i+1}$ for any i (as by definition $B_{i-1} \subseteq B_i$ implies that $B_{i-1}^{\mathbb{Q}} \subseteq B_i^{\mathbb{Q}}$ i.e. $B_i \subseteq B_{i+1}$). As $\mathbb{Q}$ is of algebraic length one we can use Proposition 3.2 to infer that $\{a\}^{\cdot\rightarrow\cdot} \subseteq B_1$ and further that $\{a\}^{\cdot\xrightarrow{k}\cdot} \subseteq B_k$ for any k. These facts together imply that

$$\bigcup_{i=0}^{k} \{a\}^{\cdot\xrightarrow{i}\cdot} \subseteq B_k$$

and, as the digraph is finite, we can find a cycle in one of the B_k's. Take $\mathbb{P}$ to be the $\mathbb{Q}$ concatenated with itself sufficiently many times to witness the claim. □

Claim 3.8. Let H be a weak component of $\mathbb{G}$ of algebraic length one, then there exists $a \in H$ and a fence $\mathbb{F}$ such that $\{a\}^{\mathbb{F}} = H$.

Proof. Let us choose $a \in H$ and $\mathbb{P}'$ as provided by Claim 3.7. Set B to be the set of elements of $\{a\}^{\mathbb{P}'}$ which belong to some cycle fully contained in $\{a\}^{\mathbb{P}'}$. Proposition 3.2 implies that $B^{\mathbb{F}[|A|,1]}$ contains all elements reachable by $\cdot \xrightarrow{i} \cdot$ or $\cdot \xleftarrow{i} \cdot$ (for any i), from any element of B. Indeed if such a c is reachable from $b \in B$ by $\cdot \xleftarrow{i} \cdot$ then it is reachable by $\cdot \xleftarrow{|A|} \cdot$ from some $b' \in B$ and further by $\mathbb{F}[|A|, 1]$ from some $b'' \in B$. In the other case $b \xrightarrow{i} c$ for some $b \in B$. There obviously exists d such that $d \xleftarrow{|A|} c$ and since $b \xrightarrow{i} c \xrightarrow{|A|} d$ we have some $j \leq |A|$ and $b \xrightarrow{j} d$. Thus there exists $b' \in B$ with $b' \xrightarrow{|A|} d$ and c is reachable by $\mathbb{F}[|A|, 1]$ from b'.

For every element c in H we can find $b_0, b_1, \ldots, b_{|A|} = c$ such that each b_j, $j \neq |A|$, is in a cycle B_j where $B_0 \subseteq B$, and $b_0 \xrightarrow{i_0} b_1 \xleftarrow{i_1} b_2 \xrightarrow{i_2} b_3 \leftarrow \ldots b_{|A|}$ for some $i_0, i_1, \ldots, i_{|A|-1}$. The reasoning above shows that B_j is contained in $B_{j-1}^{\mathbb{F}[|A|,1]}$ (for all $1 \leq j < |A|$) and $b_{|A|}$ belongs to $B_{|A|-1}^{\mathbb{F}[|A|,1]}$, therefore $B^{\mathbb{F}[|A|,|A|]} = H$.

Thus, for an appropriate path $\mathbb{P}$ we have a connected to every element of H by $\mathbb{P}$. The second item of Proposition 3.2 provides b and an initial segment $\mathbb{Q}$ of a fence $\mathbb{F}$ such that b is connected to every element from H by $\mathbb{Q}$. Let $\mathbb{S}$ denote the remaining part of the fence $\mathbb{F}$. Then $\{b\}^{\mathbb{F}} = (\{b\}^{\mathbb{Q}})^{\mathbb{S}} = H^{\mathbb{S}} = H$ and the claim is proved. □

The remaining part of the proof splits into two cases: in the first case the algebra $\mathbf{A}$ has an absorbing subuniverse in a weak component of algebraic length one and in the second it doesn't. Let us focus on the first case and define $I \lhd \mathbf{A}$ contained in a weak component (denoted by H) of algebraic length one of $\mathbb{G}$.

Claim 3.9. There is a fence $\mathbb{F}$ such that $I^{\mathbb{F}} = H$.

Proof. Let a and $\mathbb{F}'$ be provided by Claim 3.8. We put $\mathbb{F}$ to be a concatenation of $\mathbb{F}'$ with itself. Since $a \in I^{\mathbb{F}'}$, then $I^{\mathbb{F}} = H$. □

Let $\mathbb{P}$ be the longest initial segment of $\mathbb{F}$ (provided by Claim 3.9) such that $I^{\mathbb{P}} \neq H$. Put $S = I^{\mathbb{P}}$. By multiple application of Lemma 2.5 we infer that S is a subuniverse of $\mathbf{A}$ and that $S \lhd \mathbf{A}$. The definition of S implies that $S^{\cdot \to \cdot} = H \supseteq S$ or $S^{\cdot \leftarrow \cdot} = H \supseteq S$, and therefore, by Proposition 3.2, S contains a cycle. Thus the smooth part of $\mathbb{G}_{|S}$, denoted by S', is non-empty and, by Lemma 3.3, it absorbs $\mathbf{A}$. If the digraph $\mathbb{G}_{|S'}$ has algebraic length one and is weakly connected, then we use the inductive assumption:

- either $\mathbb{G}_{|S'}$ has no absorbing subuniverses in a weak component of algebraic length one; in such a case, as it is weakly connected, it has no absorbing subuniverses at all — therefore $S' \lhd\lhd \mathbf{A}$ and the inductive assumption provides a loop in S', or
- $\mathbb{G}_{|S'}$ has an absorbing subuniverse; then it has a loop in $J \lhd\lhd S'$ and, as $J \lhd\lhd \mathbf{A}$, the theorem is proved.

Therefore to conclude the first case of the theorem it remains to prove

Claim 3.10. $\mathbb{G}_{|S'}$ is a weakly connected digraph of algebraic length 1.

Proof. Assume that S' absorbs $\mathbf{A}$ with respect to t of arity k and let m, n be natural numbers such that every two vertices of H are connected via the (m,n)-fence (implied by Claim 3.8) denoted by $\mathbb{F}$. We will show that any two vertices $a, b \in S'$ are connected via the (m, nk)-fence in the digraph $\mathbb{G}_{|S'}$.

As the digraph $\mathbb{G}_{|S'}$ is smooth, a is connected to a via $\mathbb{F}$ and b is connected to b via $\mathbb{F}$ (by the first item of Proposition 3.2). Let $f : \mathbb{F} \to S'$ and $g : \mathbb{F} \to S'$ be the corresponding digraph homomorphisms. Moreover, a is connected to b via $\mathbb{F}$ in the digraph $\mathbb{G}$ and we take the corresponding homomorphism $h : \mathbb{F} \to \mathbb{G}$. For every $i = 0, 1, \ldots, k-1$ we consider the following matrix with k rows and $2nm+1$ columns: To the first $(k-i-1)$ rows we write f-images of the vertices of $\mathbb{F}$, to the $(k-i)$th row we write h-images, and to the last i rows we write g-images. We apply the term operation t to columns of this matrix. Since $E \leq \mathbf{A}^2$ we obtain a homomorphism from $\mathbb{F}$ to $\mathbb{G}$ which realizes a connection from

$$t(\underbrace{a, a, \ldots, a}_{(k-i)}, \underbrace{b, b, \ldots, b}_{i})$$

to

$$t(\underbrace{a, a, \ldots, a}_{(k-i-1)}, \underbrace{b, b, \ldots, b}_{(i+1)}).$$

Moreover, since all but one member of each column are elements of S' and $S' \lhd \mathbf{A}$, we actually get a homomorphism $\mathbb{F} \to S'$. By joining these homomorphisms for $i = 0, 1, \ldots, k-1$ we obtain that $a = t(a, a, \ldots, a)$ is connected to $b = t(b, b, \ldots, b)$ via the (m, nk)-fence in S'.

As $S' \subseteq H$ all the elements of S' are connected in H, and, using the paragraph above, also in S'. Moreover we can take two elements $a, b \in S'$ such that $a \to b$. As a is connected to b via a (m, nk)-fence in S' the algebraic length of $\mathbb{G}_{|S'}$ is one. □

It remains to prove the case of Theorem 3.6 when there is no absorbing subuniverse in any weak component of $\mathbb{G}$ of algebraic length one. We choose such a component and call it H. By Claim 3.8 there is an $a \in H$ and $\mathbb{F}$ such that $H = \{a\}^{\mathbb{F}}$. Since $\{a\}$ is a subuniverse, multiple application of Lemma 2.5 (as above) shows that H is a subuniverse as well. If $H \subsetneq A$ we are done by the inductive assumption. Therefore $H = A$ and there is no absorbing subuniverse in $\mathbf{A}$.

Let k be minimal such that there exists m and $a \in A$ with $\{a\}^{\mathbb{F}[k,m]} = A$. This implies that $E^{\circ k} \leq_S A \times A$ is linked and, as there is no absorbing subuniverse in $\mathbf{A}$, Theorem 2.3

implies that $E^{\circ k} = A \times A$. In particular the digraph $\mathbb{G}$ is strongly connected. Choose any $a \in A$ and consider the fence $\mathbb{F}[k-1, m']$ for m' large enough so that $B = \{a\}^{\mathbb{F}[k-1,m']} = \{a\}^{\mathbb{F}[k-1,m'+1]}$. The set B is a proper subset of A (by minimality of k) and it is a subuniverse of $\mathbf{A}$ (by Lemma 2.5 again). It suffices to prove that the smooth part of $\mathbb{G}_{|B}$ (which is a subuniverse by Lemma 3.3) has algebraic length 1.

Claim 3.11. The smooth part of $\mathbb{G}_{|B}$, denoted by B', is non-empty and has algebraic length one.

Proof. Note that, by definition of B, $B^{\mathbb{F}[k-1,1]} = B$.

Let b be an arbitrary element of B. As $\mathbb{G}$ is smooth we can find $c \in A$ such that $b \xrightarrow{k-1} c$. Since $E^{\circ k} = A \times A$ we get $b \xrightarrow{k} c$. Consider the first element b_1 on this path: $b \to b_1$ and $b_1 \in B$ as $b \xrightarrow{k-1} c \xleftarrow{k-1} b_1$. Therefore $b \to b_1$ in $\mathbb{G}_{|B}$. We have shown that $B^{\cdot\leftarrow\cdot} \supseteq B$. By Proposition 3.2 the smooth part of B is non-empty.

To show that $\mathbb{G}_{|B'}$ has algebraic length one we pick arbitrary $b, b' \in B'$ such that $b \xrightarrow{k-1} b'$ in $G_{|B'}$. As $E^{\circ k} = A \times A$ we have $b \xrightarrow{k} b'$ in $\mathbb{G}$. All the vertices on the path $b \xrightarrow{k} b'$ are in B, because $B^{\mathbb{F}[k-1,1]} = B$ and b' is in the smooth part of $G_{|B}$. Since b, b' are in B', the whole path falls in B'. This gives a path of algebraic length one connecting b to b in B' which proves the claim. $\square$

4. Cyclic terms in Taylor varieties

In the final section we prove our second main result – a characterization of Taylor varieties as the varieties possessing a cyclic term.

Theorem 4.1. *Let $\mathcal{V}$ be an idempotent variety generated by a finite algebra $\mathbf{A}$ then the following are equivalent.*
- *$\mathcal{V}$ is a Taylor variety;*
- *$\mathcal{V}$ (equivalently the algebra $\mathbf{A}$) has a cyclic term;*
- *$\mathcal{V}$ (equivalently the algebra $\mathbf{A}$) has a cyclic term of arity p, for every prime $p > |A|$.*

The proof uses the Absorption Theorem and its corollaries, and Theorem 3.6. This result is then applied to restate the Algebraic Dichotomy Conjecture, and to give short proofs of Theorem 1.3 and the dichotomy theorem for undirected graphs [HN90]. At the very end of the section we provide more information about possible arities of cyclic terms of a finite algebra.

4.1. **Proof of Theorem 4.1.**

As every cyclic term is a Taylor term, Theorem 4.1 will follow immediately when we prove:

Theorem 4.2. *Let $\mathbf{A}$ be a finite algebra in a Taylor variety and let p be a prime such that $p > |A|$. Then $\mathbf{A}$ has a p-ary cyclic term operation.*

As in the proofs of partial results [BKM$^+$09, BK10], the proof of Theorem 4.2 is based on studying cyclic relations:

Definition 4.3. An n-ary relation R on a set A is called *cyclic*, if for all $a_0, \dots, a_{n-1} \in A$

$$(a_0, a_1, \dots, a_{n-1}) \in R \quad \Rightarrow \quad (a_1, a_2, \dots, a_{n-1}, a_0) \in R.$$

The following lemma from [BKM+09] gives a connection between cyclic operations and cyclic relations.

Lemma 4.4. *For a finite, idempotent algebra* $\mathbf{A}$ *the following are equivalent:*

- $\mathbf{A}$ *has a* k*-ary cyclic term operation;*
- *every nonempty cyclic subalgebra of* $\mathbf{A}^k$ *contains a constant tuple.*

Proof. Assume first that $\mathbf{A}$ has a k-ary cyclic term operation t and consider an arbitrary tuple $\mathbf{a} = (a_0, a_1, \dots, a_{k-1})$ in a cyclic subalgebra $\mathbf{R}$ of $\mathbf{A}^k$. We denote by $\sigma(\mathbf{a})$, $\sigma^2(\mathbf{a})$, …, $\sigma^{k-1}(\mathbf{a})$ the cyclic shifts of $\mathbf{a}$, that is $\sigma(\mathbf{a}) = (a_1, a_2, \dots, a_{k-1}, a_0)$, $\sigma^2(\mathbf{a}) = (a_2, a_3, \dots, a_{k-1}, a_0, a_1)$, …, $\sigma^{k-1}(\mathbf{a}) = (a_{k-1}, a_0, a_1, \dots, a_{k-2})$. As R is cyclic, all these shifts belong to R. By applying t to the tuples $\mathbf{a}$, $\sigma(\mathbf{a})$, …, $\sigma^{k-1}(\mathbf{a})$ coordinatewise we get the tuple

$$(t(a_0, a_1, \dots, a_{k-1}), t(a_1, a_2, \dots, a_{k-1}, a_0), \dots, t(a_{k-1}, a_0, a_1, \dots, a_{k-2})),$$

which belongs to R, since R is a subuniverse of $\mathbf{A}^k$. But t is a cyclic operation, therefore this tuple is constant.

To prove the converse implication, we assume that every nonempty cyclic subalgebra of $\mathbf{A}^k$ contains a constant tuple. For a k-ary operation $t \in \mathrm{Clo}(\mathbf{A})$ we define $S(t) \subseteq A^k$ to be the set of all $\mathbf{a} \in A^k$ such that $t(\mathbf{a}) = t(\sigma(\mathbf{a})) = \cdots = t(\sigma^{k-1}(\mathbf{a}))$. Let t be such that $|S(t)|$ is maximal.

If $S(t) = A^k$, then the term operation t is cyclic and we are done. Assume the contrary, that is, there exists a tuple $\mathbf{a} \in A^k$ such that $t(\mathbf{a}) = t(\sigma(\mathbf{a})) = \cdots = t(\sigma^{k-1}(\mathbf{a}))$ fails. Consider the tuple $\mathbf{b} = (b_0, b_1, \dots, b_{k-1})$ defined by $b_i = t(\sigma^i(\mathbf{a}))$, $0 \le i < k$, and let $B = \{\mathbf{b}, \sigma(\mathbf{b}), \dots, \sigma^{k-1}(\mathbf{b})\}$.

We claim that the subalgebra $\mathbf{C} = \mathrm{Sg}_{\mathbf{A}^k}(B)$ of $\mathbf{A}^k$ is cyclic. Indeed, every tuple $\mathbf{c} \in C$ can be written as $\mathbf{c} = s(\mathbf{b}, \sigma(\mathbf{b}), \dots, \sigma^{k-1}(\mathbf{b}))$ for some term s. Then the element $s(\sigma(\mathbf{b}), \sigma^2(\mathbf{b}), \dots, \sigma^{k-1}(\mathbf{b}), \mathbf{b})$ of $\mathbf{C}$ is equal to $\sigma(\mathbf{c})$.

According to our assumption, the algebra $\mathbf{C}$ contains a constant tuple. It follows that there exists a k-ary term $s \in \mathrm{Clo}(\mathbf{A})$ such that $\mathbf{b} \in S(s)$. Now consider the term r defined by

$$r(x_0, x_1, \dots, x_{k-1}) = s(t(x_0, x_1, \dots, x_{k-1}), t(x_1, \dots, x_{k-1}, x_0), \dots, t(x_{k-1}, x_0, x_1, \dots, x_{k-2})).$$

We claim that $S(t) \subseteq S(r)$, but also that $\mathbf{a} \in S(r)$. This would clearly be a contradiction with the maximality of $|S(t)|$. Let $\mathbf{x} \in S(t)$. Then

$$r(\sigma^i(\mathbf{x})) = s(t(\sigma^i(\mathbf{x})), t(\sigma^{i+1}(\mathbf{x})), \dots, t(\sigma^{i-1}(\mathbf{x}))) = s(t(\mathbf{x}), t(\mathbf{x}), \dots, t(\mathbf{x})) = t(\mathbf{x})$$

for all i, so $\mathbf{x} \in S(r)$. On the other hand,

$$r(\sigma^i(\mathbf{a})) = s(t(\sigma^i(\mathbf{a})), t(\sigma^{i+1}(\mathbf{a})), \dots, t(\sigma^{i-1}(\mathbf{a}))) = s(b_i, b_{i+1}, \dots, b_{i-1}) = s(\sigma^i(\mathbf{b})),$$

which is constant for all i by the choice of s. Therefore $\mathbf{a} \in S(r)$ and the contradiction is established. ☐

For the rest of the proof of Theorem 4.2 we fix a prime number p, we fix a Taylor variety $\mathcal{V}$ and we consider a minimal counterexample to the theorem with respect to the size of A. Thus $\mathbf{A}$ is a finite algebra in $\mathcal{V}$, $p > |A|$, and for all $\mathbf{B} \in \mathcal{V}$ with $|B| < |A|$, $\mathbf{B}$ has a cyclic term of arity p, i.e., by Lemma 4.4, every nonempty cyclic subuniverse of $\mathbf{B}^p$ contains a constant tuple.

An easy reduction proving the following claim can also be found in [BKM+09].

Claim 4.5. $\mathbf{A}$ is simple.

Proof. Suppose that $\mathbf{A}$ is not simple, and α is a nontrivial congruence of $\mathbf{A}$.

To apply Lemma 4.4 we focus on an arbitrary cyclic subalgebra $\mathbf{R}$ of $\mathbf{A}^p$. Our first objective is to find a tuple in $\mathbf{R}$ with all elements congruent to each other modulo α. Let us choose any tuple $(a_0, \dots, a_{k-1}) \in R$ and let $c(x_0, \dots, x_{k-1})$ be the operation of $\mathbf{A}$ which gives rise to the cyclic operation of $\mathbf{A}/\alpha$ (such an operation exists from the minimality assumption). Therefore $c(a_0, \dots, a_{k-1}), c(a_1, \dots, a_{k-1}, a_0), \dots$ all lie in one congruence block of α as the results of these evaluations are equal in $\mathbf{A}/\alpha$. Now we apply the term $c(x_0, \dots, x_{k-1})$ in $\mathbf{R}$ to $(a_0, \dots, a_{k-1}), (a_1, \dots, a_{k-1}, a_0), \dots$ and obtain the tuple $(c(a_0, \dots, a_{k-1}), c(a_1, \dots, a_{k-1}, a_0), \dots)$ in $\mathbf{R}$ with all the coordinates in the same congruence block.

Let C be a congruence block of α such that $C^p \cap R \neq \emptyset$. It is easy to see that in such a case $C^p \cap R$ is a (nonempty) cyclic subuniverse of $\mathbf{C}^p$. As the block C has a cyclic operation of arity p then, again by Lemma 4.4, we obtain a constant in $C^p \cap R$ and the claim is proved. □

From Lemma 4.4 it follows that there exists a cyclic subalgebra $\mathbf{R}$ of $\mathbf{A}^p$ containing no constant tuple. We fix such a subalgebra $\mathbf{R}$. Let $\mathbf{R}_k$, $k = 1, 2, \dots, p$, denote the projection of $\mathbf{R}$ to the first k coordinates, that is

$$R_k = \{(a_0, a_1, \dots, a_{k-1}) : (a_0, \dots, a_{p-1}) \in R\}.$$

Note that, from the cyclicity of R, it follows that for any i we have

$$R_k = \{(a_i, a_{i+1}, \dots, a_{i+k-1}) : (a_0, \dots, a_{p-1}) \in R\},$$

where indices are computed modulo p. In the next claim we show that R is subdirect in $\mathbf{A}^p$.

Claim 4.6. $R_1 = A$.

Proof. The projection of R to any coordinate is a subalgebra of $\mathbf{A}$. From the cyclicity of R it follows that all the projections are equal, say to B. The set B is a subuniverse of $\mathbf{A}$ and if it is a proper subset of A, then $R \leq_S \mathbf{B}^p$ contains a constant tuple by the minimality assumption, a contradiction. □

We will prove the following two claims by induction on $n = 1, 2, \dots, p$. Note that for $n = 1$ both claims are valid and that property (P1) for $n = p$ contradicts the absence of a constant tuple in R.

(P1) There exists $\mathbf{I} \lhd\!\lhd \mathbf{A}$ such that $\mathbf{I}^n \lhd\!\lhd \mathbf{R}_n$.

(P2) If $\mathbf{I}_1, \dots, \mathbf{I}_n \lhd\!\lhd \mathbf{A}$ and $(I_1 \times \dots \times I_n) \cap R_n \neq \emptyset$, then $\mathbf{I}_1 \times \dots \times \mathbf{I}_n \lhd\!\lhd R_n$.

We assume that both (P1) and (P2) hold for some $n \in \{1, \dots, p-1\}$ and we aim to prove these properties for $n+1$. We fix $\mathbf{I} \lhd\!\lhd \mathbf{A}$ such that $\mathbf{I}^n \lhd\!\lhd \mathbf{R}_n$ guaranteed by (P1). Let

$$S = \{((a_0, \dots, a_{n-1}), a_n) : (a_0, \dots, a_n) \in R_{n+1}\}$$

and let $\mathbf{S}$ denote the subalgebra of $\mathbf{A}^{n+1}$ with universe S. Thus $\mathbf{S}$ is basically $\mathbf{R}_{n+1}$, but we look at it as a (subdirect) product of two algebras $\mathbf{R}_n$ and $\mathbf{A}$: $S \leq_S \mathbf{R}_n \times \mathbf{A}$.

The aim of the next few claims is to show that S is linked. First we show, that it is enough to have a “fork”.

Claim 4.7. If there exist $\mathbf{a} \in \mathbf{R}_n$ and $b, b' \in A$, $b \neq b'$ such that $(\mathbf{a}, b), (\mathbf{a}, b') \in S$, then S is linked.

Proof. Let $k = |A|$. We define a binary relation $\sim$ on A by putting $b \sim b'$ if and only if there exist tuples $\mathbf{a}^1, \dots, \mathbf{a}^k \in R_n$ and elements $b = c_0, c_1, \dots, c_k = b' \in A$ such that for every $i \in \{1, 2, \dots, k\}$ we have

$$(\mathbf{a}^i, c_{i-1}), (\mathbf{a}^i, c_i) \in S.$$

The relation $\sim$ is clearly reflexive and symmetric. It is also transitive as we have chosen k big enough. It follows immediately from the definition that $\sim$ is a subuniverse of $\mathbf{A}^2$.

Therefore $\sim$ is a congruence of $\mathbf{A}$. Moreover, from the assumption of the claim it follows that it is not the smallest congruence (as $b \sim b'$ for $b \neq b'$). Since, by Claim 4.5, $\mathbf{A}$ is simple, then $\sim$ is the full relation on A and therefore S is linked. □

The next claim shows that S is linked in case that $\mathbf{A}$ has no proper absorbing subuniverse.

Claim 4.8. If $I = A$ then S is linked.

Proof. From (P1) we have $R_n = A^n$. If there are $(a_0, \dots, a_{p-1}), (b_0, \dots, b_{p-1}) \in R$ such that $a_i \neq b_i$ for some i and $a_0 = b_0$, $a_1 = b_1$, …, $a_{i-1} = b_{i-1}$, $a_{i+1} = b_{i+1}$, …, $a_{n-1} = b_{n-1}$, then, by cyclically shifting these tuples, we obtain tuples $(a'_0, a'_1, \dots, a'_{p-1})$ and $(b'_0, b'_1, \dots, b'_{p-1})$ such that $a'_0 = b'_0$, …, $a'_{n-1} = b'_{n-1}$, and $a_n \neq b_n$. Then Claim 4.7 proves that S is linked.

In the other case, tuples in R are determined by the first n projections, thus $|R| = |R_n| = |A|^n$. Consider the mapping $\sigma : R \to R$ sending a tuple $(a_0, \dots, a_{p-1}) \in R$ to its cyclic shift $(a_1, \dots, a_{p-1}, a_0) \in R$. Clearly, σ is a permutation of R satisfying $\sigma^p = \mathrm{id}$. Now p is a prime number and $|R| = |A|^n$ is not divisible by p (as $p > |A|$), therefore σ has a fixed point, that is, a constant tuple. A contradiction. □

The harder case is when $I \neq A$. We need two more auxiliary claims.

Claim 4.9. If $I \neq A$ then there exists $\mathbf{J} \lhd\!\lhd \mathbf{A}$ such that $I \neq J$ and $(I^n \times J) \cap R_{n+1} \neq \emptyset$.

Proof. Observe that $I^p \cap R$ is a cyclic subuniverse of $\mathbf{I}^p$ without a constant tuple. Therefore, by minimality, the intersection $I^p \cap R$ is empty. On the other hand $I^n \cap R_n \neq \emptyset$ by (P1), so that there exists a greatest number k, $n \leq k < p$, such that $(I^k \times A^{p-k}) \cap R$ is nonempty. Consider the set

$$X = \{a : (a_0, \dots, a_{k-1}, a) \in R_{k+1}, \quad a_0, \dots, a_{k-1} \in I\}.$$

It is easy to check that X is an absorbing subuniverse of $\mathbf{A}$. As $I^{k+1} \cap R_{k+1}$ is empty, X is disjoint from I. Let J be a minimal absorbing subuniverse of $\mathbf{X}$. We have $J \lhd\!\lhd \mathbf{A}$ (as $J \lhd\!\lhd \mathbf{X} \lhd \mathbf{A}$), $I \neq J$ and $(I^k \times J) \cap R_{k+1} \neq \emptyset$. We take a tuple in R whose projection to the first $(k+1)$ coordinates lies in $I^k \times J$, and shift it $(k - n)$ times to the left (recall that $k - n \geq 0$). This tuple shows that $(I^n \times J) \cap R_{n+1}$ is nonempty. □

Similarly we can show that there exists a minimal absorbing subalgebra $\mathbf{J}'$ of $\mathbf{A}$ distinct from I such that $(J' \times I^n) \cap R_{n+1}$ is nonempty.

We consider the following two subsets of $A \times A$.

$$\begin{aligned} F &= \{(a, b) : \exists\ (a, c_1, \dots, c_{n-1}, b) \in R_{n+1}\} \\ E &= \{(a, b) : \exists\ (a, c_1, \dots, c_{n-1}, b) \in R_{n+1} \text{ and } \forall i\ c_i \in I\} \end{aligned}$$

Let V_1 and V_2 denote the projections of E to the first and the second coordinate, so that $E \subseteq_S V_1 \times V_2$.

Claim 4.10. E is a subuniverse of $\mathbf{A}^2$, is linked and subdirect in $V_1 \times V_2$ and $V_1, V_2 \lhd \mathbf{A}$.

Proof. It is straightforward to check that E and F are subuniverses of $\mathbf{A}^2$, that $\mathbf{E} \lhd \mathbf{F}$ and that $\mathbf{V}_1, \mathbf{V}_2 \lhd \mathbf{A}$, where $\mathbf{E}, \mathbf{F}$ denote the subalgebras of $\mathbf{A}^2$ with universes E, F and $\mathbf{V}_1, \mathbf{V}_2$ denote the subalgebras of $\mathbf{A}$ with universes V_1, V_2. From Claim 4.6 we know that $F \leq_S A \times A$.

Similarly as in the proof of Claim 4.7 we will show that F is linked. Let $k = |A|$ and let us define a congruence $\sim$ on $\mathbf{A}$ by putting $b \sim b'$ if and only if there are $a_1, a_2, \dots, a_k, b = b_0, b_1, \dots, b_k = b' \in A$ such that for all $i \in \{1, 2, \dots, k\}$

$$(a_i, b_{i-1}), (a_i, b_i) \in F.$$

The proof that $\sim$ is a congruence follows exactly as in Claim 4.7.

Take an arbitrary tuple $(a_0, \dots, a_{p-1}) \in R$. As p is greater than $|A|$ we can find indices $i \neq j$ such that $a_i = a_j$. There exists k such that $a_{i+kn} \neq a_{j+kn}$ (indices computed modulo p), otherwise (as p is a prime number) the tuple would be constant. It follows that there exist i', j' such that $a_{i'} = a_{j'}$ and $a_{i'+n} \neq a_{j'+n}$. The pairs $(a_{i'}, a_{i'+n})$ and $(a_{j'}, a_{j'+n})$ are in F (by shifting $(a_0, \dots, a_{p-1})$), therefore $\sim$ is not the smallest congruence. Since $\mathbf{A}$ is simple, $\sim$ is the full congruence on $\mathbf{A}$, thus F is linked. By Proposition 2.15.(i), E is linked as well. □

Now we can finally show that S is linked.

Claim 4.11. S is linked.

Proof. From Claim 4.9 and the remark following it we know that $(a, b'), (a', b) \in E$ for some $a, b \in I, a' \in J', b' \in J$, $J, J' \lhd\!\lhd \mathbf{A}$, $I \neq J$, $I \neq J'$. As E is linked, we can find elements $a = c_0, c_1, \dots, c_{2i} = a'$ such that $c_0, c_2, \dots, c_{2i} \in V_1$, $c_1, c_3, \dots, c_{2i-1} \in V_2$ and $(c_{2j}, c_{2j+1}), (c_{2j+2}, c_{2j+1}) \in E$ for all $j = 0, 1, \dots, i-1$. By Proposition 2.15.(v) (used for $\mathbf{E} \leq_S \mathbf{V}_1 \times \mathbf{V}_2$) we can assume that all the elements $c_0, \dots c_{2i}$ lie in minimal absorbing subuniverses of $\mathbf{V}_1$ or $\mathbf{V}_2$ (which are also minimal absorbing subuniverses of $\mathbf{A}$, since $V_1, V_2 \lhd \mathbf{A}$). It follows that there exist $w \in W \lhd\!\lhd \mathbf{V}_1$ and $u \in U \lhd\!\lhd \mathbf{V}_2, v \in V \lhd\!\lhd \mathbf{V}_2$ such that $(w, u), (w, v) \in E$, $U \neq V$. Therefore there exist $a_1, \dots, a_{n-1}, a'_1, \dots, a'_{n-1} \in I$ such that $(w, a_1, \dots, a_{n-1}, u), (w, a'_1, \dots, a'_{n-1}, v) \in R_{n+1}$.

From the induction hypotheses (P2) we know that $W \times I^{n-1} \lhd\!\lhd R_n$. Also $V \lhd\!\lhd \mathbf{A}$ and $((W \times I^{n-1}) \times V) \cap S \neq \emptyset$. By Proposition 2.15.(ii), $((W \times I^{n-1}) \times V) \cap S \leq_S (W \times I^{n-1}) \times V$. In particular, there exists $v' \in V$ such that $(w, a_1, \dots, a_{n-1}, v') \in R_{n+1}$. Now recall that $(w, a_1, \dots, a_{n-1}, u) \in R_{n+1}$ and observe that u and v' are distinct, since they lie in different minimal absorbing subuniverses. Then S is linked by Claim 4.7. □

We are ready to prove (P2) for $n+1$.

Claim 4.12. (P2) holds for $n+1$.

Proof. Let $\mathbf{I}_1, \dots, \mathbf{I}_{n+1}$ be absorbing subalgebras of $\mathbf{A}$ such that $(I_1 \times \dots \times I_{n+1}) \cap R_{n+1} \neq \emptyset$. Now S is a linked subdirect subuniverse of $\mathbf{R}_n \times \mathbf{A}$, $I_1 \times \dots \times I_n$ is a minimal absorbing subuniverse of $\mathbf{R}_n$ (from the induction hypotheses (P2)), $I_{n+1} \lhd\!\lhd \mathbf{A}$ and $((I_1 \times \dots \times I_n) \times I_{n+1}) \cap S \neq \emptyset$. By Proposition 2.15.(iii), $(I_1 \times \dots \times I_n) \times I_{n+1}$ is a minimal absorbing subuniverse of $\mathbf{S}$ and thus $I_1 \times \dots \times I_{n+1}$ is a minimal absorbing subuniverse of $\mathbf{R}_{n+1}$. □

To prove (P1) for $n+1$ we define a digraph on the vertex set R_n by putting

$$((a_0, \dots, a_{n-1}), (a_1, \dots, a_n)) \in H$$

whenever $(a_0, \dots, a_n) \in R_{n+1}$. We want to apply Theorem 3.6 to obtain a loop of the digraph $\mathbb{G} = (R_n, H)$ in a minimal absorbing subuniverse of $\mathbf{R}_n$.

Observe that H is a subuniverse of $\mathbf{R}_n^2$. Next we show that I^n is contained in a weak component of $\mathbb{G}$.

Claim 4.13. Any two elements of I^n are in the same weak component of the digraph $\mathbb{G}$.

Proof. The set $X = \{x : (a_0, \dots, a_{n-1}, x) \in R_{n+1}, \quad a_0, \dots, a_{n-1} \in I\}$ is an absorbing subuniverse of $\mathbf{A}$. Let X_0 be a minimal absorbing subuniverse of the algebra $\mathbf{X}$ with universe X. We have found $X_0 \triangleleft \mathbf{A}$ such that $(I^n \times X_0) \cap R_{n+1} \neq \emptyset$. Similarly we can find $X_1, X_2, \dots, X_{n-1}$ such that $(I^{n-i} \times X_0 \times X_1 \times \dots \times X_i) \cap R_{n+1} \neq \emptyset$ for all $i = 0, 1, \dots, n-1$. From (P2) for $n+1$ (Claim 4.12) it follows that $I^{n-i} \times X_0 \times X_1 \times \dots \times X_i \subseteq R_{n+1}$ for all i. Now choose arbitrary elements $x_i \in X_i$ and take any tuple $(b_0, \dots, b_{n-1}) \in I^n$. Since, for all $i = 0, 1, \dots, n-1$, the tuple $(b_i, \dots, b_{n-1}, x_0, x_1, \dots, x_i)$ belongs to R_{n+1}, the vertices $(b_i, \dots, b_{n-1}, x_0, \dots, x_{i-1})$ and $(b_{i+1}, \dots, b_{n-1}, x_0, \dots, x_i)$ are in the same weak component of $\mathbb{G}$. Therefore the vertex $(b_0, \dots, b_{n-1})$, which was an arbitrarily chosen vertex in I^n, is in the same weak component as the vertex $(x_0, \dots, x_{n-1})$. ☐

The last assumption of Theorem 3.6 is proved in the next claim.

Claim 4.14. The weak component of $\mathbb{G}$ containing I^n has algebraic length 1.

Proof. Let $b \in I$ be arbitrary. As E is linked, $b \in V_1$ can be E-linked to $b \in V_2$, i.e. there exist $b = c_0, c_1, \dots, c_{2i}$ such that $(c_{2j}, c_{2j+1}), (c_{2j+2}, c_{2j+1}) \in E$ for all $j = 0, \dots, i-1$ and $(c_{2i}, b) \in E$. By Proposition 2.15.(v) we can assume that these elements lie in minimal absorbing subuniverses of $\mathbf{A}$. Property (P2) for $n+1$ (Claim 4.12) proves that $(c_{2j}, b, \dots, b, c_{2j+1}), (c_{2j+2}, b, \dots, b, c_{2j+1}) \in R_{n+1}$ for all $j = 0, \dots, i-1$ and $(c_{2i}, b, \dots, b, b) \in R_{n+1}$. This gives rise to a $(1, j)$-fence connecting, in $\mathbb{G}$, the tuple $(c_0 = b, \dots, b)$ to the tuple $(c_{2i}, b, \dots, b)$. As $((c_{2i}, b, \dots, b), (b, \dots, b)) \in H$ we showed that the algebraic length of the weak component containing I^n is one. ☐

By Theorem 3.6 there exists a loop inside a minimal absorbing subuniverse K of $\mathbf{R}_n$. Since the projection J of K to the first coordinate is a minimal absorbing subuniverse of $\mathbf{A}$, we actually get an element $a \in J \triangleleft\!\triangleleft \mathbf{A}$ such that $(a, \dots, a) \in R_{n+1}$. Now (P1) follows from (P2) and the proof of Theorem 4.2 is concluded.

4.2. **Consequences of Theorem 4.1.** First we restate the hardness criterion in Theorem 1.4 and the Algebraic Dichotomy Conjecture of Bulatov, Jeavons and Krokhin. These statements are equivalent to the original ones by Theorem 4.1 and Lemma 4.4.

Theorem 4.15. *Let $\mathbb{A}$ be a core relational structure and let p be a prime number greater than the size of the universe of $\mathbb{A}$. If there exists a nonempty positively primitively defined cyclic p-ary relation without a constant tuple then* CSP($\mathbb{A}$) *is NP-complete.* ☐

The Algebraic Dichotomy Conjecture. *Let $\mathbb{A}$ be a a core relational structure. Let p be a prime number greater than the size of the universe of $\mathbb{A}$. If every nonempty positively primitively defined cyclic p-ary relation has a constant tuple then* CSP($\mathbb{A}$) *is solvable in polynomial time. Otherwise it is NP-complete.*

As a second consequence we reprove the dichotomy theorem of Hell and Nešetřil [HN90]. It follows immediately from the Smooth Theorem from Section 3, but the following proof is an elegant way of presenting it.

Corollary 4.16 (Hell and Nešetřil [HN90])**.** *Let $\mathbb{G}$ be an undirected graph without loops. If $\mathbb{G}$ is bipartite then* CSP($\mathbb{G}$) *is solvable in polynomial time. Otherwise it is NP-complete.*

Proof. Without loss of generality we can assume that $\mathbb{G}$ is a core. If the graph $\mathbb{G}$ is bipartite then it is a single edge and CSP($\mathbb{G}$) is solvable in polynomial time. Assume now that $\mathbb{G}$ is not bipartite — therefore there exists a cycle $a \xrightarrow{2k+1} a$ of odd length in $\mathbb{G}$. As vertex a is in a 2-cycle (i.e. an undirected edge) we can find a path $a \xrightarrow{i(2k+1)+j2} a$ for any non-negative numbers i and j. Thus, for any number $l \geq 2k$ we have $a \xrightarrow{l} a$. Let p be any prime greater than $\max\{2k, |A|\}$ and t be any p-ary polymorphism of $\mathbb{G}$. Let $a = a_0 \to a_1 \to \cdots \to a_{p-1} \to a$. Then

$$t(a_0, \ldots, a_{p-1}) \to t(a_1, \ldots, a_{p-1}, a_0)$$

and, if t were a cyclic operation we would have

$$t(a_0, \ldots, a_{p-1}) = t(a_1, \ldots, a_{p-1}, a_0)$$

which implies a loop in $\mathbb{G}$. This contradiction shows that $\mathbb{G}$ has no cyclic polymorphism for some prime greater than the size of the vertex set which, by Theorem 4.1, implies that the associated variety is not Taylor and therefore, by Theorem 1.4, CSP($\mathbb{G}$) is NP-complete.

Equivalently one can consider the relation

$$R = \{(a_0, \ldots, a_{p-1}) : a_0 \to a_1 \to a_2 \to \cdots \to a_{p-1} \to a_0\},$$

where p is chosen as above. It is easy to see that R is a cyclic, positively primitively defined nonempty relation without a constant tuple and therefore CSP($\mathbb{G}$) is NP-complete by Theorem 4.15. □

Finally, we observe that the weak near-unanimity characterization of Taylor varieties (Theorem 1.3) is a consequence of Theorem 4.1:

Corollary 4.17 (Maroti and McKenzie [MM08]). *For a locally finite idempotent variety $\mathcal{V}$ the following are equivalent.*

- *$\mathcal{V}$ is a Taylor variety;*
- *$\mathcal{V}$ has a weak near-unanimity term.*

Proof. In the case that $\mathcal{V}$ is finitely generated, the theorem is an immediate consequence of Theorem 4.1. In the general case the proof can be done by a standard universal algebraic argument — we apply Theorem 4.1 to the *free algebra* on two generators. □

As opposed to the previous theorem the assumption in Theorem 4.1 that $\mathcal{V}$ is finitely generated cannot be relaxed to locally finite [BKM+09].

It was observed by Matt Valeriote [Val] that Sigger's characterization of Taylor varieties [Sig10] is also an easy corollary of Theorem 4.1. The proof will appear elsewhere.

4.3. **Arities of cyclic terms.**

Let $\mathbf{A}$ be a finite algebra and let $C(\mathbf{A})$ be the set of arities of cyclic operations of $\mathbf{A}$ i.e.:

$$C(\mathbf{A}) = \{n : \mathbf{A} \text{ has a cyclic term of arity } n\}.$$

The following simple proposition was proved in [BKM+09].

Proposition 4.18 ([BKM+09]). *Let $\mathbf{A}$ be a finite algebra let m, n be natural numbers. Then the following are equivalent.*

(i) *$m, n \in C(\mathbf{A})$;*
(ii) *$mn \in C(\mathbf{A})$.*

This implies that $C(\mathbf{A})$ is fully determined by its prime elements. There are algebras in Taylor varieties with no cyclic terms of arities smaller than their size [BKM+09]. However the following simple lemma provides, under special circumstances, additional elements in $C(\mathbf{A})$. Its proof follows the lines of the proof of Claim 4.5.

Lemma 4.19. *Let* $\mathbf{A}$ *be a finite, idempotent algebra and* α *be a congruence of* $\mathbf{A}$*. If* $\mathbf{A}/\alpha$ *and every* α*-block in* A *have cyclic operation of arity* k *then so does* $\mathbf{A}$*.* □

This leads to the following observation.

Corollary 4.20. *Let* $\mathbf{A}$ *be a finite, idempotent algebra in Taylor variety. Let* $0_A = \alpha_0 \subseteq \cdots \subseteq \alpha_n = 1_A$ *be an increasing sequence of congruences on* $\mathbf{A}$*. If* p *is a prime number such that, for every* $i \geq 1$*, every class of* α_i *splits into less than* p *classes of* α_{i-1} *then* $\mathbf{A}$ *has a* p*-ary cyclic term.* □

References

[Bar11] Libor Barto. The dichotomy for conservative constraint satisfaction problems revisited. In *Proceedings of the 26th Annual IEEE Symposium on Logic in Computer Science, LICS 2011, June 21-24, 2011, Toronto, Ontario, Canada*, pages 301–310. IEEE Computer Society, 2011.

[BD06] Andrei Bulatov and Víctor Dalmau. A simple algorithm for Mal′tsev constraints. *SIAM J. Comput.*, 36(1):16–27 (electronic), 2006.

[BJH90] Jørgen Bang-Jensen and Pavol Hell. The effect of two cycles on the complexity of colourings by directed graphs. *Discrete Appl. Math.*, 26(1):1–23, 1990.

