All or Nothing: Toward a Promise Problem Dichotomy for Constraint Problems

Lucy Ham and Marcel Jackson(✉)

Department of Mathematics and Statistics, La Trobe University, Melbourne 3086, Australia
leham@students.latrobe.edu.au, m.g.jackson@latrobe.edu.au

Abstract. We show that intractability of the constraint satisfaction problem over a fixed finite constraint language can, in all known cases, be replaced by an infinite hierarchy of intractable promise problems of increasingly disparate promise conditions. The instances are guaranteed to either have no solutions at all, or to be k-robustly satisfiable (for any fixed k), meaning that every "reasonable" partial instantiation on k variables extends to a solution.

Keywords: Constraint satisfaction problem · Dichotomy · Robust satisfiability · Promise problem · Quasivariety · Universal horn class

1 Introduction

In the constraint satisfaction problem (CSP) we are given a domain A, a list of relations $\mathscr{R}$ on A and a finite set V of variables, in which various tuples of variables have been constrained by the relations in $\mathscr{R}$. The fundamental *satisfaction* question is to decide whether there is a function $\phi : V \to A$ such that $(\phi(v_1), \ldots, \phi(v_n)) \in r$ whenever $\langle (v_1, \ldots, v_n), r \rangle$ is a constraint (and $r \in \mathscr{R}$ is of arity n). Many computational problems are expressible in this framework, even in the nonuniform case, where the domain A and relations $\mathscr{R}$ are fixed. Such *fixed template* CSPs have received particular attention in theoretical investigations: examples include the SAT variants considered by Schaefer [40], graph homomorphism problems such as in the Hell-Nešetřil dichotomy [22] as well as list-homomorphism problems and conservative CSPs [9]. Feder and Vardi [14] generated particular attention on the theoretical analysis of computational complexity of fixed template CSPs, by tying the complexity of fixed finite template CSPs precisely to those complexities to be found in the largest logically definable class for which they were unable to prove that Ladner's Theorem holds. This motivated their famous *dichotomy conjecture*: is it true that a fixed finite template CSP is either solvable in polynomial time or is NP-complete?

M. Jackson—The second author was supported by ARC grants FT120100666 and DP1094578.

J.C. Beck (Ed.): CP 2017, LNCS 10416, pp. 139–156, 2017.
DOI: 10.1007/978-3-319-66158-2_10

A pivotal development in the efforts toward a possible proof of the dichotomy conjecture was the introduction of universal algebraic methods. This provided fresh tools to build tractable algorithms, and to build reductions for hardness, as well as an established mathematical landscape in which to formulate conjectures on complexity. The method is fundamental to Bulatov's classification of 3-element CSPs [8], of the Dichotomy Theorem for conservative CSPs [9], for homomorphism problems on digraphs without sources and sinks [6], in the classification of when a CSP is solvable by generalised Gaussian elimination [23], and of when a CSP is solvable by a local consistency check algorithm [5], among others. The *algebraic dichotomy conjecture* (ADC) of [11] refines the Feder-Vardi conjecture by speculating the precise boundary between $\mathtt{P}$ and $\mathtt{NP}$, in terms of the presence of certain algebraic properties. The ADC has been verified in each of the aforementioned tractability classifications.

The present article shows that $\mathtt{NP}$-completeness results obtained via the algebraic method also imply the $\mathtt{NP}$-completeness of a strong promise problem. The NO instances are those for which there is no solution, but the YES instances are instances for which any "reasonable" partial assignment on k variables can extend to a solution. "Reasonable" here means subject to some finite set of local, necessary conditions. The All or Nothing Theorem (ANT) below proves the $\mathtt{NP}$-completeness of this promise problem for any integer $k \geq 0$ and in any intractable CSP covered by the algebraic method. The promise conditions include satisfaction as a special case, and complement the promise condition on NO instances provided by the PCP Theorem [4] (at least ε proportion of the constraints must fail, for some $\varepsilon > 0$). We are also able to prove a dichotomy theorem by showing that for sufficiently large k, our promise problem is solvable in $\mathtt{AC}^0$ if and only if the CSP is of bounded width (in the sense of Barto and Kozik [5]) and otherwise is hard for the complexity class $\mathtt{Mod}_p(\mathtt{L})$ for some prime p.

A second contribution of the article is to connect the model-theoretic notion of *quasivariety* to the concept of *implied constraints*. Identifying various kinds of implied constraints is a central method employed in constraint solvers [36], and the proliferation of implied constraints is associated with phase transitions in randomly generated constraint problems [37]. We explain how the absence of implied constraints corresponds to membership in the quasivariety generated by the template. Intuitively, it seems quite unlikely that the problem of recognising "no implied constraints" can be approached using the algebraic method, because there is no obvious reduction between constraint languages $\mathscr{R}_1$ and $\mathscr{R}_2$ when $\mathscr{R}_1 \subsetneq \mathscr{R}_2$. Despite this intuition, the strength of the promise in the ANT enables us to show that whenever the algebraic method shows hardness of $\mathrm{CSP}(\mathscr{R})$, then there is no polynomial time algorithm to distinguish constraint instances with no solution, from those that have no implied constraints with respect to $\mathscr{R}$. We can also use our bounded width dichotomy to obtain the most general nonfinite axiomatisability result known for finitely generated quasivarieties. A further important corollary is a promise problem extension of Hell and Nešetřil's well-known dichotomy for simple graphs [22].

More generally, the ability to extend all reasonable partial assignments holds potential for a wide range of applications, with recent examples including minimal networks [18], quantum mechanics [3], and semigroup theory [24].