[BJK05] Andrei Bulatov, Peter Jeavons, and Andrei Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM J. Comput.*, 34(3):720–742 (electronic), 2005.

[BK09a] Libor Barto and Marcin Kozik. Congruence distributivity implies bounded width. *SIAM Journal on Computing*, 39(4):1531–1542, 2009.

[BK09b] Libor Barto and Marcin Kozik. Constraint satisfaction problems of bounded width. In *FOCS'09: Proceedings of the 50th Symposium on Foundations of Computer Science*, pages 595–603, 2009.

[BK10] Libor Barto and Marcin Kozik. Cyclic terms for $SD_\vee$ varieties revisited. *Algebra Universalis*, 64(1-2):137–142, 2010.

[BKJ00] Andrei A. Bulatov, Andrei A. Krokhin, and Peter Jeavons. Constraint satisfaction problems and finite algebras. In *Automata, languages and programming (Geneva, 2000)*, volume 1853 of *Lecture Notes in Comput. Sci.*, pages 272–282. Springer, Berlin, 2000.

[BKKR69] V. G. Bodnarčuk, L. A. Kalužnin, V. N. Kotov, and B. A. Romov. Galois theory for Post algebras. I, II. *Kibernetika (Kiev)*, (3):1–10; ibid. 1969, no. 5, 1–9, 1969.

[BKM+09] Libor Barto, Marcin Kozik, Miklós Maróti, Ralph McKenzie, and Todd Niven. Congruence modularity implies cyclic terms for finite algebras. *Algebra Universalis*, 61(3):365–380, 2009.

[BKN09] Libor Barto, Marcin Kozik, and Todd Niven. The CSP dichotomy holds for digraphs with no sources and no sinks (a positive answer to a conjecture of Bang-Jensen and Hell). *SIAM J. Comput.*, 38(5):1782–1802, 2008/09.

[BKN08] Libor Barto, Marcin Kozik, and Todd Niven. Graphs, polymorphisms and the complexity of homomorphism problems. In *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 789–796, New York, NY, USA, 2008. ACM.

[BS81] Stanley N. Burris and H. P. Sankappanavar. *A course in universal algebra*, volume 78 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1981.

[Bul03] Andrei A. Bulatov. Tractable conservative constraint satisfaction problems. In *18th IEEE Symposium on Logic in Computer Science (LICS 2003), 22-25 June 2003, Ottawa, Canada, Proceedings*, pages 321–. IEEE Computer Society, 2003.

[Bul06] Andrei A. Bulatov. A dichotomy theorem for constraint satisfaction problems on a 3-element set. *J. ACM*, 53(1):66–120 (electronic), 2006.

[Bul11] Andrei A. Bulatov. Complexity of conservative constraint satisfaction problems. *ACM Trans. Comput. Logic*, 12:24:1–24:66, July 2011.

[Dal06] Víctor Dalmau. Generalized majority-minority operations are tractable. *Log. Methods Comput. Sci.*, 2(4):4:1, 14, 2006.

[FV99] Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory. *SIAM J. Comput.*, 28(1):57–104 (electronic), 1999.

[Gei68] David Geiger. Closed systems of functions and predicates. *Pacific J. Math.*, 27:95–100, 1968.

[HM88] David Hobby and Ralph McKenzie. *The structure of finite algebras*, volume 76 of *Contemporary Mathematics*. American Mathematical Society, Providence, RI, 1988.

[HN90] Pavol Hell and Jaroslav Nešetřil. On the complexity of H-coloring. *J. Combin. Theory Ser. B*, 48(1):92–110, 1990.

[IMM$^+$07] Pawel M. Idziak, Petar Markovic, Ralph McKenzie, Matthew Valeriote, and Ross Willard. Tractability and learnability arising from algebras with few subpowers. In *LICS*, pages 213–224. IEEE Computer Society, 2007.

[JCG97] Peter Jeavons, David Cohen, and Marc Gyssens. Closure properties of constraints. *J. ACM*, 44(4):527–548, 1997.

[KS09] Gábor Kun and Mario Szegedy. A new line of attack on the dichotomy conjecture. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 725–734, New York, NY, USA, 2009. ACM.

[MM08] Miklós Maróti and Ralph McKenzie. Existence theorems for weakly symmetric operations. *Algebra Universalis*, 59(3-4):463–489, 2008.

[Sch78] Thomas J. Schaefer. The complexity of satisfiability problems. In *Conference Record of the Tenth Annual ACM Symposium on Theory of Computing (San Diego, Calif., 1978)*, pages 216–226. ACM, New York, 1978.

[Sig10] Mark H. Siggers. A strong mal'cev condition for locally finite varieties omitting the unary type. *Algebra Universalis*, 64(1-2):15–20, 2010.

[Tay77] Walter Taylor. Varieties obeying homotopy laws. *Canad. J. Math.*, 29(3):498–527, 1977.

[Val] Matt Valeriote. personal communication.

Appendix D – Congruence distributive finitely related algebras

Canad. J. Math. Vol. **65** (1), 2013 pp. 3–21
http://dx.doi.org/10.4153/CJM-2011-087-3

Finitely Related Algebras in Congruence Distributive Varieties Have Near Unanimity Terms

Libor Barto

Abstract. We show that every finite, finitely related algebra in a congruence distributive variety has a near unanimity term operation. As a consequence we solve the near unanimity problem for relational structures: it is decidable whether a given finite set of relations on a finite set admits a compatible near unanimity operation. This consequence also implies that it is decidable whether a given finite constraint language defines a constraint satisfaction problem of bounded strict width.

1 Introduction

Since the beginning of the systematic study of universal algebras in the 1930's it has been recognized that an important class of invariants of algebras and classes of algebras are their congruence lattices. Particularly widely studied objects are congruence distributive varieties, *i.e.,* equationally definable classes of algebras whose congruence lattices are distributive (see Section 2 for definitions).

We call an algebra in a congruence distributive variety a CD algebra. Examples of CD algebras include lattices, and, more generally, algebras that have a near unanimity term operation. These operations have also attracted a great deal of attention, not only in universal algebra, but also in graph theory and, recently, in computer science in connection with the constraint satisfaction problem (CSP), where, for instance, near unanimity operations characterize CSPs of bounded strict width [11].

Every finite algebra is, in some sense, determined by a set of relations. We call an algebra finitely related if this set of relations can be chosen to be finite. A useful corollary of a classical result of Baker and Pixley [2] is that every algebra with a near unanimity term operation is finitely related. Our main result provides a partial converse.

Theorem 1.1 *Every finite, finitely related CD algebra has a near unanimity term operation.*

A special case of this theorem for algebras determined by posets was conjectured in [10,23]. An affirmative answer was given in [24] for bounded posets and in [19] in

Received by the editors May 3, 2011; revised October 14, 2011.
Published electronically December 24, 2011.
Research supported by the Grant Agency of the Czech Republic under the grant No. 201/09/P223 and by the Ministry of Education of the Czech Republic under the grant No. MSM 0021620839 and No. MEB 040915.
AMS subject classification: **08B05**, 08B10.
Keywords: congruence distributive variety, Jónsson operations, near unanimity operation, finitely related algebra, constraint satisfaction problem.

full generality. Another special case of the theorem, namely, for algebras determined by a reflexive undirected graph, was proved in [18]. The general version is commonly referred to as the Zádori conjecture, although it has been never stated in a journal paper, perhaps because of scant evidence.

What made this result possible is the connection between the constraint satisfaction problem and universal algebra discovered in [8, 15]. The interaction between these areas is very fruitful in both directions. On one hand, universal algebra has brought a deeper understanding and strong results about the CSP. On the other hand, the CSP has motivated much of the recent work in universal algebra and opened new research directions. This is nicely illustrated by the main result of [3] (Theorem 5.7 in this paper). This theorem contributed to the study of local consistency methods for the CSP (and was an important step toward the full characterization of applicability of local consistency methods given in [4]), and it is also one of the two main ingredients of the proof of our main, purely algebraic result.

We remark that none of the assumptions of Theorem 1.1 is superfluous. In [24], Zádori provides an example of an infinite, bounded poset that determines a CD algebra with no near unanimity term operation. A simple example of a finite CD algebra with no near unanimity operation is the two element set $\{0, 1\}$ together with the implication regarded as a binary operation. Finally, the algebra determined by the complete loopless graph with three vertices does not have any near unanimity operation (it actually has no idempotent operations other than projections).

Of independent interest is a corollary of the main theorem (Corollary 7.1), which gives an affirmative answer to the near unanimity problem for relational structures. It is decidable whether a given set of relations on a finite set admits a compatible near unanimity operation. This consequence is discussed in more detail in Section 7.

1.1 Organization of the Paper

In Section 2 we recall basic notions and results about algebras and relational structures. In Section 3 we show that it is enough to deal with algebras determined by at most binary relations. In Section 4 we associate with such an algebra an instance of the CSP whose solutions are term operations of that algebra. The definitions and results about CSP instances that we require are stated in Section 5, where we also prove the main theorem. The main new tool is only stated in this section; its proof covers Section 6. Finally, in Section 7 we discuss consequences and open problems.

2 Preliminaries

In this section we recall universal algebraic notions and results that will be needed throughout the paper. This material, except for the notion of a Jónsson ideal, is covered in any standard reference on universal algebra, for example, [9].

2.1 Algebras and Varieties

An *n-ary operation* on a set A is a mapping $f : A^n \to A$. In this paper we assume that all operations are finitary, *i.e., n* is a natural number. An operation is *idempotent* if it satisfies the identity $f(a, a, \dots, a) = a$, *i.e.*, this equation holds for every $a \in A$. An operation of arity at least three is called a *near unanimity operation*, if it satisfies the identity

$$f(a, a, \dots, a, b, a, a, \dots, a) = a$$

for every position of b in the tuple.

An *algebra* is a pair $\mathbf{A} = (A, \mathcal{F})$, where A is a set, called the *universe* of $\mathbf{A}$, and $\mathcal{F}$ is a set (possibly indexed) of operations on A. We use a boldface letter to denote an algebra and the same letter in plain type to denote its universe. An algebra is *idempotent* if all of its operations are idempotent. Two algebras are *similar* if their operations are indexed by the same set and corresponding operations have the same arities.

A *term operation* of $\mathbf{A}$ is an operation that can be obtained from operations in $\mathbf{A}$ using composition and the projection operations. The set of all term operations of $\mathbf{A}$ is denoted by $\mathrm{Clo}(\mathbf{A})$. Most structural properties of an algebra (such as subalgebras, congruences, automorphisms, *etc.*,) depend only on the set of term operations rather than on a particular choice of the basic operations.

There are three fundamental operations on algebras of a fixed similarity type: forming subalgebras, factor algebras, and products.

A subset B of the universe of an algebra $\mathbf{A}$ is called a *subuniverse* if it is closed under all operations (equivalently term operations) of $\mathbf{A}$. Given a subuniverse B of $\mathbf{A}$ we can form the algebra $\mathbf{B}$ by restricting all the operations of $\mathbf{A}$ to the set B. In this situation we say that $\mathbf{B}$ is a *subalgebra* of $\mathbf{A}$ and we write $B \leq \mathbf{A}$ or $\mathbf{B} \leq \mathbf{A}$.

The *product* of algebras $\mathbf{A}_1, \dots, \mathbf{A}_n$ is the algebra with the universe equal to $A_1 \times \cdots \times A_n$ and with operations computed coordinatewise. The product of n copies of an algebra $\mathbf{A}$ is denoted by $\mathbf{A}^n$. A subalgebra (or a subuniverse) of a product of $\mathbf{A}$ is called a *subpower* of $\mathbf{A}$.

An equivalence relation $\sim$ on the universe of an algebra $\mathbf{A}$ is a *congruence* if it is a subalgebra of $\mathbf{A}^2$. The corresponding *factor algebra* $\mathbf{A}/\sim$ has, as its universe, the set of $\sim$-blocks and operations that are defined using arbitrarily chosen representatives. The set of congruences of $\mathbf{A}$ forms a lattice, called the *congruence lattice of* $\mathbf{A}$.

A *variety* is a class of similar algebras closed under forming sublagebras, products (possibly infinite), factor algebras, and isomorphic copies. A fundamental theorem of universal algebra, due to G. Birkhoff, states that a class of similar algebras is a variety if and only if this class can be defined via a set of identities.

2.2 Relational Structures

An *n-ary relation* on a set A is a subset of A^n (again, n is always finite in this article). A relational structure is a pair $\mathbb{A} = (A, \mathcal{R})$, where A is the *universe of* $\mathbb{A}$ and $\mathcal{R}$ is a set of relations on A. We use blackboard bold letters to denote relational structures.

We say that an operation $f \colon A^n \to A$ is *compatible* with a relation $R \subseteq A^m$ (or, R is *preserved* by f) if the tuple

$$\big(f(a_1^1, a_1^2, \dots, a_1^n), f(a_2^1, a_2^2, \dots, a_2^n), \dots, f(a_m^1, a_m^2, \dots, a_m^n)\big)$$

belongs to R whenever $(a^i_1, a^i_2, \ldots, a^i_m) \in R$ for all $i \leq n$. In other words, f is compatible with R, if R is a subpower of the algebra $(A, \{f\})$.

An operation compatible with all relations of a relational structure $\mathbb{A}$ is a *polymorphism* of $\mathbb{A}$. The set of n-ary polymorphisms of $\mathbb{A}$ is denoted by $\mathrm{Pol}_n(\mathbb{A})$, and the set of all polymorphisms of $\mathbb{A}$ is denoted by $\mathrm{Pol}(\mathbb{A})$. This set of operations is closed under composition and contains the projection operations. On the other hand, every set of operations on a finite set closed under projections and composition can be obtained in this way.

Theorem 2.1 ([6, 13]) *For every finite algebra* $\mathbf{A}$ *there exists a relational structure* $\mathbb{A}$ *(with the same universe) such that* $\mathrm{Pol}(\mathbb{A}) = \mathrm{Clo}(\mathbf{A})$.

An algebra is called *finitely related* if finitely many relations suffice to determine $\mathrm{Clo}(\mathbf{A})$.

Definition 2.2 An algebra $\mathbf{A}$ is said to be *finitely related*, if there exists a relational structure $\mathbb{A}$ with finitely many relations such that $\mathrm{Pol}(\mathbb{A}) = \mathrm{Clo}(\mathbf{A})$.

By a classic result of Baker and Pixley [2], every algebra with a near unanimity term operation is finitely related. More generally, every algebra with few subpowers is finitely related [5] (see Subsection 7.3).

2.3 CD Algebras

Definition 2.3 A variety is called *congruence distributive*, if all algebras in it have distributive congruence lattices. A *CD algebra* is an algebra in a congruence distributive variety.

A theorem of Jónsson [16] characterizes CD algebras using operations satisfying certain identities.

Definition 2.4 A sequence $p_0, p_1, \ldots, p_s$ of ternary operations on a set A is called a *Jónsson chain*, if the following identities are satisfied:

$$\begin{aligned}
&p_0(a,b,c) = a, \\
&p_s(a,b,c) = c, \\
&p_i(a,b,a) = a && \text{for all } i \leq s, \\
&p_i(a,a,b) = p_{i+1}(a,a,b) && \text{for all even } i < s, \\
&p_i(a,b,b) = p_{i+1}(a,b,b) && \text{for all odd } i < s.
\end{aligned}$$

Theorem 2.5 ([16]) *An algebra* $\mathbf{A}$ *has a Jónsson chain of term operations if and only if* $\mathbf{A}$ *is a CD algebra.*

Example Every algebra with a near unanimity term operation t is a CD algebra.

This can be shown, for instance, by constructing a Jónsson chain:

$$
\begin{aligned}
p_0(a,b,c) &= a \\
p_1(a,b,c) &= t(a,a,\dots,a,b,c) \\
p_2(a,b,c) &= t(a,a,\dots,a,c,c) \\
p_3(a,b,c) &= t(a,a,\dots,a,b,c,c) \\
p_4(a,b,c) &= t(a,a,\dots,a,c,c,c) \\
&\dots .
\end{aligned}
$$

A useful notion for studying CD algebras is a Jónsson ideal.

Definition 2.6 Let $\mathbf{A}$ be a CD algebra with Jónsson chain of term operations $p_0, p_1, \dots, p_s$. A subuniverse B of $\mathbf{A}$ is a *Jónsson ideal*, if $p_i(b_1, a, b_2) \in B$ for every $a \in A$, $b_1, b_2 \in B$ and every $i \leq n$.

Every one element subuniverse of a CD algebra is its Jónsson ideal. Therefore, if $\mathbf{A}$ is an idempotent CD algebra, then every singleton is a Jónsson ideal of $\mathbf{A}$.

3 Reduction to Binary Structures

In this section we show that to prove the main result it is enough to consider algebras determined by *binary relational structures*, *i.e.*, relational structures with at most binary relations. This will make the presentation technically easier.

Proposition 3.1 *Let $\mathbb{A}$ be a relational structure whose relations all have arity at most k. Then there exists a binary relational structure $\bar{\mathbb{A}}$ with universe $\bar{A} = A^k$ such that*

$$\mathrm{Pol}(\bar{\mathbb{A}}) = \{\bar{f} : f \in \mathrm{Pol}(\mathbb{A})\},$$

where $\bar{f}$ is defined (if f is n-ary) by

$$\bar{f}\big((a_1^1, a_2^1, \dots, a_k^1), (a_1^2, \dots, a_k^2), \dots, (a_1^n, \dots, a_k^n)\big) = \big(f(a_1^1, a_1^2, \dots, a_1^n), f(a_2^1, \dots, a_2^n), \dots, f(a_k^1, \dots, a_k^n)\big).$$

Proof First we replace every relation R in $\mathbb{A}$ with arity $l < k$ by the k-ary relation $R \times A^{k-l}$. This clearly does not change the set of polymorphisms, therefore we may assume that every relation in $\mathbb{A}$ has arity precisely k.

Next we introduce the relations in $\bar{\mathbb{A}}$. For every k-ary relation R (on A) in $\mathbb{A}$ we include in $\bar{\mathbb{A}}$ the unary relation R (on $\bar{A} = A^k$), and for every pair $1 \leq i, j \leq k$ we add a binary relation σ_{ij} defined by

$$\big((a_1, \dots, a_k), (b_1, \dots, b_k)\big) \in \sigma_{ij} \quad \text{if and only if} \quad a_i = b_j.$$

It is straightforward to check that $\bar{f} \in \mathrm{Pol}(\bar{\mathbb{A}})$ for every polymorphism f of $\mathbb{A}$.

To prove the converse inclusion, let $h\colon (A^k)^n \to A^k$ be a polymorphism of $\bar{\mathbb{A}}$ and let $h_1, \dots, h_k$ be its components, that is,

$$h\big((a_1^1,\dots,a_k^1),(a_1^2,\dots,a_k^2),\dots,(a_1^n,\dots,a_k^n)\big) = \Big(h_1\big((a_1^1,\dots,a_k^1),(a_1^2,\dots,a_k^2),\dots,\dots(a_1^n,\dots,a_k^n)\big),\dots, h_k\big((a_1^1,\dots,a_k^1),(a_1^2,\dots,a_k^2),\dots,(a_1^n,\dots,a_k^n)\big)\Big).$$

For any i, $1 \le i \le k$, the relation σ_{ii} is compatible with h. Therefore for any elements $a_1^1, a_2^1, \dots, a_1^2, \dots a_k^n, b_1^1, \dots, b_k^n \in A$ such that, for all $1 \le l \le n$, $a_i^l = b_i^l$ we have

$$h_i\big((a_1^1,\dots,a_k^1),(a_1^2,\dots,a_k^2),\dots,(a_1^n,\dots,a_k^n)\big) = h_i\big((b_1^1,\dots,b_k^1),(b_1^2,\dots,b_k^2),\dots,(b_1^n,\dots,b_k^n)\big).$$

In other words, $h_i((a_1^1,\dots),\dots)$ depends only on $a_i^1, a_i^2, \dots, a_i^n$, and thus there exists an n-ary operation f_i on A such that

$$h_i((a_1^1,\dots,a_k^1),(a_1^2,\dots,a_k^2),\dots,(a_1^n,\dots,a_k^n)) = f_i(a_i^1,a_i^2,\dots,a_i^n).$$

Now we use the relations σ_{ij} for $i \neq j$. For any $a_1^1, \dots, a_k^n$ we choose arbitrarily $b_1^1, \dots, b_k^n$ so that $a_i^l = b_j^l$ (for all $1 \le l \le n$). As σ_{ij} is compatible with h, it follows that $f_i(a_i^1,\dots,a_i^n) = f_j(b_j^1,\dots,b_j^n) = f_j(a_i^1,\dots,a_i^n)$. Therefore, $f_1 = f_2 = \dots = f_k$.

We have shown that $h = \bar{f}$ for certain n-ary operation f on A. As each relation R of $\mathbb{A}$ is compatible with h it follows that f is a polymorphism of $\mathbb{A}$. ■

The proposition implies that Theorem 1.1 follows from the following theorem, to be proved later.

Theorem 3.2 *If $\mathbb{A}$ is a finite binary relational structure such that $(A, \mathrm{Pol}(\mathbb{A}))$ is a CD algebra, then $\mathbb{A}$ has a near unanimity polymorphism.*

Proof of Theorem 1.1 assuming Theorem 3.2 Let $\mathbf{A}$ be a finite, finitely related CD algebra and let $\mathbb{A}$ be a relational structure with finitely many relations (say all of them have arity at most k) such that $\mathrm{Pol}(\mathbb{A}) = \mathrm{Clo}(\mathbf{A})$. Let $\bar{\mathbb{A}}$ be the relational structure from the previous proposition. Then $\bar{\mathbf{A}} = (A, \mathrm{Pol}(\bar{\mathbb{A}}))$ is a CD algebra, since $\bar{p}_0, \dots, \bar{p}_s$ is a Jónsson chain of $\bar{\mathbf{A}}$ whenever $p_0, \dots, p_s$ is a Jónsson chain of $\mathbf{A}$. By Theorem 3.2, $\bar{\mathbb{A}}$ has a near unanimity polymorphism h. Using Proposition 3.1 again we have $h = \bar{f}$ for some polymorphism f of $\mathbb{A}$, and f is clearly a near unanimity operation. ■

4 CSP Instance Associated with a Binary Relational Structure

Definition 4.1 An *instance of the constraint satisfaction problem* (CSP) is a triple $P = (V, A, \mathcal{C})$ with

- V a nonempty, finite set of *variables*,
- A a nonempty, finite *domain*,
- $\mathcal{C}$ a finite set of *constraints*, where each constraint is a pair $C = (\mathbf{x}, R)$ with
  - $\mathbf{x}$ a tuple of variables of length n, called the *scope of* C, and
  - R an n-ary relation on A, called the *constraint relation of* C.

Let $\mathbf{A}$ be a finite idempotent algebra. An *instance of the CSP over* $\mathbf{A}$, denoted by CSP($\mathbf{A}$), is an instance such that all constraint relations are subpowers of $\mathbf{A}$.

A *solution* to an instance P is a function $f\colon V \to A$ such that, for each constraint $C = (\mathbf{x}, R) \in \mathcal{C}$, the tuple $f(\mathbf{x})$ belongs to R.

Remark 4.2 The CSP is often parametrized by relational structures: an instance whose constraint relations are in a relational structure $\mathbb{A}$ is called an instance of CSP($\mathbb{A}$). It was proved in [15] that the computational complexity of deciding whether an instance of CSP($\mathbb{A}$) has a solution is fully determined, at least when $\mathbb{A}$ has finitely many relations, by the algebra $\mathbf{A} = (A, \mathrm{Pol}(\mathbb{A}))$. Moreover, Bulatov, Jeavons, and Krokhin proved in [8] that the complexity depends only on the variety generated by $\mathbf{A}$ (*i.e.,* the smallest variety containing $\mathbf{A}$). These results are at the heart of the connection between universal algebra and the CSP mentioned in the introduction.

For simplicity we will formulate our definitions and results for a special type of CSP instance with a single binary constraint for each pair of variables, although most of the material can be generalized.

Definition 4.3 An instance $P = (V, A, \mathcal{C})$ of the CSP is called a *simple binary instance* if

- $\mathcal{C} = \{((x_1, x_2), R^P_{x_1,x_2}) : x_1, x_2 \in V\}$,
- $R^P_{x_2,x_1} = {R^P_{x_1,x_2}}^{-1}$ $(= \{(b, a) : (a, b) \in R^P_{x_1,x_2}\})$ for every $x_1, x_2 \in V$, and
- $R^P_{x,x} \subseteq \{(a, a) : a \in A\}$ for every $x \in V$.

We omit the superscript P if the instance is clear from the context.

A simple binary instance can be drawn as a $|V|$-partite graph in the following way. Each part is a copy of A, one for each variable $x \in V$ (the parts are now commonly referred to as potatoes), and elements of R_{x_1,x_2} are edges between the corresponding copies of A. Solutions then correspond to cliques with V vertices (with one vertex in each part).

To every binary relational structure $\mathbb{A}$ and natural number n we can associate, in a natural way, a simple binary instance $P(\mathbb{A}, n)$ of CSP($(A, \mathrm{Pol}(\mathbb{A}))$) whose solutions are precisely the n-ary polymorphisms of $\mathbb{A}$.

Definition 4.4 Let $\mathbb{A}$ be a binary relational structure and let $n \geq 2$ be a natural number. The instance $P(\mathbb{A}, n) = (V, A, \mathcal{C})$ is defined by

- $V = A^n$
- $R^P_{(a_1,\dots,a_n),(b_1,\dots,b_n)} = \{(t(a_1, \dots, a_n), t(b_1, \dots, b_n)) : t \in \mathrm{Pol}_n(\mathbb{A})\}$

Note that $P(\mathbb{A}, n)$ is indeed an instance of CSP($(A, \mathrm{Pol}(\mathbb{A}))$).

Proposition 4.5 *For every binary relational structure $\mathbb{A}$, the set of solutions of $P(\mathbb{A}, n)$ is equal to* $\mathrm{Pol}_n(\mathbb{A})$.

Proof It is clear that every n-ary polymorphism of $\mathbb{A}$ is a solution of $P(\mathbb{A}, n)$.

Let f be a solution to $P(\mathbb{A}, n)$. We have to show that every relation R of $\mathbb{A}$ is preserved by f, but this is easy. If R is binary and n-tuples $(a_1, \dots, a_n), (b_1, \dots, b_n) \in A^n$ are such that $(a_i, b_i) \in R$ for each $1 \leq i \leq n$, then $(f(a_1, \dots, a_n), f(b_1, \dots, b_n)) \in R_{(a_1,\dots,a_n),(b_1,\dots,b_n)}$ (as f is a solution), therefore

$$\big(f(a_1, \dots, a_n), f(b_1, \dots, b_n)\big) = \big(t(a_1, \dots, a_n), t(b_1, \dots, b_n)\big)$$

for some $t \in \mathrm{Pol}_n(\mathbb{A})$. Since t is a polymorphism, the right hand side is an element of R. The proof for a unary relation R can be done, for instance, by using this reasoning for the relation $R \times A$. ∎

We are interested in near unanimity polymorphisms, solutions f of $P(\mathbb{A}, n)$ satisfying the additional conditions $f(a, a, \dots, a, b, a, a, \dots, a) = a$ (for any $a, b \in A$ and any position of b in the tuple). Therefore the following notion of a restriction of an instance will be useful.

Definition 4.6 Let $P = (V, A, \mathcal{C})$ be a simple binary instance of CSP and let $\mathcal{J} = \{J_x : x \in V\}$ be a family of subsets of A. By the *restriction of P to $\mathcal{J}$* we mean the simple binary instance $P_{|\mathcal{J}} = (V, A, \mathcal{C}')$ with

$$R^{P_{|\mathcal{J}}}_{x_1,x_2} = R^{P}_{x_1,x_2} \cap (J_{x_1} \times J_{x_2})$$

for every $x_1, x_2 \in V$.

To find an n-ary polymorphism of a binary relational structure $\mathbb{A}$, we will consider the instance $P = P(\mathbb{A}, n)$ and its restriction to the family $\mathcal{J} = \{J_x : x \in V\}$, where $J_{(a,a,\dots,a,b,a,a,\dots,a)} = \{a\}$ (for every $a, b \in A$ and every position of b in the tuple), and $J_{(a_1,\dots,a_n)} = A$ otherwise. With this choice, the set of solutions of $P_{|\mathcal{J}}$ coincides with the set of n-ary near unanimity polymorphisms of $\mathbb{A}$. We show that this set is nonempty in two steps. First we prove that $P_{|\mathcal{J}}$ contains a subinstance that is "consistent enough", and then we apply a result from [3] saying that such instances always have a solution.

In the next section we introduce the required consistency notions.

5 Consistency Notions, Proof of Theorem 3.2

5.1 $(1, 2)$-systems

Definition 5.1 Let $P = (V, A, \mathcal{C})$ be a simple binary instance and let $\{R_x : x \in V\}$ be a family of nonempty subsets of A. We say that P is a $(1, 2)$*-system with unary projections* $\{R_x : x \in V\}$, if, for any $x_1, x_2 \in V$, the projection of R_{x_1,x_2} to the first coordinate is equal to R_{x_1}. (It follows that the projection to the second coordinate is equal to R_{x_2}.)

If, moreover, $\mathbf{A}$ is an algebra and P is an instance of $\mathrm{CSP}(\mathbf{A})$ we say that P is a $(1, 2)$*-system over* $\mathbf{A}$.

Observe that if P is a $(1,2)$-system over $\mathbf{A}$, then each R_x is a subuniverse $\mathbf{A}$ (since the set R_x is equal to the projection of $R_{x,x}$ to the first coordinate). In this case we denote the subalgebra of $\mathbf{A}$ with universe R_x by $\mathbf{R}_x$.

Also note that the instance $P(\mathbb{A}, n)$ introduced in Section 4 is always a $(1,2)$-system with unary projections $\{R_x : x \in V\}$, where

$$R_{(a_1,\ldots,a_n)} = \{t(a_1,\ldots,a_n) : t \in \mathrm{Pol}_n(\mathbb{A})\}.$$

When a simple binary instance P is drawn as a multipartite graph (see the note after Definition 4.3), then P is a $(1,2)$-system if and only if, for every pair x_1, x_2 of variables, every vertex $a \in R_{x_1}$ is adjacent to at least one vertex from R_{x_2} and to no vertex outside R_{x_2} (in particular, vertices outside the sets R_x are isolated).

Whether an instance has a restriction that is a (nonempty) $(1,2)$-system can be decided using trees.

Definition 5.2 Let $P = (V, A, \mathcal{C})$ be a simple binary instance. A *P-tree* T is a tree (*i.e.,* an undirected connected graph without loops or cycles) whose vertices are labeled by variables in V. The vertex set of T is denoted by $\mathrm{vert}(T)$ and the label of a vertex $v \in \mathrm{vert}(T)$ by $\mathrm{lbl}(v)$.

A *realization* of a P-tree T in P is a mapping $r\colon \mathrm{vert}(T) \to A$ such that $(r(v_1), r(v_2)) \in R_{\mathrm{lbl}(v_1),\mathrm{lbl}(v_2)}$ whenever v_1, v_2 are adjacent vertices of T. For a vertex v of T we put

$$T[v] = \{r(v) : r \text{ is a realization of } T \text{ in } P\}.$$

If P is a $(1,2)$-system with unary projections $\{R_x : x \in V\}$, then every P-tree is clearly realizable. Moreover, for every P-tree T and every vertex v of T, we have $T[v] = R_{\mathrm{lbl}(v)}$. The following proposition provides a converse to this observation.

Proposition 5.3 *Let $P = (V, A, \mathcal{C})$ be a simple binary instance over an algebra $\mathbf{A}$. If every P-tree is realizable in P then, for every $x \in V$, the set*

$$R_x = \bigcap_{\substack{T \text{ is a } P\text{-tree} \\ v \in \mathrm{vert}(T) \\ \mathrm{lbl}(v) = x}} T[v]$$

is nonempty and $P_{|\{R_x : x \in V\}}$ is a $(1,2)$-system over $\mathbf{A}$.

Proof Since A is a finite set, each R_x can be obtained by intersecting the sets $T[v]$ for only finitely many P-trees T. Moreover, there exists a single tree T_x with vertex v_x labeled by x such that $R_x = T_x[v_x]$. We take the disjoint union of the finite collection of trees and identify the vertices v to a single vertex. It follows that R_x is nonempty for every $x \in V$.

Next we prove that $P_{|\{R_x : x \in V\}}$ is a $(1,2)$-system. It is enough to show that for every $x_1, x_2 \in V$ and every $a_1 \in R_{x_1}$ there exists $a_2 \in R_{x_2}$ such that $(a_1, a_2) \in R_{x_1,x_2}$. Consider the P-tree T constructed from T_{x_2} by adding a vertex v_1 adjacent to v_{x_2} with label x_1. This P-tree has a realization r such that $r(v_1) = a_1$ (since $R_{x_1} \subseteq T[v_1]$). Now we can put $a_2 = r(v_{x_2})$, because

$$\big(r(v_1), r(v_2)\big) \in R_{x_1,x_2} \quad \text{and} \quad r(v_2) \in T[v_{x_2}] \subseteq T_{x_2}[v_{x_2}] = R_{x_2}.$$

Finally, we have to show that $P_{|\{R_x : x \in V\}}$ is an instance of CSP($\mathbf{A}$). It is clearly enough to prove that R_x $(= T_x[v_x])$ is a subuniverse of $\mathbf{A}$ for every $x \in V$. But this is a straightforward consequence of the definitions: for any P-tree T, any operation t of $\mathbf{A}$ (say, k-ary) and any k-tuple of realizations $r_1, \dots, r_k$ of T in P, the mapping r defined by $r(v) = t(r_1(v), \dots, r_k(v))$ is a realization of T in P (as R_{x_1,x_2} is a subuniverse of $\mathbf{A}^2$ for every $x_1, x_2 \in V$). ■

Remark 5.4 The family $\mathcal{R} = \{R_x : x \in V\}$ from the previous proposition is actually the largest family such that $P_{|\mathcal{R}}$ is a $(1, 2)$-system. Also observe that if some P-tree is not realizable, then no such a family exists.

5.2 $(2, 3)$-systems

A $(2, 3)$-system is a $(1, 2)$-system such that every edge extends to a triangle:

Definition 5.5 A $(1, 2)$-system $P = (V, A, \mathcal{C})$ is called a $(2, 3)$-*system* if for every $x_1, x_2, x_3 \in V$ and every $(a_1, a_2) \in R_{x_1,x_2}$ there exists $a_3 \in A$ such that $(a_1, a_3) \in R_{x_1,x_3}$ and $(a_2, a_3) \in R_{x_2,x_3}$.

Examples of $(2, 3)$-system include the instances $P(\mathbb{A}, n)$.

The following theorem is the main new ingredient for the proof of the Zádori conjecture. It is proved in Section 6.

Theorem 5.6 *Let $P = (V, A, \mathcal{C})$ be a $(2, 3)$-system with unary projections $\{R_x : x \in V\}$ over a CD algebra $\mathbf{A}$ and let $\mathcal{J} = \{J_x : x \in V\}$ be a family of (nonempty) subsets of A such that each J_x is a Jónsson ideal of $\mathbf{R}_x$. If all P-trees with at most $4^{8^{|A|}}$ vertices are realizable in $P_{|\mathcal{J}}$, then all P-trees are realizable in $P_{|\mathcal{J}}$.*

The core result of [3] states that *every* $(2, 3)$-*system over a CD algebra has a solution* ([3, Theorem 5.2]). We will need a refinement proved (although not explicitly stated) in the same article.

Theorem 5.7 *Let $P = (V, A, \mathcal{C})$ be a $(2, 3)$-system with unary projections $\{R_x : x \in V\}$ over a CD algebra $\mathbf{A}$ and let $\mathcal{J} = \{J_x : x \in V\}$ be a family of (nonempty) subsets of A such that each J_x is a Jónsson ideal of $\mathbf{R}_x$. If $P_{|\mathcal{J}}$ is a $(1, 2)$-system, then $P_{|\mathcal{J}}$ has a solution.*

Remark 5.8 The method used to prove [3, Theorem 5.2] was the following. If $\mathcal{J}$ satisfies the assumptions (such families are called absorbing systems in [3]) and some of the sets J_x have more than one element, then it is possible ([3, Lemma 6.9]) to find a family $\mathcal{J}' = \{J'_x : x \in V\}$ such that $\mathcal{J}'$ satisfies the same conditions, $J'_x \subseteq J_x$ and at least one of these inclusions is proper. In this way we eventually get a solution to $\mathcal{P}_{|\mathcal{J}}$.

More recently, the result that every $(2, 3)$-system over a CD algebra $\mathbf{A}$ has a solution was generalized in two directions. First, a weaker consistency notion than $(2, 3)$-system is enough to guarantee a solution. It suffices to assume that the instance is a so-called *Prague strategy* (see [4]). A more "modern" proof of Theorem 5.7 would be to show that $P_{|\mathcal{J}}$ is a Prague strategy (which is not hard).

A weaker condition can also be put on the algebra. It is enough to assume that **A** lies in a meet semi-distributive variety (actually, for an idempotent finite algebra **A**, the statement "every $(2,3)$-system (or Prague strategy) over **A** has a solution" is equivalent to "**A** lies in a meet semi-distributive variety" [4]).

5.3 Proof of the Zádori Conjecture

We are ready to prove the main theorem. As discussed in Section 3, it is enough to prove Theorem 3.2.

Theorem 3.2 *If $\mathbb{A}$ is a finite binary relational structure such that* $(A, \mathrm{Pol}(\mathbb{A}))$ *is a CD algebra, then $\mathbb{A}$ has a near unanimity polymorphism.*

Proof Let $p_0, \dots, p_s$ be a Jónsson chain of operations of the algebra $(A, \mathrm{Pol}(\mathbb{A}))$. Let **A** be the algebra with universe A whose operations are *idempotent* polymorphisms of $\mathbb{A}$. Since p_i's are idempotent, $p_0, \dots, p_s$ is a Jónsson chain for the algebra **A**.

Let n be a natural number greater than $4^{8^{|A|}}$ and let $P = P(\mathbb{A}, n)$. It was observed above that P is a $(2,3)$-system over **A** with unary projections $\{R_x : x \in V\}$.

Let $\mathcal{J} = \{J_x : x \in V\}$, where $J_{(a,a,\dots,a,b,a,a,\dots,a)} = \{a\}$ (for every $a, b \in A$ and every position of b in the tuple), and $J_{(a_1,\dots,a_n)} = A$ otherwise. Since **A** is idempotent, each J_x is a Jónsson ideal of $\mathbf{R}_x$. As discussed in Section 4, the solutions to the instance $Q = P_{|\mathcal{J}}$ are n-ary near unanimity polymorphisms of $\mathbb{A}$; therefore, it is enough to show that Q has a solution.

First we observe that every P-tree T with at most $n-1$ vertices is realizable in Q. Indeed, the variables are n-tuples and T has less than n vertices; therefore, there exists a natural number i (with $1 \leq i \leq n$) such that b is *not* on the i-th position of any tuple $x = (a, a, \dots, a, b, a, a, \dots a)$, $a \neq b$ which is a label of a vertex of T. Then the mapping assigning a_i to a vertex of label $(a_1, \dots, a_n)$ is a realization of T in Q.

By Theorem 5.6 every tree is realizable in Q.

Proposition 5.3 (applied to the simple binary instance Q) gives us a system $\mathcal{J}' = \{J'_x : x \in V\}$ such that $Q_{|\mathcal{J}'}$ is a $(1,2)$-system with unary projections $\mathcal{J}'$.

For every $x \in V$, J'_x is a Jónsson ideal of $\mathbf{R}_x$. Indeed, in the proof of Proposition 5.3 we have shown that $J'_x = T_x^Q[v_x]$ for certain tree T_x and its vertex x. If $a_1, a_2 \in J'_x$ and $b \in R_x$, then there exists a realization r_1 (resp. r_2) of T_x in Q such that $r_1(v_x) = a_1$ (resp. $r_2(v_x) = a_2$), and, since P is a $(1,2)$-system, there exists a realization r_3 of T_x in P such that $r_3(v_x) = b$. Now we apply the Jónsson term operation p_i to r_1, r_2, r_3 (in the same way as in the last paragraph of the proof of Proposition 5.3), and we get a realization r of T_x in P such that $r(v_x) = p_i(a_1, b, a_2)$. From the assumption that $J_{x'}$ is a Jónsson ideal of $\mathbf{R}_{x'}$ (for every $x' \in V$), it follows that r is a realization in Q. Therefore, $p_i(a_1, b, a_2) \in T_x^Q[v_x] = J'_x$.

Finally, P is a $(2,3)$-system, $\mathcal{J}'$ is formed by Jónsson ideals of appropriate R_x's and $P_{|\mathcal{J}'}$ $(= Q_{|\mathcal{J}'})$ is a $(1,2)$-system; thus, by Theorem 5.7, $P_{|\mathcal{J}'}$ has a solution, which is of course also a solution to $P_{|\mathcal{J}}$. ∎

6 Proof of Theorem 5.6

Theorem 5.6 *Let $P = (V, A, \mathcal{C})$ be a $(2,3)$-system with unary projections $\{R_x : x \in V\}$ over a CD algebra $\mathbf{A}$ and let $\mathcal{J} = \{J_x : x \in V\}$ be a family of subsets of A such that each J_x is a Jónsson ideal of $\mathbf{R}_x$. If all P-trees with at most $4^{8^{|A|}}$ vertices are realizable in $P_{|\mathcal{J}}$, then all P-trees are realizable in $P_{|\mathcal{J}}$.*

We argue by contradiction. We take a tree that is not realizable in $P_{|\mathcal{J}}$, and we eventually obtain a configuration (a tuple $(\mathbf{B}, L, U, E, F, a, b)$) that will contradict the following auxiliary result. In this lemma we look at the binary relations E, F on B as digraphs.

Lemma 6.1 *Let $\mathbf{B}$ be a finite CD algebra and let $U, L \subseteq B$, $E, F \leq \mathbf{B}^2$, $a, b \in B$ be such that*

- *E is a Jónsson ideal of F;*
- *U is disjoint from L;*
- *$a \in U$, $b \in L$, $(a, b) \in F$;*
- *the digraph $E \cap U^2$ has no sources (that is, for all $c \in U$ there exists $d \in U$ such that $(d, c) \in E$);*
- *the digraph $E \cap L^2$ has no sinks (that is, for all $c \in L$ there exists $d \in L$ such that $(c, d) \in E$).*

Then there exist $c \in U$ and $d \in B \setminus U$ such that $(c, d) \in E$.

Proof We take a counterexample to the lemma and fix a Jónsson chain $p_0, p_1, \ldots, p_s$ of term operations of $\mathbf{B}$. We may assume that $\mathbf{B}$ is idempotent, otherwise we can replace $\mathbf{B}$ by the algebra $(B, \{p_0, p_1, \ldots, p_s\})$.

Let us quickly sketch the proof on a smallest choice that does not satisfy the conclusion:

$$B = \{1,2\}, \quad U = \{1\}, \quad L = \{2\}, \quad E = \{(1,1),(2,2)\},$$
$$F = \{(1,1),(2,2),(1,2)\}, \quad a = 1, \quad b = 2.$$

The first step of the proof is to transform our counterexample into a form closer to this simplest one. Next we prove that E must at least contain the edge $(2,1)$, and finally we show that in this case we would have a directed path from 1 to 2 in the digraph E.

Since $E \cap U^2$ has no sources, we can find a sequence $a = a_1, a_2, \ldots$ of elements in U such that $(a_{i+1}, a_i) \in E$ for all i. As U is finite, there are positive numbers k and l such that $a_k = a_{k+l}$. Similarly, we find a sequence $b = b_1, b_2, \ldots$ of elements in L such that $(b_i, b_{i+1}) \in E$ and positive numbers k' and l' such that $b_{k'} = b_{k'+l'}$. Let m be a natural number greater than or equal to $k + k' - 1$ and divisible by l and l', let

$$E' = \underbrace{E \circ E \circ \cdots \circ E}_{m\text{-times}}, \quad F' = \underbrace{F \circ F \circ \cdots \circ F}_{m\text{-times}}, \quad U' = U, \quad \mathbf{B}' = \mathbf{B}, \quad a' = a_k,$$

where $\circ$ denotes the composition of relations defined by

$$S \circ S' = \{(s, s'') : \exists s' \in B \ (s, s') \in S, (s', s'') \in S'\}$$

and let $b' \in \{b_{k'}, b_{k'+1}, \ldots, b_{k'+l'}\}$ be an element such that there exists a directed path in the digraph F from a' to b' of length m (we can take the element of appropriate distance from a' on the path $a' = a_k, a_{k-1}, \ldots, a_1, b_1, b_2, \ldots, b_{k'}, b_{k'+1}, \ldots, b_{k'+l'} = b_{k'}, b_{k'+1}, \ldots$).

These new sets $\mathbf{B}', E', F', U'$ and elements a', b' have the following properties.

- *$\mathbf{B}'$ is a CD algebra, $E' \leq F' \leq \mathbf{B}'^2$, E' is a Jónsson ideal of F'.* This is straightforward. (That E', F' are subalgebras follows from a more general fact that any relation positively primitively defined from subpowers is a subpower, but it is easy to check the claims directly.)
- *$a' \in U'$, $b' \in B' \setminus U'$, $(a', a') \in E'$, $(b', b') \in E'$, $(a', b') \in F'$.* We have chosen b' so that there exists a directed path in F of length m between a' and b', thus $(a', b') \in F'$. Since a' (resp. b') are in a closed path of length l (resp. l') and this length divides m, it follows that $(a', a'), (b', b') \in E'$.
- *There do not exist $c \in U'$, $d \in B' \setminus U'$ such that $(c, d) \in E'$.* Otherwise there is a path in E from c to d in E, which is impossible as there is no edge from U' to $B' \setminus U'$.

We will show that it is impossible to find $\mathbf{B}', U', E', F', a', b'$ satisfying the conditions above. For contradiction, assume that $\mathbf{B}', U', E', F', a', b'$ satisfy these three conditions and $|B'|$ is the smallest possible.

The minimality assumption has some useful consequences.

- *$(c, c) \in E'$ for any $c \in B'$.* Otherwise the following choice would form a smaller counterexample: $B'' = \{c : (c, c) \in E'\}$, $U'' = U' \cap B''$, $E'' = E' \cap B''^2$, $F'' = F' \cap B''^2$, $a'' = a'$, $b'' = b'$. That B'' is a subuniverse of $\mathbf{B}'$ is straightforward to check (and it again follows from the general fact about positive primitive definitions of subpowers).
- *$(a', c) \in F'$ for any $c \in B'$.* Otherwise we could take $B'' = \{c : (a', c) \in F'\}$ and restrict all the sets to B'' as above, *i.e.,* $U'' = U' \cap B''$, $E'' = E' \cap B''^2$, $F'' = F' \cap B''^2$, $a'' = a'$, $b'' = b'$. Note that we need idempotency to show that B'' is a subuniverse of $\mathbf{B}'$ (B'' is defined using F' and the subuniverse $\{a'\}$ of $\mathbf{B}'$).
- *$(c, d) \in F'$ for any $c \in U'$, $d \in B' \setminus U'$.* Otherwise we take $B'' = \{e : (e, d) \in F'\}$, $a'' = a'$, $b'' = d$ and, again, restrict U', E', F' to B''. From the first item it follows that $(b'', b'') \in E''$, and the second item implies that $(a'', b'') \in F''$.

Now we consider the sequence

$$
\begin{aligned}
a' = p_1(a', a', b'),\ & p_1(a', b', b') = p_2(a', b', b'),\ p_2(a', a', b') \\
&= p_3(a'a', b'), \ldots, p_{s'}(a', b', b') = b',
\end{aligned}
$$

where $s' = s$ if s is odd and $s' = s - 1$ if s is even.

As $(a', a'), (b', b') \in E'$, $(a', b') \in F'$, and E' is a Jónsson ideal of F', it follows that the first pair of elements of this sequence is in E'. Similarly, the second pair is in E'^{-1}, the third pair in E', and so on. Thus we have a "fence" in E' from a' to b', and, since we are assuming that there are no $c \in U', d \in B' \setminus U'$ such that $(c, d) \in E'$, there must exist elements $c \in U'$ and $d \in B' \setminus U'$ such that $(d, c) \in E'$.

We have $(c, c), (d, d), (d, c) \in E'$ and $(c, d) \in F'$. It follows that

$$c = p_1(c, c, d), p_1(c, d, d) = p_2(c, d, d), p_2(c, c, d) = p_3(c, c, d), \ldots, d$$

is a sequence where all the pairs are in E'. This contradicts the assumption that there is no element in U' that is E'-related to an element outside U'. ■

For the remainder of this section we fix P, R_x's, and $\mathcal{J}$ satisfying the hypotheses of Theorem 5.6, and we assume that there exists a tree that is not realizable in $P_{|\mathcal{J}}$.

To obtain a configuration contradicting the previous lemma we will first transform our non-realizable tree to a tree whose every vertex has degree 1 or 3 and that has no realization in P with leaves realized in $P_{|\mathcal{J}}$. We require the following definition.

Definition 6.2 Let T be a P-tree and let S be a subset of vertices of T. A realization r of T in P is called an *S-realization* if $r(v) \in J_{\mathrm{lbl}(v)}$ for every $v \in S$.

For a vertex v of T we define

$$T_S[v] = \{r(v) : r \text{ is an } S\text{-realization of } T \text{ in } P\}$$

The set S from the definition will often be the set of all leaves of T, which we denote by leaves(T).

Lemma 6.3 *There exists a P-tree T such that*

- *the degree of any vertex of T is* 1 *or* 3*;*
- *T has no* leaves(T)*-realization;*
- *T has a S-realization for every proper subset S of* leaves(T).

Proof We start with a P-tree T that is not realizable in $P_{|\mathcal{J}}$. To every inner vertex (*i.e.,* a vertex of degree greater than one) we add an adjacent vertex with the same label. Since $R_{x,x}$ is a subset of the equality relation, any realization maps the new leaf to the same element of A as the inner vertex. It follows that the new tree is not leaves(T)-realizable.

In a similar way we can modify the tree so that all the vertices have degree at most 3. If a vertex v has degree at least 4, we can split it into two adjacent vertices v_1, v_2 with the same label in such a way that v_1 is adjacent to 2 of the original neighbors of v and v_2 is adjacent to the remaining neighbors. Clearly, v_1 and v_2 have smaller degree than v; therefore, we can repeat this splitting procedure until we obtain a tree whose every vertex has degree at most 3 and that is not leaves(T)-realizable.

Let T be such a tree with minimal number of vertices.

Now we show that T has no vertex of degree 2. Suppose otherwise, that is, there is a vertex v with precisely two neighbors v_1, v_2. The tree T' obtained by removing the vertex v and adding the edge $v_1 - v_2$ is smaller than T, therefore T' has a leaves(T')-realization r'. As $(r'(v_1), r'(v_2)) \in R_{\mathrm{lbl}(v_1),\mathrm{lbl}(v_2)}$ and P is a $(2,3)$-system, there exists $a \in A$ such that $(r'(v_1), a) \in R_{\mathrm{lbl}(v_1),\mathrm{lbl}(v)}$ and $(r'(v_2), a) \in R_{\mathrm{lbl}(v_2),\mathrm{lbl}(v)}$ (This is the only place in this section where we use the assumption that P is a $(2,3)$-system. For the rest it would suffice to assume that P is a $(1,2)$-system.) It follows that the extension r of the mapping r' by $r(v) = a$ is a leaves(T)-realization of T, a contradiction.

It remains to show that T is S-realizable for every proper subset S of leaves(T), but this is easy. If we remove a leaf outside S, the remaining tree is S-realizable (from the minimality of T), and this realization can be extended to an S-realization of T as P is a $(1,2)$-system. ■

For the remainder of the proof we fix a P-tree T with the properties stated in the previous lemma.

Lemma 6.4 *T contains a path of length at least* $2 \cdot 8^{|A|}$.

Proof It can be easily computed that a tree, which has all vertices of degree at most 3 and which does not contain any path with more than k vertices, has size at most 2^k (this is a crude estimate, one computes that the most accurate estimate is $3 \cdot 2^{k/2} - 2$ for even k and $2^{(k+3)/2} - 2$ for odd $k > 1$).

Since T has more than $4^{8^{|A|}}$ vertices by our assumption (smaller P-trees are even realizable in $P_{|\mathcal{J}}$), the claim follows. ■

We fix a subpath $v_1, v_2, \dots, v_m$ of T, where $m \geq 2 \cdot 8^{|A|}$. We define subsets S_i of leaves(T), $i = 1, 2, \dots, m$, as follows. A leaf of T is in S_i if and only if the shortest path from this leaf to v_i contains neither v_{i-1} nor v_{i+1}. (For v_1 only the vertex v_2 is considered. If v_1 is a leaf, then $S_1 = \{v_1\}$. Similarly for v_m.) In other words, we straighten the line $v_1, \dots, v_m$ and shake the tree. Then S_i is the set of leaves below v_i.

The next lemma will enable us to find the sought after configuration.

Lemma 6.5 *There exist natural numbers k, l such that*

- $1 \leq k, l \leq m, k \leq l + 2$;
- $T_{S_k}[v_k] = T_{S_l}[v_l]$;
- $T_{S_1 \cup S_2 \cup \dots \cup S_k}[v_k] = T_{S_1 \cup S_2 \cup \dots S_l}[v_l] \neq \emptyset$;
- $T_{S_k \cup S_{k+1} \cup \dots \cup S_m}[v_k] = T_{S_l \cup S_{l+1} \cup \dots \cup S_m}[v_l] \neq \emptyset$.

Proof There is at least $m/2 - 1 \geq 8^{|A|} - 1$ even numbers less than m. For each such number i we consider the triple

$$\left(T_{S_i}[v_i], T_{S_1 \cup \dots \cup S_i}[v_i], T_{S_i \cup \dots \cup S_m}[v_i] \right)$$

of subsets of A (note that these subsets are nonempty by the third item of Lemma 6.3). There are less than $(2^{|A|} - 1)^3 < 8^{|A|} - 2$ possible triples, therefore, by the pigeonhole principle, there exist distinct k, l with the same associated triples and the lemma follows.

Again, the estimates we used are very rough. For instance, the second and third sets in the triple are disjoint subsets of the first subset. This significantly reduces the number of possibilities, etc. ■

Let

$$Q_1 = S_1 \cup S_2 \cup \dots \cup S_k, \quad Q_2 = S_k \cup \dots \cup S_l, \quad Q_3 = S_l \cup \dots \cup S_m.$$

Now we define

$$
\begin{aligned}
B &= T_{S_k}[v_k] = T_{S_l}[v_l],\\
L &= T_{Q_2 \cup Q_3}[v_k] = T_{Q_3}[v_l],\\
U &= T_{Q_1}[v_k] = T_{Q_1 \cup Q_2}[v_l],\\
E &= \{(r(v_k), r(v_l)) : r \text{ is a } Q_2\text{-realization of } T\},\\
F &= \{(r(v_k), r(v_l)) : r \text{ is a } (S_k \cup S_l)\text{-realization of } T\},\\
(a,b) &= (r(v_k), r(v_l)) \ \text{ for a chosen } (Q_1 \cup Q_3)\text{-realization } r \text{ of } T.
\end{aligned}
$$

Since $k \leq l-2$ and $S_{k+1} \neq \varnothing$ (by the first item of Lemma 6.3), $Q_1 \cup Q_3$ is a proper subset of leaves(T); therefore, T has a $(Q_1 \cup Q_3)$-realization by the third item of Lemma 6.3, and the definition of a and b makes sense. This choice satisfies all the assumptions of Lemma 6.1:

- *B is a subuniverse of* $\mathbf{A}$. It follows directly from the definitions (see the last paragraph of the proof of Proposition 5.3). Let $\mathbf{B}$ be the subalgebra of $\mathbf{A}$ with universe B.
- *$E, F \leq \mathbf{B}^2$, E is a Jónsson ideal of F.* This is also straightforward. That E is a Jónsson ideal of F follows from the assumption that J_x is a Jónsson ideal of $\mathbf{R}_x$ for every $x \in Q_2$.
- *U and L are disjoint.* Suppose $c \in U \cap L$. Since $U = T_{Q_1}[v_k]$, there exists a Q_1-realization r_1 of T such that $r_1(v_k) = c$. Similarly, since $L = T_{Q_2 \cup Q_3}[v_k]$, there exists a $(Q_2 \cup Q_3)$-realization r_2 of T such that $r_2(v_k) = c$. The realizations r_1 and r_2 can be joined in the following way. We put $r(v) = r_1(v)$ for vertices v whose shortest path to v_k does not contain v_{k+1}, and $r(v) = r_2(v)$ for the other vertices. Now r is a $(Q_1 \cup Q_2 \cup Q_3)$-realization of T. But $Q_1 \cup Q_2 \cup Q_3$ is the set of all leaves of T, a contradiction (see the second item of Lemma 6.3).
- *$a \in U, b \in L, (a,b) \in F$.* The element a is defined as $r(v_k)$ for a $(Q_1 \cup Q_3)$-realization r of T. Since $Q_1 \subseteq Q_1 \cup Q_3$, we have $T_{Q_1 \cup Q_3}[v_k] \subseteq T_{Q_1}[v_k] = U$; therefore, $a = r(v_k) \in U$. Similarly, $b \in L$ follows from $b = r(v_l)$, $Q_3 \subseteq Q_1 \cup Q_3$ and $L = T_{Q_3}[v_l]$, and $(a,b) \in F$ follows from $S_k \cup S_l \subseteq Q_1 \cup Q_3$.
- *$E \cap U^2$ has no sources, $E \cap L^2$ has no sinks.* Let c be an arbitrary element of U. Since $U = T_{Q_1 \cup Q_2}[v_l]$, there exists a $(Q_1 \cup Q_2)$-realization r of T such that $r(v_l) = c$. But r is also a Q_2-realization of T, hence $(r(v_k), r(v_l)) \in E$. The element $d = r(v_k)$ lies in $T_{Q_1 \cup Q_2}[v_k] \subseteq T_{Q_1}[v_k] = U$. We can analogically show that $E \cap L^2$ has no sinks: any $c \in L$ is equal to $r(v_k)$ for a $(Q_2 \cup Q_3)$-realization r of T, and $r(v_l) \in T_{Q_3}[v_l] = L$.
- *There do not exist $c \in U$ and $d \in B \setminus U$ such that $(c,d) \in E$.* If $c \in U = T_{Q_1}[v_k]$ and $(c,d) \in E$, then there exists a Q_1-realization r_1 of T and a Q_2-realization r_2 of T such that $r_1(v_k) = c = r_2(v_k)$ and $r_2(v_l) = d$. When we join r_1 and r_2 in the same way as in the proof that U and L are disjoint, we get a $(Q_1 \cup Q_2)$-realization r of T such that $r(v_l) = d$. But $U = T_{Q_1 \cup Q_2}[v_l]$, thus $d \in U$.

The last property contradicts Lemma 6.1, and this concludes the proof of Theorem 5.6. ■

7 Conclusion

7.1 Decidability of Near Unanimity for Relational Structures

As a corollary of the main theorem we obtain an affirmative answer to the near unanimity problem for relations.

Corollary 7.1 *It is decidable whether a finite relational structure with finitely many relations has a near unanimity polymorphism.*

Proof It is enough to decide whether the given relational structure has a Jónsson chain of polymorphisms. This can be decided as follows. We first compute the set P of all ternary idempotent polymorphisms satisfying $p(a,b,a) = a$ and then compute the graph whose vertices are idempotent binary operations having f and g the vertices of an edge if and only if there exist $p_1, p_2 \in P$ such that for all a, b,

$$f(a,b) = p_1(a,a,b), p_1(a,b,b) = p_2(a,b,b) \quad \text{and} \quad g(a,b) = p_2(a,a,b).$$

A Jónsson chain exists if and only if π_1 is connected to π_2, where π_i is the idempotent binary operation that is the projection operation on the i-th coordinate. ■

It was shown in [22] that the corresponding decision problem for algebras (that is, does a given finite algebra with finitely many operations have a near unanimity term operation?) is decidable. This was a surprising development after undecidability results about closely related questions [21].

The naive algorithm described in the proof of Corollary 7.1 runs in exponential time.

Open Problem 7.2 *Determine the computational complexity of deciding whether a finite relational structure with finitely many relations has a near unanimity polymorphism.*

There exist polynomial time algorithms for finite posets [17] and for finite reflexive undirected graphs [18].

The complexity of the same problem for algebras is also unknown. There exists a polynomial time algorithm for deciding whether a finite idempotent algebra (with finitely many operations) is a CD algebra, and the same problem without assuming idempotency is exponential time complete [12].

7.2 Arities

Our proof gives some upper bound on the minimal arity of a near unanimity polymorphism; namely, a binary relational structure $\mathbb{A}$ either has a near unanimity polymorphism of arity $4^{8^{|A|}} + 1$ or has none. From the reduction presented in Section 3 it follows that for a relational structure whose relations have maximum arity k an upper bound is $4^{8^{|A|^k}} + 1$. We have used quite rough estimates in a couple of places; however, this proof most likely cannot provide a better upper bound than doubly exponential.

For finite algebras with finitely many operations the upper bound also exists, but is tremendously large and is not even computed in [22].

Therefore we have the following open problem.

Open Problem 7.3 *Give a better upper bound for the minimal arity of a near unanimity polymorphism (resp. term operation) for relational structures with finitely many relations (resp. finite algebras).*

7.3 Valeriote's Conjecture

The most important open problem related to this work is the Valeriote conjecture (also known as the Edinburgh conjecture [7]).

Conjecture 7.4 *Every finite, finitely related algebra in a congruence modular variety has few subpowers.*

Congruence modularity is a widely studied generalization of congruence distributivity. An algebra $\mathbf{A}$ has *few subpowers* if the logarithm of the number of subalgebras of $\mathbf{A}^n$ is bounded by a polynomial in n. This property was defined and its importance in the CSP demonstrated in [5,14]. Examples of algebras with few supbowers include algebras with a Maltsev operation (*e.g.,* groups, rings, modules) and algebras with a near unanimity operation. It is known [5, 20] that every finite CD algebra with few subpowers has a near unanimity term. Therefore, a positive solution to the Valeriote conjecture would imply the main result of this paper. It would also have deep consequences in the complexity of constraints.

A converse to the conjecture generalizing the Baker–Pixley result [2] was proved recently. E. Aichinger, P. Mayr, and R. McKenzie [1] have shown that every finite algebra with few subpowers is finitely related.

References

[1] E. Aichinger, P. Mayr, and R. McKenzie, *On the number of finite algebraic structures.* arxiv:1103.2265.

[2] K. A. Baker and A. F. Pixley, *Polynomial interpolation and the Chinese remainder theorem for algebraic systems.* Math. Z. **143**(1975), no. 2, 165–174. http://dx.doi.org/10.1007/BF01187059

[3] L. Barto and M. Kozik, *Congruence distributivity implies bounded width.* SIAM J. Comput. **39**(2009/10), no. 4, 1531–1542. http://dx.doi.org/10.1137/080743238

[4] ———, *Constraint satisfaction problems of bounded width.* In: 2009 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009), IEEE Computer Soc., Los Alamitos, CA, 2009, pp. 595–603.

[5] J. Berman, P. Idziak, P. Marković, R. McKenzie, M. Valeriote, and R. Willard, *Varieties with few subalgebras of powers.* Trans. Amer. Math. Soc. **362**(2010), no. 3, 1445–1473. http://dx.doi.org/10.1090/S0002-9947-09-04874-0

[6] V. G. Bodnarčuk, L. A. Kalužnin, V. N. Kotov, and B. A. Romov, *Galois theory for Post algebras. I, II.* (Russian), Kibernetika (Kiev) **1969**, no. 3, 1–10; ibid. **1969**, no. 5, 1–9.

[7] S. Bova, H. Chen, and M. Valeriote, *Generic expression hardness results for primitive positive formula comparison.* 38th International Colloquium on Automata, Languages and Programming (ICALP), Zürich, Switzerland, 2011.

[8] A. Bulatov, P. Jeavons, and A. Krokhin, *Classifying the complexity of constraints using finite algebras.* SIAM J. Comput. **34**(2005), no. 3, 720–742. http://dx.doi.org/10.1137/S0097539700376676

[9] S. N. Burris and H. P. Sankappanavar, *A course in universal algebra.* Graduate Texts in Mathematics, 78, Springer-Verlag, New York-Berlin, 1981.

[10] B. A. Davey, *Monotone clones and congruence modularity.* Order **6**(1990), no. 4, 389–400. http://dx.doi.org/10.1007/BF00346133

[11] T. Feder and M. Y. Vardi, *The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory.* SIAM J. Comput. **28**(1999), no. 1, 57–104. http://dx.doi.org/10.1137/S0097539794266766

[12] R. Freese and M. A. Valeriote, *On the complexity of some Maltsev conditions.* Internat. J. Algebra Comput. **19**(2009), no. 1, 41–77. http://dx.doi.org/10.1142/S0218196709004956
[13] D. Geiger, *Closed systems of functions and predicates.* Pacific J. Math. **27**(1968), 95–100.
[14] P. Idziak, P. Marković, R. McKenzie, M. Valeriote, and R. Willard, *Tractability and learnability arising from algebras with few subpowers.* In: Proceedings of the Twenty-Second Annual IEEE Symposium on Logic in Computer Science (LICS 2007), IEEE Computer Society Press, 2007, pp. 213–222.
[15] P. Jeavons, D. Cohen, and M. Gyssens, *Closure properties of constraints.* J. ACM **44**(1997), no. 4, 527–548. http://dx.doi.org/10.1145/263867.263489
[16] B. Jónsson, *Algebras whose congruence lattices are distributive.* Math. Scand. **21**(1968), 110–121.
[17] G. Kun and C. Szabó, *Order varieties and monotone retractions of finite posets.* Order **18**(2001), no. 1, 79–88. http://dx.doi.org/10.1023/A:1010681409599
[18] B. Larose, C. Loten, and L. Zádori, *A polynomial-time algorithm for near-unanimity graphs.* J. Algorithms **55**(2005), no. 2, 177–191. http://dx.doi.org/10.1016/j.jalgor.2004.04.011
[19] B. Larose and L. Zádori, *Algebraic properties and dismantlability of finite posets.* Discrete Math. **163**(1997), no. 1–3, 89–99. http://dx.doi.org/10.1016/0012-365X(95)00312-K
[20] P. Marković and R. McKenzie, *Few subpowers, congruence distributivity and near-unanimity terms.* Algebra Universalis **58**(2008), no. 2, 119–128. http://dx.doi.org/10.1007/s00012-008-2049-1
[21] M. Maróti, *On the (un)decidability of a near-unanimity term.* Algebra Universalis **57**(2007), no. 2, 215–237. http://dx.doi.org/10.1007/s00012-007-2037-x
[22] M. Maróti, *The existence of a near-unanimity term in a finite algebra is decidable.* J. Symbolic Logic **74**(2009), no. 3, 1001–1014. http://dx.doi.org/10.2178/jsl/1245158096
[23] ———, *Monotone clones, residual smallness and congruence distributivity.* Bull. Austral. Math. Soc. **41**(1990), no. 2, 283–300. http://dx.doi.org/10.1017/S0004972700018104
[24] L. Zádori, *Monotone Jónsson operations and near unanimity functions.* Algebra Universalis **33**(1995), 216–236. http://dx.doi.org/10.1007/BF01190934

Department of Mathematics and Statistics, McMaster University, Hamilton, ON

and

Department of Algebra, Charles University, Prague, Czech Republic
e-mail: libor.barto@gmail.com

Appendix E – Conservative CSPs

2011 26th Annual IEEE Symposium on Logic in Computer Science

The Dichotomy for Conservative Constraint Satisfaction Problems Revisited

Libor Barto
Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada
and
Department of Algebra, Charles University, Prague, Czech Republic
Email: libor.barto@gmail.com

***Abstract*—A central open question in the study of non-uniform constraint satisfaction problems (CSPs) is the dichotomy conjecture of Feder and Vardi stating that the CSP over a fixed constraint language is either NP-complete, or tractable. One of the main achievements in this direction is a result of Bulatov (LICS'03) confirming the dichotomy conjecture for conservative CSPs, that is, CSPs over constraint languages containing all unary relations. Unfortunately, the proof is very long and complicated, and therefore hard to understand even for a specialist. This paper provides a short and transparent proof.**

***Keywords*-constraint satisfaction problem; list homomorphism problem; conservative algebra; dichotomy theorem;**

INTRODUCTION

The constraint satisfaction problem (CSP) provides a common framework for many theoretical problems in computer science as well as for many real-life applications. An instance of the CSP consists of a number of variables and constraints imposed on them and the objective is to determine whether variables can be evaluated in such a way that all the constraints are met. The CSP can also be expressed as the problem of deciding whether a given conjunctive formula is satisfiable, or as the problem of deciding whether there exists a homomorphism between two relational structures.

The general CSP is NP-complete, however certain natural restrictions on the form of the constraints can ensure tractability. This paper deals with so called *non-uniform* CSP — the same decision problem as the ordinary CSP, but the set of allowed constraint relations is fixed. A central open problem in this area is the dichotomy conjecture of Feder and Vardi [1] stating that, for every finite, fixed set of constraint relations (a fixed *constraint language*), the CSP defined by it is NP-complete or solvable in polynomial time, i.e. the class of CSPs exhibits a dichotomy.

Most of recent progress toward the dichotomy conjecture has been made using the algebraic approach to the CSP [2], [3], [4]. The main achievements include the algorithm for CSPs with "Maltsev constraints" [5] (which was substantially simplified in [6] and generalized in [7], [8]), the characterization of CSPs solvable by local consistency methods [9], [10], the dichotomy theorem for CSPs over a three element domain [11] (which generalizes the Boolean CSP dichotomy theorem [12]) and the dichotomy theorem for conservative CSPs [13].

The last result proves the dichotomy conjecture of Feder and Vardi for the CSP over any template which contains all unary relations. In other words, this Bulatov's theorem proves the dichotomy for the CSPs, in which we can restrict the value of each variable to an arbitrary subset of the domain (that is why the conservative CSPs are sometimes called list CSPs, or, in homomorphism setting, list homomorphism problems). This result is of major importance in the area, but, unfortunately, the proof is very involved (the full paper has 80 pages and it has not yet been published), which makes the study of possible generalizations and further research harder.

This paper provides a new, shorter and more natural proof. It relies on techniques developed and successfully applied in [14], [15], [16], [9], [17], [18].

Related work

The complexity of list homomorphism problems has been studied by combinatorial methods, e.g., in [19], [20]. A structural distinction between tractable and NP-complete list homomorphism problem for digraphs was found in [21]. A finer complexity classification for the list homomorphism problem for graphs was given in [22]. The conservative case is also studied for different variants of the CSP, see, e.g., [23], [24].

Organization of the paper

In Section I we define the CSP and its non-uniform version. In Section II we introduce the necessary notions concerning algebras and the algebraic approach to the CSP. In Section III we collect all the necessary ingredients. One of them is a reduction to minimal absorbing subuniverses, details are provided in Section V. Also the core algebraic result is just stated in this section and its proof covers Section VI. In Section IV we formulate the algorithm for tractable conservative CSPs and prove its correctness.

I. CSP

An *n-ary relation* on a set A is a subset of the n-th cartesian power A^n of the set A.

Definition I.1. *An* instance of the constraint satisfaction problem *(CSP) is a triple $P = (V, A, \mathcal{C})$ with*

The research was supported by the Grant Agency of the Czech Republic, grant 201/09/P223 and by the Ministry of Education of the Czech Republic, grants MSM 0021620839 and MEB 040915.

1043-6871/11 $26.00 © 2011 IEEE
DOI 10.1109/LICS.2011.25

- V *a nonempty, finite set of* variables*,*
- A *a nonempty, finite* domain*,*
- $\mathcal{C}$ *a finite set of* constraints*, where each constraint is a pair* $C = (\mathbf{x}, R)$ *with*
  - $\mathbf{x}$ *a tuple of distinct variables of length* n*, called the* scope of C*, and*
  - R *an* n*-ary relation on* A*, called the* constraint relation of C.

  The question is whether there exists a solution *to* P*, that is, a function* $f : V \to A$ *such that, for each constraint* $C = (\mathbf{x}, R) \in \mathcal{C}$*, the tuple* $f(\mathbf{x})$ *belongs to* R*.*

For purely technical reasons we have made a nonstandard assumption that the scope of a constraint contains distinct variables. This clearly does not change the complexity modulo polynomial-time reductions.

In the non-uniform CSP we fix a domain and a set of allowed constraints:

Definition I.2. *A* constraint language Γ *is a set of relations on a finite set* A*. The* constraint satisfaction problem over Γ*, denoted* $\mathrm{CSP}(\Gamma)$*, is the subclass of the CSP defined by the property that any constraint relation in any instance must belong to* Γ.

The following dichotomy conjecture was originally formulated in [1] only for finite constraint languages. The known results suggest that even the following stronger version might be true.

Conjecture I.3. *For every constraint language* Γ*,* $\mathrm{CSP}(\Gamma)$ *is either tractable, or NP-complete.*

Our main theorem, first proved by Bulatov [13], confirms the dichotomy conjecture for conservative CSPs:

Definition I.4. *A constraint language* Γ *on* A *is called* conservative*, if* Γ *contains all unary relations on* A *(i.e., all subsets of* A*).*

Theorem I.5. *For every conservative constraint language* Γ*,* $\mathrm{CSP}(\Gamma)$ *is either tractable, or NP-complete.*

II. Algebra and CSP

A. Algebraic preliminaries

An n-ary operation on a set A is a mapping $f : A^n \to A$. An operation f is called *cyclic*, if $n \geq 2$ and $f(a_1, a_2, \ldots, a_n) = f(a_2, a_3, \ldots, a_n, a_1)$ for any $a_1, a_2, \ldots, a_n \in A$. A ternary operation m is called *Maltsev*, if $f(a, a, b) = f(b, a, a) = b$ for any $a, b \in A$.

A *signature* is a finite set of symbols with natural numbers (the *arities*) assigned to them. An *algebra* of a signature Σ is a pair $\mathbf{A} = (A, (t^{\mathbf{A}})_{t\in\Sigma})$, where A is a set, called the *universe* of $\mathbf{A}$, and $t^{\mathbf{A}}$ is an operation on A of arity $\mathrm{ar}(t)$. We use a boldface letter to denote an algebra and the same letter in the plain type to denote its universe. We omit the superscripts of operations as the algebra is always clear from the context.

A *term operation* of $\mathbf{A}$ is an operation which can be obtained from operations in $\mathbf{A}$ using composition and the projection operations. The set of all term operations of $\mathbf{A}$ is denoted by $\mathrm{Clo}(\mathbf{A})$.

There are three fundamental operations on algebras of a fixed signature Σ: forming subalgebras, factoralgebras and products.

A subset B of the universe of an algebra $\mathbf{A}$ is called a *subuniverse*, if it is closed under all operations (equivalently term operations) of $\mathbf{A}$. Given a subuniverse B of $\mathbf{A}$ we can form the algebra $\mathbf{B}$ by restricting all the operations of $\mathbf{A}$ to the set B. In this situation we say that $\mathbf{B}$ is a *subalgebra* of $\mathbf{A}$ and we write $B \leq \mathbf{A}$ or $\mathbf{B} \leq \mathbf{A}$. We call the subuniverse B (or the subalgebra $\mathbf{B}$) *proper* if $\emptyset \neq B \neq A$.

We define the *product* of algebras $\mathbf{A}_1, \ldots, \mathbf{A}_n$ to be the algebra with the universe equal to $A_1 \times \cdots \times A_n$ and with operations computed coordinatewise. The product of n copies of an algebra $\mathbf{A}$ is denoted by $\mathbf{A}^n$.

An equivalence relation $\sim$ on the universe of an algebra $\mathbf{A}$ is a *congruence*, if it is a subalgebra of $\mathbf{A}^2$. The corresponding *factor algebra* $\mathbf{A}/\sim$ has, as its universe, the set of $\sim$-blocks and operations are defined using (arbitrary chosen) representatives. Every algebra $\mathbf{A}$ has two *trivial congruences*: the diagonal congruence $\sim = \{(a, a) : a \in A\}$ and the full congruence $\sim = A \times A$. A congruence is *proper*, if it is not equal to the full congruence. A congruence is *maximal*, if the only coarser congruence of $\mathbf{A}$ is the full congruence.

For a finite algebra $\mathbf{A}$ the class of all factor algebras of subalgebras of finite powers of $\mathbf{A}$ will be denoted by $\mathrm{V_{fin}}(\mathbf{A})$.

An operation $f : A^n \to A$ is *idempotent*, if $f(a, a, \ldots, a) = a$ for any $a \in A$. An operation $f : A^n \to A$ is *conservative*, if $f(a_1, a_2, \ldots, a_n) \in \{a_1, a_2, \ldots, a_n\}$ for any $a_1, a_2, \ldots, a_n \in A$. An algebra is *idempotent* (resp. *conservative*), if all operations of $\mathbf{A}$ are idempotent (resp. conservative). In other words, an algebra is idempotent (resp. conservative), if all one-element subsets of A (resp. all subsets of A) are subuniverses of $\mathbf{A}$.

B. Algebraic approach

An operation $f : A^n \to A$ is *compatible* with a relation $R \subseteq A^m$ if the tuple

$$(f(a_1^1, a_1^2, \ldots, a_1^n), f(a_2^1, a_2^2, \ldots, a_2^n), \ldots, f(a_m^1, a_m^2, \ldots, a_m^n))$$

belongs to R whenever $(a_1^i, a_2^i, \ldots, a_m^i) \in R$ for all $i \leq n$.

An operation compatible with all relations in a constraint language Γ is a *polymorphism* of Γ. The set A together with all polymorphisms of Γ is the *algebra of polymorphisms of* Γ, it is denoted $\mathrm{Pol}\,\Gamma$, or often just $\mathbf{A}$ (we formally define the signature of $\mathbf{A}$ to be identical with the set of its operations). Note that every relation in Γ is a subalgebra of a power of $\mathbf{A}$. The set of all subalgebras of powers of $\mathbf{A}$ is denoted by $\mathrm{Inv}\,\mathbf{A}$.

In the following discussion we assume, for simplicity, that Γ contains all singleton unary relations (it is known that CSP can be reduced to CSP over such a constraint language). Observe that in such a case the algebra $\mathbf{A}$ is idempotent. Moreover, if Γ is conservative, then $\mathbf{A}$ is conservative as well.

Already the first results on the algebraic approach to CSP [2], [3], [4] show that $\mathbf{A}$ fully determines the computational complexity of $\mathrm{CSP}(\Gamma)$, at least for finite constraint languages. Moreover, a borderline between tractable and NP-complete CSPs was conjectured in terms of the algebra of polymorphisms: if there exists a two-element factor algebra of a subalgebra of $\mathbf{A}$ whose every operation is a projection, then $\mathrm{CSP}(\Gamma)$ is NP-complete, otherwise $\mathrm{CSP}(\Gamma)$ is tractable. The hardness part of this algebraic dichotomy conjecture is known [3], [4]:

Theorem II.1. *Let Γ be a constraint language containing all singleton unary relations, and let $\mathbf{A} = \mathrm{Pol}\,\Gamma$. If $\mathbf{A}$ has a subalgebra with a two-element factor algebra whose every operation is a projection, then $\mathrm{CSP}(\Gamma)$ is NP-complete.*

The algebras, which satisfy this necessary (and conjecturally sufficient) condition for tractability, are called *Taylor* algebras, that is, $\mathbf{A}$ is Taylor if no two-element factor algebra of a subalgebra $\mathbf{A}$ has projection operations only. We will use the following characterization of Taylor algebras from [17], [18] although the characterization in terms of weak near-unanimity operations [25] would suffice for our purposes.

Theorem II.2. *Let $\mathbf{A}$ be a finite idempotent algebra and let $p > |A|$ be a prime number. The following are equivalent.*

- $\mathbf{A}$ *is a Taylor algebra.*
- $\mathbf{A}$ *has a cyclic term operation of arity p.*

In view of Theorem II.1, the dichotomy for conservative CSPs will follow when we prove:

Theorem II.3. *Let $\mathbf{A}$ be a finite conservative Taylor algebra. Then $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$ is tractable.*

A polynomial time algorithm for solving $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$, where $\mathbf{A}$ is a finite conservative Taylor algebra, is presented in Section IV.

III. Ingredients

The building blocks of our algorithm are the (k,l)-minimality algorithm (Subsection III-B), a reduction to minimal absorbing subuniverses (Subsection III-D) and the algorithm for Maltsev instances (Subsection III-F). Subsection III-A and Subsection III-C cover necessary notation.

The main new algebraic tool for proving correctness is stated in Subsection III-E and this is the place where we make essential use of the assumption that the constraint language is conservative (the result is not true in general). This theorem enables us to show that partial solutions of certain restricted instances can be glued together to obtain a solution of a larger instance. If one of these smaller instances does not have a solution then we can delete some elements from the constraint relations. In this place we use conservativity the second time, it ensures that the new relations will still be in the constraint language.

A. Projections and restrictions

Tuples are denoted by boldface letters and their elements are indexed from 1, for instance $\mathbf{a} = (a_1, a_2, \dots, a_n)$. For an n-tuple $\mathbf{a}$ and a tuple $\mathbf{k} = (k_1, \dots, k_m)$ of elements of $\{1, 2, \dots, n\}$ we define the *projection of* $\mathbf{a}$ *to* $\mathbf{k}$ by

$$\mathbf{a}|_{\mathbf{k}} = (a_{k_1}, a_{k_2}, \dots, a_{k_m}).$$

For a subset $K \subseteq \{1, 2, \dots, n\}$ we put $\mathbf{a}|_K = \mathbf{a}|_{\mathbf{k}}$, where $\mathbf{k}$ is the list of elements of K in the ascending order.

The *projection of a set $R \subseteq A_1 \times \dots A_n$ to $\mathbf{k}$ (resp. K)* is defined by

$$R|_{\mathbf{k}} = \{\mathbf{a}|_{\mathbf{k}} : \mathbf{a} \in R\} \quad (\text{resp. } R|_K = \{\mathbf{a}|_K : \mathbf{a} \in R\}).$$

The set R is *subdirect in* $A_1 \times \cdots \times A_n$ (denoted by $R \subseteq_S A_1 \times \cdots \times A_n$), if $R|_{\{i\}} = A_i$ for all $i = 1, \dots, n$. If, moreover, $\mathbf{A}_1, \dots, \mathbf{A}_n$ are algebras of the same signature and R is a subalgebra of their product, we write $R \leq_S \mathbf{A}_1 \times \cdots \times \mathbf{A}_n$.

Let $P = (V, A, \mathcal{C})$ be an instance of the CSP. The *projection of a constraint $C = ((x_1, \dots, x_n), R) \in \mathcal{C}$ to a tuple of variables $(x_{k_1}, \dots, x_{k_m})$* is the relation

$$C|_{(x_{k_1}, \dots, x_{k_m})} = \{(a_{k_1}, \dots, a_{k_m}) : \mathbf{a} \in R\}.$$

Finally, we introduce two types of restrictions of a CSP instance. In the variable restriction we delete some of the variables and replace the constraints with appropriate projections, in the domain restriction we restrict the value of some of the variables to specified subsets of the domain.

The *variable restriction of P to a subset $W \subseteq V$* is the instance $P|_W = (W, A, \mathcal{C}')$, where $\mathcal{C}'$ is obtained from $\mathcal{C}$ by replacing each constraint $C = (\mathbf{x}, R) \in \mathcal{C}$ with $(\mathbf{x} \cap W, C|_{\mathbf{x} \cap W})$, where $\mathbf{x} \cap W$ is the subtuple of $\mathbf{x}$ formed by the variables belonging to W.

The *domain restriction of P to a system $\mathcal{E} = \{E_x : x \in W\}$* of subsets of A indexed by $W \subseteq V$ is the instance $P|_{\mathcal{E}} = (V, A, \mathcal{C}')$, where $\mathcal{C}'$ is obtained from $\mathcal{C}$ by replacing each constraint $C = ((x_1, \dots, x_n), R) \in \mathcal{C}$ with $C' = ((x_1, \dots, x_n), \{\mathbf{a} \in R : \forall i \ x_i \in W \Rightarrow a_i \in E_{x_i}\})$. If $\mathbf{A}$ is a conservative algebra, then every domain restriction of an instance of $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$ is an instance of $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$ (because all subset of A are subalgebras of $\mathbf{A}$).

B. (k,l)-minimality

The first step in our algorithm will be to ensure a certain kind of local consistency. The following notion is the most convenient for our purposes.

Definition III.1. *Let $l \geq k > 0$ be natural numbers. An instance $P = (V, A, \mathcal{C})$ of the CSP is* (k,l)-minimal*, if:*

- *Every at most l-element tuple of distinct variables is the scope of some constraint in $\mathcal{C}$,*
- *For every tuple $\mathbf{x}$ of at most k variables and every pair of constraints C_1 and C_2 from $\mathcal{C}$ whose scopes contain all variables from $\mathbf{x}$, the projections of the constraints C_1 and C_2 to $\mathbf{x}$ are the same.*

A (k,k)-minimal instance is also called k-minimal.

For fixed k, l there is an obvious polynomial time algorithm for transforming an instance of the CSP to a (k,l)-minimal instance with the same set of solutions: First we add dummy constraints to ensure that the first condition is satisfied and then we gradually remove those tuples from the constraint relations which falsify the second condition (see [13] for a more detailed discussion). It is a folklore fact (which is in the literature often used without mentioning) that an instance of $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$ is in this way transformed to an instance of $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$, that is, the constraint relations of the new instance are still members of $\mathrm{Inv}\,\mathbf{A}$. See the discussion after Definition III.3 in [9], where an argument is given for a similar consistency notion.

If an instance P is (at least) 1-minimal, then, for each variable $x \in V$, there is a unique constraint whose scope is (x). We denote its constraint relation by S_x^P, i.e. $((x), S_x^P) \in \mathcal{C}$. Then the projection of any constraint whose scope contains x to (x) is equal to S_x^P. If, moreover, P is an instance of $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$, the set S_x^P is a subuniverse of $\mathbf{A}$ and we denote the corresponding subalgebra of $\mathbf{A}$ by $\mathbf{S}_x^P$.

If an instance is 2-minimal, then we have a unique constraint $((x,x'), S_{(x,x')}^P)$ for each pair of distinct variables $x, x' \in V$, and $C|_{(x,x')} = S_{(x,x')}^P$ for any constraint C whose scope contains x and x'. We formally define $S_{(x,x)}^P = \{(a,a) : a \in S_x^P\}$.

$(2,3)$-minimal instances have the following useful property.