2 Constraints and Implied Constraints

Since Feder and Vardi [14] it has been standard to reformulate the fixed template CSP over domain A and finite language $\mathscr{R}$ as a homomorphism problem between model-theoretic structures. The template is a relational structure $\mathbb{A} = \langle A, \mathscr{R}^A \rangle$ (with $\mathscr{R}$ a relational signature), as is the instance $(V, \mathscr{C})$ (where $\mathscr{C}$ is the list of constraints) where the variable set V is the universe, and with each $r \in \mathscr{R}$ being interpreted as the relation on V equal to the set of tuples constrained to r in the set $\mathscr{C}$. Thus each individual constraint $\langle (v_1, \ldots, v_n), r \rangle$ becomes a membership of a tuple $(v_1, \ldots, v_n)$ in the relation r^V on V. We refer to $(v_1, \ldots, v_n) \in r^V$ as a *hyperedge*. The *constraint satisfaction problem for* $\mathbb{A}$, which we denote by CSP($\mathbb{A}$) is the problem of deciding membership in the class of finite structures admitting homomorphism into $\mathbb{A}$. Throughout the article, $\mathbb{A}$ will be the default notation for a CSP template of signature $\mathscr{R}$ (both assumed to be finite) and $\mathbb{B}$ for a general (finite) $\mathscr{R}$-structure. We let arity($\mathscr{R}$) denote the maximal arity of any relation in $\mathscr{R}$.

A nonhyperedge $(v_1, \ldots, v_n) \notin r^B$, where $r \in \mathcal{R} \cup \{=\}$ (of arity n) satisfies the *separation condition* if there is a homomorphism $\phi : \mathbb{B} \to \mathbb{A}$ with $(\phi(v_1), \ldots, \phi(v_n)) \notin r^{\mathbb{A}}$. When a nonhyperedge fails the separation property we say that it is an *implied constraint*, as every homomorphism into $\mathbb{A}$ places it within the corresponding relation of $\mathbb{A}$; equivalently, $(v_1, \ldots, v_n)$ can be added to r^B without changing the set of possible solutions for $\mathbb{B}$ with respect to $\mathbb{A}$. We say that $\mathbb{B}$ *satisfies the separation condition* (w.r.t. $\mathbb{A}$), or *has no implied constraints* if no nonhyperedge is an implied constraint. Note that if $\mathbb{B}$ is a NO instance of CSP($\mathbb{A}$), then every nonhyperedge is implied (including equalities between distinct elements).

The separation condition for $\mathbb{B}$ is widely known to be equivalent to the property that $\mathbb{B}$ lies in the *quasivariety* generated by $\mathbb{A}$: the class of isomorphic copies of induced substructures of direct powers of $\mathbb{A}$; see Maltsev [34] and Gorbunov [17], but also [24, Theorem 2.1] and [26, §2.1,2.2] for the CSP interpretations and generalizations. The one-element total structure $\mathbf{1}_{\mathscr{R}}$, with no nonhyperedges, satisfies the separation condition vacuously. If we wish to exclude $\mathbf{1}_{\mathscr{R}}$ we arrive at the *universal Horn class* generated by $\mathbb{A}$ (which excludes the zeroth power from "direct powers"). We let $\mathsf{Q}(\mathbb{A})$ denote the quasivariety of $\mathbb{A}$ and $\mathsf{Q}^+(\mathbb{A})$ the universal Horn class of $\mathbb{A}$. Membership in $\mathsf{Q}(\mathbb{A})$ is the problem of deciding if an input has no implied constraints, which we denote by $\mathrm{CSP}_\infty(\mathbb{A})$. Membership in $\mathsf{Q}^+(\mathbb{A})$ is essentially the same as $\mathrm{CSP}_\infty(\mathbb{A})$ because $\mathsf{Q}(\mathbb{A})$ and $\mathsf{Q}^+(\mathbb{A})$ differ on at most the structure $\mathbf{1}_{\mathscr{R}}$.

Problem: $\mathrm{CSP}_\infty(\mathbb{A})$ (no implied constraints)
Instance: a finite $\mathscr{R}$-structure $\mathbb{B}$.
Question: for every nonhyperedge $(v_1, \ldots, v_n) \notin r^B$, is there a homomorphism into $\mathbb{B}$ taking $(v_1, \ldots, v_n) \notin r^B$ to a nonhyperedge $(a_1, \ldots, a_n) \notin r^A$ of $\mathbb{A}$?

The case of no implied *equalities* is considered in Ham [20,21], with a complete tractability classification in the case of Boolean constraint languages.

3 Primitive Positive Formulæ and Robust Satisfiability

Definition 1. *An* atomic formula *is an expression of the form* $(x_1, \ldots, x_n) \in r$ *for some* $r \in \mathscr{R}$ *or* $x = y$. *A* primitive positive formula (*abbreviated to* pp-formula) *is a formula obtained from a conjunction of atomic formulæ by existentially quantifying some variables. A pp-formula* $\phi(x_1, \ldots, x_n)$ *with free variables* $x_1, \ldots, x_n$ *defines an n-ary relation* r_ϕ, *which in any* $\mathscr{R}$*-structure* $\mathbb{C}$ *is interpreted as the solution set of* ϕ. *If* $\mathscr{F}$ *is a set of pp-formulæ, then* $\mathbb{C}_\mathscr{F}$ *denotes* $\langle C; \{r_\phi^C \mid \phi \in \mathscr{F}\}\rangle$.

We let $\mathrm{pp}(\mathscr{R})$ be the set of all pp-formulæ (over some fixed countably infinite set of variables) in $\mathscr{R}$ and let $\mathrm{pp}(\mathbb{C})$ denote the set $\{r_\phi^C \mid \phi \in \mathrm{pp}(\mathscr{R})\}$ of all relations on C that are pp-definable from the fundamental relations of $\mathbb{C}$.

Let $\mathbb{A}, \mathbb{B}$ be $\mathscr{R}$-structures and $\mathscr{F} \subseteq \mathrm{pp}(\mathscr{R})$. For a subset $S \subseteq B$, a function $\nu : S \to A$ is $\mathscr{F}$*-compatible* if it is a homomorphism from the substructure $\mathbb{S}$ of $\mathbb{B}_\mathscr{F}$ to $\mathbb{A}_\mathscr{F}$. A function $\nu : S \to A$ extends to a homomorphism precisely when it is $\mathrm{pp}(\mathscr{R})$-compatible [24, Lemma 3.1], so restricting $\mathscr{F}$ to a fixed finite subset of $\mathrm{pp}(\mathscr{R})$ is the natural local condition for extendability.

Definition 2 [21]. *Let* $\mathscr{F}$ *be a finite set of pp-formulæ in* $\mathscr{R}$ *and let* $\mathbb{A}$ *be a fixed finite* $\mathscr{R}$*-structure. For a finite* $\mathscr{R}$*-structure* $\mathbb{B}$, *we say that* $\mathbb{B}$ *is* $(k, \mathscr{F})$-robustly satisfiable (*with respect to* $\mathbb{A}$) *if* $\mathbb{B}$ *is a YES instance of* $\mathrm{CSP}(\mathbb{A})$ *and for every* k*-element subset* S *of* B *and every* $\mathscr{F}$*-compatible assignment* $\nu : S \to A$, *there is a solution to* $\mathbb{B}$ *extending* ν. *The structure* $\mathbb{B}$ *is* $(\leq k, \mathscr{F})$*-robustly satisfiable if it is* $(\ell, \mathscr{F})$*-robustly satisfiable for every* $\ell \leq k$.

Note that $(0, \mathscr{F})$-robust satisfiability coincides with satisfiability. Intuitively, $(k, \mathscr{F})$-robust satisfiability is a very strong condition on an instance. For example, a graph is $(2, \mathscr{F})$-robustly 2-colorable if every $\mathscr{F}$-compatible 2-coloring of any 2 vertices extends to a full 2-coloring. It is an easy exercise to show that a $(2, \mathscr{F})$-robustly 2-colorable graph must have diameter at most m, where m is the number of variables appearing in $\mathscr{F}$.

In [7], the case of $(k, \varnothing)$-robust satisfiability is considered for SAT-related probems using the notation $\widehat{\mathbf{U}^k}$; this appears in the context of phase transitions and implied constraints. The concept of $(k, \varnothing)$-robust satisfiability is called k*-supersymmetric* in Gottlob [18], where it is used to show that there is no polynomial time solver for a minimal constraint network. If $\mathscr{P}$ denotes the conjunction-free pp-formulæ, then $(k, \mathscr{P})$-robust satisfiability is the "k-robust satisfiability" concept introduced in Abramsky, Gottlob and Kolaitis [3], where (for $k = 3$ in 3SAT) it is applied to show the intractability of detecting local hidden-variable models in quantum mechanics. Jackson [24] showed the NP-completeness of a promise problem form of $(2, \mathscr{P})$-robust satisfiability for positive 1-in-3SAT, and used it to solve a 20+ year old problem in semigroup theory [2, Problem 4], itself motivated by issues in formal languages. These examples involve very technical case-checking arguments. A more unified algebraic approach was very recently initiated by Ham [20,21], who classified the tractability of $(2, \mathscr{F})$-robust satisfiability (for some $\mathscr{F}$) in the case of Boolean constraint languages.

4 Primitive Positive Definability and Polymorphisms

When $\mathscr{R}$ is pp-definable from a set of relations $\mathscr{S}$ on a set A then there is logspace reduction from $\mathrm{CSP}(\langle A; \mathscr{R}\rangle)$ to $\mathrm{CSP}(\langle A; \mathscr{S}\rangle)$. This fundamental idea was primarily developed through the work of Cohen, Jeavons and others [27–31], though aspects appear in proof of Schaefer's original dichotomy for Boolean CSPs [40].

There is a well-known Galois correspondence between sets of relations on a set A and the sets of operations on A; see [16]. The link is via *polymorphisms*, which are homomorphisms from the direct product $\mathbb{A}^n$ to $\mathbb{A}$. In other words, for each relation $r \in \mathscr{R}$ (with arity k, say), if we are given an $k \times n$ matrix of entries from A, with each column being a k-tuple in r, then applying the polymorphism f to each row produces a k-tuple of outputs that also must lie in r. We let $\mathrm{Pol}(\mathbb{A})$ denote the family of all polymorphisms of the relational structure $\mathbb{A}$. For finite A we have $\mathrm{Pol}(\langle A; \mathscr{R}\rangle) \subseteq \mathrm{Pol}(\langle A; \mathscr{S}\rangle)$ if and only if $\mathscr{S} \subseteq \mathrm{pp}(\langle A, \mathscr{R}\rangle)$, so that pp-definability is captured by polymorphisms. We now list some conditions on polymorphisms that we will use; see an article such as [25] for a survey of other conditions that play a role in understanding the complexity of CSP complexity.

- An n-ary operation $w : A^n \to A$ on a set A is a *weak near unanimity* operation (or *WNU*) if it satisfies $w(x, x, \ldots, x) = x$ (idempotence) and $w(y, x, \ldots, x) = w(x, y, \ldots, x) = \cdots = w(x, x, \ldots, y)$ for all x, y. A weak near unanimity operation is *near unanimity* (NU) if it additionally satisfies $w(y, x, \ldots, x) = x$.
- If the condition of being idempotent is dropped, we refer to a *quasi WNU*, and a *quasi NU* respectively.

We mention that most algebraic approaches use the assumption that the template $\mathbb{A}$ is a *core*, meaning that it has no proper retracts. We now list a selection of pertinent results and conjectures that are expressed in the language of polymorphisms.

The fundamental conjecture on fixed template CSP complexity is the following refinement of Feder and Vardi's original.

Algebraic Dichotomy Conjecture (ADC) 3 [11]. *Let* $\mathbb{A}$ *be a finite core relational structure of finite relational signature. If* $\mathbb{A}$ *has a WNU polymorphism then* CSP($\mathbb{A}$) *is tractable. If* $\mathbb{A}$ *has no WNU polymorphism then* CSP($\mathbb{A}$) *is* **NP***-complete.*

The final sentence in the conjecture is proved already in [11] (with the WNU condition we state established in [35]), with completeness with respect to first order reductions established in [33]. There are no counterexamples to the conjecture amongst known classifications, and recently several purported proofs have been claimed [10,39,42] (we do not assume these as verified).

In the following, *bounded width* corresponds to solvability by way of a local consistency check algorithm, while strict width is a restricted case of this, where every family of locally consistent partial solutions extends to a solution.