Lemma III.2. *Let P be a $(2,3)$-minimal instance and let $x, x', x'' \in V$. Then for any $(a,a') \in S_{(x,x')}^P$ there exists $a'' \in A$ such that $(a,a'') \in S_{(x,x'')}^P$ and $(a',a'') \in S_{(x',x'')}^P$.*

Proof: Let $C \in \mathcal{C}$ be a constraint with the scope (x,x',x''), say $C = ((x,x',x''), R)$. The projection of C to (x,x') is equal to $S_{(x,x')}^P$, therefore there exists $a'' \in A$ such that $(a,a',a'') \in R$. This element satisfies the conclusion of the lemma. ∎

C. Walking with subsets

Let $R \subseteq A_1 \times A_2$ and let $B \subseteq A_1$. We define

$$R^+[B] = \{c \in A_2 : \exists\, b \in B \quad (b,c) \in R\}.$$

A *qoset* is a set A together with a quasi-ordering on A, i.e. a reflexive, transitive (binary) relation $\leq$ on A. The blocks of the induced equivalence $\sim$, given by $a \sim b$ iff $a \leq b \leq a$, are called *components* of the qoset. A component C is *maximal*, if $a \sim c$ for any $a \in A, c \in C$ such that $c \leq a$. A *subqoset* is a subset of A together with $\leq$ restricted to A.

For a 1-minimal instance $P = (V, A, \mathcal{C})$ we introduce a qoset $\mathrm{Qoset}(P)$ as follows. The elements are all the pairs (x,B), where $x \in V$ and B is a subset of S_x^P. We put $(x,B) \leq (x',B')$, if there exists a constraint $C \in \mathcal{C}$ whose scope contains $\{x,x'\}$ such that ${C|_{(x,x')}}^+[B] = B'$. The ordering of the qoset $\mathrm{Qoset}(P)$ is the transitive closure of $\leq$.

If the instance P is $(2,3)$-minimal, the components of $\mathrm{Qoset}(P)$ are nicely behaved:

Proposition III.3. *Let $P = (V, A, \mathcal{C})$ be a $(2,3)$-minimal instance of the CSP and let (x,B) and (x',B') be two elements of the same component of the qoset $\mathrm{Qoset}(P)$. Then ${S_{(x,x')}^P}^+[B] = B'$. In particular, if $x = x'$, then $B = B'$.*

Proof: Let $(x,B) = (x_1,B_1), (x_2,B_2), \ldots, (x_k,B_k) = (x',B')$ be a sequence of elements of $\mathrm{Qoset}(P)$ such that ${S_{(x_i,x_{i+1})}^P}^+[B_i] = B_{i+1}$ for all $i = 1, \ldots, k-1$. From Lemma III.2 it follows that

$${S_{(x_1,x_{i+1})}^P}^+[B_1] \subseteq {S_{(x_i,x_{i+1})}^P}^+\left[{S_{(x_1,x_i)}^P}^+[B_1]\right],$$

therefore ${S_{(x,x')}^P}^+[B] \subseteq B'$. Similarly, ${S_{(x',x)}^P}^+[B'] \subseteq B$. For each $b' \in B'$ $(\subseteq S_{x'}^P)$ there exists $b \in A$ such that $(b,b') \in {S_{(x,x')}^P}^+[B]$. This element b has to belong to B (since ${S_{(x',x)}^P}^+[B'] \subseteq B$), which proves the inclusion ${S_{(x,x')}^P}^+[B] \supseteq B'$. ∎

Let P be a $(2,3)$-minimal instance and $\mathcal{E} = \{E_x : x \in V\}$ be a system of subsets of A such that $E_x \subseteq S_x^P$ for each $x \in V$. A $(P,\mathcal{E})$-*strand* is a maximal subset W of V such that all the pairs (x, E_x), $x \in W$ belong to the same component of the qoset $\mathrm{Qoset}(P)$. The name of this concept is justified by the previous proposition: For example, any solution $f : V \to A$ to P with $f(x) \in E_x$ for *some* $x \in W$ satisfies $f(x) \in E_x$ for *all* $x \in W$.

D. Absorbing subuniverses

Definition III.4. *Let $\mathbf{A}$ be a finite idempotent algebra and $t \in \mathrm{Clo}(\mathbf{A})$. We say that a subalgebra $\mathbf{B}$ of $\mathbf{A}$ is an* absorbing subalgebra of $\mathbf{A}$ with respect to t *if, for every $k \leq \mathrm{ar}(t)$ and every $a_1, \ldots, a_{\mathrm{ar}(t)} \in A$ such that $a_i \in B$ for all $i \neq k$, we have $t(a_1, \ldots, a_{\mathrm{ar}(t)}) \in B$.*

We say that $\mathbf{B}$ is an absorbing subalgebra of $\mathbf{A}$*, or that $\mathbf{B}$* absorbs $\mathbf{A}$ *(and write $\mathbf{B} \triangleleft \mathbf{A}$), if there exists $t \in \mathrm{Clo}(\mathbf{A})$ such that $\mathbf{B}$ is an absorbing subalgebra of $\mathbf{A}$ with respect to t.*

We say that $\mathbf{A}$ is an absorption free algebra*, if it has no proper absorbing subalgebras.*

We also speak about absorbing subuniverses i.e. universes of absorbing subalgebras.

Definition III.5. *If $\mathbf{B} \triangleleft \mathbf{A}$ and no proper subalgebra of $\mathbf{B}$ absorbs $\mathbf{A}$, we call $\mathbf{B}$ a* minimal absorbing subalgebra *of $\mathbf{A}$ (and write $\mathbf{B} \triangleleft\!\triangleleft\, \mathbf{A}$).*

Alternatively, we can say that $\mathbf{B}$ is a minimal absorbing subalgebra of $\mathbf{A}$, if $\mathbf{B} \triangleleft \mathbf{A}$ and $\mathbf{B}$ is an absorption free algebra. Equivalence of these definitions follows from transitivity of $\triangleleft$ (see Proposition III.2 in [17]).

Algorithm 1 finds, for a given $(2,3)$-minimal instance P of the CSP, a domain restriction Q of P which is 1-minimal and satisfies $\mathbf{S}_x^Q \triangleleft\!\triangleleft\, \mathbf{S}_x^P$ for any $x \in V$.

The algorithms uses a subqoset $\mathrm{AbsQoset}(P)$ of $\mathrm{Qoset}(P)$ formed by the elements (x,B) such that B is a proper absorbing subuniverse of $\mathbf{S}_x^P$.

Theorem III.6. *Algorithm 1 is correct and, for a fixed idempotent algebra $\mathbf{A}$, works in polynomial time.*

Proof: The qoset $\mathrm{AbsQoset}(P)$ contains at most $2^{|A|}|V|$ elements, therefore its maximal component can be found in a polynomial time. In each while loop at least one of the sets S_x^P becomes smaller, thus the while loop is repeated at most $|V||A|$ times, and the algorithm is therefore polynomial.

Fig. 1. Algorithm 1: Minimal absorbing subuniverses

Input: $(2,3)$-minimal instance $P = (V, A, \mathcal{C})$ of $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$
Output: $\mathcal{E} = \{E_x : x \in V\}$ such that $E_x \lhd\!\lhd \mathbf{S}_x^P$, $x \in V$, and $P|_{\mathcal{E}}$ is 1-minimal

1: **while** some S_x^P has a proper absorbing subuniverse **do**
2: find a maximal component $\mathcal{F} = \{(x, E_x) : x \in W\}$ of the qoset $\mathrm{AbsQoset}(P)$
3: $P := P|_{\mathcal{F}}$
4: **return** $\{S_x^P : x \in V\}$

The correctness follows from a slightly generalized results from [9] (the generalized version will be in [10]): In the beginning of the while loop, P is so called *Prague strategy*. An analogue of Proposition III.3 remains valid for Prague strategies (Lemma IV.10 in [9], Lemma V.5 part (iii) in Section V), in particular, for each variable $x \in V$, there is at most one element (x, E_x) in the maximal component, therefore the definition of $\mathcal{F}$ in step 2 makes sense. Finally, the restriction of P to $\mathcal{F}$ is again a Prague strategy (Theorem IV.15 in [9], Lemma V.6 in Section V). The details are in Section V. ■

The presented algorithm as well as the main algorithm require knowledge of absorbing subuniverses of a given algebra and its subalgebras. We do not need to provide algorithm for this because the algebra is fixed. Actually, we do not even know if it is possible. See remarks in Section VII.

E. Rectangularity

The core result for proving correctness of our algorithm for conservative CSPs is the "Rectangularity Theorem". We state the theorem here, its proof spans Section VI.

We need one more notion. Let $A_1, \dots, A_n, B_1, \dots, B_n$ be sets such that $B_i \subseteq A_i$ and let $R \subseteq_S A_1 \times \dots \times A_n$. We define a quasi-ordering $\preceq$ on the set $\{1, 2, \dots, n\}$ by

$$i \preceq j \quad \text{if} \quad R|_{(i,j)}{}^{+}[B_i] \subseteq B_j.$$

Components of this qoset are called (R, B)-*strands*.

Theorem III.7. *Let $\mathbf{A}$ be a finite Taylor algebra, let $\mathbf{A}_1, \dots, \mathbf{A}_n, \mathbf{B}_1, \dots, \mathbf{B}_n \in \mathrm{V}_{\mathrm{fin}}(\mathbf{A})$ be conservative algebras such that $\mathbf{B}_i \lhd\!\lhd \mathbf{A}_i$ for all i, let $R \leq_S \mathbf{A}_1 \times \dots \times \mathbf{A}_n$ and assume that $R \cap (B_1 \times \dots \times B_n) \neq \emptyset$. Then a tuple $\mathbf{a} \in B_1 \times \dots \times B_n$ belongs to R whenever $\mathbf{a}|_K \in R|_K$ for each (R, B)-strand K.*

F. Maltsev instances

Our final ingredient is the polynomial time algorithm by Bulatov and Dalmau [6] for the CSPs over constraint languages with a Maltsev polymorphism. Their algorithm can be used without any change in the following setting:

Theorem III.8. *[6] Let $\mathbf{A}$ be a finite algebra with a ternary term operation m. Then there is a polynomial time algorithm which correctly decides every 1-minimal instance P of $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$ such that, for every variable x, m is a Maltsev operation of $\mathbf{S}_x^P$.*

IV. ALGORITHM

The algorithm for conservative CSPs is in Figure 2. It uses a subqoset $\mathrm{NafaQoset}(P)$ of the qoset $\mathrm{Qoset}(P)$ formed by the elements (x, B) such that $B \subseteq S_x^P$ and $\mathbf{B}$ has a proper absorbing subalgebra (where $\mathbf{B}$ stands for the subalgebra of $\mathbf{A}$ with universe B).

Fig. 2. Algorithm 2 for solving $CSP(\mathrm{Inv}\,\mathbf{A})$ for conservative $\mathbf{A}$

Input: Instance $P = (V, A, \mathcal{C})$ of $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$
Output: "YES" if P has a solution, "NO" otherwise

1: Transform P to a $(2,3)$-minimal instance with the same solution set
2: **if** some subalgebra of S_x^P has a proper absorbing subalgebra **then**
3: Find a maximal component $\mathcal{D} = \{D_x : x \in W\}$ of $\mathrm{NafaQoset}(P)$
4: $Q := (P|_W)|_{\mathcal{D}}$
5: $\mathcal{E}$:= the result of Algorithm 1 for the instance Q
6: **for** each $(Q, \mathcal{E})$-strand U **do**
7: Use this algorithm for the instance $(Q|_U)|_{\{E_x : x \in U\}}$
8: **if** no solution exists **then**
9: $\mathcal{F} := \{S_x^P - E_x : x \in U\}$
10: $P := P|_{\mathcal{F}}$
11: **goto** step 1
12: $\mathcal{F} := \{S_x^P - (D_x - E_x) : x \in W\}$
13: $P := P|_{\mathcal{F}}$
14: **goto** step 1
15: Use the algorithm for Maltsev instances (Theorem III.8)

Theorem IV.1. *If $\mathbf{A}$ is a conservative finite algebra, then Algorithm 2 is correct and works in polynomial time.*

Proof: By induction on k we show that the algorithm works in polynomial time for all instances such that $|S_x^P| \leq k$. The base case of the induction is obvious: if every S_x^P is at most one-element, then the algorithm proceeds directly to Step 15 (where the algorithm answers YES iff every S_x^P is one-element).

Step 1 can be done in polynomial time as discussed in Subsection III-B. In Step 3 the qoset has size at most $2^{|A|}|V|$, therefore its maximal component can be found in polynomial time. Step 5 is polynomial according to Theorem III.6. There are at most $|V|$ repetitions of the **for** cycle in Step 6. Step 7 is polynomial by the induction hypothesis, since every E_x is a minimal absorbing subuniverse of $\mathbf{D}_x$ $(= \mathbf{S}_x^Q)$ and $\mathbf{D}_x$ has, as a member of $\mathrm{NafaQoset}(P)$, a proper absorbing subuniverse. Before we return to Step 1 (either in Step 11 or in Step 14) at least one of the sets S_x^P becomes strictly smaller. It follows that there are at most $|A||V|$ returns to the first step. Finally, the last step is polynomial by Theorem III.8.

Now we show the correctness of the algorithm.

First, we observe that no solution is lost in Step 10. As the pairs (x, E_x), $x \in U$ are in one component of the qoset $\mathrm{Qoset}(Q)$ and the instance Q is the restriction of P to elements of the same component of $\mathrm{Qoset}(P)$, it follows that all the pairs (x, E_x), $x \in U$ lie in the same component of $\mathrm{Qoset}(P)$. Therefore, if $f : V \to A$ is a solution to P such that $f(x) \in E_x$ for some $x \in U$, then $f(x) \in E_x$ for all $x \in U$ (see Proposition III.3 and the discussion bellow). But the restriction of such a function f to the set U would be a solution to the instance $(Q|_U)|_{\{E_x : x \in U\}}$, thus we would not get to this step. We have shown that in Step 10 every solution to P misses all the sets E_x, $x \in U$, and hence we do not lose any solution when we restrict P to $\mathcal{F}$.

Next, we show that if P has a solution before Step 13, then the restricted instance $P|_{\mathcal{F}}$ has a solution as well. If $f : V \to A$ is a solution to P such that $f(x) \notin D_x$ for some $x \in W$, then $f(x) \notin D_x$ for all $x \in W$, because (x, D_x), $x \in W$ are in the same component of $\mathrm{Qoset}(P)$ and we can use Proposition III.3 as above. In this case f is a solution to the restricted instance. Now we assume that f is a solution to P such that $f(x) \in D_x$ for all $x \in W$. For each $(Q, \mathcal{E})$-strand U let $g_U : U \to A$ be a solution to the instance $(Q|_U)|_{\{E_x : x \in U\}}$. Let $h : V \to A$ be the mapping satisfying $h|_{V-W} = f|_{V-W}$ and $h|_U = g_U$ for each $(Q, \mathcal{E})$-strand U. We claim that this mapping is a solution to the instance $P|_{\mathcal{F}}$.

Clearly, $h(x) \in S_x^P - (D_x - E_x)$ for every $x \in W$.

We define D_x for $x \in V - W$ by $D_x = {S_{(y,x)}^P}^+[D_y]$, where y is an arbitrarily chosen element of W. The definition of D_x does not depend on the choice of y: Let $y, y' \in W$ and take an arbitrary $a \in {S_{(y,x)}^P}^+[D_y]$. From the choice of a it follows that there is $b \in D_y$ such that $(b, a) \in S_{(y,x)}^P$. Lemma III.2 provides us with an element $b' \in A$ such that $(b', a) \in S_{(y',x)}^P$ and $(b, b') \in S_{(y,y')}^P$. The latter fact together with Proposition III.3 implies $b' \in D_{y'}$, therefore $a \in {S_{(y',x)}^P}^+[D_{y'}]$. We have proved the inclusion ${S_{(y,x)}^P}^+[D_y] \subseteq {S_{(y',x)}^P}^+[D_{y'}]$, the opposite inclusion is proved similarly.

We put $E_x = D_x$ for $x \in V - W$. Let $\mathbf{D}_x$ (resp. $\mathbf{E}_x$) denote the subalgebra of $\mathbf{A}$ with universe D_x (resp. E_x), $x \in V$. For any $x \in V - W$ and $y \in W$, the pair (x, D_x) is greater than or equal to (y, D_y) in the qoset $\mathrm{Qoset}(P)$. Since $\mathcal{D}$ is a maximal component and $x \notin W$, it follows that D_x is outside the qoset $\mathrm{NafaQoset}(P)$ and thus $\mathbf{D}_x$ has no proper absorbing subuniverse. Therefore $E_x \lhd\!\lhd \mathbf{D}_x$ for all $x \in V$ (for $x \in W$ it follows from the fact that $\mathcal{E}$ is the result of Algorithm 1).

Now we are ready to show that h is a solution to P, i.e. h satisfies all the constraints in $\mathcal{C}$. So, let $C = (\mathbf{x}, R) \in \mathcal{C}$, $\mathbf{x} = (x_1, \dots, x_n)$ be an arbitrary constraint. For each $i \in V$ let $\mathbf{A}_i = \mathbf{D}_{x_i}$ and $\mathbf{B}_i = \mathbf{E}_{x_i}$, let $\mathbf{a} = (h(x_1), \dots, h(x_n))$, and let $L = \{l_1, \dots, l_k\} := \{i : x_i \in W\}$. By the choice of D_xs, the relation R is subdirect in $A_1 \times \cdots \times A_n$. Since $Q|_{\mathcal{E}}$ is 1-minimal (it is the result of Algorithm 1), the projection of C to $(x_{l_1}, \dots, x_{l_k})$ has a nonempty intersection with $B_{l_1} \times \cdots \times B_{l_k}$. By the choice of E_x, $x \in V - W$ it follows that the relation R has a nonempty intersection with $B_1 \times \cdots \times B_n$. For any $i \in L$ and $j \in \{1, \dots, n\} - L$ we have ${R|_{(j,i)}}^+[B_j] = A_i \not\subseteq B_i$, therefore no element of L is in the same (R, B)-strand as an element outside L. Moreover, $i, j \subseteq L$ are in the same (R, B)-strand if and only if x_i, x_j are in the same $(Q, \mathcal{E})$-strand, since $R|_{(i,j)} = S_{(x_i,x_j)}^P$. It follows that $\mathbf{a}|_K \in R|_K$ for each (R, B)-strand $K \subseteq L$, and the same is of course true for each (R, B)-strand $K \subseteq \{1, 2, \dots, n\} - L$ as $f|_{V-W} = h|_{V-W}$. We have checked all the assumptions of Theorem III.7, which gives us $\mathbf{a} \in R$. In other words, h satisfies the constraint C.

From the fact that $\mathbf{A}$ is conservative it easily follows that after both Step 10 and Step 13 the restricted instance is still an instance of $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$.

Finally, we prove that P satisfies the assumptions of Theorem III.8 when we get to Step 15. Note that at this point we know that no subalgebra of $\mathbf{S}_x^P$ has a proper absorbing subalgebra. Let t be a cyclic term operation of the algebra $\mathbf{A}$ (guaranteed by Theorem II.2). If $t(a, a, \dots, a, b) = a$ for some $x \in V$, $a, b \in S_x^P$, then $t(a, a, \dots, a, b) = t(a, a, \dots, a, b, a) = \cdots = t(b, a, a, \dots, a)$, and hence $\{a\}$ is an absorbing subuniverse of $\{a, b\}$ with respect to t, a contradiction. Therefore, as $\mathbf{A}$ is conservative, $t(a, a, \dots, a, b) = b = t(b, a, a, \dots, a)$ for any $x \in V, a, b \in S_x^P$. Now the term operation $m(x, y, z) = t(x, y, y, \dots, y, z)$ satisfies the assumptions of Theorem III.8 and the proof is concluded. ■

V. Prague strategies

This section fills the gaps in the proof of Theorem III.6.

Definition V.1. *Let $P = (V, A, \mathcal{C})$ be a 1-minimal instance of the CSP. A* pattern *in P is a tuple $(x_1, C_1, x_2, C_2, \dots, C_{n-1}, x_n)$, where $x_1, \dots, x_n \in V$ and, for every $i = 1, \dots, n-1$, C_i is a constraint whose scope contains $\{x_i, x_{i+1}\}$. The pattern w is* closed with base x, *if $x_1 = x_n = x$. We define $[\![w]\!] = \{x_1, \dots, x_n\}$.*

A sequence $a_1, \dots, a_n \in A$ is a realization *of w in P, if $(a_i, a_{i+1}) \in C_i|_{(x_i, x_{i+1})}$ for any $i \in \{1, \dots, n-1\}$. We say that two elements $a, a' \in A$ are* connected *via w (in P), if there exists a realization $a = a_1, a_2, \dots, a_{n-1}, a_n = a'$ of the pattern w.*

For two patterns $w = (x_1, C_1, \dots, x_n)$, $w' = (x'_1, C'_1, \dots, x'_m)$ with $x_n = x'_1$ we define their concatenation by $wv = (x_1, C_1, \dots, x_n, C'_1, \dots, x'_m)$. We write w^k for a k-fold concatenation of a closed pattern w with itself.

Definition V.2. *A 1-minimal instance $P = (V, A, \mathcal{C})$ is a* Prague strategy, *if for every $x \in V$, every pair of closed patterns v, w in P with base x such that $[\![v]\!] \subseteq [\![w]\!]$, and every $a, a' \in S_x^P$ connected via the pattern v in P, there exists a natural number k such that a is connected to a' via the pattern w^k.*

First we show that every $(2, 3)$-minimal instance is a Prague strategy. We need an auxiliary lemma.

Lemma V.3. *Let $P = (V, A, \mathcal{C})$ be a $(2, 3)$-minimal instance, let $x, x' \in V$ and let $w = (x = x_1, C_1, x_2, \dots, x_n = x')$ be a pattern. Then a is connected to a' via w in P for any $a, a' \in A$ such that $(a, a') \in S_{(x,x')}^P$.*

Proof: Using Lemma III.2 we obtain $a_2 \in A$ such that $(a, a_2) \in S_{(x_1,x_2)}^P$ and $(a_2, a') \in S_{(x_2,x_n)}^P$. The element a_2 is the second (after a) element of a realization of the pattern w.

Similarly, there exists an element $a_3 \in A$ such that $(a_2, a_3) \in S^P_{(x_2,x_3)}$ and $(a_3, a') \in S^P_{(x_3,x_n)}$. Repeated applications of this reasoning produce a realization of the pattern w connecting a to a'. ■

Lemma V.4. *Every $(2,3)$-minimal instance is a Prague strategy.*

Proof: Let $x \in V$, let v, w be closed patterns in P with base x such that $[\![v]\!] \subseteq [\![w]\!]$, and let $a, a' \in S^P_x$ be elements connected via $v = (x_1, \dots, x_n)$. Let $a = a_1, \dots a_n = a'$ be a realization of v. Since x_2 appears in w there exists an initial part of w, say w', starting with x and ending with x_2. Since $(a, a_2) \in S^P_{(x,x_2)}$ we use Lemma V.3 to connect a to a_2 via w'. Since x_3 appears in w there exists w'' such that $w'w''$ is an initial part of w^2 and such that w'' ends in x_3. Since $(a_2, a_3) \in S^P_{(x_2,x_3)}$ we use Lemma V.3 again to connect a_2 to a_3 via the pattern w''. Now a_1 and a_3 are connected via the pattern $w'w''$. By continuing this reasoning we obtain the pattern w^k (for some k) connecting a to a'. ■

Part (iii) of the following lemma generalizes Proposition III.3.

Lemma V.5. *Let $P = (V, A, \mathcal{C})$ be a 1-minimal instance. The following are equivalent.*

- (i) *P is a Prague strategy.*
- (ii) *For every $x \in V$, every pair of closed patterns v, w in P with base x such that $[\![v]\!] \subseteq [\![w]\!]$, and every $a, a' \in S^P_x$ connected via the pattern v in P, there exists a natural number m such that, for all $k \geq m$, the elemenents a, a' are connected via the pattern w^k;*
- (iii) *For every two elements $(x, B), (x', B')$ in the same component of the qoset $\mathrm{Qoset}(P)$ and every constraint $C \in \mathcal{C}$ whose scope contains $\{x, x'\}$, we have $C|_{(x,x')}{}^+[B] = B'$.*

Proof: Trivially $(ii) \implies (i)$. We do not need the implication $(iii) \implies (i)$ in this paper, therefore we omit the proof (see [9]).

For $(i) \implies (ii)$ it is clearly enough to prove the claim for $a = a'$. To do so, we obtain (using (i)) a natural number p such that a is connected to a via w^p. Let b be an element of A such that a is connected to b via w and b is connected to a via w^{p-1}. We use the property (i) for a, b and the pattern w^p to find a natural number q such that a is connected to b via w^{pq}. From the facts that a is connected to a via w^p and also via w^{pq+p-1} (as a is connected to b via w^{pq} and b to a via w^{p-1}) we get that a is connected to a via $w^{ip+j(pq+p-1)}$ for arbitrary i, j. Since p and $pq + p - 1$ are coprime, the claim follows.

For $(i) \implies (iii)$ let $(x, B) = (x_1, B_1)$, (x_2, B_2), $\dots$, $(x_n, B_n) = (x', B') = (x_{n+1}, B'_1)$, (x_{n+2}, B'_2), $\dots$, $(x_m, B_m) = (x, B)$ be a sequence of elements of $\mathrm{Qoset}(P)$ and $C_1, \dots, C_{m-1} \in \mathcal{C}$ be constraints such that $C_i|_{(x_i,x_{i+1})}{}^+[B_i] = B_{i+1}$ for every $i = 1, \dots, m-1$.

Assume that there exists $a, a' \in A$ such that $(a, a') \in C|_{(x,x')}$ and $a' \in B'$ while $a \notin B$. We can find an element $b \in B$ such that b is connected to a' via the pattern $(x_1, C_1, \dots, x_n)$. The elements b, a are connected via the pattern $(x_1, C_1, \dots, C_{n-1}, x_n, C, x_1)$, therefore, by (i), they must be connected via a power of the pattern $(x_1, C_1, \dots, C_{m-1}, x_m)$, which contradicts $a \notin B$. This contradiction shows that $C|_{(x',x)}{}^+[B'] \subseteq B$. Similarly $C|_{(x,x')}{}^+[B] \subseteq B$ and the proof can be finished as in Proposition III.3. ■

The following lemma covers the last gap.

Lemma V.6. *Let $P = (V, A, \mathcal{C})$ be an instance of $\mathrm{CSP}(\mathrm{Inv}\,\mathbf{A})$ which is a Prague strategy and let $\mathcal{F} = \{(x, E_x) : x \in W\}$ be a maximal component of the qoset $\mathrm{AbsQoset}(P)$. Then the restriction $Q = P|_{\mathcal{F}}$ is a Prague strategy.*

Proof: It is easy to see that, for any $x, x' \in V$, any $B \triangleleft \mathbf{S}^P_x$ and any constraint C whose scope contains $\{x, x'\}$, the set $C|_{(x,x')}{}^+[B]$ is an absorbing subuniverse of $\mathbf{S}^P_{x'}$ (with respect to the same term operation of $\mathbf{A}$). Therefore $C|_{(x,x')}{}^+[E_x] = S^P_{x'}$ whenever $x \in W$ and $x' \in V - W$. From this fact and Lemma V.5 part (iii) it follows that Q is 1-minimal.

To prove that Q is a Prague strategy let v and $w = (x = x_1, C_1, x_2, \dots, C_{n-1}, x_n = x)$ be closed patterns with base x such that $[\![v]\!] \subseteq [\![w]\!]$ and let $a, a' \in S^Q_x$ be elements connected via v in Q. Let t be a k-ary term operation providing the absorptions $E_x \triangleleft \mathbf{S}^P_x$. By Lemma V.5 part (ii) we can find a natural number m such that any two elements b, b', which are connected in P via some closed pattern v' with base x such that $[\![v']\!] \subseteq [\![w]\!]$, are connected via w^m.

We form a matrix with k rows and $(km(n-1)+1)$ columns. The i-th row is formed as follows. We find a realization (1) of the pattern $w^{(i-1)m}$ connecting a to an element b in Q. This is possible since Q is 1-minimal. (For $i = 1$ we consider the empty sequence.) Then we find a realization (3) of the pattern $w^{(k-i)m}$ connecting some element b' to a' in Q. Finally we find a realization (2) of the pattern w^m connecting b to b' in the strategy P (which is possible by the last sentence in the previous paragraph). Finally we join the realizations (1),(2),(3). When we apply the operation t to the columns of this matrix, we get a realization of the pattern w^{km} connecting $a = t(a, \dots, a)$ to $a' = t(a', \dots, a')$ in Q, which finishes the proof. ■

VI. Proof of Theorem III.7

For the entire section we fix a finite idempotent Taylor algebra $\mathbf{A}$.

Two absorptions can be provided by different term operations. A simple trick can unify them:

Lemma VI.1. *Let $\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_1, \mathbf{B}_2 \in \mathrm{V_{fin}}(\mathbf{A})$ and $\mathbf{B}_1 \triangleleft \mathbf{A}_1, \mathbf{B}_2 \triangleleft \mathbf{A}_2$. Then there exists a term operation t of $\mathbf{A}$ such that both absorptions are with respect to the operation t. (More precisely, $\mathbf{B}_1$ absorbs $\mathbf{A}_1$ with respect to $t^{\mathbf{A}_1}$ and $\mathbf{B}_2$ absorbs $\mathbf{A}_2$ with respect to $t^{\mathbf{A}_2}$.)*

Proof: If $\mathbf{B}_i$ is an absorbing subalgebra of $\mathbf{A}_i$ with respect to an n_i-ary operation t_i, $i = 1, 2$, then the n_1n_2-ary operation defined by $t(a_1, \dots, a_{n_1n_2}) = t_1(t_2(a_1, \dots, a_{n_2}), t_2(a_{n_2+1}, \dots), \dots)$ satisfies the conclusion. ■

The main tool for proving Theorem III.7 is the Absorption Theorem (Theorem III.6. in [17]). We require a definition of a linked subdirect product:

Definition VI.2. *Let $R \subseteq_S A_1 \times A_2$. We say that two elements $a, a' \in A_1$ are R-*linked *via $c_0, c_1, \dots, c_{2n}$, if $a = c_0, c_{2n} = a'$ and $(c_{2i}, c_{2i+1}) \in R$ and $(c_{2i+2}, c_{2i+1}) \in R$ for all $i = 0, 1, \dots, n-1$.*

We say that R is linked*, if any two elements $a, a' \in A_1$ are R-linked.*

Theorem VI.3. *[17], [18] Let $\mathbf{A}_1, \mathbf{A}_2 \in \mathrm{V}_{\mathrm{fin}}(\mathbf{A})$ be absorption free algebras and let $R \leq_S \mathbf{A}_1 \times \mathbf{A}_2$ be linked. Then $R = A_1 \times A_2$.*

We will need the following consequence.

Lemma VI.4. *Let $\mathbf{A}_1, \mathbf{A}_2 \in \mathrm{V}_{\mathrm{fin}}(\mathbf{A})$ be absorption free algebras, let $R \leq_S \mathbf{A}_1 \times \mathbf{A}_2$ and let α_1 be a maximal congruence of $\mathbf{A}_1$. Then either $\{R^+[C] : C$ is an α_1-block$\}$ is the set of blocks of a maximal congruence α_2 of $\mathbf{A}_2$, or $R^+[C] = A_2$ for every α_1-block C.*

Proof: If the sets $R^+[C]$ are disjoint, then they are blocks of an equivalence on A_2, and it is straightforward to check that this equivalence is indeed a maximal congruence of $\mathbf{A}_2$.

In the other case we consider the factor algebra $\mathbf{A}_1' = \mathbf{A}_1/\alpha_1$ and the subdirect subalgebra $R' = \{([a_1]_{\alpha_1}, a_2) : (a_1, a_2) \in R\}$ of $\mathbf{A}_1 \times \mathbf{A}_2$. Since α_1 is maximal, the algebra $\mathbf{A}_1'$ has only trivial congruences. Also, $\mathbf{A}_1'$ is an absorption free algebra, because the preimage of any absorbing subalgebra $\mathbf{C} \leq \mathbf{A}_1'$ is an absorbing subalgebra of $\mathbf{A}_1$.

We define a congruence $\sim$ on $\mathbf{A}_1'$ by $[a_1] \sim [a_2]$, if $[a_1], [a_2]$ are R'-linked. As not all of the sets $R^+[C]$ are disjoint, $\sim$ is not the diagonal congruence, therefore $\sim$ must be the full congruence, and it follows that R' is linked. By Theorem VI.3 $R' = A_1' \times A_2$. In other words, $R^+[C] = A_2$ for every α_1-block C. ■

Links are absorbed to absorbing subuniverses:

Lemma VI.5. *Let $\mathbf{A}_1, \mathbf{A}_2 \in \mathrm{V}_{\mathrm{fin}}(\mathbf{A})$, let $R \leq_S \mathbf{A}_1 \times \mathbf{A}_2$, let $\mathbf{B}_1 \lhd \mathbf{A}_1$, $\mathbf{B}_2 \lhd \mathbf{A}_2$ and let $S = R \cap (B_1 \times B_2)$ be subdirect in $B_1 \times B_2$. Then every pair $b_1, b_1' \in B_1$ of R-linked elements is also S-linked.*

Proof: By Lemma VI.1 there exists a term operation t such that both absorptions are with respect to t. Let $b_1, b_1' \in B_1$ be arbirary. Since S is subdirect, there exist $b_2, b_2' \in B_2$ such that $(b_1, b_2), (b_1', b_2') \in S$. Let b_1, b_1' be R-linked via $c_0, c_1, \dots, c_{2n}$. Now the following sequence S-links b_1 to b_1':

$$b_1 = t(b_1 = c_0, b_1, \dots, b_1), t(c_1, b_2, \dots, b_2), t(c_2, b_1, \dots, b_1),$$
$$\dots, t(b_1' = c_{2n}, b_1, \dots, b_1), t(b_2', c_1, b_2, \dots, b_2), \dots,$$
$$t(b_1', b_1', b_1, \dots, b_1), t(b_2', b_2', c_1, b_2, \dots, b_2), \dots, \dots, \dots,$$
$$t(b_1', \dots, b_1') = b_1'.$$

■

A subalgebra of a conservative absorption free algebra which hits all blocks of a proper congruence is absorption free:

Lemma VI.6. *Let $\mathbf{A}_1 \in \mathrm{V}_{\mathrm{fin}}(\mathbf{A})$ be a conservative absorption free algebra and let α be a proper congruence of $\mathbf{A}_1$. Then any subalgebra $\mathbf{B}$ of $\mathbf{A}_1$ which has a nonempty intersection with every α-block is an absorption free algebra.*

Proof: For a contradiction, consider a proper absorbing subuniverse C of $\mathbf{B}$. Let $D_1, \dots, D_k$ be all the α-blocks whose intersections with B and C are equal and let $E_1, \dots, E_l$ be the remaining α-blocks which intersect C nonempty.

We claim that, for every $m \leq l$, the set $F = D_1 \cup \dots \cup D_k \cup E_1 \cup \dots \cup E_m$ is an absorbing subuniverse of $\mathbf{A}$: Let t be a term operation providing the absorption $C \lhd \mathbf{B}$ and let $\mathbf{a}$ be a tuple of elements in A with all the coordinates in F with the exception of, say, a_i. We take any tuple $\mathbf{b}$ such that $b_j \; \alpha \; a_j$ for all coordinates j, $b_i \in B - C$ and $b_j \in C$ for all $j \neq i$. As $C \lhd \mathbf{B}$, $t(\mathbf{b})$ is an element of C and, due to conservativity, $t(\mathbf{b}) \in F$. Therefore $t(\mathbf{a}) \in F$ as this element is α-congruent to $t(\mathbf{b})$.

For an appropriate choice of $m \leq l$, F is a proper nonempty subset of $\mathbf{A}$ and $F \lhd \mathbf{A}$, a contradiction. ■

A subdirect product of conservative absorption free algebras is absorption free:

Lemma VI.7. *Let $\mathbf{R} \leq_S \mathbf{A}_1 \times \mathbf{A}_2 \times \dots \times \mathbf{A}_n$, where every $\mathbf{A}_i \in \mathrm{V}_{\mathrm{fin}}(\mathbf{A})$ is a conservative absorption free algebra. Then $\mathbf{R}$ is an absorption free algebra.*

Proof: We take a minimal counterexample to the lemma in the following sense: We assume that the lemma holds true for every smaller n, and also for every $\mathbf{R}' \leq_S \mathbf{A}_1' \times \dots \times \mathbf{A}_n'$ such that $|A_i'| \leq |A_i|$, $i = 1, \dots, n$, where at least one inequality is strict. We can assume that no A_i is one-element, otherwise we can employ the minimality assumption and use the lemma for the projection to the remaining coordinates.

Let S be a proper absorbing subuniverse of $\mathbf{R}$. It is easily seen that the projection of S to any coordinate i is an absorbing subuniverse of $\mathbf{A}_i$, thus S is subdirect. Let α_1 be a maximal congruence of $\mathbf{A}_1$.

For every $i \in \{1, 2, \dots, n\}$ we have two possibilities (see Lemma VI.4):

(i) $\{R|_{(1,i)}{}^+[C] : C$ is an α_1-block$\}$ are blocks of a maximal congruence α_i of $\mathbf{A}_i$
(ii) $R|_{(1,i)}{}^+[C] = A_i$ for every α_1-block C

Let G (resp. W) denote the set of is for which the first (resp. the second) possibility takes place. By using Lemma VI.4 again, we get that $R|_{(i,j)}{}^+[C] = A_j$ for any $i \in G$, $j \in W$ and any α_i-block C.

We take an arbitrary tuple $(a_1, \dots, a_n) \in R$ and we aim to show that this tuple belongs to S as well. The proof splits into two cases.

Assume first that for every $i \in G$, α_i is the diagonal congruence. Let $A_j' = R|_{(1,j)}{}^+[\{a\}]$, $j = 1, 2, \dots, n$, let $R' = R \cap (A_1' \times \dots \times A_n')$ and let $S' = S \cap (A_1' \times \dots \times A_n')$. Note that A_i' is one element (for $i \in G$) or equal to A_i (for $i \in W$), and S' absorbs R'. Therefore $R' = S'$ (by the minimality assumption) and hence $(a_1, \dots, a_n) \in S$.

Now assume that some α_i, $i \in G$ is not the diagonal congruence. Take a proper subset B of A_i which contains a_i and which intersects all α_i-blocks nonempty. Let $A'_j = R|_{(i,j)}{}^+[B]$, $j = 1, \ldots, n$, and let R', S' be as in the previous paragraph. By Lemma VI.6 every $\mathbf{A}'_j$, $j \in G$ is an absorption free algebra, and $\mathbf{A}'_j = A_j$ for $j \in W$ is absorption free as well. Now, by the minimality assumption, $S' = R'$, hence $(a_1, \ldots, a_n) \in S$. ■

The following lemma proves a special case of Theorem III.7. Note that we do not require $\mathbf{A}_2$ to be conservative.

Lemma VI.8. *Let $\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_1, \mathbf{B}_2 \in \mathrm{V}_{\text{fin}}(\mathbf{A})$ be algebras such that $\mathbf{A}_1$ is conservative, $\mathbf{B}_1 \lhd\!\!\lhd \mathbf{A}_1$ and $\mathbf{B}_2 \lhd\!\!\lhd \mathbf{A}_2$. Let $R \leq_S \mathbf{A}_1 \times \mathbf{A}_2$. If $R \cap (B_1 \times B_2) \neq \emptyset$ and there exists a pair $(a_1, b_2) \in R$ such that $a_1 \in A_1 - B_1$ and $b_2 \in B_2$, then $B_1 \times B_2 \subseteq R$.*

Proof: Let $S = R \cap (B_1 \times B_2)$. As before, the projection of S to the first (resp. second) coordinate is an absorbing subuniverse of $\mathbf{B}_1$ (resp. $\mathbf{B}_2$), and, by the assumption, S is nonempty, therefore $S \leq_S B_1 \times B_2$. Let $b_1 \in B_1$ be such that $(b_1, b_2) \in R$. We define a congruence on $\mathbf{A}_1$ by putting $c \sim d$, if c and d are R-linked.

Let C denote the set of all the elements of B_1 which are not R-linked to b_1. If C is empty, then, by Lemma VI.5, S is linked and therefore $S = B_1 \times B_2$ by Theorem VI.3.

Otherwise, C is a proper subuniverse of $\mathbf{B}_1$ and we claim that $C \lhd \mathbf{B}_1$: Let t be a term operation providing the absorption $B_1 \lhd \mathbf{A}_1$ and let $\mathbf{c}$ be a tuple of elements of B_1 with all the coordinates but one, say c_i, in C. Let $\mathbf{d}$ be the tuple defined by $d_i = a_1$, and $d_j = c_j$ for $j \neq i$. As $d_i \sim c_i$ for all i we have $t(\mathbf{c}) \sim t(\mathbf{d})$. But $t(\mathbf{d})$ lies inside C (as B_1 absorbs $\mathbf{A}_1$ and $\mathbf{A}_1$ is conservative), hence also $t(\mathbf{c}) \in C$.

We have found a proper absorbing subuniverse C of $\mathbf{B}_1$, a contradiction. ■

The next lemma generalizes the previous one. Recall the definition of the quasi-ordering $\preceq$ introduced in Subsection III-E.

Lemma VI.9. *Let $\mathbf{A}_1, \ldots, \mathbf{A}_n, \mathbf{B}_1, \ldots, \mathbf{B}_n \in \mathrm{V}_{\text{fin}}(\mathbf{A})$ be algebras such that $\mathbf{A}_1, \ldots, \mathbf{A}_{n-1}$ are conservative and $\mathbf{B}_i \lhd\!\!\lhd \mathbf{A}_i$ for all $i = 1, \ldots, n$. Let $R \leq_S \mathbf{A}_1 \times \cdots \times \mathbf{A}_n$ and assume $R \cap (B_1 \times \cdots \times B_n) \neq \emptyset$. If $\{1, 2, \ldots, n-1\}$ is an (R, B)-strand and there exists a tuple $(a_1, a_2, \ldots, a_{n-1}, b_n) \in R$ such that $b_n \in B_n$ and $a_i \in A_i - B_i$ for some (equivalently every) $i \in \{1, 2, \ldots, n-1\}$, then every tuple $\mathbf{c} \in B_1 \times \cdots \times B_n$ such that $\mathbf{c}|_{\{1,2,\ldots,n-1\}} \in R|_{\{1,2,\ldots,n-1\}}$ belongs to R.*

Proof: We take a minimal counterexample in the same sense as in Lemma VI.7, i.e., we assume that the lemma holds if n is smaller and also if some A_i is smaller.

We may assume that all B_is are at least two-element and let us also assume that if some of the algebras $\mathbf{B}_1, \ldots, \mathbf{B}_n$ has a nontrivial congruence, then $\mathbf{B}_1$ has a nontrivial congruence (otherwise we just change the indices).

Let α_1 be a maximal congruence of $\mathbf{B}_1$. By applying Lemma VI.4 as in the proof of Lemma VI.7 we get that, for each $i \in \{1, 2, \ldots, n-1\}$, either $R|_{(1,i)}{}^+[C] = B_i$ for every α_1-block C, or $\{R|^+_{(1,i)}[C] : C \text{ is an } \alpha\text{-block } \}$ are blocks of a maximal congruence α_i on $\mathbf{B}_i$.

Let $\mathbf{c} \in B_1 \times \cdots \times B_n$ be an arbitrary tuple such that $\mathbf{c}|_{\{1,2,\ldots,n-1\}} \in R|_{\{1,2,\ldots,n-1\}}$.