Theorem 4. *Let* $\mathbb{A}$ *be a finite core relational structure of finite relational signature.*

1. (*Feder and Vardi* [14].) CSP($\mathbb{A}$) *has strict width if and only if* $\mathbb{A}$ *has an NU polymorphism.*
2. (*Barto and Kozik* [5].) CSP($\mathbb{A}$) *has bounded width if and only if* $\mathbb{A}$ *has a* 3*-ary WNU* w_3 *and a* 4*-ary WNU* w_4 *such that* $w_3(y,x,x) = w_4(y,x,x,x)$ *holds for all* x, y.

5 Main Results

Recall that a *promise problem* consists of a pair of disjoint languages (Y, N). The question is conditional: given the promise that an instance lies in $Y \cup N$, decide if it lies in Y; see [19] for example.

The main result (ANT) concerns the following promise problem, which simultaneously extends CSP($\mathbb{A}$), $\mathrm{CSP}_\infty(\mathbb{A})$, robust-CSP($\mathbb{A}$) [3], SEP($\mathbb{A}$) [20,21] and others. In the title line, k is a non-negative integer and $\mathscr{F}$ is a finite set of pp-formulæ in the signature of $\mathbb{A}$.

Promise problem: $(Y_{(k,\mathscr{F}),\mathsf{Q}}, N_{\mathrm{CSP}})$ for $\mathbb{A}$.

YES: $\mathbb{B}$ is $(k, \mathscr{F})$-robustly satisfiable with respect to $\mathbb{A}$ and has no implied constraints.

NO: $\mathbb{B}$ is a no instance of CSP($\mathbb{A}$).

We will let $Y_{(k,\mathscr{F})}$ denote the YES promise but where "no implied constraints" is omitted.

All or Nothing Theorem (ANT) 5. *Let* $\mathbb{A}$ *be a finite core relational structure in finite signature* $\mathscr{R}$.

1. (*Everything is easy.*) *If* CSP($\mathbb{A}$) *is tractable then so also is deciding both* $\mathrm{CSP}_\infty(\mathbb{A})$ *and* $(k, \mathscr{F})$*-robust satisfiability, for any* k *and any finite set* $\mathscr{F}$ *of pp-formulæ.*

2. (*Nothing is easy.*) *If* $\mathbb{A}$ *has no WNU, then for all* k *there exists a finite set of pp-formulæ* $\mathscr{F}$ *such that* $(Y_{(k,\mathscr{F}),\mathtt{Q}}, N_{\mathrm{CSP}})$ *is* NP*-complete for* $\mathbb{A}$ *with respect to first order reductions.*

Remark 6. *The ANT shows that the ADC is equivalent to the ostensibly far stronger dichotomy statement*: *either there is a WNU and* (1) *holds, or there is no WNU and* $(Y_{(k,\mathscr{F}),\mathtt{Q}}, N_{\mathrm{CSP}})$ *is* NP*-complete for some finite family of formulæ* $\mathscr{F}$*.*

As an example, the ANT shows that or all k there exists an $\mathscr{F}$ such that it is NP-hard to distinguish the $(k,\mathscr{F})$-robustly 3-colorable graphs from those that are not 3-colorable at all.

The following result gives a dichotomy within tractable complexity classes.

Theorem 7. *Let* $\mathbb{A}$ *be a finite core relational structure in finite signature* $\mathscr{R}$*.*

1. *If* CSP($\mathbb{A}$) *has bounded width, then there exists* n *such that for all* $k \geq n$ *and for all finite sets of pp-formulæ* $\mathscr{F}$*, the promise problem* $(Y_{(k,\mathscr{F}),\mathtt{Q}}, N_{\mathrm{CSP}})$ *lies in* $\mathtt{AC}^0$*.* (*If* CSP($\mathbb{A}$) *has strict width, then the class of* $(k,\mathscr{F})$*-robustly satisfiable instances is itself first order definable.*)
2. *If* CSP($\mathbb{A}$) *does not have bounded width then for some prime number* p *and for all* k *there exists an* $\mathscr{F}$ *such that* $(Y_{(k,\mathscr{F}),\mathtt{Q}}, N_{\mathrm{CSP}})$ *is* $\mathtt{Mod}_p(\mathtt{L})$*-hard.*

Recall that the $\mathtt{Mod}_p(\mathtt{L})$ class contains L and hence properly contains $\mathtt{AC}^0$; the precise relationship with NL is unknown. Thus Theorem 7 shows that in contrast to CSP($\mathbb{A}$) (see [1,33]), one cannot get L-completeness, nor NL-completeness for $(Y_{(k,\mathscr{F}),\mathtt{Q}}, N_{\mathrm{CSP}})$ over $\mathbb{A}$ unless there are unexpected collapses between L, NL and $\mathtt{Mod}_p(\mathtt{L})$ for various p. For example: while graph 2-colorability is L-complete, deciding $(k,\mathscr{F})$-robust 2-colorability is first-order when $k \geq 2$ (and for any $\mathscr{F}$).

The complexity of CSP($\mathbb{A}$) is determined by the core retract of $\mathbb{A}$, but this is not true for $(k,\mathscr{F})$-robust satisfiability and quasivariety membership; see [24] and [20,21] for example. The following results however apply regardless of whether $\mathbb{A}$ itself is a core.

Corollary 8. *Let* $\mathbb{A}$ *be finite relational structure of finite signature.*

– *If* $\mathbb{A}$ *has no quasi WNU polymorphism then* $\mathrm{CSP}_\infty(\mathbb{A})$ *is* NP*-complete,*
– *If the core retract of* $\mathbb{A}$ *fails to have bounded width then* Q($\mathbb{A}$) *is not finitely axiomatisable in first order logic, even amongst finite structures.*

The second statement is equivalent to the absence of quasi WNUs satisfying the conditions of Theorem 4(2). Similar statements to Corollary 8 hold for problems intermediate to CSP($\mathbb{A}$) and $\mathrm{CSP}_\infty(\mathbb{A})$, such as the SEP($\mathbb{A}$) of [21] and the problem of detecting if no variable is nontrivially forced to take a fixed value: variables with implicitly fixed values have been called the "backbone" or "frozen variables"; see [32] for example.

The following corollary simultaneously covers the original Hell-Nešetřil Dichotomy for simple graphs and a corresponding quasivariety dichotomy; again it does not assume cores.

Gap Dichotomy for Simple Graphs 9. *Let $\mathbb{G}$ be a finite simple graph.*

1. *If $\mathbb{G}$ is bipartite, then deciding* CSP($\mathbb{G}$) *and deciding membership in the quasivariety of $\mathbb{G}$ are both tractable.*
2. *Otherwise, the following promise problem is* NP*-complete with respect to first order reductions and for finite input graph $\mathbb{H}$:*
   Yes $\mathbb{H}$ *is in* $\mathsf{Q}^+(\mathbb{G})$*.*
   No $\mathbb{H}$ *has no homomorphisms into* $\mathbb{G}$*.*

We also complete a line of investigation initiated by Beacham and Culberson [7], by identifying the threshold value for k in the intractability of $(k, \varnothing)$-robust satisfiability for nSAT; see Theorem 19 below.

To complete this section we give an overview of how the proof of the ANT develops across the remaining sections. Part (1) of ANT is a quite straightforward and is given in Sect. 12. The proof of ANT part (2) mimics the proof that CSP($\mathbb{A}$) is NP-complete when $\mathbb{A}$ has no WNU. Every step involves substantial difficulties in establishing that the promise $(Y_{(k,\mathscr{F}),\mathsf{Q}}, N_{\text{CSP}})$ can be carried through for some suitably constructed $\mathscr{F}$. There are five main steps which are developed as separate sections once we have introduced some further preliminary development. The various stages of the proof are unified in Sect. 12, where an outline of the proof of Theorem 7 can also be found. Section 13 gives some ideas for future work, including an example demonstrating the limits to which the NO promise provided by the PCP Theorem can be incorporated in the ANT.

6 Preliminary Development: $\mathscr{F}$-Types and Claw Formulæ

We now establish some useful preliminary constructions relating to pp-formulæ and $(k, \mathscr{F})$-robustness. Throughout, $\mathbb{A}$ and $\mathscr{F} \subseteq \text{pp}(\mathscr{R})$ are fixed and $\mathbb{B}$ is an input $\mathscr{R}$-structure; all are finite.

Let $x_1, \ldots, x_n$ denote the free variables in some pp-formula $\phi(x_1, \ldots, x_n) \in \mathscr{F}$ and let k be a nonnegative integer. For any function $\iota : \{x_1, \ldots, x_n\} \to \{x_1, \ldots, x_k\}$ we let $\phi^\iota(x_1, \ldots, x_k)$ denote the formula $\phi(\iota(x_1), \ldots, \iota(x_n))$. We let $\mathscr{F}_k$ denote the set of all formulæ obtained in this way. This is the standard way that high arity formulæ can produce lower arity ones, and the following is immediate.

Lemma 10. *Let $\mathbb{A}$ and $\mathbb{B}$ be $\mathscr{R}$-structures and consider a subset $\{b_1, \ldots, b_k\}$ of B. A function $\nu\colon \{b_1, \ldots, b_k\} \to A$ is $\mathscr{F}$-compatible if and only if for every $\phi(x_1, \ldots, x_k) \in \mathscr{F}_k$, if $\mathbb{B} \models \phi(b_1, \ldots, b_k)$ then $\mathbb{A} \models \phi(\nu(b_1), \ldots, \nu(b_k))$.*

The following gives a natural restriction of the model theoretic “k-type” to pp-formulæ.

Definition 11. *Let $\mathscr{R}$ be a finite relational signature and $\mathscr{F}$ a set of pp-formulæ in $\mathscr{R}$.*

1. *A $(k, \mathscr{F})$*-type *is any finite conjunction of distinct k-ary formulæ in $\mathscr{F}_k$ over $x_1, \ldots, x_k$. The set of all $(k, \mathscr{F})$-types is denoted by* $\text{type}_k(\mathscr{F})$*.*

2. The $(k, \mathscr{F})$-type *of a tuple* $\overrightarrow{b} \in B^k$ *is the conjunction* $\bigwedge_{\substack{\phi(\overrightarrow{x}) \in \mathscr{F}_k, \\ \mathbb{B} \models \phi(\overrightarrow{b})}} \phi(\overrightarrow{x})$.

3. For $\ell \leq k$ *we let* $\mathscr{F}|_\ell$ *denote* $\{\exists x_{\ell+1} \ldots \exists x_k \ \tau(x_1, \ldots, x_k) \mid \tau \in type_k(\mathscr{F})\}$.

The following follows immediately from Lemma 10 and the definition of $(k, \mathscr{F})$-types.

Lemma 12. *Let* $\mathbb{A}$ *and* $\mathbb{B}$ *be* $\mathscr{R}$*-structures and consider a subset* $\{b_1, \ldots, b_k\}$ *of* B. *A partial map* $\nu \colon \{b_1, \ldots, b_k\} \to A$ *is* $\mathscr{F}$*-compatible if and only if* $\mathbb{A} \models \tau(\nu(b_1), \ldots, \nu(b_k))$, *where* τ *is the* $(k, \mathscr{F})$ *type of* $(b_1, \ldots, b_k)$.

The next lemma has a straightforward proof. We assume that $|B| \geq k$, though minor amendment to the definition of $\mathscr{F}|_i$ can accommodate smaller $|B|$.

Lemma 13. *Let* $\mathbb{A}$ *and* $\mathbb{B}$ *be finite* $\mathscr{R}$*-structures and let* $\mathscr{F}$ *be a finite set of pp-formulæ in* $\mathscr{R}$. *If* $\mathbb{B}$ *is* $(k, \mathscr{F})$*-robustly satisfiable into* $\mathbb{A}$ *and* $\ell \leq k$, *then* $\mathbb{B}$ *is* $(\ell, \mathscr{F}|_\ell)$*-robustly satisfiable. In particular, if* $\mathbb{B}$ *is* $(k, \mathscr{F})$*-robustly satisfiable for some finite set of pp-formulæ* $\mathscr{F}$, *then* $\mathbb{B}$ *is* $(\leq k, \bigcup_{0 \leq i \leq k} \mathscr{F}|_i)$*-robustly satisfiable.*