If α_1 is the diagonal congruence, then we put $D = \{c_1\} \cup (A_1 - B_1)$. Otherwise, we take an arbitrary D such that $(A_1 - B_1) \cup \{c_1\} \subseteq D \subsetneq A_1$ and D intersects every α_1-block nonempty. Let $A'_i = R|_{(1,i)}{}^+[D]$, $B'_i = B_i \cap A'_i$, $i = 1 \ldots, n$ and $R' = R \cap (A'_1 \times \cdots \times A'_n)$.

For every $i \in \{1, \ldots, n-1\}$, B'_i is an absorption free algebra, either because B'_i is a singleton, or B'_i intersects every α_i block nonempty and we can apply Lemma VI.6. From Lemma VI.8 it follows that $B_1 \times B_n \subseteq R|_{(1,n)}$, therefore $B'_n = B_n$, in particular, $(a_1, a_2, \ldots, a_{n-1}, b_n) \in R'$. Obviously B'_i is an absorbing sublagebra of $\mathbf{A}'_i$ for every $i = 1, \ldots, n$. Now $\mathbf{c} \in R'$ ($\subseteq R$) follows from the miniminality of our counterexample. ■

We are ready to prove Theorem III.7.

Theorem VI.10. *Let $\mathbf{A}_1, \ldots, \mathbf{A}_n, \mathbf{B}_1, \ldots, \mathbf{B}_n \in \mathrm{V}_{\text{fin}}(\mathbf{A})$ be conservative algebras such that $\mathbf{B}_i \lhd\!\!\lhd \mathbf{A}_i$ for all $i = 1, \ldots, n$, let $R \leq_S \mathbf{A}_1 \times \cdots \times \mathbf{A}_n$ and assume that $R \cap (B_1 \times \ldots, B_n) \neq \emptyset$. Then a tuple $\mathbf{a} \in B_1 \times \cdots \times B_n$ belongs to R whenever $\mathbf{a}|_K \in R|_K$ for each (R, B)-strand K.*

Proof: We again use the minimality assumption, i.e., we assume that the theorem holds if n is smaller, or if some A_i is smaller. We can assume that there are at least two (R, B)-strands and that $|B_i| > 1$ for all $i = 1, \ldots, n$. Let $\mathbf{a} \in B_1 \times \cdots \times B_n$ be a tuple such that $\mathbf{a}|_K \in R|_K$ for each (R, B)-strand K, but $\mathbf{a} \notin R$. Note that $\mathbf{a}|_L \in R|_L$ for every proper subset L of $\{1, 2, \ldots, n\}$, because of the minimality assumption – we can apply the theorem to $R|_L$.

Let D be a $\preceq$-minimal (R, B)-strand and let $l \notin D$. Since $l \not\preceq D$, there exists a tuple $\mathbf{c} \in R$ such that $c_l \in B_l$ and $c_i \notin B_i$ for all $i \in D$. Let $E = \{i \in \{1, \ldots, n\} : c_i \notin B_i\} - D$ and $F = \{i \in \{1, \ldots, n\} : c_i \in B_i\}$. Clearly, E and F are unions of (R, B)-strands.

Our aim now is to find a tuple $\mathbf{c}' \in R$ such that $c'_i \in A_i - B_i$ for all $i \in D$, and $c'_i \in B_i$ for all $i \notin D$. If $E = \emptyset$, we can take $\mathbf{c}' = \mathbf{c}$, so suppose otherwise. We consider the following subset of R:

$$R' = \{\mathbf{b} \in R : b_i \in \{a_i\} \cup (A_i - B_i) \text{ for all } i \in D\}.$$

For all $i \in \{1, \ldots, n\}$, let $A'_i = R'|_{\{i\}}$. Let $B'_i = \{a_i\}$ for all $i \in D$, and $B'_i = B_i$ for $i \notin D$.

We have $B'_i \subseteq A'_i$ for every $i \notin D$: for any $b \in B_i$ we apply the theorem for $R|_{D \cup \{i\}}$ to obtain a tuple $\mathbf{e} \in R$ such that $e_i = b$ and $e_j = d_j$ for every $j \in D$.

We know that $\mathbf{a}|_{D \cup E} \in R|_{D \cup E}$, therefore $a|_E \in R'|_E$. Similarly, $a|_F \in R'|_F$.

Observe that any $i \in E$, $j \in F$ are in different (R', B')-strands, since $\mathbf{c} \in R'$. Therefore, the theorem, used for R' and the minimal absorbing subuniverses B'_i of A'_i, proves $\mathbf{a}|_{E \cup F} \in R'|_{E \cup F}$. Let $\mathbf{c}'$ be a tuple from R' with $\mathbf{c}'|_{E \cup F} = \mathbf{a}|_{E \cup F}$. The tuple $\mathbf{c}'$ cannot be equal to $\mathbf{a}$ as $\mathbf{a} \notin R$, therefore $c'_i \in A_i - B_i$ for all $i \in D$, $c'_i \in B_i$ for $i \notin D$.

Now, when we have the sought after tuple $\mathbf{c}'$, we can finish the proof by applying Lemma VI.9 for the following choice: $n' = |D|+1$; $\mathbf{A}'_i = \mathbf{A}_{d_i}$ and $\mathbf{B}'_i = \mathbf{B}_{d_i}$ for $i = 1, \ldots, l$, where $D = \{d_1, \ldots, d_l\}$; $\mathbf{A}'_{n'} = R|_{E\cup F}$; $\mathbf{B}'_{n'} = S|_{E\cup F}$, where $S = \{\mathbf{b} \in R : b_i \in B_i \text{ for all } i \in E \cup F\}$; R' is equal to R viewed as a subset of $A'_1 \times \cdots \times A'_{n'}$; and $(a'_1, a'_2, \ldots, a'_{n'-1}, b'_n) = (c_{d_1}, c_{d_2}, \ldots, c_{d_l}, \mathbf{c}'|_{E\cup F})$. All the assumptions are satisfied, the only nontrivial fact is that $\mathbf{B}'_{n'}$ is absorption free and this follows from Lemma VI.7. ■

VII. Conclusion

We have presented a new, simple algorithm for solving tractable CSPs over conservative languages. We believe that this simplification can help in the final attack on the dichotomy conjecture. No effort has been made to optimize the algorithm, we have not computed its time complexity and we have not compared the complexity with the algorithm of Bulatov. This can be a topic of further research. We note that some reductions can be done using a trick from [26], it would be interesting to see whether this trick can improve the running time.

As mentioned before, our algorithm uses absorbing subuniverses of (subalgebras of) the fixed algebra $\mathbf{A}$ and we do not know if they can be found algorithmically. The dual, relational, version of this problem is also interesting.

Open problem VII.1. *Is the following problem decidable? On input we are given a finite algebra $\mathbf{A}$ with finitely many operations (resp. a finite constraint language Γ on a finite set A) and a subset B of A, and we are asking whether B is an absorbing subuniverse of $\mathbf{A}$ (resp. $\operatorname{Pol}\Gamma$).*

Finally, we remark that our algorithm can be slightly modified so that we would consider only some of the absorptions, namely, the absorptions with respect to a fixed cyclic operation and the absorptions enforced by Theorem VI.3. From the proof of this theorem it can be seen that such absorptions can be found algorithmically.

Acknowledgment

I would like to thank Marcin Kozik, Miklós Maróti and Ralph McKenzie for helpful discussions and the referees for useful comments.

References

[1] T. Feder and M. Y. Vardi, "The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory," *SIAM J. Comput.*, vol. 28, no. 1, pp. 57–104 (electronic), 1999. [Online]. Available: http://dx.doi.org/10.1137/S0097539794266766

[2] P. Jeavons, D. Cohen, and M. Gyssens, "Closure properties of constraints," *J. ACM*, vol. 44, no. 4, pp. 527–548, 1997. [Online]. Available: http://dx.doi.org/10.1145/263867.263489

[3] A. A. Bulatov, A. A. Krokhin, and P. Jeavons, "Constraint satisfaction problems and finite algebras," in *Automata, languages and programming (Geneva, 2000)*, ser. Lecture Notes in Comput. Sci. Berlin: Springer, 2000, vol. 1853, pp. 272–282.

[4] A. Bulatov, P. Jeavons, and A. Krokhin, "Classifying the complexity of constraints using finite algebras," *SIAM J. Comput.*, vol. 34, no. 3, pp. 720–742 (electronic), 2005. [Online]. Available: http://dx.doi.org/10.1137/S0097539700376676

[5] A. Bulatov, "Mal'tsev constraints are tractable," Computing Laboratory, University of Oxford, Oxford, UK, Tech. Rep. PRG-RR-02-05, 2002.

[6] A. Bulatov and V. Dalmau, "A simple algorithm for Mal′tsev constraints," *SIAM J. Comput.*, vol. 36, no. 1, pp. 16–27 (electronic), 2006. [Online]. Available: http://dx.doi.org/10.1137/050628957

[7] V. Dalmau, "Generalized majority-minority operations are tractable," *Log. Methods Comput. Sci.*, vol. 2, no. 4, pp. 4:1, 14, 2006.

[8] P. M. Idziak, P. Markovic, R. McKenzie, M. Valeriote, and R. Willard, "Tractability and learnability arising from algebras with few subpowers," in *LICS*. IEEE Computer Society, 2007, pp. 213–224.

[9] L. Barto and M. Kozik, "Constraint satisfaction problems of bounded width," in *FOCS'09: Proceedings of the 50th Symposium on Foundations of Computer Science*, 2009, pp. 595–603.

[10] ——, "Constraint satisfaction problems solvable by local consistency methods," in preparation.

[11] A. A. Bulatov, "A dichotomy theorem for constraint satisfaction problems on a 3-element set," *J. ACM*, vol. 53, no. 1, pp. 66–120 (electronic), 2006. [Online]. Available: http://dx.doi.org/10.1145/1120582.1120584

[12] T. J. Schaefer, "The complexity of satisfiability problems," in *Conference Record of the Tenth Annual ACM Symposium on Theory of Computing (San Diego, Calif., 1978)*. New York: ACM, 1978, pp. 216–226.

[13] A. A. Bulatov, "Tractable conservative constraint satisfaction problems," *Logic in Computer Science, Symposium on*, vol. 0, p. 321, 2003.

[14] L. Barto, M. Kozik, and T. Niven, "Graphs, polymorphisms and the complexity of homomorphism problems," in *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*. New York, NY, USA: ACM, 2008, pp. 789–796.

[15] ——, "The CSP dichotomy holds for digraphs with no sources and no sinks (a positive answer to a conjecture of Bang-Jensen and Hell)," *SIAM J. Comput.*, vol. 38, no. 5, pp. 1782–1802, 2008/09. [Online]. Available: http://dx.doi.org/10.1137/070708093

[16] L. Barto and M. Kozik, "Congruence distributivity implies bounded width," *SIAM Journal on Computing*, vol. 39, no. 4, pp. 1531–1542, 2009. [Online]. Available: http://link.aip.org/link/?SMJ/39/1531/1

[17] ——, "New conditions for Taylor varieties and CSP," *Logic in Computer Science, Symposium on*, vol. 0, pp. 100–109, 2010.

[18] ——, "Absorbing subalgebras, cyclic terms and the constraint satisfaction problem," submitted.

[19] T. Feder, P. Hell, and J. Huang, "List homomorphisms and circular arc graphs," *Combinatorica*, vol. 19, pp. 487–505, 1999.

[20] ——, "Bi-arc graphs and the complexity of list homomorphisms," *J. Graph Theory*, vol. 42, pp. 61–80, January 2003. [Online]. Available: http://portal.acm.org/citation.cfm?id=1380681.1380685

[21] P. Hell and A. Rafiey, "The dichotomy of list homomorphisms for digraphs," *CoRR*, vol. abs/1004.2908, 2010.

[22] L. Egri, A. Krokhin, B. Larose, and P. Tesson, "The complexity of the list homomorphism problem for graphs," in *27th International Symposium on Theoretical Aspects of Computer Science (STACS 2010)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), J.-Y. Marion and T. Schwentick, Eds., vol. 5. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2010, pp. 335–346. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2010/2467

[23] R. Takhanov, "A dichotomy theorem for the general minimum cost homomorphism problem," in *27th International Symposium on Theoretical Aspects of Computer Science (STACS 2010)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), J.-Y. Marion and T. Schwentick, Eds., vol. 5. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2010, pp. 657–668. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2010/2493

[24] V. Kolmogorov and S. Zivny, "The complexity of conservative finite-valued csps," *CoRR*, vol. abs/1008.1555, 2010.

[25] M. Maróti and R. McKenzie, "Existence theorems for weakly symmetric operations," *Algebra Universalis*, vol. 59, no. 3-4, pp. 463–489, 2008. [Online]. Available: http://dx.doi.org/10.1007/s00012-008-2122-9

[26] M. Maróti, "Tree on top of maltsev," preprint.

Appendix F – Near unanimity in NL

Near Unanimity Constraints Have Bounded Pathwidth Duality

Libor Barto
Mathematics & Statistics Department
McMaster University
Hamilton, Ontario L8S 4L8, Canada
and
Dept. of Algebra, Charles University
Praha 8, Czech Republic
Email: libor.barto@gmail.com

Marcin Kozik
Theoretical Computer Science Department
Faculty of Mathematics & Computer Science
Jagiellonian University
Kraków 30-348, Poland
Email: kozik@tcs.uj.edu.pl

Ross Willard
Pure Mathematics Department
University of Waterloo
Waterloo, Ontario N2L 3G1, Canada
Email: rdwillar@uwaterloo.ca

***Abstract*—We show that if a finite relational structure has a near unanimity polymorphism, then the constraint satisfaction problem with that structure as its fixed template has bounded pathwidth duality, putting the problem in nondeterministic logspace. This generalizes the analogous result of Dalmau and Krokhin for majority polymorphisms and lends further support to a conjecture suggested by Larose and Tesson.**

***Keywords*-constraint satisfaction; polymorphism; near unanimity; pathwidth duality; linear datalog; absorption.**

I. Introduction

The *constraint satisfaction problem* (CSP) is a well-known and important protocol for declaring combinatorial problems arising from artificial intelligence [31]. It is also the source of deep research problems in theoretical computer science. In particular, Feder and Vardi [18] identified *fixed-template* versions of CSP as worthy of study and formulated their famous *CSP Dichotomy Conjecture*: for every template (i.e., finite relational structure) $\mathbf{B}$, the problem $\mathrm{CSP}(\mathbf{B})$ is either NP-complete or solvable in polynomial time. Considerable progress towards resolving this conjecture has been achieved during the last 12 years, in part because of the success of the so-called "algebraic approach" championed by Jeavons (e.g. [23], [11]). In this approach templates are classified according to their "polymorphisms," i.e., multi-variable functions that preserve the relations of the template; these functions connect CSP to universal algebra and its toolboxes and perspectives.

An important illustration of the power of the algebraic approach is the recent characterization by the first two authors of templates having *bounded width*. These are structures $\mathbf{B}$ for which $\mathrm{CSP}(\mathbf{B})$ can be solved in polynomial time by a standard local consistency checking algorithm. An "obvious" obstruction to having bounded width is the structure having relations which encode linear equations over some additive abelian group [18]. The algebraic perspective gives a precise, though technical, description ([29], [28]) of the class of templates which omit the obvious obstruction. In [30] a simple characterization of this class in terms of polymorphisms was given, and in [6] this characterization was used to show that every member of the class does indeed have bounded width.

Having bounded width can be characterized in many equivalent ways, including $\mathrm{CSP}(\mathbf{B})$ (here identified with the class of finite structures that admit a homomorphism to $\mathbf{B}$) having *bounded treewidth duality* (see [21], [18], [25], [10], [20], [29]), or the complement class $\neg\mathrm{CSP}(\mathbf{B})$ being definable in the logic Datalog [18]. (*Datalog* is a relational query language whose salient feature is its ability to formulate least-fixed-point recursive definitions [33], [1].)

A related property is that of $\mathrm{CSP}(\mathbf{B})$ having bounded *pathwidth* duality; this more restrictive property puts $\mathrm{CSP}(\mathbf{B})$ in the complexity class NL ([14], [15]) and has several equivalent formulations [15], including $\neg\mathrm{CSP}(\mathbf{B})$ being definable in *linear* Datalog (in which only non-branching recursion is permitted [1]). The "obvious" obstruction to bounded pathwidth duality, in addition to linear equations over an abelian group, is Horn 3-SAT ([2], [13]). Again, universal algebra gives a precise characterization of the class of templates which omit both obstructions [27], and in light of the available evidence (especially [26]) it is natural (as [27] noted) to conjecture that every template in this class has bounded pathwidth duality.

From the algebraic perspective, four reasonable intermediate steps on the journey to verifying this latter conjecture are:

1) Verify the conjecture on the 2-element domain.
2) Prove it for templates having a *majority* polymorphism.
3) Prove it for templates having a *near unanimity* polymorphism in $d+1$ variables for some $d \geq 2$.
4) Prove it for templates having *Jónsson* polymorphisms.

Step 1 was accomplished in [27], [2]. Step 2 was solved by Dalmau and Krokhin [16], who proved that if $\mathbf{B}$ has a majority polymorphism then $\mathrm{CSP}(\mathbf{B})$ has bounded pathwidth duality; they also posed Steps 3 and 4 as next steps. (In fact, Step 2 is the first case of Step 3, i.e. with $d = 2$.) The property stated in Step 3 has been called the *d-mapping property* by Feder and Vardi [18], who also showed that $\mathbf{B}$ having this property is equivalent to $\mathbf{B}$ having *bounded strict width* (implying that solutions to $\mathrm{CSP}(\mathbf{B})$ can be found by a greedy algorithm).

In this paper we verify Step 3. That is, we show (Theorem 7) that if a template $\mathbf{B}$ has a near unanimity polymorphism then

1043-6871/12 $26.00 © 2012 IEEE
DOI 10.1109/LICS.2012.24

$\mathrm{CSP}(\mathbf{B})$ has bounded pathwidth duality and hence is in NL. By a result of the first author [4], this also verifies Step 4.

Our proof is inspired by and follows to some extent the proof in [16] for the case $d = 2$. However the details are rather more complicated. In addition, we need (and establish) a surprising new algebraic fact about absorption (Theorem 6) which may be of independent interest to universal algebraists.

The plan of this paper is the following. In section II we summarize the background needed regarding constraint satisfaction problems and templates, bounded pathwidth duality, and algebra. The new algebraic result (Theorem 6) is stated at the end of subsection II-C but its proof is deferred until the end of section III. In section III the main result (Theorem 7) is stated and quickly reduced to the "binary" case; then in subsections III-B and III-C the binary case is proved using Theorem 6; finally in subsection III-D Theorem 6 is proved.

II. Basic Definitions and Tools

A. Structures and Constraint Satisfaction Problems

Everything in this subsection before Definition 1 is standard.

A (*relational*) *vocabulary* is any set of *relation symbols*, each of which is assigned an integer $n \geq 1$ called the *arity* of the symbol. In this paper all relational vocabularies are finite.

If τ is a vocabulary, a τ*-structure* is an object $\mathbf{B}$ consisting of a non-empty set B (the *universe* of $\mathbf{B}$) and, for each relation symbol $R \in \tau$ of arity n, an n-ary *relation* $R^{\mathbf{B}}$ on B, i.e., a subset $R^{\mathbf{B}} \subseteq B^n$. The relations $R^{\mathbf{B}}$ ($R \in \tau$) are the *basic relations* of $\mathbf{B}$. A structure is *finite* if its universe is finite, and is *binary* if every symbol in its vocabulary has arity 1 or 2.

Given two structures $\mathbf{A}, \mathbf{B}$ with the same vocabulary τ, a *homomorphism* from $\mathbf{A}$ to $\mathbf{B}$ is a function $h : A \to B$ which preserves basic relations; that is, for all $R \in \tau$ of arity n and for all $a_1, \ldots, a_n \in A$, if $(a_1, \ldots, a_n) \in R^{\mathbf{A}}$ then $(h(a_1), \ldots, h(a_n)) \in R^{\mathbf{B}}$. $\mathrm{Hom}(\mathbf{A}, \mathbf{B})$ denotes the set of homomorphisms from $\mathbf{A}$ to $\mathbf{B}$. We write $\mathbf{A} \to \mathbf{B}$ to assert $\mathrm{Hom}(\mathbf{A}, \mathbf{B}) \neq \varnothing$.

Given a finite τ-structure $\mathbf{B}$, the *constraint satisfaction problem with fixed template* $\mathbf{B}$ ("homomorphism version") is the decision problem, denoted $\mathrm{CSP}(\mathbf{B})$, which takes as input an arbitrary finite τ-structure $\mathbf{A}$ and asks whether $\mathbf{A} \to \mathbf{B}$.

Given a τ-structure $\mathbf{B}$ with universe B and a subset $X \subseteq B$, the *substructure induced by* $\mathbf{B}$ *on* X is the τ-structure $\mathbf{B}{\restriction}_X$ with universe X and relations defined by $R^{\mathbf{B}\restriction_X} = R^{\mathbf{B}} \cap X^n$ for each n-ary $R \in \tau$.

Given sets $A_1, \ldots, A_n$ and $X \subseteq A_1 \times \cdots \times A_n$, $\mathrm{proj}_i(X)$ denotes the projection of X onto coordinate i. We say that X is *subdirect* and write $X \subseteq_{sd} A_1 \times \cdots \times A_n$ if $\mathrm{proj}_i(X) = A_i$ for all $1 \leq i \leq n$.

Given a τ-structure $\mathbf{B}$, a relation $S \subseteq B^n$ is $\wedge$*-atomic definable over* $\mathbf{B}$ if there exists a formula $\varphi(x_1, \ldots, x_n)$ in the language of first-order logic with equality and vocabulary τ such that (i) φ is a conjunction of *atomic formulas* (assertions "$(x_{i_1}, \ldots, x_{i_n}) \in R$" or "$x_i = x_j$") and (ii) S is defined by φ, i.e., $\{\mathbf{b} \in B^n : \mathbf{B} \models \varphi(\mathbf{b})\} = S$. We say that S is *primitive positive* (or *pp*)*-definable over* $\mathbf{B}$ if, for some $m \geq 0$, S is the projection onto the first n coordinates of an $(n+m)$-ary $\wedge$-atomic-definable relation. If $\mathbf{B}$ is a binary structure, we say that the set of basic relations of $\mathbf{B}$ is *closed under* $\wedge$*-atomic definitions* provided every at-most-2-ary relation on B which is $\wedge$-atomic definable over $\mathbf{B}$ is already a basic relation of $\mathbf{B}$.

The remaining notions in this subsection are not standard.

Definition 1. Suppose $\mathbf{B}$ is a finite binary τ-structure and A is a finite non-empty set. A *potato system over* $\mathbf{B}$ *with domain* A is an indexed system $\mathcal{P} = (P_a, E_{a,b} : a, b \in A)$ satisfying the following. For all $a, b \in A$:

1) P_a is a 1-ary basic relation of $\mathbf{B}$.
2) $E_{a,b}$ is a 2-ary basic relation of $\mathbf{B}$.
3) $E_{a,b} \subseteq P_a \times P_b$.
4) $E_{a,a} = \{(x, x) : x \in P_a\}$.
5) $E_{b,a} = \{(y, x) : (x, y) \in E_{a,b}\}$.

Potato systems over $\mathbf{B}$ are similar to (1,2)-systems defined in [6]; they differ in that we do not require $E_{a,b} \subseteq_{sd} P_a \times P_b$.

Definition 2. Given a potato system $\mathcal{P} = (P_a, E_{a,b} : a, b \in A)$ over the finite τ-structure $\mathbf{B}$, the *structure associated to* $\mathcal{P}$ is the τ-structure $\mathbf{A}$ with universe A and basic relations defined as follows:

1) (For 1-ary $R \in \tau$): $R^{\mathbf{A}} := \{a \in A : P_a = R^{\mathbf{B}}\}$.
2) (For 2-ary $R \in \tau$): $R^{\mathbf{A}} := \{(a, b) \in A^2 : E_{a,b} = R^{\mathbf{B}}\}$.

A τ-structure is $\mathbf{B}$*-reduced* if it is the structure associated to some potato system over $\mathbf{B}$.

Lemma 1. *Suppose* $\mathbf{B}$ *is a finite binary* τ*-structure whose set of basic relations is closed under* $\wedge$*-atomic definitions. For every finite* τ*-structure* $\mathbf{A}$ *there exists a* $\mathbf{B}$*-reduced* τ*-structure* $\mathbf{A}^\circ$ *having the same domain as* $\mathbf{A}$ *and which satisfies the following: for all* $X \subseteq A$*,* $\mathrm{Hom}(\mathbf{A}{\restriction}_X, \mathbf{B}) = \mathrm{Hom}(\mathbf{A}^\circ{\restriction}_X, \mathbf{B})$.

B. Bounded Pathwidth Duality

In this section we present the facts we need about pathwidth duality. The following three definitions are from [14], [15].

Definition 3. Let $\mathbf{B}$ be a finite τ-structure. A set $\mathcal{O}$ of finite τ-structures is an *obstruction set for* $\mathrm{CSP}(\mathbf{B})$ if for all finite τ-structures $\mathbf{A}$, $\mathbf{A} \not\to \mathbf{B}$ if and only if there exists $\mathbf{C} \in \mathcal{O}$ with $\mathbf{C} \to \mathbf{A}$.

Obstruction sets are useful when their members are simple. One way they can be simple is by having bounded pathwidth.

Definition 4. A finite τ-structure $\mathbf{C}$ has *pathwidth at most* (j, k) if there is a sequence $\mathcal{I} = (I_0, \ldots, I_N)$ of subsets of C such that:

1) $|I_t| \leq k$ for all t, and $|I_t \cap I_{t+1}| \leq j$ for all $t < N$.
2) $I_i \cap I_j \subseteq I_\ell$ for all $0 \leq i \leq \ell \leq j \leq N$.
3) For every $R \in \tau$ of arity n and every $(c_1, \ldots, c_n) \in R^{\mathbf{C}}$ there exists $t \leq N$ with $\{c_1, \ldots, c_n\} \subseteq I_t$.

The sequence $\mathcal{I}$ is called a (j, k)*-path decomposition* of $\mathbf{C}$.

Definition 5. Let $\mathbf{B}$ be a finite τ-structure, let $\mathcal{O}$ be a set of finite τ-structures, and let $0 \leq j \leq k$.

1) $\mathcal{O}$ has *pathwidth at most* (j, k) if every $\mathbf{C} \in \mathcal{O}$ has pathwidth at most (j, k).

2) $\mathrm{CSP}(\mathbf{B})$ has *(j,k)-pathwidth duality* if $\mathrm{CSP}(\mathbf{B})$ has an obstruction set of pathwidth at most (j,k).
3) $\mathrm{CSP}(\mathbf{B})$ has *bounded pathwidth duality* if $\mathrm{CSP}(\mathbf{B})$ has (j',k')-pathwidth duality for some $0 \leq j' \leq k'$.

The characterization of bounded pathwidth duality that will be most useful to us in this paper is one involving the following variation of Dalmau's "pebble relation games" [15].

Definition 6. Suppose $\mathbf{A},\mathbf{B}$ are finite τ-structures.

1) A *solo play of the (j,k)-PR game on* $(\mathbf{A},\mathbf{B})$ is a finite sequence $\mathcal{I} = (I_0, I_1, \ldots, I_N)$ of subsets of A satisfying
   a) $|I_t| \leq k$ for all $t \leq N$.
   b) For all $t < N$, either $I_{t+1} \subseteq I_t$ or $I_t \subset I_{t+1}$. If the latter, then $|I_t| \leq j$.
2) Given a solo play $\mathcal{I} = (I_0, \ldots, I_N)$ of the (j,k)-PR game on $(\mathbf{A},\mathbf{B})$, the *resulting relations* $H_0, H_1, \ldots, H_N$ are defined recursively as follows:
   a) $H_0 = \mathrm{Hom}(\mathbf{A}{\restriction}_{I_0}, \mathbf{B})$.
   b) If $t < N$ and $I_{t+1} \subseteq I_t$, then $H_{t+1} = H_t{\restriction}_{I_{t+1}}$.
   c) If $t < N$ and $I_t \subset I_{t+1}$, then $H_{t+1} = \{h \in \mathrm{Hom}(\mathbf{A}{\restriction}_{I_{t+1}}, \mathbf{B}) : h{\restriction}_{I_t} \in H_t\}$.
3) We write $\mathbf{A} \looparrowright_{j,k} \mathbf{B}$ to mean that for every solo play $\mathcal{I} = (I_0, \ldots, I_N)$ of the (j,k)-PR game on $(\mathbf{A},\mathbf{B})$, the final resulting relation H_N is non-empty.

Solo plays and their resulting relations correspond to plays of Dalmau's two-player pebble relation game [15] where Spoiler chooses each set I_t and the resulting relation H_t is Duplicator's maximum allowable response. In particular, $\mathbf{A} \looparrowright_{j,k} \mathbf{B}$ if and only if Duplicator has a *strict* (in Dalmau's sense) winning strategy for the two-player pebble relation game played on $(\mathbf{A},\mathbf{B})$. Thus:

Proposition 2. ([15, Theorem 5 and Claim 1, p. 15]) *Let $\mathbf{B}$ be a finite τ-structure and $j \leq k$. The following are equivalent:*

1) *$\mathrm{CSP}(\mathbf{B})$ has (j,k)-pathwidth duality.*
2) *For all finite τ-structures $\mathbf{A}$, if $\mathbf{A} \looparrowright_{j,k} \mathbf{B}$ then $\mathbf{A} \to \mathbf{B}$.*

Combining Proposition 2 with Lemma 1 we get:

Corollary 3. *Suppose $\mathbf{B}$ is a binary τ-structure whose set of basic relations is closed under $\wedge$-atomic definitions. For any $0 \leq j \leq k$, the following are equivalent:*

1) *$\mathrm{CSP}(\mathbf{B})$ has (j,k)-pathwidth duality.*
2) *For all finite $\mathbf{B}$-reduced τ-structures $\mathbf{A}$, if $\mathbf{A} \looparrowright_{j,k} \mathbf{B}$ then $\mathbf{A} \to \mathbf{B}$.*

C. Algebra

In this section we summarize the algebraic background needed in this paper. More in-depth treatments may be found in [12], [8]. Everything preceding Definition 8 is standard.

Given a non-empty set A, an *operation* on A is any function $\phi : A^n \to A$ for some $n \geq 1$; n is the *arity* of ϕ. An operation ϕ is *idempotent* if it satisfies the equation $\phi(x, x, \ldots, x) = x$ for all $x \in A$. A 3-ary operation $\phi : A^3 \to A$ is a *majority operation on A* provided it is idempotent and satisfies the equations $\phi(y,x,x) = \phi(x,y,x) = \phi(x,x,y) = x$ for all $x, y \in A$. More generally, an n-ary operation $\phi : A^n \to A$ for $n \geq 3$ is a *near unanimity* (or *NU*) *operation on* A provided it is idempotent and for all $1 \leq i \leq n$ it satisfies

$$\phi(\underbrace{x,\ldots,x}_{i-1}, y, \underbrace{x,\ldots,x}_{n-i}) = x \qquad \text{for all } x,y \in A.$$

An *algebraic vocabulary* is any set (possibly infinite) of *operation symbols*, each of which has an assigned *arity* $n \geq 1$. If τ is an algebraic vocabulary, an *algebra of type* τ is an object $\mathbb{A}$ consisting of a non-empty set A (the universe) and, for each each operation symbol $\mathsf{f} \in \tau$ of arity n, an n-ary operation $\mathsf{f}^{\mathbb{A}}$ on A. The operations $\mathsf{f}^{\mathbb{A}}$ ($\mathsf{f} \in \tau$) are the *basic operations* of $\mathbb{A}$. An algebra is *finite* if its universe is finite, and is *idempotent* if each of its basic operations is idempotent.

Suppose $\mathbb{A}$ is an algebra of type τ and $X \subseteq A$. X is a *subuniverse* of $\mathbf{A}$ if X is closed under every basic operation of $\mathbf{A}$; that is, if $\mathsf{f}^{\mathbb{A}}(X^n) \subseteq X$ for all n-ary $\mathsf{f} \in \tau$. We denote this by $X \leq \mathbb{A}$; if in addition $X \subseteq_{sd} A$ then we write $X \leq_{sd} \mathbb{A}$. If $\varnothing \neq X \leq \mathbb{A}$, the *subalgebra of $\mathbb{A}$ with universe X* is the algebra $\mathbb{X}$ of type τ whose operations are given by $\mathsf{f}^{\mathbb{X}} = \mathsf{f}^{\mathbb{A}}{\restriction}_X$.

For every $X \subseteq A$ there is a unique smallest subuniverse of $\mathbb{A}$ containing X, which is denoted $\mathrm{Sg}^{\mathbb{A}}(X)$.

We use the following device: any set $\mathcal{F}$ of operations on A can be considered as an algebraic vocabulary in the obvious way, making $(A, \mathcal{F})$ an algebra of type $\mathcal{F}$.

Two algebras are *similar* if they have the same vocabulary. The *product* of any number of similar algebras is defined naturally, that is, by defining operations coordinatewise. We sometimes use the same notation, e.g. ϕ, for both an operation symbol and its interpretations in similar algebras $\mathbb{A}, \mathbb{B}$, etc.

Definition 7. Given a structure $\mathbf{A}$ with universe A, an n-ary relation R on A, and an m-ary operation ϕ on A, we say that:

1) *ϕ preserves R* provided for all $\mathbf{a}_1, \ldots, \mathbf{a}_m \in R$, if $\mathbf{c}_1, \ldots, \mathbf{c}_n \in A^m$ are the columns of the matrix whose rows are $\mathbf{a}_1, \ldots, \mathbf{a}_m$, then $(\phi(\mathbf{c}_1), \ldots, \phi(\mathbf{c}_n)) \in R$.
2) ϕ is a *polymorphism* of $\mathbf{A}$ if ϕ preserves every basic relation of $\mathbf{A}$.

Let $\mathrm{Pol}(\mathbf{A})$ denote the set of all polymorphisms of $\mathbf{A}$. The *polymorphism algebra* of $\mathbf{A}$ is the algebra $\mathbb{A} = (A, \mathrm{Pol}(\mathbf{A}))$ (of type $\mathrm{Pol}(\mathbf{A})$). It is a well-known fact (e.g., [19], [9], [24]) that if R is an arbitrary nonempty n-ary relation on A, then R is pp-definable over $\mathbf{A}$ if and only if $R \leq \mathbb{A}^n$.

The next definition slightly extends a notion from [7], [6].

Definition 8. Suppose $\mathbb{A}$ is an algebra, $B, C \leq \mathbb{A}$, and ϕ is an n-ary operation of $\mathbb{A}$. We say that *B absorbs C with respect to ϕ* provided the following condition holds: for all $1 \leq i \leq n$, all $b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n \in B$ and all $c \in C$, $\phi(b_1, \ldots, b_{i-1}, c, b_{i+1}, \ldots, b_n) \in B$. We write $B \lhd_\phi C$ to mean $B \subseteq C$ and B absorbs C with respect to ϕ.

Here are two easy facts about absorption.

Lemma 4. *Suppose ϕ is an operation of the algebra $\mathbb{A}$.*

1) *If $B \lhd_\phi C \leq \mathbb{A}$ then $B \cap D \lhd_\phi C \cap D$ for any $D \leq \mathbb{A}$.*

2) *If $\mathbb{A}$ is idempotent, then ϕ is an NU operation if and only if $\{a\} \lhd_\phi A$ for all $a \in A$.*

The following claim is a good exercise, or can be extracted from [7, Lemma 2.5].

Lemma 5. *Suppose $\mathbb{B}_1, \mathbb{C}_1$ are similar algebras, ϕ is an operation symbol in their common vocabulary, $S \leq \mathbb{B}_1 \times \mathbb{C}_1$ with $\mathrm{proj}_1(S) = B_1$, and $C_0 \lhd_\phi C_1$. Define $B_0 = \{b \in B_1 : \exists c \in C_0 \text{ with } (b,c) \in S\}$. Then $B_0 \lhd_\phi B_1$.*

We need one new result about absorption, whose proof will be postponed until subsection III-D. Given an algebra $\mathbb{D}$ and an integer $n \geq 2$, define $0_D^{(n)} = \{(b, b, \ldots, b) : b \in D\} \subseteq D^n$, the set of constant n-tuples over D. Note that $0_D^{(n)} \leq \mathbb{D}^n$.

Definition 9. Let $\mathbb{D}$ be an algebra, ϕ an operation of $\mathbb{D}$, and $b \in D$. We call b an *absorption constant for $\mathbb{D}$ with respect to ϕ* provided, for all $n \geq 2$ and every $R \leq_{sd} \mathbb{D}^n$, if R absorbs $0_D^{(n)}$ with respect to ϕ then $(b, b, \ldots, b) \in R$.

Theorem 6. *Let $\mathbb{D}$ be a finite algebra and ϕ an idempotent operation of $\mathbb{D}$. There exists an absorption constant for $\mathbb{D}$ with respect to ϕ.*

III. Main Result

A. Statement and Reduction to the Binary Case

The main result of this paper is the following.

Theorem 7. *Suppose the finite τ-structure $\mathbf{B}$ has a $(d+1)$-ary NU polymorphism for some $d \geq 2$. Then $\mathrm{CSP}(\mathbf{B})$ has bounded pathwidth duality and hence is in NL.*

The rest of the paper is devoted to proving this theorem. In this subsection we will reduce it to the case of binary structures. The first reduction is a variant of one step in the proof of the well-known "CSP reduction to digraphs" [18, Theorem 11] (see also e.g. the proof of [5, Theorem 4.4]). It applies to any structure.

Lemma 8. *Suppose τ is a vocabulary, $n \geq 1$ is an integer such that every relation symbol in τ has arity at most $2n$, $\mathbf{B}$ is a finite τ-structure, and $\mathbb{B}$ is its polymorphism algebra. There exists a binary structure $\mathbf{B}^{(n)}$ with universe B^n such that:*

1) *The polymorphism algebra of $\mathbf{B}^{(n)}$ is $\mathbb{B}^n$.*
2) *For any $0 \leq j \leq k$, if $\mathrm{CSP}(\mathbf{B}^{(n)})$ has (j,k)-pathwidth duality, then $\mathrm{CSP}(\mathbf{B})$ has (jn, kn)-pathwidth duality.*

The next reduction applies only to structures with an NU polymorphism. It is the obvious and straightforward generalization of [16, Lemma 2] for majority polymorphisms.

Lemma 9. *Suppose $\mathbf{B}$ is a finite τ-structure with universe B and a $(d+1)$-ary NU polymorphism for some $d \geq 2$. Let $s = \max(\{arity(R) : R \in \tau\} \cup \{d\})$. There exists a vocabulary τ_d and a τ_d-structure $\mathbf{B}_d$ with universe B satisfying:*

1) *Every relation symbol in τ_d has arity at most d.*
2) *$\mathbf{B}$ and $\mathbf{B}_d$ have the same polymorphisms.*
3) *If $\mathrm{CSP}(\mathbf{B}_d)$ has (j,k)-pathwidth duality, then $\mathrm{CSP}(\mathbf{B})$ has $(k, k+s-d)$-pathwidth duality.*

Lemmas 8 and 9 reduce the task of proving Theorem 7 to proving it for the special case of binary structures. In subsections III-C and III-D we will verify this special case by proving the following.

Proposition 10. *Suppose $d \geq 2$ and $\mathbf{B}$ is a binary structure having a $(d+1)$-ary NU polymorphism. Let $k = |B|$, $c = \lfloor \log_3(2d-3) \rfloor + 2$, and $p = 2c^k - k - 1$. Then $\mathrm{CSP}(\mathbf{B})$ has $(p, p+1)$-pathwidth duality. If $d = 2$ then $\mathrm{CSP}(\mathbf{B})$ has $(2k, 2k+1)$-pathwidth duality.*

Hence we get the following sharpening of Theorem 7.

Corollary 11. *Suppose the finite τ-structure $\mathbf{B}$ has a $(d+1)$-ary NU polymorphism for some $d \geq 2$. Let k, p be defined as in Proposition 10, let $q = \lceil d/2 \rceil (p+1)$, and let s be defined as in Lemma 9. Then $\mathrm{CSP}(\mathbf{B})$ has $(q, q+s-d)$-pathwidth duality.*

Proof of Theorem 7 and Corollary 11: Given $\mathbf{B}$, Let $\mathbf{B}_d$ be the structure defined in Lemma 9, let $e = \lceil d/2 \rceil$, and let $(\mathbf{B}_d)^{(e)}$ be the structure obtained from $\mathbf{B}_d$ via Lemma 8. Both $\mathbf{B}_d$ and $(\mathbf{B}_d)^{(e)}$ inherit $(d+1)$-ary NU polymorphisms from $\mathbf{B}$. As $(\mathbf{B}_d)^{(e)}$ is binary, $(\mathbf{B}_d)^{(e)}$ has $(p, p+1)$-pathwidth duality by Proposition 10; hence $\mathbf{B}$ has $(q, q+s-d)$-pathwidth duality by Lemmas 8 and 9. ■

B. A-Trees

Proposition 10 will be proved via an intricate analysis of realizations of certain trees in τ-structures. In this subsection we define these trees and state some facts about them that will be needed in subsection III-C.

Following [15], [16], if $G = (V, E)$ is an undirected graph, we denote by $\mathrm{pw}(G)$ the least k for which G has pathwidth (j, k) for some $j \leq k$, and call $\mathrm{pw}(G)$ the *pathwidth* of G. (Note that $\mathrm{pw}(G)$ is 1 greater than the usual graph-theoretic measure of the pathwidth of G as defined in [32].)

Definition 10. Let A be a non-empty set. An *A-tree* is a pair (T, χ) where $T = (V, E)$ is a tree (i.e., a connected undirected graph with no cycles) and χ is a coloring of the vertices of T by elements of A (i.e., $\chi : V \to A$).

Definition 11. Let $T_0, \ldots, T_n$ be trees on disjoint vertex sets. A *tree composition* of $T_0, \ldots, T_n$ is any tree T that can be constructed from the union of $T_0, \ldots, T_n$ by identifying some vertices among the leaves of $T_0, \ldots, T_n$ (enough to connect the graph, but not so many as to introduce cycles).

If T is a tree composition of $T_0, \ldots, T_n$, then the vertices of T which were formed by identifying leaves of the original trees are called the *composition vertices* of T. The subtrees of T corresponding to the original trees $T_0, \ldots, T_n$ are called the *components* of T. Among the components of T we distinguish the *leaf components*, which are those which had at most one leaf identified in the construction of T. (Note that, strictly speaking, the notions in this paragraph are not invariants of T itself, but of a particular composition that produces T. Whenever discussing tree compositions we shall assume that such trees come equipped with a record of a particular composition that produced it.)

Definition 12. Suppose $T = (V, E)$ is a tree composition and $L \subseteq A$. An A-coloring $\chi : V \to A$ is said to *lie over L* if χ sends all composition vertices of T into L.

Definition 13. Suppose $d \geq 2$.

1) $\mathcal{T}_d^0$ is the set of all trees consisting of exactly one edge.
2) For $i \geq 1$, $\mathcal{T}_d^{i+1}$ is the set of all tree compositions having components from $\mathcal{T}_d^i$ and at most d leaf components.