Recall from Sect. 4 that when $\mathscr{R}$ is pp-definable from a set of relations $\mathscr{S}$ on a set A then there is logspace reduction from $\mathrm{CSP}(\langle A; \mathscr{R}^A \rangle)$ to $\mathrm{CSP}(\langle A; \mathscr{S}^A \rangle)$. Assume then that each relation symbol $r \in \mathscr{R}$ has been matched to some fixed defining $\mathscr{S}$-formula $\rho_r(x_1, \ldots, x_n)$ of the same arity n as r:

$$\exists y_1 \ldots \exists y_m \bigwedge_{1 \leq i \leq k} \alpha_i(x_{i,1}, \ldots, x_{i,n_i}, y_{i,1}, \ldots, y_{i,m_i}), \qquad (\dagger)$$

where each α_i is an atomic formula in $\mathscr{S} \cup \{=\}$, and $\bigcup_{1 \leq i \leq k} \{x_{i,1}, \ldots, x_{i,n_i}\} = \{x_1, \ldots, x_n\}$ and $\bigcup_{1 \leq i \leq k} \{y_{i,1}, \ldots, y_{i,m_i}\} = \{y_1, \ldots, y_m\}$. Let $\rho_r^\flat$ denote the underlying open formula obtained from ρ_r by removing quantifiers: variables of $\rho_r^\flat$ that are quantified in ρ_r will be called *existential variables* (or $\exists$-variables: the y_i in $\dagger$) and the other variables will be referred to as *open variables.*

Each pp-formula $\psi(x_1, \ldots, x_\ell)$ in the signature $\mathscr{R}$ is equivalent to a pp-formula $\psi^{\mathscr{S}}(x_1, \ldots, x_\ell)$ in the signature $\mathscr{S}$: replace each conjunct in ψ—an atomic formula $r(x_1, \ldots, x_n)$ in for some $r \in \mathscr{R}$—by the defining formula ρ_r as in $\dagger$, and then apply the usual logical rules for moving quantifiers to the front (including renaming quantified variables where necessary).

Definition 14. *Let* $\mathscr{S}$ *define* $\mathscr{R}$ *by pp-formulæ* $\{\rho_r \mid r \in \mathscr{R}\} \subseteq \mathrm{pp}(\mathscr{S})$. *Let* k, ℓ *be fixed non-negative integers and* $\mathscr{F}$ *a finite set of pp-formulæ in* $\mathscr{R}$. *A* claw formula *for* $\mathscr{F}$ *of arity* k *and bound* ℓ *is any pp-formula in* $\mathscr{S}$ *of the form constructed in the third step below:*

1. *(The* talon.*) Let* γ *denote any conjunction* $\bigwedge_{1 \leq i \leq k'} \rho_{r_i}^\flat$, *where* $r_i \in \mathscr{R}$ *and* $k' \leq k$. *We allow some identification between open variables, but not between existential variables.*

2. (*The* wrist.) *Let σ be an $(\ell', \mathscr{F})$-type in $\mathscr{R}$ for some $\ell' \leq \ell$. Some of the ℓ' free variables in σ may be identified with open variables in γ, but not with existential variables.*
3. (*The* claw.) *Existentially quantify all but k of the unquantified variables in the conjunction $\gamma \wedge \sigma^{\mathscr{S}}$.*

7 Step 1. Reflection

Definition 15. *Let $\mathbb{A}$, k and $\mathscr{F}$ be fixed. For an input $\mathscr{R}$-structure $\mathbb{B}$, let $\mathbb{B}^{\downarrow}$ be the result of adjoining all hyperedges to $\mathbb{B}$ that are implied by $\mathscr{F}$-compatible assignments from subsets of $\mathbb{B}$ on at most k elements. The structure $\mathbb{B}^{\downarrow}$ will be called the* 1-step $(k, \mathscr{F})$-reflection *of $\mathbb{B}$.*

Under the promise $(Y_{(k,\mathscr{F})}, N_{\mathrm{CSP}})$ it is possible to show that there is a first order query that defines $\mathbb{B}^{\downarrow}$. The details take some effort and we omit them. To achieve the main results with respect to polynomial time reductions however, simply observe that $\mathbb{B}^{\downarrow}$ can be constructed from $\mathbb{B}$ in polynomial time, because there are only polynomially many $\mathscr{F}$-compatible assignments from subsets of size at most k. Then all that is needed is the following lemma.

Lemma 16. *If $\mathbb{B}$ is $(\leq k, \mathscr{F})$-robustly satisfiable with respect to $\mathbb{A}$, and $k \geq \mathrm{arity}(\mathscr{R})$, then $\mathbb{B}^{\downarrow}$ lies in the quasivariety of $\mathbb{A}$ and is also $(\leq k, \mathscr{F})$-robustly satisfiable. If $\mathbb{B}$ is a NO instance of $\mathrm{CSP}(\mathbb{A})$ then so also is $\mathbb{B}^{\downarrow}$.*

8 Step 2. Stability of Robustness over Primitive-Positive Reductions

We prove the following variant of the usual pp-reduction for CSPs.

Theorem 17. *Assume that $\mathbb{A}_1 = \langle A, \mathscr{R}^A \rangle$ and $\mathbb{A}_2 = \langle A, \mathscr{S}^A \rangle$ are two relational structures on the same finite set A, with $\mathscr{R}^A \subseteq \mathrm{pp}(\mathbb{A}_2)$ finite and $\ell := \mathrm{arity}(\mathscr{R})$. Let $\mathscr{F}$ be a finite set of pp-formulæ in the language of $\mathscr{R}$. Then, for any k, the standard pp-reduction of $\mathrm{CSP}(\mathbb{A}_1)$ to $\mathrm{CSP}(\mathbb{A}_2)$ takes $(\leq k\ell, \mathscr{F})$-robustly satisfiable instances of $\mathrm{CSP}(\mathbb{A}_1)$ to $(k, \mathscr{G})$-robustly satisfiable instances of $\mathrm{CSP}(\mathbb{A}_2)$, where $\mathscr{G}$ denotes the k-ary claw formulæ for $\mathscr{F}$ of bound $k\ell$.*

First briefly recall the precise nature of the "standard reduction" described in Theorem 17. Recall that each $r \in \mathscr{R}$ corresponds to an $\mathscr{S}$ formula ρ_r, as in (†). For an instance $\mathbb{B} = \langle B; \mathscr{R}^B \rangle$ of $\mathrm{CSP}(\mathbb{A}_1)$, an instance $\mathbb{B}^\sharp$ of $\mathrm{CSP}(\mathbb{A}_2)$ is constructed in the following way. For each hyperedge $(b_1, \dots, b_n) \in r$ in $\mathbb{B}$ (and adopting the generic notation of †), new elements $c_1, \dots, c_m$ are added to the universe of B, and the hyperedge $(b_1, \dots, b_n) \in r$ is replaced by the hyperedges $\alpha_i(b_{i,1}, \dots, b_{i,n_k}, c_{i,1}, \dots, c_{i,m_i})$ for each $i = 1, \dots, k$. Note that new elements $c_1, \dots, c_m$ are introduced for every instance of a hyperedge. The new elements will be referred to as *existential elements* (or $\exists$-elements), and for any $D \subseteq B^\sharp$

we let $D_\exists$ denote $\exists$-elements in D. Elements of B will be referred to as *open elements*, and we write D_B for $D \cap B = D \backslash D_\exists$.

It is easy to see that there is a homomorphism from $\mathbb{B}$ to $\mathbb{A}_1$ if and only if there is one from $\mathbb{B}^\sharp$ to $\mathbb{A}_2$: this is the usual logspace CSP reduction, which is a first order reduction when none of the ρ_r formulæ involve equality [33]. Now assume that $\mathbb{B}$ is $(\leq k\ell, \mathscr{F})$-robustly satisfiable with respect to $\mathbb{A}_1$ and consider a k-element subset $D \subseteq B^\sharp$, for which there is a $\mathscr{G}$-compatible assignment into A. The following arguments will refer back to the 3-step construction of claw formulæ in Definition 14.

Each $c \in D_\exists$ was introduced in replacing a hyperedge of $\mathbb{B}$ in signature $\mathscr{R}$ by a *family* of hyperedges in the signature $\mathscr{S}$, according to the pp-definition as in †. Each element of $D_\exists$ appears in at most one such family of $\mathscr{S}$-hyperedges, so the number of these, k', is at most $|D_\exists| \leq k$. Observe that these hyperedge families correspond to an interpretation of a conjunction γ of k' many formulæ as in step 1 of Definition 14: there is no identification of $\exists$-elements, but there may be of open elements. Each of these families involves at most ℓ open elements, so that at most $k' \times \ell$ open elements appear in these hyperedge families. Let O_B denote these elements. Because $k' + |D_B| \leq |D_\exists| + |D_B| = k$ and $|O_B| \leq k'\ell$, we have $|O_B \cup D_B| \leq k'\ell + |D_B| \leq k\ell$. Let σ denote the $(\ell', \mathscr{F})$-type of $O_B \cup D_B$ in $\mathbb{B}$, as in the second step of Definition 14. (Here we treat $O_B \cup D_B$ as a tuple ordered in any fixed way.) Observe that some elements b of D_B may also lie in O_B, and we will assume then that the variable in σ corresponding to b has been identified with the variable in γ corresponding to b. Let U be the set of all unquantified variables in $\gamma \wedge \sigma^{\mathscr{S}}$ that do not correspond to elements of D. The claw formula $\exists U\ \gamma \wedge \sigma^{\mathscr{S}}$ is in $\mathscr{G}$ and is satisfied by $\mathbb{B}^\sharp$ at D (again, arbitrarily treated as a tuple). Hence $\exists U\ \gamma \wedge \sigma^{\mathscr{S}}$ is preserved by ν. In particular then, in $\mathbb{A}_2$ we can find values for the variables corresponding to the elements of O_B that witness the satisfaction of $\exists U\ \gamma \wedge \sigma^{\mathscr{S}}$ at $\nu(D)$. Let $\nu' : O_B \cup D \to A$ be the extension of ν obtained by giving elements of $O_B \backslash D_B$ these witnessing values. Because σ is the $(\ell', \mathscr{F})$-type of $O_B \cup D_B$, it follows from Lemma 12 that $\nu'|_{O_B \cup D_B}$ is $\mathscr{F}$-compatible, so by the assumed $(\leq k\ell, \mathscr{F})$-robust satisfiability of $\mathbb{B}$ it follows that $\nu'|_{O_B \cup D_B}$ extends to a homomorphism ν^+ from $\mathbb{B}$ to $\mathbb{A}_1$. By the usual pp-reduction, ν^+ extends to a homomorphism $\nu^\sharp$ from $\mathbb{B}^\sharp$ to $\mathbb{A}_2$. Now $\nu^\sharp$ agrees with ν on D_B, but also, we may assume that it agrees with ν on $D_\exists$, because the values given O_B by ν' (and hence $\nu^\sharp$) were such that γ held. Thus we have extended ν to a homomorphism, as required.

9 Step 3. $(k, \varnothing)$-Robustness of $(3k+3)$SAT

Gottlob [18, Lemma 1] showed that the standard Yes/No decision problem 3SAT reduces to the promise $(Y_{(k,\varnothing)}, N_{\mathrm{CSP}})$ for $(3k+3)$SAT. For the sake of completeness of our sketch, we recall the basic idea. The construction is to replace in a 3SAT instance $\mathbb{B}$, each element b by $2k+1$ copies $b_1, \ldots, b_{2k+1}$ and then each clause $(b \vee c \vee d)$ by all $\binom{2k+1}{k+1}^3$ clauses of the form $(b_{i_1} \vee \cdots \vee b_{i_{k+1}} \vee c_{i'_1} \vee \cdots \vee c_{i'_{k+1}} \vee d_{i''_1} \vee \cdots \vee d_{i''_{k+1}})$ where the i_j, i'_j, i''_j are from $\{1, \ldots, 2k+1\}$. No assignment on k elements covers all of the $k+1$ copies of any element in a clause it appears, which enables the flexibility for such assignments to always extend to a solution, provided (and only when) $\mathbb{B}$ is a YES instance. We omit the details showing that this can be achieved via a first-order query.

10 Step 4. $(k, \mathscr{F})$-Robustness of 3SAT

We now establish the following theorem by reduction from the result in Step 3. Critically, the value of k is arbitrary, but the constraint language (3SAT) has fixed arity 3.