Note that $\mathcal{T}_d^0 \subseteq \mathcal{T}_d^1 \subseteq \mathcal{T}_d^2 \subseteq \cdots$. Note also that for $d = 2$ we have $\mathcal{T}_2^1 = \mathcal{T}_2^2 = \cdots = \{\text{all paths}\}$; hence every tree in $\bigcup_i \mathcal{T}_2^i$ has pathwidth at most 2. For $d > 2$ we have the following:

Lemma 12. *Suppose $d \geq 3$ and $i \geq 1$. The trees in $\mathcal{T}_d^i$ have bounded pathwidth. More precisely, every tree $T \in \mathcal{T}_d^i$ satisfies $\mathrm{pw}(T) \leq c^i + c^{i-1} - 1$ where $c = \lfloor \log_3(2d-3) \rfloor + 2$.*

Definition 14. Suppose A is a non-empty set, $L \subseteq A$, $d \geq 2$, and $i \geq 0$. $\mathcal{T}_d^i(A)$ denotes the set of all A-trees (T, χ) where $T \in \mathcal{T}_d^i$. If $i > 0$ we also define $\mathcal{T}_d^i(A, L) = \{(T, \chi) \in \mathcal{T}_d^i(A) : \chi \text{ lies over } L\}$.

Note that for $i > 0$ we have $\mathcal{T}_d^i(A, A) = \mathcal{T}_d^i(A)$, and $L \subseteq L'$ implies $\mathcal{T}_d^i(A, L) \subseteq \mathcal{T}_d^i(A, L')$. Also note that $\mathcal{T}_d^i(A) = \mathcal{T}_d^{i+1}(A, \varnothing)$ for all i.

We will use A-trees in the following way. Suppose $\mathbf{B}$ is a binary τ-structure whose set of basic relations is closed under $\wedge$-atomic definitions. Let $\mathcal{P} = (P_a, E_{a,b} : a, b \in A)$ be a potato system over $\mathbf{B}$ indexed by A. Given an A-tree (T, χ) with $T = (V, E)$, a *realization of (T, χ) in $\mathcal{P}$* is a function $f : V \to B$ satisfying the following:

1) $f(u) \in P_{\chi(u)}$ for all $u \in V$.
2) $(f(u), f(v)) \in E_{\chi(u),\chi(v)}$ for all $\{u, v\} \in E$.

The proof of the next lemma is straightforward.

Lemma 13. *Suppose $\mathbf{B}$ and $\mathcal{P}$ are as above, (T, χ) is an A-tree with $T = (V, E)$, $U = \{u_1, \ldots, u_n\}$ is a non-empty subset of V, and for each $v \in V$ we have an associated subset B_v of B which is pp-definable over $\mathbf{B}$. Then the following relations are pp-definable over $\mathbf{B}$:*

1) $R := \{(f(u_1), \ldots, f(u_n)) \in B^n : f$ *is a realization of* (T, χ) *in* $\mathcal{P}$ *and* $f(v) \in B_v$ *for all* $v \in V\}$.
2) $S := \{b \in B : (b, \ldots, b) \in R\}$.

C. Proof of Proposition 10

In this subsection we prove Proposition 10, modulo the proof of Theorem 6. Much of the argument in this subsection is inspired by the proof of [16, Theorem 1].

Let $\mathbf{B}$ be a binary τ-structure having a $(d+1)$-ary NU polymorphism ϕ. Because adding (arbitrary) relations to the vocabulary of $\mathbf{B}$ cannot decrease the pathwidth of obstruction sets for its constraint satisfaction problem, we may assume that the set of basic relations of $\mathbf{B}$ contains every ϕ-invariant unary and binary relation on B. In particular, the set of basic relations of $\mathbf{B}$ contains all 1-element subsets of B and is closed under $\wedge$-atomic definitions. Let k, c, p be defined as in the statement of Proposition 10, and let $r = p$ if $d \geq 3$ while $r = 2k$ if $d = 2$. Let $\mathbf{A}$ be a $\mathbf{B}$-reduced τ-structure such that $\mathbf{A} \looparrowright_{r,r+1} \mathbf{B}$. By Corollary 3, to prove Proposition 10 it suffices to show that $\mathbf{A} \to \mathbf{B}$.

Let $\mathcal{P} = (P_a, E_{a,b} : a, b \in A)$ be the potato-system to which $\mathbf{A}$ is associated. For each $a \in A$, define a sequence of "levels" $P_a^0 \supseteq P_a^1 \supseteq P_a^2 \supseteq \cdots$ within P_a as follows:

1) $P_a^0 = P_a$.
2) If $i \geq 0$ and $(T, \chi) \in \mathcal{T}_d^{i+1}(A)$ with $T = (V, E)$, then $P_a^{i+1}(T, \chi)$ is the set of $b \in P_a^i$ for which there exists a realization f of (T, χ) in $\mathcal{P}$ such that
   a) f maps V into level i; i.e., $f(u) \in P_{\chi(u)}^i$ for all $u \in V$.
   b) f maps each a-labelled vertex to b; i.e., $f(u) = b$ for all vertices $u \in V$ such that $\chi(u) = a$.
3) $P_a^{i+1} = \bigcap \{P_a^{i+1}(T, \chi) : (T, \chi) \in \mathcal{T}_d^{i+1}(A)\}$.

Lemma 13 implies that each set P_a^i is pp-definable over $\mathbf{B}$.

Claim 1. *$P_a^k \neq \varnothing$ for all $a \in A$.*

Proof sketch: For $i \geq 0$ define $g(i) = 2i$ if $d = 2$ and $g(i) = (c+1)(c-1)^{-1}(c^i - 1) - i$ otherwise. We claim that for all $i \geq 0$ and all $a \in A$, there exists a solo play $\mathcal{I}(i, a) = (I_0, I_1, \ldots, I_n)$ of the $(g(i), g(i)+1)$-PR game on $(\mathbf{A}, \mathbf{B})$ such that $a \in I_j$ for all $j \leq n$, $I_n = \{a\}$, and the final resulting relation of $\mathcal{I}(i, a)$ is contained in P_a^i. This will suffice, since $\mathbf{A} \looparrowright_{r,r+1} \mathbf{B}$ and $g(k) \leq r$ imply that the final resulting relation of each $\mathcal{I}(k, a)$ is non-empty.

The proof of the claim is by induction on i. If $i = 0$, then we can choose $\mathcal{I}(0, a) = (I_0)$ where $I_0 = \{a\}$. Now assume that $i \geq 0$ and the claim has been verified for i. For each $a' \in A$ inductively choose and fix a solo play $\mathcal{I}(a')$ of the $(g(i), g(i)+1)$-PR game each of whose sets contains a', whose final set is $\{a'\}$, and whose final resulting relation is contained in $P_{a'}^i$. Fix $a \in A$.

Call a solo play $\mathcal{I} = (I_0, \ldots, I_n)$ of the $(g(i+1), g(i+1)+1)$-PR game on $(\mathbf{A}, \mathbf{B})$ an *a-play* if $a \in I_j$ for all $j \leq n$ and $I_n = \{a\}$. If $\mathcal{I}, \mathcal{J}$ are a-plays with final resulting relations R, S respectively, then the concatenation of $\mathcal{I}$ and $\mathcal{J}$ is also an a-play and its final resulting relation is contained in $R \cap S$. Thus it suffices to show that for every $(T, \chi) \in \mathcal{T}_d^{i+1}(A)$ there exists an a-play $\mathcal{J}_a^{T,\chi}$ whose final resulting relation is contained in $P_a^{i+1}(T, \chi)$. Fix $(T, \chi) \in \mathcal{T}_d^{i+1}(A)$ with $T = (V, E)$.

1) Using Lemma 12 and the comment preceding it, let $(J_0, \ldots, J_m)$ be a $(t, t+1)$-path decomposition of T where $t = 1$ if $d = 2$ and $t = c^{i+1} + c^i - 2$ otherwise.
2) Let $\{a_1^0, \ldots, a_{k_0}^0\}$ be an enumeration of $\chi(J_0)$. For each $1 \leq j \leq m$, let $\{a_1^j, \ldots, a_{k_j}^j\}$ be an enumeration of $\chi(J_j) \setminus \chi(J_{j-1})$.
3) For each $0 \leq j \leq m$ and $1 \leq \ell \leq k_j$, let $\mathcal{I}_a^*(j, \ell)$ be the play defined as follows: if $\mathcal{I}(a_\ell^j) = (I_0, I_1, \ldots, I_t)$ then $\mathcal{I}_a^*(j, \ell) = (I_0^*, I_1^*, \ldots, I_t^*)$ where $I_u^* = I_u \cup \chi(J_j) \cup \{a\}$. (Note in particular that the final set is $I_t^* = \chi(J_j) \cup \{a\}$.) Also define $\mathcal{I}_a^*(j, 0)$ to be the 1-step play $(\chi(J_j) \cup \{a\})$.
4) For each $0 \leq j \leq m$ let $\mathcal{I}_a^*(j)$ be the concatenation of the plays $\mathcal{I}_a^*(j, 0), \mathcal{I}_a^*(j, 1), \ldots, \mathcal{I}_a^*(j, k_j)$.

5) Let $\mathcal{J}_a^{T,\chi}$ be the concatenation of $\mathcal{J}(a), \mathcal{J}_a^*(0), \ldots, \mathcal{J}_a^*(m)$, and the 1-step play $(\{a\})$.

Suppose $0 \leq j \leq m$ and $0 \leq \ell \leq k_j$. Let $\mathcal{J}^\star$ denote the initial segment of $\mathcal{J}_a^{T,\chi}$ consisting of $\mathcal{J}(a)$, $\mathcal{J}_a^{T,\chi}(0), \ldots, \mathcal{J}_a^{T,\chi}(j-1)$ and that portion of $\mathcal{J}_a^{T,\chi}(j)$ up to and including $\mathcal{J}_a^*(j,\ell)$. Note that the last set of $\mathcal{J}^*$ is $\chi(J_j) \cup \{a\}$. Let $V^* = J_0 \cup \cdots \cup J_j$, let $C = \{a_1^j, \ldots, a_\ell^j\}$, and let T^* be the subtree of T with vertex set V^*. It can be shown inductively that if h is in the final resulting relation of $\mathcal{J}^*$ then $h(a) \in P_a^i$ and there exists a realization f of the A-tree $(T^*, \chi{\restriction}_{T^*})$ such that

1) f maps $J_0 \cup \cdots \cup J_{j-1}$ into level i;
2) f maps $J_j \cap \chi^{-1}(\{a_1^j, \ldots, a_\ell^j\})$ into level i;
3) $f(u) = h(\chi(u))$ for all $u \in J_j$.
4) f maps all a-labelled vertices of T^* to $h(a)$.

In particular, if h is in the penultimate resulting relation of $\mathcal{J}_a^{T,\chi}$ then $h(a) \in P_a^{i+1}(T,\chi)$; hence the final resulting relation of $\mathcal{J}_a^{T,\chi}$ is contained in $P_a^{i+1}(T,\chi)$.

It remains to check that every set in $\mathcal{J}_a^{T,\chi}$ has size at most $g(i+1)+1$. Note that any set in $\mathcal{J}_a^{T,\chi}$ of maximal size is of the form $I_u \cup \chi(J_j) \cup \{a\}$ where I_u is a set in $\mathcal{J}(a')$ for some $a' \in \chi(J_j)$. We have $|J_j| \leq t+1$ and $|I_u| \leq g(i)+1$ and $|I_u \cap \chi(J_j)| \geq 1$. Thus $|I_u \cup \chi(J_j) \cup \{a\}| \leq g(i)+t+2 = g(i+1)+1$ as required. ∎

Thus for each $a \in A$ we have a chain $P_a^0 \supseteq P_a^1 \supseteq \cdots \supseteq P_a^k$ of $k+1$ non-empty subsets of B, where $|B| = k$. Hence there exists $r < k$ such that $P_a^r = P_a^{r+1}$. Let r_a be the least r with this property. Enumerate A as $\{a_0, a_1, \ldots, a_N\}$ so that $r_{a_0} \geq r_{a_1} \geq \cdots$ and for each $j \leq N$ define $\mathrm{rank}(j) = r_{a_j}$.

Our aim is to construct a homomorphism $h : \mathbf{A} \to \mathbf{B}$, and we will do this by inductively defining $h(a_0), h(a_1)$, etc. At stage i we will have defined $h(a_0), h(a_1), \ldots, h(a_i)$. In this context we will consider certain A-trees and their realizations in $\mathcal{P}$. Suppose (T,χ) is an A-tree and f is a realization of (T,χ) in $\mathcal{P}$. If Λ_T is the set of leaves of T and $U \subseteq \Lambda_T$, then we say that f is *fixed on U up to i* if for all $u \in U$, if $\chi(u) \in \{a_0, \ldots, a_i\}$ then $f(u) = h(\chi(u))$. We say that f is *fixed up to i* if it is fixed on Λ_T up to i.

The inductive property that we will establish at stage i is the following:

1) $h(a_j) \in P_{a_j}^{\mathrm{rank}(j)}$ for all $j \leq i$.
2) Let $r = \mathrm{rank}(i)$ and $L = \{a_j : j \leq N, \mathrm{rank}(j) = r\}$. Then for every A-tree $(T,\chi) \in \mathcal{T}_d^{r+1}(A,L)$ with $T = (V,E)$ there exists a realization of (T,χ) in $\mathcal{P}$ which sends V into level r and is fixed up to i.

This will suffice since at stage N we will have fully defined a function $h : A \to B$ satisfying property 1) above; moreover, as $\mathcal{T}_d^{r+1}(A,L) \supseteq \mathcal{T}_d^0(A)$, property 2) above can be applied as follows: for any $a_i, a_j \in A$, if we let (T,χ) be the 2-vertex A-tree $\{u,v\}$ with $\chi(u) = a_i$ and $\chi(v) = a_j$, and if f is a realization of (T,χ) in $\mathcal{P}$ which is fixed up to N, then $(h(a_i), h(a_j)) = (f(u), f(v)) \in E_{\chi(u),\chi(v)} = E_{a_i,a_j}$, proving h is a homomorphism from $\mathbf{A}$ to $\mathbf{B}$.

At stage 0 we can define $h(a_0)$ to be any element of $P_{a_0}^r$ where $r = \mathrm{rank}(0)$. This works as $P_{a_0}^r = P_{a_0}^{r+1}$, so the definition of $P_{a_0}^{r+1}$ and the fact that $\mathcal{T}_d^{r+1}(A,L) \subseteq \mathcal{T}_d^{r+1}(A)$ imply property 2) above.

Assume that we have finished stage $i-1$ and want to define $h(a_i)$. Define $r^* = \mathrm{rank}(i-1)$, $L^* = \{a_j : \mathrm{rank}(j) = r^*\}$, $r = \mathrm{rank}(i)$, and $L = \{a_j : \mathrm{rank}(j) = r\}$.

Claim 2. *For every A-tree $(T,\chi) \in \mathcal{T}_d^{r+1}(A,L)$ with $T = (V,E)$ there exists a realization of (T,χ) in $\mathcal{P}$ which maps V into level r and is fixed up to $i-1$.*

Proof: If $r^* = r$, then $L^* = L$ and the claim follows directly from the inductive property at stage $i-1$. If on the other hand $r^* > r$, then use the fact that $\mathcal{T}_d^{r+1}(A,L) \subseteq \mathcal{T}_d^{r^*+1}(A,\varnothing) \subseteq \mathcal{T}_d^{r^*+1}(A,L^*)$ and the inductive property. ∎

Definition 15. Let $\mathcal{T}^*$ denote the set of all A-trees (T,χ) where T is an (arbitrary) tree composition of component trees from $\mathcal{T}_d^r$ and χ lies over L. (Thus $\mathcal{T}_d^{r+1}(A,L) \subseteq \mathcal{T}^*$.)

Claim 3. *Suppose $(T,\chi) \in \mathcal{T}^*$ with $T = (V,E)$.*

1) *For every leaf u of T such that $\chi(u) = a \in L$, and for every $b \in P_a^r$, there exists a realization f of (T,χ) in $\mathcal{P}$ which maps V into level r and satisfies $f(u) = b$.*
2) *There exists a realization of (T,χ) in $\mathcal{P}$ which maps V into level r and is fixed up to $i-1$.*

Proof: (1) Let $T_0 = (V_0, E_0)$ be the component of T containing u, let $\chi_0 = \chi{\restriction}_{V_0}$, and note that $(T_0, \chi_0) \in \mathcal{T}_d^{r+1}(A)$. Because $a \in L$ we have $P_a^r = P_a^{r+1}$. Thus if $b \in P_a^r$ then by definition of P_a^{r+1} there exists a realization f_0 of (T_0, χ_0) in $\mathcal{P}$ which sends V_0 into level r and maps u to b. Consider a composition vertex u' of T which lies in V_0 and let $a' = \chi(u')$ and $b' = f(u')$. We have $a' \in L$ (as χ lies over L) and $b' \in P_{a'}^r$ by our choice of f; hence $b' \in P_{a'}^{r+1}$. If we now let $T_1 = (V_1, E_1)$ be another component of T containing u' and put $\chi_1 = \chi{\restriction}_{V_1}$, then we can repeat the previous argument to get a realization f_1 of (T_1, χ_1) in $\mathcal{P}$ which sends V_1 into level r and maps u' to b'; in particular, $f_1(u') = f_0(u')$. Repeating this process one component of T at a time, we can eventually construct the required realization f of (T,χ).

(2) Let Λ_T denote the set of leaves of T. It will suffice to prove that, for every $U \subseteq \Lambda_T$, there exists a realization of (T,χ) in $\mathcal{P}$ which maps V into level r and is fixed on U up to $i-1$. We do this by induction on $|U|$. Assume first that $|U| \leq d$. Let $T' = (V', E')$ be the smallest subtree of T which is a tree composition of (some of) the components of T and contains U, and let $\chi' = \chi{\restriction}_{T'}$. Also let $C_1, \ldots, C_m$ be the connected components of $T \setminus T'$. Observe that for each C_j there exists a composition vertex u_j of T such that

1) u_j is either a composition vertex or a leaf of T'.
2) The subtree of T with vertex set $C_j \cup \{u_j\}$ is a tree composition of (some of the) components of T which are not components of T'. Call this subtree T_j and let $\chi_j = \chi{\restriction}_{T_j}$. Note that $(T_j, \chi_j) \in \mathcal{T}^*$.

T' has at most d leaf components and so $(T', \chi') \in \mathcal{T}_d^{r+1}(A,L)$. By Claim 2 there exists a realization f' of $(T', \chi{\restriction}_{T'})$ which maps V' into level r and is fixed up to $i-1$; in particular, f' is fixed on U up to $i-1$. By applying part

(1) to the A-trees (T_j, χ_j) we can extend f' to the desired realization f of (T, χ).

Assume next that $|U| > d$. Choose distinct leaves $u_0, u_1, \ldots, u_d \in U$ and for each $j \leq d$ let $U_j = U \setminus \{u_j\}$. By induction, there exist realizations f_j of (T, χ) in $\mathcal{P}$ which map V into level r and are such that f_j is fixed on U_j up to $i-1$. Let $f : V \to B$ be defined by

$$f(v) = \phi(f_0(v), f_1(v), \ldots, f_d(v)).$$

Because ϕ is a polymorphism of $\mathbf{B}$, f is a realization of (T, χ) which sends V into level r. It remains to check that f is fixed on U up to $i-1$; this follows from the NU property of ϕ. ■

Claim 4. *There exists a non-empty set $D \subseteq P^r_{a_i}$ such that (i) D is pp-definable over $\mathbf{B}$, and (ii) for every $(T, \chi) \in \mathcal{T}^*$ with $T = (V, E)$, if Δ is the set of all leaves u of T with $\chi(u) = a_i$, then for every $u \in \Delta$ and every $b \in D$ there exists a realization f of (T, χ) in $\mathcal{P}$ satisfying:*

1) *f sends V into level r and is fixed up to $i-1$.*
2) *$f(v) \in D$ for every $v \in \Delta$.*
3) *$f(u) = b$.*

Proof: Suppose no such set D exists. Let $D_0 = P^r_{a_i}$. As D_0 is a non-empty, is pp-definable over $\mathbf{B}$, but does not satisfy the statement of the Claim, there must exist an A-tree $(T_0, \chi_0) \in \mathcal{T}^*$ with $T_0 = (V_0, E_0)$, whose set of leaves u with $\chi_0(u) = a_i$ is Δ_0, and there must exist $u_0 \in \Delta_0$ and $b_0 \in D_0$, such that for all realizations f of (T_0, χ_0) in $\mathcal{P}$, if f sends V_0 into level r and is fixed up to $i-1$, then $f(u_0) \neq b_0$. Define

$$D_1 = \{f(u_0) : f \text{ is realization of } (T_0, \chi_0) \text{ in } \mathcal{P}, \text{ sends } V_0 \text{ into level } r, \text{ and is fixed up to } i-1\}.$$

D_1 is non-empty by Claim 3(2), is pp-definable over $\mathbf{B}$ by Lemma 13, satisfies $D_1 \subseteq D_0$ because f maps V_0 into level r, and satisfies $D_1 \neq D_0$ because $b_0 \in D_0 \setminus D_1$.

Again as D_1 does not satisfy the statement of the Claim, there must exist an A-tree $(T_1, \chi_1) \in \mathcal{T}^*$ with $T_1 = (V_1, E_1)$, whose set of leaves u with $\chi_1(u) = a_i$ is Δ_1, and there must exist $u_1 \in \Delta_1$ and $b_1 \in D_1$, such that for all realizations f of (T_1, χ_1) in $\mathcal{P}$, if f sends V_1 into level r, sends Δ_1 into D_1, and is fixed up to $i-1$, then $f(u_1) \neq b_1$. Let $(T^\circ_1, \chi^\circ_1)$ be the A-tree with $T^\circ_1 = (V^\circ_1, E^\circ_1)$ obtained by

1) starting with (T_1, χ_1);
2) gluing to every leaf $u \in \Delta_1$ a copy (T^u_0, χ^u_0) of (T_0, χ_0) at the vertex u_0; that is, u and the image of u_0 in T^u_0 are identified;
3) adding a new leaf u°_1 with an edge to u_1, and defining $\chi^\circ_1(u^\circ_1) = a_i$.

Note that (i) $(T^\circ_1, \chi^\circ_1) \in \mathcal{T}^*$, and (ii) if f is a realization of $(T^\circ_1, \chi^\circ_1)$ in $\mathcal{P}$ which sends V°_1 into level r and is fixed up to $i-1$, then $f\restriction_{T_1}$ is a realization of (T_1, χ_1) which sends V_1 into level r, is fixed up to $i-1$, and sends Δ_1 into D_1; hence $f(u_1) \neq b_1$. As $(f(u^\circ_1), f(u_1)) \in E_{a_i,a_i} = \{(x, x) : x \in P_{a_i}\}$, this implies $f(u^\circ_1) = f(u_1) \neq b_1$. Thus if we define

$$D_2 = \{f(u^\circ_1) : f \text{ is realization of } (T^\circ_1, \chi^\circ_1) \text{ in } \mathcal{P}, \text{ sends } V^\circ_1 \text{ into level } r, \text{ and is fixed up to } i-1\},$$

then again D_2 is non-empty, is pp-definable over $\mathbf{B}$, and is a proper subset of D_1. By repeating this process we get an infinite strictly decreasing sequence $D_0 \supsetneq D_1 \supsetneq D_2 \supsetneq \cdots$ which is impossible. ■

We are now ready to define $h(a_i)$. Let $\mathbb{B}$ be the algebra (B, ϕ) and let $\mathbb{D}$ be its subalgebra $(D, \phi^{\mathbb{D}})$ where D is given by Claim 4. Choose any absorption constant β for $\mathbb{D}$ with respect to $\phi^{\mathbb{D}}$ (at least one exists by Theorem 6) and define $h(a_i) = \beta$. We claim that this choice achieves stage i. What needs to be shown is that for every A-tree $(T, \chi) \in \mathcal{T}^{r+1}_d(A, L)$ with $T = (V, E)$ there exists a realization of (T, χ) in $\mathcal{P}$ which:

$$(\dagger) \quad \begin{cases} 1. & \text{sends } V \text{ into level } r, \text{ and} \\ 2. & \text{is fixed up to } i; \text{ that is, is fixed up to } i-1 \\ & \quad \text{and sends every } a_i\text{-labelled leaf to } \beta. \end{cases}$$

Let (T, χ) be such an A-tree. Let $\Delta = \{u_1, \ldots, u_n\}$ be an enumeration of the set of leaves u of T satisfying $\chi(u) = a_i$, and let $\Gamma = \{v_1, \ldots, v_m\}$ be an enumeration of the set of leaves u of T satisfying $\chi(u) \in \{a_0, a_1, \ldots, a_{i-1}\}$. Define two subsets of B^{n+m}:

$$S^+ = \{(f(u_1), \ldots, f(u_n), f(v_1, \ldots, f(v_m)) : f \text{ is a realization of } (T, \chi) \text{ in } \mathcal{P} \text{ sending } V \text{ into level } r\},$$

$$R^+ = \{(f(u_1), \ldots, f(u_n), f(v_1, \ldots, f(v_m)) : f \text{ is a realization of } (T, \chi) \text{ in } \mathcal{P} \text{ sending } V \text{ into level } r \text{ and which is fixed up to } i-1\}.$$

Let B_0 and B_1 be the projections onto the first n coordinates of R^+ and S^+ respectively, and let $R = B_0 \cap D^n$ and $S = B_1 \cap D^n$. R^+ and S^+ are pp-definable over $\mathbf{B}$ by Lemma 13 and so are subuniverses of $\mathbb{B}^{n+m}$; we consider them as subuniverses of $\mathbb{B}^n \times \mathbb{B}^m$. Define $\mathbb{C}_1 = \mathbb{B}^m$ and $C_0 = \{\mathbf{c}\} \subseteq C_1$ where $\mathbf{c} = (h(\chi(v_1)), \ldots, h(\chi(v_m)))$; then $S^+ \leq \mathbb{B}_1 \times \mathbb{C}_1$ and $B_0 = \{\mathbf{x} \in B_1 : \exists \mathbf{y} \in C_0 \text{ with } (\mathbf{x}, \mathbf{y}) \in S^+\}$. Note that $C_0 \lhd_\phi C_1$ by Lemma 4(2) (because $\mathbb{C}_1$ is idempotent and $\phi^{\mathbb{C}_1}$ is NU); hence $B_0 \lhd_\phi B_1$ by Lemma 5 and so $R \lhd_\phi S$ by Lemma 4(1). R is subdirect in D^n for the following reason: for any $1 \leq j \leq n$ and $b \in D$, Claim 4 gives a realization f of (T, χ) which sends V into level r, is fixed up to $i-1$, and satisfies $f(u_\ell) \in D$ for all $1 \leq \ell \leq n$ and $f(u_j) = b$. This f puts b into $\mathrm{proj}_j(R)$, proving $R \subseteq_{sd} D^n$. And S contains all constant tuples (over D), for the following reason: if $b \in D$, then because $D \subseteq P^r_{a_i} = P^{r+1}_{a_i}$ and $(T, \chi) \in \mathcal{T}^{r+1}_d(A)$, there exists a realization f of (T, χ) which maps V into level r and maps every a_i-labelled leaf to b. This f puts $(b, b, \ldots, b)$ into S, as claimed.

In summary, R and S are subuniverses of $\mathbb{D}^n$; R is subdirect in D^n; S contains all constant tuples; and $R \lhd_\phi S$. As β is an absorption constant for $\mathbb{D}$ with respect to $\phi^{\mathbb{D}}$, it follows that the constant tuple $(\beta, \ldots, \beta)$ is in R. This witnesses the existence of the desired realization of (T, χ) satisfying $(\dagger)$, completing the proof that stage i can be achieved in the construction of $h : \mathbf{A} \to \mathbf{B}$, and thus completes the proof of Proposition 10.

D. Proof of Theorem 6

Theorem 6 (restated). *Let $\mathbb{D}$ be a finite algebra and ϕ an idempotent operation of $\mathbb{D}$. There exists an absorption constant* *for $\mathbb{D}$ with respect to ϕ.*

Proof: We may assume that $\mathbb{D} = (D, \phi)$. We argue by induction on $|D|$. The claim is clearly true if $|D| = 1$, so we may assume $|D| \geq 2$. For $n \geq 2$ let $\mathcal{A}_n = \{R \leq_{sd} \mathbb{D}^n : R$ absorbs $0_D^{(n)}$ with respect to $\phi\}$ and let $\mathcal{A} = \bigcup_{n=2}^{\infty} \mathcal{A}_n$. We must prove the existence of $b \in D$ such that $(b, \ldots, b) \in R$ for all $R \in \mathcal{A}$.

Let $\mathcal{P}$ be the set of all subuniverses of $\mathbb{D}$ (including $\varnothing$ and D) and let $\mathcal{P}^+ = \mathcal{P} \setminus \{\varnothing, D\}$. Observe that $\mathcal{P}^+ \neq \varnothing$, as $\mathcal{P}^+$ contains all the 1-element subsets of D. Suppose there exists $S \in \mathcal{P}^+$ such that, for all $n \geq 2$, we have $R \cap S^n \subseteq_{sd} S^n$ for all $R \in \mathcal{A}_n$. Let $\mathbb{S}$ be the subalgebra of $\mathbb{D}$ with universe S and note that $\phi\restriction_S$ is an idempotent operation of $\mathbb{S}$. Hence by the inductive hypothesis, $\mathbb{S}$ has an absorption constant b with respect to $\phi\restriction_S$. For any $n \geq 2$ and $R \in \mathcal{A}_n$, let $\overline{R} = R \cap S^n$; then $\overline{R} \leq_{sd} \mathbb{S}^n$ by our choice of S and $\overline{R}$ absorbs $0_S^{(n)}$ with respect to $\phi\restriction_S$ by Lemma 4(1); hence $(b, \ldots, b) \in \overline{R} \subseteq R$. Thus b is an absorption constant for $\mathbb{D}$, completing the proof.

Hence it suffices to prove that there exists $S \in \mathcal{P}^+$ such that, for all $n \geq 2$, we have $R \cap S^n \subseteq_{sd} S^n$ for all $R \in \mathcal{A}_n$. Let τ be the following algebraic vocabulary: for each $n \geq 2$, $R \in \mathcal{A}_n$ and $1 \leq i \leq n$, let $\mathsf{f}_{R,i}$ be an $(n-1)$-ary operation symbol and let $\tau = \{\mathsf{f}_{R,i} : n \geq 2, R \in \mathcal{A}_n, 1 \leq i \leq n\}$. Now define $\mathbb{P}$ to be the algebra with universe $\mathcal{P}$ and vocabulary τ where, for $n \geq 2$, $R \in \mathcal{A}_n$, $1 \leq i \leq n$ and $C_1, \ldots, C_{n-1} \in \mathcal{P}$,

$$\mathsf{f}_{R,i}^{\mathbb{P}}(C_1, \ldots, C_{n-1}) = \{x \in D : \exists c_1 \in C_1, \ldots, \exists c_{n-1} \in C_{n-1} \text{ such that } (c_1, \ldots, c_{i-1}, x, c_i, \ldots, c_{n-1}) \in R\}.$$

For convenience, we also let $\mathsf{f}_R^{\mathbb{P}}$ denote $\mathsf{f}_{R,1}^{\mathbb{P}}$ whenever $R \in \mathcal{A}$. Note that

1) $\{\mathsf{f}_{R,i}^{\mathbb{P}} : n \geq 2, R \in \mathcal{A}_n, 1 \leq i \leq n\} = \{\mathsf{f}_R^{\mathbb{P}} : R \in \mathcal{A}\}$. That is, every basic operation of $\mathbb{P}$ can be written as $\mathsf{f}_R^{\mathbb{P}}$ for some $R \in \mathcal{A}$.
2) Every operation of $\mathbb{P}$ is monotone with respect to $\subseteq$; that is, if $R \in \mathcal{A}_n$ and $B_i, C_i \in \mathcal{P}$ with $B_i \subseteq C_i$ for $1 \leq i \leq n-1$, then $\mathsf{f}_R^{\mathbb{P}}(B_1, \ldots, B_{n-1}) \subseteq \mathsf{f}_R^{\mathbb{P}}(C_1, \ldots, C_{n-1})$.
3) If $R = 0_D^{(3)}$, then $R \in \mathcal{A}_3$ and $\mathsf{f}_R^{\mathbb{P}}(C_1, C_2) = C_1 \cap C_2$. That is, $\cap$ is a basic operation of $\mathbb{P}$.
4) $\varnothing$ is a "zero" element for $\mathbb{P}$; that is, for all $n \geq 2$, $R \in \mathcal{A}_n$, and $C_1, \ldots, C_{n-1} \in \mathcal{P}$, if $\varnothing \in \{C_1, \ldots, C_{n-1}\}$ then $\mathsf{f}_R^{\mathbb{P}}(C_1, \ldots, C_{n-1}) = \varnothing$.
5) $\mathsf{f}_R^{\mathbb{P}}(D, \ldots, D) = D$ for all $R \in \mathcal{A}$.

Define a quasi-ordering $\preccurlyeq$ on $\mathcal{P}$ as follows: $C_1 \preccurlyeq C_2$ if and only if C_1 is an element of the subuniverse of $\mathbb{P}$ generated by $\{C_2, D\}$, i.e., $C_1 \in \mathrm{Sg}^{\mathbb{P}}(\{C_2, D\})$. Also define $C_1 \sim C_2$ to mean $C_1 \preccurlyeq C_2$ and $C_2 \preccurlyeq C_1$. Thus $\sim$ is an equivalence relation on $\mathcal{P}$ and $\preccurlyeq$ naturally induces a partial ordering $\leq$ of the set $\mathcal{P}/{\sim}$ of $\sim$-equivalence classes. Note that $\{\varnothing\}$ and $\{D\}$ are $\sim$-classes; furthermore, $\{D\}$ is the unique minimum element of $(\mathcal{P}/{\sim}, \leq)$, and $\{\varnothing\}$ is a minimal element of the poset obtained from $(\mathcal{P}/{\sim}, \leq)$ by deleting $\{D\}$. If we delete both $\{D\}$ and $\{\varnothing\}$, we obtain the poset $\mathcal{Q} := (\mathcal{P}^+/{\sim}, \leq)$.

Choose and fix a minimal element M of $\mathcal{Q}$. Let $M_{\min}$ denote the set of members of M which are minimal with respect to $\subseteq$. Also let $X = \bigcup\{C : C \in M_{\min}\}$.

Claim 5.

1) *Suppose $n \geq 2$, $R \in \mathcal{A}_n$, $C_1, \ldots, C_{n-1} \in M \cup \{D\}$, and let $A = \mathsf{f}_R^{\mathbb{P}}(C_1, \ldots, C_{n-1})$. If $A \neq \varnothing$, then $A \in M \cup \{D\}$; hence there exists $C \in M_{\min}$ with $C \subseteq A$.*
2) *For all $C_1 \in M_{\min}$ and $C_2 \in M \cup \{D\}$, if $C_1 \cap C_2 \neq \varnothing$ then $C_1 \subseteq C_2$.*

Proof: (1) With $n, R, C_1, \ldots, C_{n-1}, A$ as in the statement of the Claim, assume $A \notin \{\varnothing, D\}$, i.e., $A \in \mathcal{P}^+$, and pick any $C \in M$. Since $C_i \preccurlyeq C$ for each i, we get $A \preccurlyeq C$, so $A \in M$ as $C/{\sim} = M$ is a minimal member of $\mathcal{Q}$.

(2) This follows from (1) and the fact that $\cap$ is a basic operation of $\mathbb{P}$. ■

Claim 6. *For all $n \geq 2$ and $R \in \mathcal{A}_n$, $R \cap X^n \subseteq_{sd} X^n$.*

Proof: We will show $\mathrm{proj}_1(R \cap X^n) = X$, the argument for the other coordinates being similar. Pick any $C \in M_{\min}$. We will show, by induction on $1 \leq i \leq n$, that there exist $C_j \in M_{\min}$ for $2 \leq j \leq i$ such that $C \subseteq \mathsf{f}_R^{\mathbb{P}}(C_2, \ldots, C_i, D, \ldots, D)$. When $i = 1$ the claim is simply that $C \subseteq \mathsf{f}_R^{\mathbb{P}}(D, \ldots, D)$, which is true by a previous observation. Assume now that $1 \leq i < n$ and the claim has been verifed for i and must be proved for $i+1$. Thus we have $C_2, \ldots, C_i \in M_{\min}$ such that

$$C \subseteq \mathsf{f}_R^{\mathbb{P}}(C_2, \ldots, C_i, D, \ldots, D). \tag{1}$$

Equation (1) implies $\mathsf{f}_{R,i+1}^{\mathbb{P}}(C, C_2, \ldots, C_i, D, \ldots, D) \neq \varnothing$; hence by Claim 5 we can choose $C_{i+1} \in M_{\min}$ with

$$C_{i+1} \subseteq \mathsf{f}_{R,i+1}^{\mathbb{P}}(C, C_2, \ldots, C_i, D, \ldots, D). \tag{2}$$

Similarly, equation (2) implies

$$C \cap \mathsf{f}_R^{\mathbb{P}}(C_2, \ldots, C_i, C_{i+1}, D, \ldots, D) \neq \varnothing$$

and hence $C \subseteq \mathsf{f}_R^{\mathbb{P}}(C_2, \ldots, C_{i+1}, D, \ldots, D)$ by Claim 5, as desired, which completes the inductive argument. When $i = n$ this gives $C \subseteq \mathsf{f}_R^{\mathbb{P}}(C_2, \ldots, C_n) \subseteq \mathsf{f}_R^{\mathbb{P}}(X, \ldots, X)$, which implies $C \subseteq \mathrm{proj}_1(R \cap X^n)$. As $C \in M_{\min}$ was arbitrary, this proves $\mathrm{proj}_1(R \cap X^n) = X$. ■

Corollary 7. *Suppose $R \in \mathcal{A}_n$ and $B_1, \ldots, B_{n-1} \in \mathcal{P}$. Let $B = \mathsf{f}_R^{\mathbb{P}}(B_1, \ldots, B_{n-1})$.*

1) *For all $(b_1, \ldots, b_{n-1}) \in B_1 \times \cdots \times B_{n-1}$ and $b \in D$, if $(b, b_1, \ldots, b_{n-1}) \in R$ then $b \in B$.*
2) *For all $b \in B$ there exists $(b_1, \ldots, b_{n-1}) \in B_1 \times \cdots \times B_{n-1}$ such that $(b, b_1, \ldots, b_{n-1}) \in R$.*
3) *For all $x \in X$ there exists $(x_1, \ldots, x_{n-1}) \in X^{n-1}$ such that $(x, x_1, \ldots, x_{n-1}) \in R$.*

Proof: (1) and (2) follow from the definition of $\mathsf{f}_R^{\mathbb{P}}$, while (3) follows from Claim 6. ■

Observe that if $M_{\min}$ contains just one member C, then $X = C \in \mathcal{P}^+$ and so by Claim 6 we could choose $S = C$ and the proof of Theorem 6 would be complete. For the remainder of the proof, let $M_{\min} = \{C_1, \ldots, C_k\}$ be an enumeration of $M_{\min}$ and assume for the sake of contradiction that $k \geq 2$.

Definition 16. An *evaluated term tree* for $\mathbb{P}$ consists of a finite ordered directed tree $T = (V, E)$, an assignment to each non-leaf node v of an operation symbol $\mathsf{f}_R \in \tau$ whose arity equals the number of children of v, and a map $A : V \to \mathcal{P}$ satisfying the following recursive condition: if v is a non-leaf node, $v_1, \ldots, v_t$ are its children (listed in increasing order), and f_R is the operation symbol assigned to v, then

$$A(v) = \mathsf{f}_R^{\mathbb{P}}(A(v_1), \ldots, A(v_t)).$$

We call $A(v)$ the *value* of the evaluated term tree at v.

Note that if T_1, T_2 are evaluated term trees for $\mathbb{P}$, v is a leaf of T_1, and the value of T_1 at v is equal to the value of T_2 at its root, then T_2 may be glued to T_1 by identifying the root of T_2 with v.

Evaluated term trees witness subuniverse generation in $\mathbb{P}$; that is, if $B_1, \ldots, B_n, C \in \mathbb{P}$, then $C \in \mathrm{Sg}^{\mathbb{P}}(B_1, \ldots, B_n)$ if and only if there exists an evaluated term tree whose root has value C and whose leaves have values in $\{B_1, \ldots, B_n, D\}$.

Claim 8. *For any $i, j \in \{1, \ldots, k\}$ with $i \neq j$ there exists an evaluated term tree $T_{i,j}$ for $\mathbb{P}$ satisfying the following:*

1) *The value at each leaf is in $\{C_i, D\}$. At least one leaf has value C_i.*
2) *The value at the root is C_j.*

Proof: The existence of a tree satisfying (2) and the first sentence of (1) follows from the fact that $C_i \preccurlyeq C_j$. That at least one leaf has value C_i follows from the fact that $\{D\}$ is a subuniverse of $\mathbb{P}$. ■

We now construct a special evaluated term tree T for $\mathbb{P}$ as follows. We start with $T_2 := T_{2,1}$. To each leaf of T_2 whose value is C_2 we glue a copy of $T_{3,2}$; call the resulting evaluated term tree T_3. Then to each leaf of T_3 whose value is C_3 we glue a copy of $T_{4,3}$, to get T_4, etc. After we have constructed T_k, we glue to each leaf whose value is C_k a copy of $T_{1,k}$, to get T. Here are the salient properties of T:

1) The value at each leaf is in $\{C_1, D\}$.
2) The value at the root is C_1.
3) For every $1 \leq i \leq k$ and every path from the root to a leaf whose value is C_1, some node on the path has value in C_i.

Let V denote the set of nodes of T, r the root, Λ the set of leaves, and Λ_1 the set of leaves whose value is C_1. Also let $A : V \to \mathbb{P}$ be the value map.

Definition 17. A map $\alpha : V \to D$ is called a *selection map* (for T). Suppose α is a selection map, u is an arbitrary node, and v is a non-leaf node of T. Let f_R be the operation symbol in τ assigned to v with $R \in \mathcal{A}_n$, let $v_1, \ldots, v_{n-1}$ be the children of v listed in increasing order, and let $\alpha(\mathbf{v}) = (\alpha(v), \alpha(v_1), \ldots, \alpha(v_{n-1}))$. We say that:

1) α *respects values at* u if $\alpha(u) \in A(u)$.
2) α *quasi-respects values at* u if $\alpha(u) \in X$.
3) α *respects relations at* v if $\alpha(\mathbf{v}) \in R$.
4) α *quasi-respects relations at* v if $\alpha(\mathbf{v}) \in R \cup 0_D^{(n)}$.

Note that, by Corollary 7(1), if α is a selection map which respects values at all C_1-valued leaves and respects relations at all non-leaf nodes, then α respects values at all nodes. In particular, $\alpha(r) \in C_1$. Conversely:

Claim 9.

1) *For all $a \in C_1$ there exists a selection map α_a which respects values at all nodes, respects relations at all non-leaf nodes, and satisfies $\alpha_a(r) = a$.*
2) *For all $a \in X$ there exists a selection map β_a which quasi-respects values at all nodes, respects relations at all non-leaf nodes, and satisfies $\beta_a(r) = a$.*
3) *For all $a \in X$ there exists a selection map γ_a which respects values at all C_1-valued leaves, quasi-respects relations at all non-leaf nodes, and satisfies $\gamma_a(r) = a$.*

Proof: (1) We inductively define α_a, starting at the root. Of course, $\alpha_a(r) = a$. Suppose now that v is a non-leaf node and $\alpha_a(v) \in A(v)$. Let f_R be the operation symbol assigned to v with $R \in \mathcal{A}_n$, and let $v_1, \ldots, v_{n-1}$ be the children of v in increasing order. As $A(v) = \mathsf{f}_R^{\mathbb{P}}(A(v_1), \ldots, A(v_{n-1}))$, by Corollary 7(2) there exists $(c_1, \ldots, c_{n-1}) \in A(v_1) \times \cdots \times A(v_{n-1})$ such that $(\alpha_a(v), c_1, \ldots, c_{n-1}) \in R$. We can thus define $\alpha_a(v_i) = c_i$ and continue inductively.

(2) is proved similarly, using Corollary 7(3).

(3) Suppose $a \in C_i$. Let V_i denote the set of nodes of T such that the path from v to the root includes a node having value C_i. We will construct γ_a so that it will inductively satisfy the following additional property: γ_a has constant value a on $V \setminus V_i$ and respects values on V_i. By one of the salient properties of T, this will imply that γ_a respects values at C_1-valued leaves.

We start by defining $\gamma_a(r) = a$. Suppose now that v is a non-leaf node and $\gamma_a(v)$ has already been defined. Let f_R be the operation assigned to v with $R \in \mathcal{A}_n$, and let $v_1, \ldots, v_{n-1}$ be the children of v in increasing order. If $v \in V_i$ then inductively we have $\gamma_a(v) \in A(v)$, so we define γ_a at the children of v exactly as in the definition of α_a; thus γ_a respects relations at v. If instead $v \in V \setminus V_i$, then inductively we have $\gamma_a(v) = a$, and we define $\gamma_a(v') = a$ for all children v' of v; thus γ_a quasi-respects relations at v. It remains to check that the inductive property is maintained by this construction. The only problem to consider is if $v \in V \setminus V_i$ but a child v' of v is in V_i. If this is the case, then it must be that v' has value C_i. As the construction assigns $\gamma_a(v') = a$ and as $a \in C_i$, there is no problem. ■

Recall that ϕ is the basic operation of $\mathbb{D}$ referenced in the statement of Theorem 6. Let m be its arity.

Claim 10. *For all $0 \leq j \leq m$,*

$$\phi(\underbrace{C_1, \ldots, C_1}_{m-j}, \underbrace{X, \ldots, X}_{j}) \subseteq C_1.$$

Proof: By induction on j. When $j = 0$ the claim is simply that $\phi(C_1, \ldots, C_1) \subseteq C_1$, which is true because C_1 is a subuniverse of $\mathbb{D}$. Suppose $j < m$ and the claim is true for j. To prove that it is true for $j + 1$, let $\ell = m - j - 1$ and assume $a_1, \ldots, a_\ell \in C_1$ and $x_0, x_1 \ldots, x_j \in X$. It suffices to prove $\phi(a_1, \ldots, a_\ell, x_0, \ldots, x_j) \in C_1$. Let

$\alpha_{a_1}, \ldots, \alpha_{a_\ell}, \beta_{x_1}, \ldots, \beta_{x_j}, \gamma_{x_0}$ be selection maps for T constructed according to Claim 9. Define $\delta : V \to D$ by

$$\delta(v) = \phi(\alpha_{a_1}(v), \ldots, \alpha_{a_\ell}(v), \gamma_{x_0}(v), \beta_{x_1}(v), \ldots, \beta_{x_j}(v)).$$

δ is a selection map. Suppose v is a C_1-valued leaf of T. Then $\alpha_{a_i}(v) \in C_1$ for all $1 \leq i \leq \ell$, $\gamma_{x_0}(v) \in C_1$, and $\beta_{x_i}(v) \in X$ for all $1 \leq i \leq j$. Thus $\delta(v) \in \phi(\underbrace{C, \ldots, C, C}_{\ell+1}, \underbrace{X, \ldots, X}_{j}) \subseteq C_1$ by the inductive assumption. This proves that δ respects values at C_1-valued leaves.

Suppose next that v is a non-leaf node of T. Let f_R be the operation symbol assigned to v with $R \in \mathcal{A}_n$, let $v_1, \ldots, v_{n-1}$ be the children of v listed in increasing order, and for any selection map η let $\eta(\mathbf{v}) = (\eta(v), \eta(v_1), \ldots, \eta(v_{n-1}))$. Then $\alpha_{a_i}(\mathbf{v}) \in R$ for $1 \leq i \leq \ell$, $\beta_{x_i}(\mathbf{v}) \in R$ for $1 \leq i \leq j$, $\gamma_{x_0}(\mathbf{v}) \in R \cup 0_D^{(n)}$, and

$$\begin{aligned}\delta(\mathbf{v}) &= \phi^{\mathbb{D}^n}(\alpha_{a_1}(\mathbf{v}), \ldots, \alpha_{a_\ell}(\mathbf{v}), \gamma_{x_0}(\mathbf{v}), \beta_{x_1}(\mathbf{v}), \ldots, \beta_{x_j}(\mathbf{v})) \\ &\in \phi^{\mathbb{D}^n}(R, \ldots, R, R \cup 0_D^{(n)}, R, \ldots, R) \quad \subseteq \quad R,\end{aligned}$$

where the last inclusion follows because R absorbs $0_D^{(n)}$ with respect to ϕ. This proves that δ respects relations at all non-leaf nodes.

It follows from the observation preceding Claim 9 that δ respects values at all nodes. In particular, δ respects values at the root, i.e., $\delta(r) = \phi(a_1, \ldots, a_\ell, x_0, \ldots, x_j) \in C_1$, which finishes the proof of the Claim. ■

In particular, Claim 10 yields $\phi(X, \ldots, X) \subseteq C_1$, while $X \subseteq \phi(X, \ldots, X)$ because ϕ is idempotent. Thus $X \subseteq C_1$, implying $X = C_1$. This contradicts our assumption that $|M_{\min}| \geq 2$ and thus completes the proof of Theorem 6. ■

IV. Conclusion

We have proved that for every finite relational structure $\mathbf{B}$ having a near-unanimity polymorphism, the corresponding constraint satisfaction problem $\mathrm{CSP}(\mathbf{B})$ has bounded pathwidth duality; equivalently, $\neg\mathrm{CSP}(\mathbf{B})$ is definable in linear Datalog, and as a consequence, $\mathrm{CSP}(\mathbf{B})$ is in the complexity class NL. This answers a question from [16].

The natural algebraic conjecture alluded to in the introduction, suggested by Larose and Tesson [27], is the following: if $\mathbf{B}$ is core and $\mathbb{B}_e$ is the idempotent reduct of the polymorphism algebra of $\mathbf{B}$, then $\mathrm{CSP}(\mathbf{B})$ has bounded pathwidth duality if and only if the variety generated by $\mathbb{B}_e$ "omits types 1, 2 and 5" in the sense of tame congruence theory. A characterization of this property in terms of idempotent polymorphisms of $\mathbf{B}$ is given by [22, Theorem 9.11]. Our result lends further support to this conjecture which, however, remains open.

Acknowledgment

L. Barto was supported by the Grant Agency of the Czech Republic, grant 201/09/P223 and by the Ministry of Education of the Czech Republic, grant MSM 0021620839. M. Kozik was supported by the Foundation for Polish Science, grant No. HOM/2008/7 (supported by MF EOG) and National Science Center grant DEC-2011/01/B/ST6/01006. R. Willard was supported by the NSERC of Canada.

References

[1] S. Abiteboul, R. Hull and V. Vianu, *Foundations of Databases*, Addison-Wesley, 1995.
[2] E. Allender, M. Bauland, N. Immerman, H. Schnoor, and H. Vollmer, The complexity of satifiability problems: Refining Schaefer's theorem, *Proc. 30th Math. Found. Comp. Sci.* (2005), 71–82.
[3] K. Baker and A. Pixley, Polynomial interpolation and the Chinese remainder theorem for algebraic systems, *Math. Z.* **143** (1975), 165–174.
[4] L. Barto, Finitely related algebras in congruence distributive varieties have near unanimity terms, *Canad. J. Math.*, published electronically Dec. 2011, doi:10.4153/CJM-2011-087-3, 19 pp.
[5] L. Barto and M. Kozik, Congruence distributivity implies bounded width, *SIAM J. Comput.* **39** (2009), 1531–1542.
[6] L. Barto and M. Kozik, Constraint satisfaction problems of bounded width, *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009)*, 595–603, IEEE Computer Soc., 2009.
[7] L. Barto and M. Kozik, Absorbing subalgebras, cyclic terms and the constraint satisfaction problem, *Logical Methods in Computer Science* **8** (1:07) (2012), 1–26.
[8] C. Bergman, *Universal Algebra*, CRC Press, Boca Raton, 2012.
[9] V.G. Bodnarčuk, L.A. Kalužnin, V.N. Kotov, and B.A. Romov, Galois theory for Post algebras. I, *Cybernetics and Systems Analysis* **5** (1969), 243–252.
[10] A. Bulatov, A graph of a relational structure and constraint satisfaction problems, *LICS '04* (2004), 448–457.
[11] A. Bulatov, A. Krokhin, and P. Jeavons, Classifying the complexity of constraints using finite algebras, *SIAM J. Comput.* **34** (2005), 720–742.
[12] S. Burris and H.P. Sankappanavar, *A Course in Universal Algebra*, Springer-Verlag, New York, 1981.
[13] S. Cook and R. Sethi, Storage requirements for deterministic polynomial time recognizable languages, *J. Compu. System Sci.* **13** (1976), 25–37.
[14] V. Dalmau, Constraint satisfaction problems in non-deteministic logarithmic space, *LNCS* **2380** (2002), 414–425.
[15] V. Dalmau, Linear datalog and bounded path duality of relational structures, *Logical Methods in Computer Science* **1** (2005), 1–32.
[16] V. Dalmau and A. Krokhin, Majority constraints have bounded pathwidth duality, *Euro. J. Combin.* **29** (2008), 821–837.
[17] J. A. Ellis, I. H. Sudborough, and J. S. Turner, The vertex separation and search number of a graph, *Inform. Comput.* **113** (1994), 50-79.
[18] T. Feder and M. Vardi, The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory, *SIAM J. Comput.* **28** (1998), 57–104.
[19] D. Geiger, Closed systems of functions and predicates, *Pacific J. Math.* **27** (1968), 95–100.
[20] P. Hell and J. Nešetřil, *Graphs and Homomorphisms*, Oxford University Press, 2004.
[21] P. Hell, J. Nešetřil, and X. Zhu, Duality and polynomial testing of tree homomorphisms, *Trans. Amer. Math. Soc.* **348** (1996), 147–156.
[22] D. Hobby and R. McKenzie, *The Structure of Finite Algebras*, American Mathematical Society, Providence, 1988.
[23] P. Jeavons, On the algebraic structure of combinatorial problems, *Theoret. Comput. Sci.* **200** (1998), 185—204.
[24] P. Jeavons, D. Cohen, and M. Gyssens, How to determine the expressive power of constraints, *Constraints* **4** (1999), 113–131.
[25] Ph. Kolaitis and M.Vardi, Conjunctive-query containment and constraint satisfaction, *J. Comput. System Sci.* **61** (2000), 302–332.
[26] A. Krokhin and B. Larose, Solving order constraints in logarithmic space, *STACS'03*, *LNCS* **2607** (2003), 379–390.
[27] B. Larose and P. Tesson, Universal algebra and hardness problems for constraint satisfaction problems, *Theoret. Comput. Sci.* **410** (2009), 1629–1647.
[28] B. Larose, M. Valeriote, and L. Zádori, Omitting types, bounded width and the ability to count, *Internat. J. Algebra Comput.* **19** (2009), 647–668.
[29] B. Larose and L. Zádori, Bounded width problems and algebras, *Algebra Universalis* **56** (2007), 439–466.
[30] M. Maróti and R. McKenzie, Existence theorems for weakly symmetric operations, *Algebra Universalis* **59** (2008), 463–489.
[31] U. Montanari, Networks of constraints: Fundamental properties and applications to picture processing, *Information Science* **7** (1974), 95–132.
[32] N. Robertson and P. Seymour, Graph minors. I. Excluding a forest, *J. Combin. Theory Ser. B* **35** (1983), 39–61.
[33] J. Ullman, *Principles of Database and Knowledge-base Systems*, Vol. 1, Computer Science Press, 1988.

Appendix G – Robust satisfiability

Robust Satisfiability of Constraint Satisfaction Problems

Libor Barto
Department of Mathematics and Statistics
McMaster Universty
and
Department of Algebra
Charles University in Prague
Sokolovská 83, 18675 Praha 8, Czech Republic
libor.barto@gmail.com

Marcin Kozik
Theoretical Computer Science Department
Faculty of Mathematics and Computer Science
Jagiellonian University
ul. Prof. St. Lojasiewicza 6
30-348 Krakow, Poland
Marcin.Kozik@uj.edu.pl

ABSTRACT

An algorithm for a constraint satisfaction problem is called robust if it outputs an assignment satisfying at least $(1 - g(\varepsilon))$-fraction of the constraints given a $(1 - \varepsilon)$-satisfiable instance, where $g(\varepsilon) \to 0$ as $\varepsilon \to 0$, $g(0) = 0$. Guruswami and Zhou conjectured a characterization of constraint languages for which the corresponding constraint satisfaction problem admits an efficient robust algorithm. This paper confirms their conjecture.

Categories and Subject Descriptors

F.2.2 [**Analysis of algorithms and problem complexity**]: Nonnumerical algorithms and problems

General Terms

Theory, Algorithms

Keywords

constraint satisfaction problem, bounded width, approximation, robust satisfiability, universal algebra

1. INTRODUCTION

The constraint satisfaction problem (CSP) provides a common framework for many theoretical problems in computer science as well as for many real-life applications. An instance of the CSP consists of a number of variables and constraints imposed on them, and the objective is to efficiently find an assignment for variables with desired properties, or at least to decide whether such an assignment exists. In the decision problem for CSP we want to decide if there is an assignment satisfying all the constraints, in Max-CSP we wish to find an assignment satisfying maximum number of constraints, in the approximation version of Max-CSP we seek for an assignment which is in some sense close to the optimal one. This paper deals with an interesting special case, robust satisfiability of the CSP: Given an instance which is almost satisfiable (say $(1 - \varepsilon)$-fraction of the constraint can be satisfied), we want to efficiently find an almost satisfying assignment (which satisfies at least $(1 - g(\varepsilon))$-fraction of the constraints, where $\lim_{\varepsilon \to 0} g(\varepsilon) = 0$).

Most of the computational problems for the CSP are hard in general, therefore we have to put some restrictions on the instance. In this paper we restrict the constraint language, that is, all constraint relations must come from a fixed, finite set of relations on the domain. Robust satisfiability was in this setting first suggested and studied in a paper by Zwick [24]. The motivation is that in certain practical situations instances might be close to satisfiable (for example, a small fraction of constraints might have been corrupted by noise) and an algorithm that is able to satisfy most of the constraints could be useful.

Zwick [24] concentrated on Boolean CSPs. He designed a semidefinite programming (SDP) based algorithm which finds $(1-O(\varepsilon^{1/3}))$-satisfying assignment for $(1-\varepsilon)$-satisfiable instances of 2-`SAT` and linear programming (LP) based algorithm which finds $(1-O(1/\log(1/\varepsilon)))$-satisfying assignment for $(1-\varepsilon)$-satisfiable instances of `Horn-`k`-Sat` (the number k refers to the maximum numbers of variables in a Horn constraint). The quantitative dependence on ε was improved for 2-`SAT` to $(1 - O(\sqrt{\varepsilon}))$ in [7]. For `CUT`, a special case of 2-`SAT`, the Goemans-Williamson algorithm [11] also achieves $(1 - O(\sqrt{\varepsilon}))$. The same dependence was proved more generally for `Unique-Games(`q`)` [6] (where q refers to the size of the domain), which improved $(1 - O(\sqrt[5]{\varepsilon}\log^{1/2}(1/\varepsilon)))$ obtained in [17]. For `Horn-2-Sat` the exponential loss can be replaced by $(1 - 3\varepsilon)$ [16] and even $(1 - 2\varepsilon)$ [12]. These bounds for `Horn-`k`-Sat` ($k \geq 3$), `Horn-2-Sat`, 2-`SAT`, and `Unique-Games(`q`)` are actually essentially optimal [17, 18, 12] assuming Khot's Unique Games Conjecture [17].

On the negative side, if the decision problem for CSP is NP-complete, then given a satisfiable instance it is NP-hard to find an assignment satisfying α-fraction of the constraints for some constant $\alpha < 1$ (see [16] for the Boolean case and [14] for the general case). In particular these problems cannot admit an efficient robust satisfiability algorithm (assuming $P \neq NP$). However NP-completeness of the decision problem is not the only obstacle for robust algorithms. In [13] Håstad proved a strikingly optimal hardness result: for `E3-LIN(`q`)` (linear equations over $\mathbb{Z}_q$ where each equation contains precisely 3 variables) it is NP-hard to find an assignment satisfying $(1/q+\varepsilon)$-fraction of the constraints given

STOC'12, May 19–22, 2012, New York, New York, USA.

an instance which is $(1-\varepsilon)$-satisfiable. Note that the trivial random algorithm achieves $1/q$ in expectation.

As observed in [24] the above results cover all Boolean CSPs, because, by Schaefer's theorem [23], `E3-LIN(`q`)`, `Horn-`k`-Sat` and `2-SAT` are essentially the only CSPs with tractable decision problem. What about larger domains? A natural property which distinguishes `Horn-`k`-Sat`, `2-SAT`, and `Unique-Games(`q`)` from `E3-LIN(`q`)` and NP-complete CSPs is bounded width [9]. Briefly, a CSP has bounded width if the decision problem can be solved by checking local consistency of the instance. These problems were characterized independently by the authors [1] and Bulatov [3]. It was proved that, in some sense, the only obstacle to bounded width is `E3-LIN(`q`)` – the same problem which is difficult for robust satisfiability. These facts motivated Guruswami and Zhou to conjecture [12] that the class of bounded width CSPs coincide with the class of CSPs admitting a robust satisfiability algorithm.

A partial answer to the conjecture for width one problems was recently independently given by Kun, O'Donnell, Tamaki, Yoshida and Zhou [19] (where they also show that width 1 characterizes problems robustly decidable by the canonical linear programming relaxation), and Dalmau and Krokhin [8] (where they also consider some problems beyond width 1). This paper confirms the Guruswami and Zhou conjecture in full generality. The proof uncovers an interesting connection between the outputs of SDP (and LP) relaxations and Prague strategies – a consistency notion crucial for the bounded width characterization [1].

2. PRELIMINARIES

Definition 1. An *instance of the* CSP is a triple $\mathcal{I} = (V, D, \mathcal{C})$ with V a finite set of *variables*, D a finite *domain*, and $\mathcal{C}$ a finite list of *constraints*, where each constraint is a pair $C = (S, R)$ with S a tuple of variables of length k, called the *scope* of C, and R an k-ary relation on D (i.e. a subset of D^k), called the *constraint relation* of C.

A finite set of relations Γ on D is called a *constraint language*. An *instance of* $\mathrm{CSP}(\Gamma)$ is an instance of the CSP such that all the constraint relations are from Γ.

An *assignment* for $\mathcal{I}$ is a mapping $F : V \to D$. We say that F *satisfies* a constraint $C = (S, R)$ if $F(S) \in R$ (where F is applied component-wise). The *value* of F, $\mathrm{Val}(F, \mathcal{I})$, is the fraction of constraints it satisfies. The *optimal value* of $\mathcal{I}$ is $\mathrm{Opt}(\mathcal{I}) = \max_{F:V\to D} \mathrm{Val}(F, \mathcal{I})$.

The *decision problem* for $\mathrm{CSP}(\Gamma)$ asks whether an input instance $\mathcal{I}$ of $\mathrm{CSP}(\Gamma)$ has a *solution*, i.e. an assignment which satisfies all the constraints. It is known [4] that if $\mathrm{CSP}(\Gamma)$ is tractable, then there exists a polynomial algorithm for finding an assignment F with $\mathrm{Val}(F, \mathcal{I}) = 1$.

Definition 2. Let Γ be a constraint language and let α, β be real numbers. An algorithm (α, β)-*approximates* $\mathrm{CSP}(\Gamma)$, if it outputs an assignment F with $\mathrm{Val}(F, \mathcal{I}) \geq \alpha$ for every instance $\mathcal{I}$ of $\mathrm{CSP}(\Gamma)$ such that $\mathrm{Opt}(\mathcal{I}) \geq \beta$.

We say that $\mathrm{CSP}(\Gamma)$ admits a *robust satisfiability algorithm* if there exists a function $g : [0,1] \to [0,1]$ such that $\lim_{\varepsilon\to 0} g(\varepsilon) = 0$, $g(0) = 0$, and a polynomial algorithm which $(1 - g(\varepsilon), 1 - \varepsilon)$-approximates $\mathrm{CSP}(\Gamma)$ for every $\varepsilon \in [0, 1]$.

Bounded width and the Guruswami-Zhou conjecture

A natural notion with distinguishes known CSPs which admit a robust satisfiability algorithm (like `Horn-`k`-Sat`, `2-SAT`, and `Unique-Games(`q`)`) from those which do not (like `E3-LIN(`q`)`, NP-complete CSPs) is *bounded width*.

Informally, $\mathrm{CSP}(\Gamma)$ has bounded width if the decision problem for $\mathrm{CSP}(\Gamma)$ can be solved by checking local consistency. More specifically, for fixed integers (k, l), the (k, l)-algorithm derives the strongest constraints on k variables which can be deduced by looking at l variables at a time. During the process we may obtain a contradiction (i.e. an empty constraint relation), in this case $\mathcal{I}$ has no solution. We say that $\mathrm{CSP}(\Gamma)$ has *width* (k, l) if this procedure is sound, that is, an instance has a solution if and only if the (k, l)-consistency algorithm does not derive a contradiction. We say that $\mathrm{CSP}(\Gamma)$ has *width* k, if it has width (k, l) for some l. Finally, we say that $\mathrm{CSP}(\Gamma)$ has *bounded width* if it has width k for some k. We refer to [9, 21, 5] for formal definitions and background.

CONJECTURE 3 (GURUSWAMI,ZHOU [12]). *$\mathrm{CSP}(\Gamma)$ admits a robust satisfiability algorithm if and only if $\mathrm{CSP}(\Gamma)$ has bounded width.*

One implication of the Guruswami-Zhou conjecture follows from known results. In [1] and [3] it was proved that `E3-LIN(`q`)` is essentially the only obstacle for bounded width – if Γ cannot "encode linear equations", then $\mathrm{CSP}(\Gamma)$ has bounded width (here we do not need to assume $\mathrm{P} \neq \mathrm{NP}$). Therefore, if $\mathrm{CSP}(\Gamma)$ does not have bounded width, then Γ can encode linear equations and, consequently, $\mathrm{CSP}(\Gamma)$ admits no robust satisfiability algorithm by Håstad's result [13] (assuming $\mathrm{P} \neq \mathrm{NP}$). Details will be presented in [8].

This paper proves the other implication:

THEOREM 4. *If $\mathrm{CSP}(\Gamma)$ has bounded width then it admits a robust satisfiability algorithm. The randomized version of this algorithm returns an assignment satisfying, in expectation, $(1 - O(\log\log(1/\varepsilon)/\log(1/\varepsilon)))$-fraction of the constraints given a $(1 - \varepsilon)$-satisfiable instance.*

LP and SDP relaxations

Essentially the only known way to design efficient approximation algorithms is through linear programming (LP) relaxations and semidefinite programming (SDP) relaxations. For instance, the robust satisfiability algorithm for `Horn-`k`-Sat` [24] uses LP relaxation while the robust satisfiability algorithms for `2-SAT` and `Unique-Games(`q`)` [24, 7] are SDP-based.

Recently, robust satisfiability algorithm was devised in [19] and independently [8] for all CSPs of width 1 (this covers `Horn-`k`-Sat`, but not `2-SAT` or `Unique-Games(`q`)`). The latter one uses a reduction to `Horn-`k`-Sat` while the former uses an LP relaxation directly. In fact, it is shown in [19] that, in some sense, LP relaxations can be used precisely for width 1 CSPs.

Our algorithm is based on the canonical SDP relaxation [22]. We will use it only for instances with unary and binary constraints (a reduction is provided in the appendix). In this case we can formulate the relaxation as follows.

Definition 5. Let Γ be a constraint language over D consisting of at most binary relations and let $\mathcal{I} = (V, D, \mathcal{C})$ be an instance of CSP(Γ) with m constraints. The goal for the *canonical SDP relaxation* of $\mathcal{I}$ is to find $(|V||D|)$-dimensional real vectors $\mathbf{x}_a, x \in V, a \in D$ maximizing

$$\frac{1}{m}\left(\sum_{(x,R)\in\mathcal{C}}\sum_{a\in R}||\mathbf{x}_a||^2 + \sum_{((x,y),R)\in\mathcal{C}}\sum_{(a,b)\in R}\mathbf{x}_a\mathbf{y}_b\right) \qquad (1)$$

subject to

(SDP1) $\mathbf{x}_a\mathbf{y}_b \geq 0$ for all $x, y \in V, a, b \in D$

(SDP2) $\mathbf{x}_a\mathbf{x}_b = 0$ for all $x \in V, a, b \in D, a \neq b$, and

(SDP3) $\sum_{a\in D}\mathbf{x}_a = \sum_{a\in D}\mathbf{y}_a,\ ||\sum_{a\in D}\mathbf{x}_a||^2 = 1$ for all $x, y \in V$.

The dot products $\mathbf{x}_a\mathbf{y}_b$ can be thought of as weights and the goal is to find vectors so that maximum weight is given to pairs (or elements) in constraint relations. It will be convenient to use the notation

$$\mathbf{x}_A = \sum_{a\in A}\mathbf{x}_a$$

for a variable $x \in V$ and a subset $A \subseteq D$, so that condition (SDP3) can be written as $\mathbf{x}_D = \mathbf{y}_D$, $||\mathbf{x}_D||^2 = 1$. The contribution of one constraint to (1) is by (SDP3) at most 1 and it is the greater the less weight is given to pairs (or elements) outside the constraint relation.

The optimal value for the sum (1), SDPOpt($\mathcal{I}$), is always at least Opt($\mathcal{I}$). There are algorithms that outputs vectors with (1) $\geq$ SDPOpt($\mathcal{I}$)$-\delta$ which are polynomial in the input size and $\log(1/\delta)$.

Polymorphisms

Much of the recent progress on the complexity of the decision problem for CSP was achieved by the algebraic approach [4]. The crucial notion linking relations and operations is a polymorphism:

Definition 6. An l-ary operation f on D is a *polymorphism* of a k-ary relation R, if

$$(f(a_1^1,\dots,a_1^l), f(a_2^1,\dots,a_2^l),\dots,f(a_k^1,\dots,a_k^l)) \in R$$

whenever $(a_1^1,\dots,a_k^1)$, $(a_1^2,\dots,a_k^2)$, $\dots$, $(a_1^l,\dots,a_k^l) \in R$.

We say that f is a polymorphism of a constraint language Γ, if it is a polymorphism of every relation in Γ. The set of all polymorphisms of Γ will be denoted by Pol(Γ)

We say that Γ is a *core*, if all its unary polymorphisms are bijections.

The complexity of the decision problem for CSP(Γ) (modulo log-space reductions) depends only on equations satisfied by operations in Pol(Γ) (see [4, 20]). Moreover, equations also determine whether CSP(Γ) has bounded width [21]. The following theorem [10] states one such an equational characterization:

THEOREM 7. *Let Γ be a core constraint language. Then the following are equivalent.*

- CSP(Γ) *has bounded width.*
- Pol(Γ) *contains a* 3*-ary operation f_1 and a* 4*-ary operation f_2 such that, for all $a, b \in D$,*

$$f_1(a,a,b) = f_1(a,b,a) = f_1(b,a,a) = \\ = f_2(a,a,a,b) = \dots = f_2(b,a,a,a)$$

  and $f_1(a,a,a) = a$.

We remark that the problem of deciding whether CSP(Γ) has bounded width, given Γ as an input, is tractable (the problem is obviously in NP).

3. PRAGUE INSTANCES

The proof of the characterization of bounded width CSPs in [1] relies on a certain consistency notion called Prague strategy. It turned out that Prague strategies are related to outputs of canonical SDP relaxations and this connection is what made our main result possible.

The notions defined below will be used only for certain types of instances and constraint languages. Therefore, in the remainder of this section we assume that Λ is a constraint language on a domain D, Λ contains only binary relations, $\mathcal{J} = (V, D, \mathcal{C}^{\mathcal{J}})$ is an instance of CSP(Λ) such that every pair of distinct variables is the scope of at most one constraint $((x,y), P^{\mathcal{J}}_{x,y})$, and if $((x,y), P^{\mathcal{J}}_{x,y}) \in \mathcal{C}^{\mathcal{J}}$ then $((y,x), P^{\mathcal{J}}_{y,x}) \in \mathcal{C}^{\mathcal{J}}$, where $P^{\mathcal{J}}_{y,x} = \{(b,a) : (a,b) \in P^{\mathcal{J}}_{x,y}\}$. (We usually omit the superscripts for $P_{x,y}$'s and $\mathcal{C}$.)

The most basic consistency notion for CSP instances is 1-minimality.

Definition 8. The instance $\mathcal{J}$ is called *1-minimal*, if there exist subsets $P^{\mathcal{J}}_x \in D$, $x \in V$ such that, for every constraint $((x,y), P^{\mathcal{J}}_{x,y})$, the constraint relation $P^{\mathcal{J}}_{x,y}$ is subdirect in $P^{\mathcal{J}}_x \times P^{\mathcal{J}}_y$, i.e. the projection of $P^{\mathcal{J}}_{x,y}$ to the first (resp. second) coordinate is equal to $P^{\mathcal{J}}_x$ (resp. $P^{\mathcal{J}}_y$).

The subset $P^{\mathcal{J}}_x$ is uniquely determined by the instance (if x is in the scope of some constraint).

Weak Prague instance

We will work with a weakening of the notion of a Prague strategy which we call a weak Prague instance. First we need to define steps and patterns.

Definition 9. A *step* (in $\mathcal{J}$) is a pair of variables (x,y) which is the scope of a constraint in $\mathcal{C}^{\mathcal{J}}$. A *pattern from x to y* is a sequence of variables $p = (x = x_1, x_2, \dots, x_k = y)$ such that every (x_i, x_{i+1}), $i = 1, \dots, k-1$ is a step.

For a pattern $p = (x_1, \dots, x_k)$ we put $-p = (x_k, \dots, x_1)$. If $p = (x_1, \dots, x_k)$, $q = (y_1, \dots, y_l)$, $x_k = y_1$ then the concatenation of p and q is the pattern $p+q = (x_1, x_2, \dots, x_k = y_1, y_2, \dots, y_k)$. For a pattern p from x to x and a natural number k, kp denotes the k-time concatenation of p with itself.

For a subset $A \subseteq D$ and a step $p = (x,y)$ we define $A + p$ to be the projection of the constraint relation $P_{x,y}$ onto the second coordinate after restricting the first coordinate to A, that is, $A + p = \{b \in D : (\exists\ a \in A)\ (a,b) \in P_{x,y}\}$. For a general pattern p, the set $A + p$ is defined step by step.

Definition 10. $\mathcal{J}$ is a *weak Prague instance* if

(P1) $\mathcal{J}$ is 1-minimal,

(P2) for every $A \subseteq P^{\mathcal{J}}_x$ and every pattern p from x to x, if $A + p = A$ then $A - p = A$, and

(P3) for any patterns p_1, p_2 from x to x and every $A \subseteq P^{\mathcal{J}}_x$, if $A + p_1 + p_2 = A$ then $A + p_1 = A$.

The instance $\mathcal{J}$ is *nontrivial*, if $P^{\mathcal{J}}_x \neq \emptyset$ for every $x \in V$.

To clarify the definition let us consider the following digraph: vertices are all the pairs (A, x), where $x \in V$ and $A \subseteq P_x^{\mathcal{J}}$, and $((A, x), (B, y))$ forms an edge iff (x, y) is a step and $A + (x, y) = B$. Condition (P3) means that no strong component contains (A, x) and (A', x) with $A \neq A'$, condition (P2) is equivalent to the fact that every strong component contains only undirected edges. Also note that 1-minimality implies $A \subseteq A + p - p$ for any pattern from x.

A simple example of a weak Prague instance (which is not a Prague strategy) is $V = \{x, y, z\}$, $D = \{0, 1\}$, $P_{x,y} = P_{x,z} = \{(0, 0), (1, 1)\}$, $P_{y,z} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

If we change $P_{y,z}$ to $\{(0, 1), (1, 0)\}$ the conditions (P1) and (P2) hold but $\{0\} + (x, y, z, x) + (x, y, z, x) = \{0\}$ and $\{0\} + (x, y, z, x) = \{1\}$.

If, on the other hand, we set $P_{y,z}$ to $\{(0, 0), (1, 0), (1, 1)\}$ then (P1) and (P3) hold while $\{0\} + (x, y, z, x) = \{0\}$, but $\{0\} - (x, y, z, x) = \{0, 1\}$.

The main result of this paper relies on the following theorem which is a slight generalization of a result in [1].

Theorem 11. *[2] If* $\mathrm{CSP}(\Lambda)$ *has bounded width and* $\mathcal{J}$ *is a nontrivial weak Prague instance of* $\mathrm{CSP}(\Lambda)$*, then* $\mathcal{J}$ *has a solution (and a solution can be found in polynomial time).*

SDP and Prague instances

We now show that one can naturally associate a weak Prague instance to an output of the canonical SDP relaxation. This material will not be used in what follows, it is included to provide some intuition for the proof of the main theorem.

Let $\mathbf{x}_a$, $x \in V$, $a \in D$ be arbitrary vectors satisfying (SDP1), (SDP2) and (SDP3). (These vectors do not need to come as a result of the canonical SDP relaxation of a CSP instance.) We define a CSP instance $\mathcal{J}$ by

$$\mathcal{J} = (V, D, \{((x, y), P_{x,y}) : x, y \in V, x \neq y\}),$$
$$P_{x,y} = \{(a, b) : \mathbf{x}_a \mathbf{y}_b > 0\},$$

and we show that it is a weak Prague instance.

The instance is 1-minimal with $P_x^{\mathcal{J}} = \{a \in D : \mathbf{x}_a \neq \mathbf{0}\}$. To prove this it is enough to verify that the projection of $P_{x,y}$ to the first coordinate is equal to $P_x^{\mathcal{J}}$. If $(a, b) \in P_{x,y}$, then clearly $\mathbf{x}_a$ cannot be the zero vector, therefore $a \in P_x^{\mathcal{J}}$. On the other hand, if $a \in P_x^{\mathcal{J}}$ then $0 < ||\mathbf{x}_a||^2 = \mathbf{x}_a \mathbf{x}_D = \mathbf{x}_a \mathbf{y}_D$ and thus at least one of the dot products $\mathbf{x}_a \mathbf{y}_b$, $b \in D$ is nonzero and $(a, b) \in P_{x,y}$.

To check (P2) and (P3) we note that, for any $x, y \in V, x \neq y$ and $A \subseteq P_x^{\mathcal{J}}$, the vector $\mathbf{y}_{A+(x,y)}$ has either a strictly greater length than $\mathbf{x}_A$, or $\mathbf{x}_A = \mathbf{y}_{A+(x,y)}$, and the latter happens iff $A + (x, y, x) = A$ (see Claim 12.3, in fact, one can check that $\mathbf{y}_{A+(x,y)}$ is obtained by adding to $\mathbf{x}_A$ an orthogonal vector whose size is greater than zero iff $A + (x, y, x) \neq A$). By induction, for any pattern p from x to y, the vector $\mathbf{y}_{A+p}$ is either strictly longer than $\mathbf{x}_A$, or $\mathbf{x}_A = \mathbf{y}_{A+p}$ and $A + p - p = A$. Now (P2) follows immediately and (P3) is also easily seen: If $A + p + q = A$ then necessarily $\mathbf{x}_A = \mathbf{x}_{A+p}$ which is possible only if $A = A + p$.

We end this section with several remarks.

- To prove property (P2) we only need to consider the lengths of the vectors. In fact, this property will be satisfied when we start with the canonical linear programming relaxation (and define the instance $\mathcal{J}$ in a similar way). This is not the case for property (P3).

- The above weak Prague instance is in fact a Prague strategy in the sense of [1]. This means that every pair of variables is the scope of a (unique) constraint and all strong components of the digraph introduced after Definition 10 are complete graphs.

- There were attempts to show that the instance $\mathcal{J}$ satisfies a still stronger consistency property – it is a $(2, 3)$-strategy. A $(2, 3)$-strategy is a 1-minimal instance such that every pair of variables is the scope of a constraint, and $P_{x,y}$ is a subset of the composition of the relations $P_{x,z}$ and $P_{z,y}$ for every x, y, z. The following example shows that it is not the case. Consider $V = \{x, y, z\}$, $D = \{0, 1\}$ and vectors $\mathbf{x}_0 = (1/2, 1/2, 0)$, $\mathbf{x}_1 = (1/2, -1/2, 0)$, $\mathbf{y}_0 = (1/4, -1/4, \sqrt{2}/4)$, $\mathbf{y}_1 = (3/4, 1/4, -\sqrt{2}/4)$, $\mathbf{z}_0 = (1/4, 1/4, \sqrt{2}/4)$, $\mathbf{z}_1 = (3/4, -1/4, \sqrt{2}/4)$. The constraint relations are then $P_{x,y} = \{(0, 1), (1, 0), (1, 1)\} = P_{y,x}$, $P_{x,z} = \{(0, 0), (0, 1), (1, 1)\} = P_{z,x}^{-1}$, $P_{y,z} = \{(0, 0), (0, 1), (1, 0), (1, 1)\} = P_{z,y}$. The pair $(0, 0) \in P_{y,z}$ is not in the composition of the relations $P_{y,x}$ and $P_{x,z}$ since there is no $a \in \{0, 1\}$ such that $(0, a) \in P_{y,x}$ and $(a, 0) \in P_{x,z}$.

- Finally, we note that if $\mathcal{I}$ is an instance of the CSP with $\mathrm{SDPOpt}(\mathcal{I}) = 1$ and we define $\mathcal{J}$ using vectors with the sum (1) equal to 1, then a solution of $\mathcal{J}$ is necessarily a solution to $\mathcal{I}$. Showing that "$\mathrm{SDPOpt}(\mathcal{I}) = 1$" implies "$\mathcal{I}$ has a solution" was suggested as a first step to prove the Guruswami-Zhou conjecture. The above example explains that it is not straightforward to achieve this goal using $(2, 3)$-strategies.

4. PROOF

The main result, Theorem 4, is a consequence of the following theorem. The reduction, derandomization and omitted details are given in the appendix.

Theorem 12. *Let* Γ *be a core constraint language over* D *containing at most binary relations. If* $\mathrm{CSP}(\Gamma)$ *has bounded width, then there exists a randomized algorithm which given an instance* $\mathcal{I}$ *of* $\mathrm{CSP}(\Gamma)$ *and an output of the canonical SDP relaxation with value at least* $1 - 1/n^{4n}$ *(where* n *is a natural number) produces an assignment with value at least* $1 - K/n$*, where* K *is a constant depending on* $|D|$*. The running time is polynomial in* m *(the number of constraints) and* n^n*.*

Proof. Let $\mathcal{I} = (V, D, \mathcal{C})$ be an instance of $\mathrm{CSP}(\Gamma)$ with m constraints and let $\mathbf{x}_a$, $x \in V$, $a \in D$ be vectors satisfying (SDP1), (SDP2), (SDP3) such that the sum (1) is at least $1 - 1/n^{4n}$. Without loss of generality we assume that $n > |D|$.

Let us first briefly sketch the idea of the algorithm. The aim is to define an instance $\mathcal{J}$ in a similar way as in the previous section ($\mathcal{J}$ is defined after Claim 12.1), but instead of all pairs with nonzero weight we only include pairs of weight greater than a threshold (chosen in Step 1). This guarantees that every solution to $\mathcal{J}$ satisfies all the constraints of $\mathcal{I}$ which do not have large weight on pairs outside the constraint relation (the bad constraints are removed in Step 3). The instance $\mathcal{J}$ (more precisely, its algebraic closure) has a solution by Theorem 11 as soon as we ensure that it is a weak Prague instance. Property (P1) is dealt with in a similar way as in [19]: We keep only constraints with a gap – all pairs have either smaller weight than the threshold, or significantly larger (Step 2). This also gives a property similar to the one in the motivating discussion in the previous section: The vector $\mathbf{y}_{A+(x,y)}$ is either significantly longer than $\mathbf{x}_A$ or these vectors are almost the same. However, large amount of small differences can add up, so we need to continue taming the instance. In Steps 4 and 5 we divide the unit ball into layers and remove some constraints so that almost the same vectors of the form $\mathbf{x}_A$, $\mathbf{y}_{A+(x,y)}$ never lie in different layers. This already guarantees property (P2). For property (P3) we use "cutting by hyperplanes" idea from [11]. We choose sufficiently many hyperplanes so that every pair $\mathbf{x}_A$, $\mathbf{x}_B$ of different vectors in the same layer is cut (the bad variables are removed in Step 7) and we do not allow almost the same vectors for different variables to cross the hyperplane (Step 8).

The description of the algorithm follows.

1. Choose $r \in \{1, 2, \ldots, n-1\}$ uniformly at random.

2. Remove from $\mathcal{C}$ all the unary constraints (x, R) such that $||\mathbf{x}_a||^2 \in [n^{-4r-4}, n^{-4r})$ for some $a \in D$ and all the binary constraints $((x, y), R)$ such that $\mathbf{x}_a\mathbf{y}_b \in [n^{-4r-4}, n^{-4r})$ for some $a, b \in D$.

3. Remove from $\mathcal{C}$ all the unary constraints (x, R) such that $||\mathbf{x}_a||^2 \geq n^{-4r}$ for some $a \notin R$ and all the binary constraints $((x, y), R)$ such that $\mathbf{x}_a\mathbf{y}_b \geq n^{-4r}$ for some $(a, b) \notin R$.

Let

$$u_1 = 2|D|^2 n^{-4r-4} \quad \text{and} \quad u_2 = n^{-4r} - u_1.$$

For two real numbers $\gamma, \psi \neq 0$ we denote by $\gamma \div \psi$ the greatest integer i such that $\gamma - i\psi > 0$ and this difference is denoted by $\gamma \mod \psi$.

4. Choose $s \in [0, u_2]$ uniformly at random.

5. Remove from $\mathcal{C}$ all the binary constraints $((x, y), R)$ such that $|\,||\mathbf{x}_A||^2 - ||\mathbf{y}_B||^2\,| \leq u_1$ and $(||\mathbf{x}_A||^2 - s) \div u_2 \neq (||\mathbf{y}_B||^2 - s) \div u_2$ for some $A, B \subseteq D$.

The remaining part of the algorithm uses the following definitions. For all $x \in V$ let

$$P_x = \{a \in D : ||\mathbf{x}_a||^2 \geq n^{-4r}\}.$$

For a vector $\mathbf{w}$ we put

$$h(\mathbf{w}) = (||\mathbf{w}||^2 - s) \div u_2$$

and

$$t(\mathbf{w}) = \left\lceil \pi(\log n)n^{2r} \min\{\sqrt{(h(\mathbf{w}) + 2)u_2}, 1\} \right\rceil.$$

We say that $\mathbf{w}_1$ and $\mathbf{w}_2$ are *almost the same* if $h(\mathbf{w}_1) = h(\mathbf{w}_2)$ and $||\mathbf{w}_1 - \mathbf{w}_2||^2 \leq u_1$.

6. Choose unit vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{\lceil \pi(\log n)n^{2n} \rceil}$ independently and uniformly at random.
7. We say that a variable $x \in V$ is *uncut* if there exists $A, B \subseteq P_x$, $A \neq B$ such that $h(\mathbf{x}_A) = h(\mathbf{x}_B)$ and $\operatorname{sgn} \mathbf{x}_A\mathbf{q}_i = \operatorname{sgn} \mathbf{x}_B\mathbf{q}_i$ for every $1 \leq i \leq t(\mathbf{x}_A)$ (in words, no hyperplane determined by the first $t(\mathbf{x}_A) = t(\mathbf{x}_B)$ vectors $\mathbf{q}_i$ cuts the vectors $\mathbf{x}_A$, $\mathbf{x}_B$). Remove from $\mathcal{C}$ all the constraints whose scope contains an uncut variable.
8. Remove from $\mathcal{C}$ all the binary constraints $((x, y), R)$ for which there exist $A \subseteq P_x, B \subseteq P_y$ such that $\mathbf{x}_A$, $\mathbf{y}_B$ are almost the same and $\operatorname{sgn} \mathbf{x}_A\mathbf{q}_i \neq \operatorname{sgn} \mathbf{y}_B\mathbf{q}_i$ for some $1 \leq i \leq t(\mathbf{x}_A)$.
9. Return a solution of $\mathcal{I}$.

CLAIM 12.1. *Expected fraction of constraints removed in steps 2, 3, 5, 7 and 8 is at most K/n for some constant K.*

Remark. The constant K depends exponentially on the size of the domain $|D|$.

PROOF. **Step 2.** For each binary constraint there are $|D|^2$ choices for $a, b \in D$ and therefore at most $|D|^2$ bad choices for r. For a unary constraint the number of bad choices is at most $|D|$. Thus the probability that a given constraint will be removed is at most $|D|^2/(n-1)$ and it follows that the expected fraction of removed constraints is at most $|D|^2/(n-1)$.

Step 3. The contribution of every removed constraint to the sum (1) is at most $1 - n^{-4r} \leq 1 - n^{-4n+4}$. If more than γ-fraction of the constraints is removed than the sum is at most $1/m((1-\gamma)m + \gamma m(1 - n^{-4n+4})) = 1 - \gamma n^{-4n+4}$. Since (1) $\geq 1 - 1/n^{4n}$, we have $\gamma \leq 1/n^4$.

Step 5. For every constraint $((x, y), R)$ and every $A, B \subseteq D$ such that $|\,||\mathbf{x}_A||^2 - ||\mathbf{y}_B||^2\,| \leq u_1$, $||\mathbf{x}_A|| \leq ||\mathbf{y}_B||$, the inequality $(||\mathbf{x}_A||^2 - s) \div u_2 < (||\mathbf{y}_B||^2 - s) \div u_2$ can be satisfied only if $(||\mathbf{y}_B||^2 - s) \mod u_2 < u_1$. The bad choices for s thus cover at most (u_1/u_2)-fraction of the interval $[0, u_2]$. As $u_1/u_2 < K_1/n^4$ (for a suitable constant K_1 depending on $|D|$), the probability of a bad choice is at most K_1/n^4. There are $4^{|D|}$ pairs of subsets $A, B \subseteq D$, therefore the probability that the constraint is removed is less than $K_1 4^{|D|}/n^4$ and so is the expected fraction of removed constraints.

Before analyzing Steps 7 and 8 let us observe that, for any vector $\mathbf{w}$ such that $1 \geq ||\mathbf{w}|| \geq n^{-4r}$,

$$\pi(\log n)n^{2r}\,||\mathbf{w}|| \leq t(\mathbf{w}) \leq 2\pi(\log n)n^{2r}\,||\mathbf{w}|| + 1.$$

The first inequality follows from

$$\sqrt{(h(\mathbf{w}) + 2)u_2} = \sqrt{u_2((||\mathbf{w}||^2 + 2u_2 - s) \div u_2)} \geq$$

$$\geq \sqrt{u_2 \frac{||\mathbf{w}||^2 + u_2 - s}{u_2}} \geq ||\mathbf{w}||$$

and the second inequality follows from

$$\sqrt{(h(\mathbf{w}) + 2)u_2} \leq \sqrt{u_2 \frac{(||\mathbf{w}||^2 + 2u_2 - s)}{u_2}} \leq$$

$$\leq \sqrt{||\mathbf{w}||^2 + 2u_2} \leq \sqrt{||\mathbf{w}||^2 + 2\,||\mathbf{w}||^2} < 2\,||\mathbf{w}||\,.$$

Step 7. Consider two different subsets A, B of P_x such that $h(\mathbf{x}_A) = h(\mathbf{x}_B)$. Suppose that $A \setminus B \neq \emptyset$, the other case is symmetric. Let θ be the angle between $\mathbf{x}_A$ and $\mathbf{x}_B$. As $\mathbf{x}_A - \mathbf{x}_{A\cap B} (= \mathbf{x}_{A\setminus B})$, $\mathbf{x}_B - \mathbf{x}_{A\cap B}$ and $\mathbf{x}_{A\cap B}$ are pairwise orthogonal, the angle θ is greater than or equal to the angle θ_A between $\mathbf{x}_A$ and $\mathbf{x}_{A\cap B}$. (Given three pairwise orthogonal vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, the angle between $\mathbf{v}_1 + \mathbf{v}_2$ and $\mathbf{v}_1 + \mathbf{v}_3$ is always greater than or equal to the angle between $\mathbf{v}_1 + \mathbf{v}_2$ and $\mathbf{v}_1$. This is a straightforward calculation using, for instance, dot products. In our situation $\mathbf{v}_1 = \mathbf{x}_{A\cap B}$, $\mathbf{v}_2 = \mathbf{x}_{A\setminus B}$ and $\mathbf{v}_3 = \mathbf{x}_{B\setminus A}$.) We have $\sin\theta_A = \left|\left|\mathbf{x}_{A\setminus B}\right|\right| / \left|\left|\mathbf{x}_A\right|\right|$. Since

$A \subseteq P_x$, we get $\left|\left|\mathbf{x}_{A\setminus B}\right|\right| \geq \sqrt{n^{-4r}} = n^{-2r}$ and then $\sin\theta_A = \left|\left|\mathbf{x}_{A\setminus B}\right|\right| / \left|\left|\mathbf{x}_A\right|\right| \geq n^{-2r}/\left|\left|\mathbf{x}_A\right|\right|$, so $\theta \geq \theta_A \geq n^{-2r}/\left|\left|\mathbf{x}_A\right|\right|$.

The probability that $\mathbf{q}_i$ does not cut $\mathbf{x}_A$ and $\mathbf{x}_B$ is thus at most $1 - n^{-2r}/\pi\left|\left|\mathbf{x}_A\right|\right|$ and the probability that none of the vectors $\mathbf{q}_1, \ldots, \mathbf{q}_{t(\mathbf{x}_A)}$ cut them is at most

$$\left(1 - \frac{n^{-2r}}{\pi\left|\left|\mathbf{x}_A\right|\right|}\right)^{t(\mathbf{x}_A)} \leq \left[\left(1 - \frac{1}{\pi n^{2r}\left|\left|\mathbf{x}_A\right|\right|}\right)^{\pi n^{2r}\left|\left|\mathbf{x}_A\right|\right|}\right]^{\log n} \leq$$

$$\leq \left(\frac{1}{2}\right)^{\log n} = \frac{1}{n}.$$

The first inequality uses that $t(\mathbf{x}_A) \geq (\log n)n^{2r}\left|\left|\mathbf{x}_A\right|\right|$ which we observed above. In the second inequality we have used that $(1-1/\eta)^{\eta} \leq 1/2$ whenever $\eta \geq 2$.

For a single variable there are at most $4^{|D|}$ choices for $A, B \subseteq P_x$, therefore the probability that x is uncut is at most $4^{|D|}/n$. The scope of every constraint contains at most 2 variables, hence the probability that a constraint is removed is at most $2\cdot 4^{|D|}/n$ and the expected fraction of the constraints removed in this step has the same upper bound.

Step 8. Assume that $((x,y),R)$ is a binary constraint and $A \subseteq P_x, B \subseteq P_y$ are such that $\mathbf{x}_A$ and $\mathbf{y}_B$ are almost the same. Let θ be the angle between $\mathbf{x}_A$ and $\mathbf{y}_B$ and θ_A be the angle between $\mathbf{y}_B$ and $\mathbf{y}_B - \mathbf{x}_A$. By the law of sines we have $\left|\left|\mathbf{x}_A\right|\right|/(\sin\theta_A) = \left|\left|\mathbf{y}_B - \mathbf{x}_A\right|\right|/(\sin(\theta)$, and

$$\theta \leq 2\sin\theta = \frac{2\left|\left|\mathbf{y}_B - \mathbf{x}_A\right|\right|}{\left|\left|\mathbf{x}_A\right|\right|}\sin(\theta_A) \leq \frac{2\left|\left|\mathbf{y}_B - \mathbf{x}_A\right|\right|}{\left|\left|\mathbf{x}_A\right|\right|} \leq \frac{2\sqrt{u_1}}{\left|\left|\mathbf{x}_A\right|\right|},$$

where the first inequality follows from $\theta \leq \pi/2$ (the difference of $\mathbf{x}_A$ and $\mathbf{y}_B$ has length at most $\sqrt{u_1}$ while both vectors have length at least $n^{-2r} > \sqrt{u_1}$). Therefore, the probability that vectors $\mathbf{x}_A$ and $\mathbf{y}_B$ are cut by some of the vectors $\mathbf{q}_i$, $1 \leq i \leq t(\mathbf{x}_A)$ is at most

$$t(\mathbf{x}_A)\frac{2\sqrt{u_1}}{\left|\left|\mathbf{x}_A\right|\right|} \leq (2\pi(\log n)n^{2r}\left|\left|\mathbf{x}_A\right|\right| + 1)\frac{2\sqrt{2|D|^2 n^{-4r-4}}}{\left|\left|\mathbf{x}_A\right|\right|} \leq$$

$$\leq K_2(\log n)n^{-2} \leq \frac{K_2}{n},$$

where K_2 is a constant. There are at most $4^{|D|}$ choices for A, B, so the probability that our constraint will be removed is less than $K_2 4^{|D|}/n$. ☐

Now we define the instance $\mathcal{J}$ and proceed to show that $\mathcal{J}$ is a weak Prague instance. Let $\mathcal{S}$ denote the set of pairs which are the scope of some binary constraint of $\mathcal{I}$ after Step 8 and let $\mathcal{S}^{-1} = \{(x,y) : (y,x) \in \mathcal{S}\}$. We put

$$\mathcal{J} = (V, D, \{((x,y), P^{\mathcal{J}}_{x,y}) : (x,y) \in \mathcal{S}\cup\mathcal{S}^{-1}\}),$$
$$P^{\mathcal{J}}_{x,y} = \{(a,b) : \mathbf{x}_a\mathbf{y}_b \geq n^{-4r}\}.$$

Claim 12.2. *The instance $\mathcal{J}$ is 1-minimal and $P^{\mathcal{J}}_x = P_x$.*

Proof. Let $(x,y) \in \mathcal{S}$ and take an arbitrary constraint $((x,y),R)$ which remained in $\mathcal{C}$.

First we prove that $P_{x,y} \subseteq P_x \times P_y$ for every $a, b \in D$. Indeed, if $(a,b) \in P_{x,y}$ then $\mathbf{x}_a\mathbf{y}_b \geq n^{-4r}$, therefore $\left|\left|\mathbf{x}_a\right|\right|^2 = \mathbf{x}_a\mathbf{x}_D = \mathbf{x}_a\mathbf{y}_D \geq n^{-4r}$, so $a \in P_x$. Similarly, $b \in P_y$.

On the other hand, if $a \in P_x$ then $n^{-4r} \leq \left|\left|\mathbf{x}_a\right|\right|^2 = \mathbf{x}_a\mathbf{y}_D$, thus there exist $b \in D$ such that $\mathbf{x}_a\mathbf{y}_b \geq n^{-4r}/|D| \geq n^{-4r-4}$ (we have used $n^4 \geq |D|$). But then $\mathbf{x}_a\mathbf{y}_b \geq n^{-4r}$, otherwise the constraint $((x,y),R)$ would be removed in Step 2. This implies that $(a,b) \in P_{x,y}$. We have shown that the projection of $P_{x,y}$ to the first coordinate contains P_x. Similarly, the second projection contains P_y, so $P_{x,y}$ is subdirect in $P_x \times P_y$. ☐

For verification of properties (P2) and (P3) the following observation will be useful.

Claim 12.3. *Let $(x,y) \in \mathcal{S}\cup\mathcal{S}^{-1}$, $A \subseteq P_x$, $B = A + (x,y)$. If $A = B + (y,x)$, then the vectors $\mathbf{x}_A$ and $\mathbf{y}_B$ are almost the same. In the other case, i.e. if $A \subsetneq B + (y,x)$, then $h(\mathbf{y}_B) > h(\mathbf{x}_A)$.*

Proof. The number $\left|\left|\mathbf{y}_B - \mathbf{x}_A\right|\right|^2$ is equal to

$$\mathbf{y}_B\mathbf{y}_B - \mathbf{x}_A\mathbf{y}_B - \mathbf{x}_A\mathbf{y}_B + \mathbf{x}_A\mathbf{x}_A =$$

$$= \mathbf{x}_D\mathbf{y}_B - \mathbf{x}_A\mathbf{y}_B - \mathbf{x}_A\mathbf{y}_B + \mathbf{x}_A\mathbf{y}_D = \mathbf{x}_{D\setminus A}\mathbf{y}_B + \mathbf{x}_A\mathbf{y}_{D\setminus B}.$$

No pair (a,b), with $a \in A$ and $b \in D\setminus B$, is in $P^{\mathcal{J}}_{x,y}$ so the dot product $\mathbf{x}_a\mathbf{y}_b$ is smaller than n^{-4r}. Then in fact $\mathbf{x}_a\mathbf{y}_b < n^{-4r-4}$ otherwise all the constraints with scope (x,y) would be removed in Step 2. It follows that the second summand is always at most $|D|^2 n^{-4r-4}$ and the first summand has the same upper bound in the case $B + (y,x) = A$.

Moreover, $\left|\left|\mathbf{y}_B\right|\right|^2 - \left|\left|\mathbf{x}_A\right|\right|^2$ is equal to

$$\mathbf{y}_B\mathbf{y}_B - \mathbf{x}_A\mathbf{x}_A = \mathbf{x}_D\mathbf{y}_B - \mathbf{x}_A\mathbf{y}_D =$$

$$= \mathbf{x}_D\mathbf{y}_B - \mathbf{x}_A\mathbf{y}_B - \mathbf{x}_A\mathbf{y}_{D\setminus B} = \mathbf{x}_{D\setminus A}\mathbf{y}_B - \mathbf{x}_A\mathbf{y}_{D\setminus B}.$$

If $B + (y,x) = A$ then we have a difference of two nonnegative numbers less than or equal $|D|^2 n^{-4r-4}$, therefore the absolute value of this expression is at most u_1. But then $h(\mathbf{x}_A) = h(\mathbf{y}_B)$, otherwise all constraint with scope (x,y) or (y,x) would be removed in Step 5. Using the previous paragraph, it follows that $\mathbf{x}_A$ and $\mathbf{y}_B$ are almost the same.

If $B + (y,x)$ properly contains A then the first summand $\mathbf{x}_{D\setminus A}\mathbf{y}_B$ is greater than or equal to n^{-4r}, so the whole expression is at least $n^{-4r} - |D|^2 n^{-4r-4} > u_2$ and thus $h(\mathbf{y}_B) > h(\mathbf{x}_A)$. ☐

Claim 12.4. *$\mathcal{J}$ is a weak Prague instance.*

Proof. **(P2).** Let $A \subseteq P_x$ and let $p = (x_1, \ldots, x_i)$ be a pattern in $\mathcal{J}$ from x to x (i.e. $x_1 = x_i = x$). By the previous claim $h(\mathbf{x}_A) = h((\mathbf{x}_i)_{A+(x_1,\ldots,x_i)}) \geq h((\mathbf{x}_{i-1})_{A+(x_1,\ldots,x_{i-1})}) \geq \cdots \geq h((\mathbf{x}_2)_{A+(x_1,x_2)}) \geq h(\mathbf{x}_A)$. It follows that all these inequalities must in fact be equalities and, by applying the claim again, we get that the vectors $(\mathbf{x}_j)_{A+(x_1,x_2,\ldots,x_j)}$ and $(\mathbf{x}_{j+1})_{A+(x_1,x_2,\ldots,x_{j+1})}$ are almost the same and, moreover, $A + (x_1, x_2, \ldots, x_{j+1}) + (x_{j+1}, x_j) = A + (x_1, x_2, \ldots, x_j)$ for every $1 \leq j < i$. Therefore $A + p - p = A$ as required.

(P3). Let $A \subseteq P_x$, let $p_1 = (x_1, \ldots, x_i), p_2$ be two patterns from x to x such that $A + p_1 + p_2 = A$ and let $B = A + p_1$. For contradiction assume $A \neq B$. The same argument as above proves that the vectors $(\mathbf{x}_j)_{A+(x_1,x_2,\ldots,x_j)}$ and $(\mathbf{x}_{j+1})_{A+(x_1,x_2,\ldots,x_{j+1})}$ are almost the same for every $1 \leq j < i$, and then $h(\mathbf{x}_A) = h(\mathbf{x}_B)$. There exists $k \leq t(\mathbf{x}_A)$ such that $\operatorname{sgn}\mathbf{x}_A\mathbf{q}_k \neq \operatorname{sgn}\mathbf{x}_B\mathbf{q}_k$, otherwise x is uncut and all constraints whose scope contains x would be removed in Step 7. But this leads to a contradiction, since $\operatorname{sgn}(\mathbf{x}_j)_{A+(x_1,\ldots,x_j)}\mathbf{q}_k = \operatorname{sgn}(\mathbf{x}_{j+1})_{A+(x_1,\ldots,x_{j+1})}\mathbf{q}_k$ for all $1 \leq j < i$, otherwise the constraints with scope (x_j, x_{j+1}) would be removed in Step 8. ☐

Observe that every solution F to $\mathcal{J}$ satisfies all the constraints which remained in $\mathcal{I}$ after Step 8: For every unary constraint (x, R) we have $P_x \subseteq R$ (from Step 3) and for every binary constraint $((x, y), R)$ we have $P_{x,y} \subseteq R$. Since we have removed at most (K/n)-fraction of the constraints from $\mathcal{C}$, the mapping F is an assignment for the original instance $\mathcal{I}$ of value at least $1 - K/n$. Also, the instance $\mathcal{J}$ is nontrivial because, for each $x \in V$, there exists at least one $a \in D$ with $||\mathbf{x}_a||^2 > 1/n^4$ (recall that we assume $n > |D|$).

The only problem is that the CSP over the constraint language of $\mathcal{J}$ (consisting of $P_{x,y}^{\mathcal{J}}$'s) does not necessarily have bounded width. This is why we form the algebraic closure $\mathcal{J}'$ of $\mathcal{J}$:

$$\mathcal{J}' = (V, D, \{((x, y), P_{x,y}^{\mathcal{J}'}) : (x, y) \in \mathcal{S} \cup \mathcal{S}^{-1}\}),$$
$$P_{x,y}^{\mathcal{J}'} = \{(f(a_1, a_2, \dots), f(b_1, b_2, \dots)) : f \in \mathrm{Pol}(\Gamma),$$
$$(a_1, b_1), (a_2, b_2), \dots \in P_{x,y}^{\mathcal{J}}\}$$

The new instance still has the property that $P_x^{\mathcal{J}'}$ (which is equal to $\{f(a_1, a_2, \dots) : f \in \mathrm{Pol}(\Gamma), a_1, a_2, \dots \in P_x\}$) is a subset of R for every unary constraint (x, R), and $P_{x,y}^{\mathcal{J}'} \subseteq R$ for every binary constraint $((x, y), R)$, since the constraint relations are preserved by every polymorphism of Γ. Moreover, every polymorphism of Γ is a polymorphism of the constraint language Λ' of $\mathcal{J}'$, therefore $\mathrm{CSP}(\Lambda')$ has bounded width (see Theorem 7 for instance; technically, Λ' does not need to be a core, but we can simply add all the singleton unary relations).

CLAIM 12.5. *The instance $\mathcal{J}'$ is a weak Prague instance.*

PROOF. See Proposition 16 in the appendix. □

Therefore $\mathcal{J}'$ (and thus $\mathcal{I}$ after Step 8) has a solution by Theorem 11. This concludes the proof.

5. OPEN PROBLEMS

The quantitative dependence of g on ε is not very far from the (UGC-) optimal bound for `Horn-k-Sat`. Is it possible to get rid of the extra $\log\log(1/\varepsilon)$?

A straightforward derandomization using a result from [15] has $g(\varepsilon) = O(\log\log(1/\varepsilon)/\sqrt{\log(1/\varepsilon)})$. How to improve it to match the randomized version?

It was observed by Andrei Krokhin that the quantitative dependence is, at least to a large extent, also controlled by the polymorphisms of the constraint language. The problems `2-SAT`, `Unique-Games(q)` suggest that majority or, more generally, near-unanimity polymorphisms could be responsible for polynomial behavior.

The simplest example of polymorphism which does not imply any known stronger property for decision CSPs other than bounded width is the 2-semilattice operation f on a three element domain $D = \{0, 1, 2\}$ defined by $f(0, 0) = f(0, 1) = f(1, 0) = 0$, $f(1, 1) = f(1, 2) = f(2, 1) = 1$, $f(2, 2) = f(2, 0), f(0, 2) = 2$. This might be a source for possible hardness results.

Finally, we believe that the connection between SDP, LP and consistency notions deserves further investigation.

6. ACKNOWLEDGMENTS

The first author was supported by the Grant Agency of the Czech Republic, grant 201/09/P223 and by the Ministry of Education of the Czech Republic, grant MSM 0021620839.

The second author was supported by the Foundation for Polish Science,grant HOM/2008/7 (supported by MF EOG), and Ministry of Science and Higher Education of Poland, grant N206 357036.

7. REFERENCES

[1] L. Barto and M. Kozik. Constraint satisfaction problems of bounded width. In *FOCS'09: Proceedings of the 50th Symposium on Foundations of Computer Science*, pages 595–603, 2009.

[2] L. Barto and M. Kozik. Constraint satisfaction problems solvable by local consistency methods. 2012. submitted.

[3] A. Bulatov. Bounded relational width. 2009. manuscript.

[4] A. Bulatov, P. Jeavons, and A. Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM J. Comput.*, 34:720–742, March 2005.

[5] A. A. Bulatov, A. Krokhin, and B. Larose. Complexity of constraints. chapter Dualities for Constraint Satisfaction Problems, pages 93–124. Springer-Verlag, Berlin, Heidelberg, 2008.

[6] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for unique games. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 205–214, New York, NY, USA, 2006. ACM.

[7] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. *ACM Trans. Algorithms*, 5:32:1–32:14, July 2009.

[8] V. Dalmau and A. Krokhin. Robust satisfiability for CSPs: algorithmic and hardness results. 2011. in preparation.

[9] T. Feder and M. Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through datalog and group theory. *SIAM J. Comput.*, 28:57–104, February 1999.

[10] R. Freese, M. Kozik, A. Krokhin, M. Maróti, R. McKenzie, and R. Willard. On Maltsev conditions associated with omitting certain types of local structures. 2011. in preparation.

[11] M. X. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.

[12] V. Guruswami and Y. Zhou. Tight bounds on the approximability of almost-satisfiable Horn SAT and exact hitting set. In D. Randall, editor, *SODA*, pages 1574–1589. SIAM, 2011.

[13] J. Håstad. Some optimal inapproximability results. *J. ACM*, 48:798–859, July 2001.

[14] P. Jonsson, A. Krokhin, and F. Kuivinen. Hard constraint satisfaction problems have hard gaps at location 1. *Theor. Comput. Sci.*, 410:3856–3874, September 2009.

[15] Z. S. Karnin, Y. Rabani, and A. Shpilka. Explicit dimension reduction and its applications. *Electronic Colloquium on Computational Complexity (ECCC)*, 16:121, 2009.

[16] S. Khanna, M. Sudan, L. Trevisan, and D. P. Williamson. The approximability of constraint satisfaction problems. *SIAM J. Comput.*, 30(6):1863–1920, 2000.

[17] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775. ACM Press, 2002.

[18] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37:319–357, April 2007.

[19] G. Kun, R. O'Donnell, S. Tamaki, Y. Yoshida, and Y. Zhou. Linear programming, width-1 CSPs and robust satisfaction. 2011. manuscript.

[20] B. Larose and P. Tesson. Universal algebra and hardness results for constraint satisfaction problems. *Theor. Comput. Sci.*, 410:1629–1647, April 2009.

[21] B. Larose and L. Zádori. Bounded width problems and algebras. *Algebra Universalis*, 56(3-4):439–466, 2007.

[22] P. Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *STOC'08*, pages 245–254, 2008.

[23] T. J. Schaefer. The complexity of satisfiability problems. In *Conference Record of the Tenth Annual ACM Symposium on Theory of Computing (San Diego, Calif., 1978)*, pages 216–226. ACM, New York, 1978.

[24] U. Zwick. Finding almost-satisfying assignments. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, STOC '98, pages 551–560, New York, NY, USA, 1998. ACM.

APPENDIX

A. REDUCTION TO CORE CONSTRAINT LANGUAGES WITH UNARY AND BINARY RELATIONS

The reduction is given in the following proposition.

PROPOSITION 13. *Let Γ be a constraint language on the domain D which contains relations of maximum arity l and such that* $\mathrm{CSP}(\Gamma)$ *has bounded width. Then there exists a core constraint language Γ' on D' containing only at most binary relations such that* $\mathrm{CSP}(\Gamma')$ *has bounded width and such that the following holds: If* $\mathrm{CSP}(\Gamma')$ *admits a robust satisfiability algorithm which is $(1-g(\varepsilon), 1-\varepsilon)$-approximating (for every ε), then* $\mathrm{CSP}(\Gamma)$ *admits a robust satisfiability algorithm which is $(1-(l+1)g(\varepsilon), 1-\varepsilon)$-approximating.*

PROOF. First we form the core of Γ: We take a unary polymorphism $f \in \mathrm{Pol}(\Gamma)$ with minimal image (with respect to inclusion) and put $\Gamma^c = \{R^c = R \cap f(D)^{\mathrm{arity}(R)} : R \in \Gamma\}$, $D^c = f(D)$. Then Γ^c is a core constraint language. It is known that $\mathrm{CSP}(\Gamma)$ has bounded width iff $\mathrm{CSP}(\Gamma^c)$ does (see [21]), therefore $\mathrm{CSP}(\Gamma^c)$ has bounded width.

Next we define the constraint language Γ'. The domain is $D' = (D^c)^l$. For every relation $R^c \in \Gamma^c$ of arity k we add to Γ' the unary relation R' defined by

$$(a_1, \ldots, a_l) \in R' \quad \text{iff} \quad (a_1, \ldots, a_k) \in R^c,$$

for every $k \leq l$ we add the binary relation

$$E_k = \{((a_1, \ldots, a_l), (b_1, \ldots, b_l)) : a_1 = b_k\},$$

and for every $(a_1, \ldots, a_l) \in D'$ we add the singleton unary relation $\{(a_1, \ldots, a_l)\}$. The singletons ensure that Γ' is a core. That $\mathrm{CSP}(\Gamma')$ has bounded width can be seen, for instance, from Theorem 7: If f_1^c, f_2^c are polymorphisms of Γ^c from this theorem, then the corresponding operations f_1', f_2' acting coordinate-wise on D' satisfy the same equations and it is straightforward to check that f_1', f_2' are polymorphisms of Γ'.

Now, let $\mathcal{I} = (V, D, \mathcal{C})$ be an instance of $\mathrm{CSP}(\Gamma)$ with $\mathrm{Opt}(\mathcal{I}) = \varepsilon$. We transform $\mathcal{I}$ to an instance $\mathcal{I}'$ of $\mathrm{CSP}(\Gamma')$ as follows. We keep the original variables and for every constraint $C = ((x_1, \ldots, x_k), R)$ in $\mathcal{C}$ we introduce a new variable x_C and add $k+1$ constraints

$$((x_C), R'), ((x_1, x_C), E_1), ((x_2, x_C), E_2), \ldots, ((x_k, x_C), E_k). \tag{2}$$

If $F : V \to D$ is an assignment for $\mathcal{I}$ of value $1-\varepsilon$ then $F^c = fF$ has at least the same value (as f preserves the constraint relations), and the assignment F' for $\mathcal{I}'$ defined by

$$\begin{aligned} F'(x) &= (F^c(x), ?, \ldots, ?) \quad \text{for } x \in V, \\ F'(x_C) &= (F^c(x_1), \ldots, F^c(x_k), ?, \ldots, ?) \\ &\quad \text{for } C = ((x_1, \ldots, x_k), R) \end{aligned}$$

(where ? stands for an arbitrary element of A) has value at least $1-\varepsilon$, since all the binary constraints in $\mathcal{I}'$ are satisfied, and the constraint (x_C, R') is satisfied whenever F satisfies C.

We run the robust algorithm for $\mathrm{CSP}(\Gamma')$ to get an assignment G' for $\mathcal{I}'$ with value at least $1 - g(\varepsilon)$, and we define $G(x)$, $x \in V$ to be the first coordinate of $G'(x)$. Note that, for any constraint C of $\mathcal{I}$, if G' satisfies all the constraints (2) then G satisfies C. Therefore the value of G is at least $1-(l+1)g(\varepsilon)$. $\square$

B. PROOF OF THEOREM 4 USING THEOREM 12

Let Γ be a core constraint language with at most binary relations (which we can assume by Proposition 13) such that $\mathrm{CSP}(\Gamma)$ has bounded width. Let $\mathcal{I}$ be an instance of $\mathrm{CSP}(\Gamma)$ with m constraints and let $1-\varepsilon = \mathrm{Opt}(\mathcal{I})$.

We first check whether $\mathcal{I}$ has a solution. This can be done in polynomial time since $\mathrm{CSP}(\Gamma)$ has bounded width. If a solution exists we can find it in polynomial time (see the note after Definition 1).

In the other case we know that $\varepsilon \geq 1/m$. We run the SDP relaxation with precision $\delta = 1/m$ and obtain vectors with the sum (1) equal to $v \geq \mathrm{SDPOpt}(\mathcal{I}) - 1/m$. Finally, we execute the algorithm provided in Theorem 12 with the following choice of n.

$$n = \left\lfloor \frac{\log \omega}{4 \log \log \omega} \right\rfloor, \qquad \text{where } \omega = \min\left\{\frac{1}{1-v}, m\right\}.$$

The assumption is satisfied, because $v \geq 1 - 1/n^{4n}$ is equivalent to $n^{4n} \leq 1/(1-v)$ and

$$n^{4n} = 2^{4n \log n} \leq 2^{4 \frac{\log \omega}{4 \log \log \omega} \log \frac{\log \omega}{4 \log \log \omega}} < 2^{\frac{\log \omega}{\log \log \omega} \log \log \omega} = \\ = \omega \leq \frac{1}{1-v}.$$

The algorithm runs in time polynomial in m as $n^n < n^{4n} \leq \omega \leq m$. To estimate the fraction of satisfied constraints, observe that $v \geq \mathrm{Opt}(\mathcal{I}) - 1/m = 1 - \varepsilon - 1/m \geq 1 - 2\varepsilon$, so $1/(1-v) \geq 1/2\varepsilon$, and also $m \geq 1/\varepsilon$, therefore $\omega \geq 1/2\varepsilon$. The fraction of satisfied constraints is at least $1 - K/n$ and

$$\frac{n}{K} \geq \frac{1}{K}\left(\frac{\log \omega}{4 \log\log \omega} - 1\right) \geq K_3 \frac{\log(1/2\varepsilon)}{\log\log(1/2\varepsilon)} \geq$$
$$\geq K_4 \frac{\log(1/\varepsilon)}{\log\log(1/\varepsilon)},$$

where K_3, K_4 are suitable constants. Therefore the fraction of satisfied constraints is at least

$$1 - O\left(\frac{\log\log(1/\varepsilon)}{\log(1/\varepsilon)}\right).$$

C. DERANDOMIZATION

We start by describing the changes in Theorem 12. The statement remains the same except the algorithm will be polynomial in m and $2^{n^2 \log^2 n}$.

The random choices in Step 1 and Step 4 can be easily avoided: In Step 1 we can try all $(n-1)$ possible choices for r and in Step 4 we can try all choices for s from some sufficiently dense finite set, for instance $\{0, u_2/n^4, 2u_2/n^4, \dots\}$. The only difference is that bad choices for s could cover a slightly bigger part of the interval than u_1/u_2 and we would get a slightly worse constant K_1.

For derandomization of Step 6 we first slightly change the constant in the definition of $t(\mathbf{w})$, say $t(\mathbf{w}) = \lceil 4(\log n) \dots \rceil$. Next we use Theorem 1.3. from [15] from which it follows that we can efficiently find a set Q of unit vectors such that

$$|Q| = (|V||D|)^{1+o(1)} 2^{O(\log^2(1/\kappa))}$$

and such that, for any vectors $\mathbf{v}$, $\mathbf{w}$ with angle θ between them, the probability that a randomly chosen vector from Q cuts $\mathbf{v}$ and $\mathbf{w}$ differs from θ/π by at most κ. We choose $\kappa = 1/n^{2n} = 1/2^{2n \log n}$, therefore

$$|Q| \leq K_5 m^{K_6}\ 2^{n^2 \log^2 n},$$

where we have used $|V| = O(m)$ which is true whenever every variable is in the scope of some constraint (we can clearly assume this without loss of generality).

Now if we choose $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_{\lceil 4(\log n) n^{2n} \rceil}$ uniformly at random from Q, the estimates derived in Steps 7 and 8 remain almost unchanged: The probability that $\mathbf{q}_i$ does not cut $\mathbf{x}_A$ and $\mathbf{x}_B$ in Step 7 is at most $1 - n^{-2r}/\pi\, ||\mathbf{x}_A|| + \kappa \leq 1 - n^{-2r}/4\, ||\mathbf{x}_A||$ (for a sufficiently large n), and the probability that vectors $\mathbf{x}_A$ and $\mathbf{y}_B$ are cut by some $\mathbf{q}_i$ in Step 8 is at most K_2'/n (for any $K_2' > K_2$).

Of course we cannot try all possible $\lceil 4(\log n) n^{2n} \rceil$-tuples of vectors from Q as there are too many. However, we can apply the method of conditional expectations – we choose the vectors one by one keeping an estimate of the expected number of constraints removed below K/n.

Finally, the proof of the deterministic version of Theorem 4 remains almost the same except we need to ensure that $2^{n^2 \log^2 n}$ is polynomial in m. Therefore we need to choose a smaller value for n, say

$$n = \left\lfloor \frac{\sqrt{\log \omega}}{\log\log \omega} \right\rfloor,$$

then the algorithm outputs an assignment satisfying at least $\left(1 - O\left(\frac{\log\log(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}}\right)\right)$-fraction of the constraints.

D. ALGEBRAIC CLOSURE OF A WEAK PRAGUE INSTANCE

Proposition 16 below justifies Claim 12.5. But first we collect some useful facts about Prague instances.

It will be convenient to replace (P2) with an alternative condition:

LEMMA 14. *Let $\mathcal{J}$ be a 1-minimal instance. Then (P2) is equivalent to the following condition.*

(P2) For every step (x, y), every $A \subseteq P_x$ and every pattern p from y to x, if $A + (x, y) + p = A$ then $A + (x, y, x) = A$.*

PROOF. (P2*) ⇒ (P2). If $p = (x = x_1, x_2, \dots, x_k = x)$ is a pattern from x to x such that $A + p = A$, then repeated application of (P2*) gives us

$$\begin{aligned}
A{+}p - p &= \\
&= [A + (x_1, x_2, \dots, x_{k-1})] + (x_{k-1}, x_k, x_{k-1}) \\
&\quad + (x_{k-1}, x_{k-2}, \dots, x_1) \\
&= A + (x_1, x_2, \dots, x_{k-1}) + (x_{k-1}, x_{k-2}, \dots, x_1) \\
&= [A + (x_1, x_2, \dots, x_{k-2})] + (x_{k-2}, x_{k-1}, x_{k-2}) \\
&\quad + (x_{k-2}, x_{k-3}, \dots, x_1) \\
&= A + (x_1, x_2, \dots, x_{k-2}) + (x_{k-2}, x_{k-3}, \dots x_1) \\
&= \dots \\
&= A,
\end{aligned}$$

where the second equality uses (P2*) for the set $A + (x_1, x_2, \dots, x_{k-1})$. The assumption of (P2*) is provided by a cyclic shift of the pattern p. The fourth equality uses (P2*) for the set $A + (x_1, \dots, x_{k-2})$ and so on.

(P2) ⇒ (P2*). By applying (P2) to the pattern $(x, y) + p$ we get $A + (x, y) + p - p + (y, x) = A$. From 1-minimality it follows that $A{+}(x, y) \subseteq A{+}(x, y){+}p{-}p$, hence $A{+}(x, y, x) = (A + (x, y)) + (y, x) \subseteq (A + (x, y) + p - p) + (y, x) = A$. The other inclusion follows again from 1-minimality. □

The next lemma shows that when we start with an element and keep adding a pattern from x to x, the process will stabilize.

LEMMA 15. *Let $\mathcal{J}$ be a weak Prague instance, $x \in V$, $a \in P_x$, and let p be a pattern from x to x. Then there exists a natural number l such that the set $[a]_p := \{a\} + lp$ satisfies $[a]_p + p = [a]_p$ and $a \in [a]_p$.*

PROOF. Because the domain is finite there exist positive integers l and l' such that $\{a\} + lp + l'p = a + l'p$. As $[a]_p + p + (l' - 1)p = [a]_p$ it follows from (P3) that $[a]_p + p = [a]_p$. By 1-minimality, a is in $\{a\} + lp - lp$ which is equal to $[a]_p$ by (P2). □

PROPOSITION 16. *Let $\mathcal{J} = (V, D, \{P_{x,y} : (x, y) \in \mathcal{S}\})$ be a weak Prague instance and let $\mathcal{F}$ be a set of operations on D. Then $\mathcal{J}' = (V, D, \{P'_{x,y} : (x, y) \in \mathcal{S}\})$, where*

$$\begin{aligned}
P'_{x,y} = \{&(f(a_1, a_2, \dots), f(b_1, b_2, \dots)) : f \in \mathcal{F}, \\
&(a_1, b_1), (a_2, b_2), \dots \in P_{x,y}\},
\end{aligned}$$

is a weak Prague instance.

Proof. It is apparent that $\mathcal{J}'$ is 1-minimal with

$$P_x^{\mathcal{J}} = P'_x := \{f(a_1, a_2, \dots) : f \in \mathcal{F},\ a_1, a_2, \dots \in P_x\}.$$

In what follows, by $A +' p$ we mean the addition computed in the instance $\mathcal{J}'$ while $A + p$ is computed in $\mathcal{J}$.

Before proving (P2*) and (P3) we make a simple observation.

Claim 16.1. *If $f \in \mathcal{F}$ is an operation of arity k, $x \in V$, p is a pattern from x, and $A_1, \dots, A_k \subseteq P_x$, $B \subseteq P'_x$ are such that $f(A_1, A_2, \dots, A_k) \subseteq B$, then $f(A_1 + p, A_2 + p, \dots A_k + p) \subseteq B +' p$.*

(By $f(A_1, \dots, A_k)$ we mean the set

$$\{f(a_1, \dots, a_k) : a_1 \in A_1, a_2 \in A_2, \dots, a_k \in A_k\}.)$$

Proof. It is enough to prove the claim for a single step $p = (x, y)$. The rest follows by induction. If $b \in f(A_1 + (x, y), \dots, A_k + (x, y))$ then there exist elements $b_1 \in A_1 + (x, y)$, $\dots$, $b_k \in A_k + (x, y)$ so that $f(b_1, b_2, \dots, b_k) = b$. As $b_i \in A_i + (x, y)$ there are elements $a_i \in A_i$ such that $(a_i, b_i) \in P_{x,y}$ for all $1 \leq i \leq k$. But then $(f(a_1, a_2, \dots, a_l)$, $f(b_1, b_2, \dots, b_k))$ is in $P'_{x,y}$ and $f(a_1, a_2, \dots, a_k) \in f(A_1, A_2, \dots, A_k) \subseteq B$, therefore $b = f(b_1, b_2, \dots, b_k) \in A +' (x, y)$. □

To prove (P2*) for $\mathcal{J}'$ let (x, y) be a step, $A \subseteq P'_x$, let p be a pattern from y to x such that $A +' (x, y) +' p = A$, and let a be an arbitrary element of $A +' (x, y, x)$. As $A +' (x, y, x) = (A +' (x, y)) +' (y, x)$, there exist $b \in A +' (x, y)$ such that $(a, b) \in P'_{x,y}$. By definition of $P'_{x,y}$, we can find $f \in \mathcal{F}$ (say, of arity k), elements $a_1, a_2, \dots, a_k$ in P_x, and $b_1, \dots, b_k$ in P_y so that $(f(a_1, a_2, \dots, a_k), f(b_1, b_2, \dots, b_k)) = (a, b)$ and $(a_i, b_i) \in P_{x,y}$ for all $1 \leq i \leq k$.

We consider the sets $[b_1]_q, [b_2]_q, \dots, [b_2]_q$ from Lemma 15 for the pattern $q = p + (x, y)$. We take l to be the maximum of the numbers for $b_1, \dots, b_k$ from this lemma, so $[b_i]_q = b_i + lq$. We get

$$\begin{aligned} a_i \in \{b_i\} + (y, x) \subseteq [b_i]_q + (y, x) &= \\ = [b_i]_q + p + (x, y) + (y, x) &= [b_i]_q + p, \end{aligned}$$

where the first step follows from $(a_i, b_i) \in P_{x,y}$, the inclusion and the first equality from Lemma 15, and the second equality from (P2*) for the instance $\mathcal{J}$ (as $([b_i]_q + p) + (x, y) + p = [b_i]_q + p$). Thus $a = f(a_1, a_2, \dots, a_k)$ is an element of

$$\begin{aligned} f([b_1]_q + p, [b_2]_q + p, \dots, [b_k]_q + p) &= \\ &= f(\{b_1\} + lq + p, \dots, \{b_k\} + lq + p) \end{aligned}$$

and this set is contained in $(A +' (x, y)) +' lq +' p = A +' (x, y) +' l(p + (x, y)) +' p = A$ by Claim 16.1 applied with $A_i = \{b_i\}$ and the pattern $lq + p$. We have shown that every element a of $A +' (x, y, x)$ lies in A. The other inclusion follows from 1-minimality.

To prove (P3) let $x \in V$, $A \subseteq P'_x$ and let p, q be patterns such that $A +' p +' q = A$. We first show that $A \subseteq A +' p$. Let $a \in P'_x$, take $f \in \mathcal{F}$, $a_1, a_2, \dots, a_k \in P_x$ such that $f(a_1, \dots, a_k) = a$, and find l so that $[a_i]_{p+q} = a_i + l(p + q)$. From (P3) for $\mathcal{J}$ and Lemma 15 it follows that $[a_i]_{p+q} + p = [a_i]_{p+q}$. By Claim 16.1,$a \in f([a_1]_{p+q}, [a_2]_{p+q}, \dots, [a_k]_{p+q}) = f([a_1]_{p+q} + p, [a_2]_{p+q} + p, \dots, [a_k]_{p+q} + p) \subseteq A +' l(p + q) +' p = A +' p$. The same argument used for $A +' p$ instead of A and the patterns $q + p, q$ instead of $p + q, p$ proves $A +' p \subseteq A +' p +' q = A$.