Theorem 18. *Fix any $k \geq 0$ and let $\mathscr{F}$ be the set of all claw formulæ for $\varnothing$ of arity k and with bound k. Then $(Y_{(k,\mathscr{F})}, N_{\mathrm{CSP}})$ for* 3SAT *is* **NP***-complete via first order reductions.*

The usual reduction of nSAT to 3SAT (as in [15] for example) is an example of a pp-reduction, because the nSAT clause relation $(x_1 \vee \cdots \vee x_n)$ (where the x_i can be negated variables if need be) is equivalent to the following pp-formula over $n-2$ clause relations of 3SAT:

$$\exists y_1 \dots \exists y_{n-4}\ (x_1 \vee x_2 \vee y_1) \wedge \left(\bigwedge_{3 \leq i \leq n-2} (\neg y_{i-2} \vee x_i \vee y_{i-1}) \right) \wedge (\neg y_{n-3} \vee x_{n-1} \vee x_n) \tag{‡}$$

As we are dealing with the standard pp-reduction, an instance $\mathbb{B}$ of $(3k+3)$SAT is satisfiable if and only if the constructed instance $\mathbb{B}^\sharp$ of 3SAT is satisfiable.

Now assume that $\mathbb{B}$ is a $(k, \varnothing)$-robustly satisfiable instance of $(3k+3)$SAT. Assume D is a k-set from $\mathbb{B}^\sharp$ and $\nu : D \to \{0, 1\}$ an $\mathscr{F}$-compatible partial assignment. As in the proof of Theorem 17, there are $k' \leq |D_\exists|$ different clause families involving elements from $D_\exists$; let F denote this set of families of clauses (each family arising by the replacement of a $(3k+3)$SAT clause by the $3k+1$ distinct 3SAT clauses). Let γ denote the conjunction of k' many pp-formulæ corresponding to these F: it is a conjunction of k' distinct copies of the underlying open formula of ‡, possibly with some of the open variables in different copies identified. Let U be the variables of γ that do not correspond to an element of D. Then $\exists U\ \gamma$ is a claw formula in the sense of Definition 14 because the only $(\ell, \varnothing)$-types (as detailed in step 2 of Definition 14) are empty formulæ. This formula $\exists U\ \gamma$ is obviously satisfied at D in $\mathbb{B}^\sharp$, so is preserved by ν. Now the proof deviates from Theorem 17. We show how to assign values to at most k of the remaining open elements of F such that any extension to a full solution on $\mathbb{B}$ extends to one for $\mathbb{B}^\sharp$ in a way consistent with the values given to $D_\exists$ by ν.

We introduce an arrow notation to help select the new open elements.

- Above the leftmost bracket of the clause family we place a right arrow $\mapsto$, and dually a $\mapsfrom$ over the rightmost bracket.
- Place a left arrow $\mapsfrom$ above a consecutive pair of brackets ")(" if the $\exists$-element immediately preceding it is given 0 by ν, and dually, $\mapsto$ if the $\exists$-element is assigned 1.

Let us say that two such arrows are *convergent* if they point toward one another. In order to extend ν to a solution, within each pair of convergent arrows, an open-literal to assign the value 1. We first give an example, consisting of a clause family, an assignment to some elements (say, $D_\exists = \{b_1, b_2, b_3\}$ and $D_B = \{a_1\}$) and the arrows placed as determined by the rules:

$$\begin{array}{l}
(\ a_1\ \neg a_2\ b_1\ )(\ \neg b_1\ a_3\ b_2\ )(\ \neg b_2\ a_4\ b_3\ )(\ \neg b_3\ a_5\ a_6\ ) \\
(\ \ 0\ \ \neg a_2\ \ 0\ )(\ \ 1\ \ a_3\ \ 1\ )(\ \ 0\ \ a_4\ \ 0\ )(\ \ 1\ \ a_5\ a_6\ ) \\
\overset{\mapsto}{(}\quad \neg a_2 \quad \overset{\leftarrow\!\dashv}{)}(\quad a_3 \quad \overset{\mapsto}{)}(\quad a_4 \quad \overset{\leftarrow\!\dashv}{)}(\quad a_5\ a_6\ \overset{\leftarrow\!\dashv}{)}
\end{array}$$

By calling on witnesses to preservation of $\mathscr{F}$ by ν we can select open literals and values (here $\nu(a_4) = 1$ and $\nu(a_2) = 0$) that are consistent with the values assigned to $D_\exists$.

In the general case: because ν preserves the claw formula $\exists U\ \gamma$, the 2-element template for 3SAT has witnesses to all quantified variables. For each pair of convergent arrows under the assignment by ν for $D_\exists$, there is a witness to one of the open variables in γ taking the value 1; only one such witness is required for each pair of convergent arrows. Let E consist of the open elements in F corresponding to the selected witnesses, and extend ν to E by giving them the witness values. Note that $|E| \leq |D_\exists|$, so that $|E \cup D_B| \leq k$. Thus $\nu|_{E \cup D_B}$ extends to a solution for $\mathbb{B}$. This solution extends to a solution for $\mathbb{B}^\sharp$ in a way that is consistent with the values given elements of $D_\exists$ by ν.

By a variation of this argument and Sect. 9, we can also obtain the following theorem, which completes one line of investigation initiated by Beacham and Culberson [7].

Theorem 19. *Let $n > 2$ and consider the problem nSAT. If $k \geq n$ then deciding $(k, \varnothing)$-robust satisfiability is in $\mathtt{AC}^0$. If $k < n$ then $(Y_{(k,\varnothing)}, N_{\mathrm{CSP}})$ is $\mathtt{NP}$-complete.*

11 Step 5: Idempotence and the Algebraic Method

A key development in the algebraic method for CSP complexity was restriction to idempotent polymorphisms [11]. We now sketch how this works for the $(Y_{(k,\mathscr{F})}, N_{\mathrm{CSP}})$ promise.

Let $\mathscr{R}_{\mathrm{Con}}$ be the signature obtained by adding a unary relation symbol $\underline{a}$ for each element a of A, and let $\mathbb{A}_{\mathrm{Con}}$ denote the structure $\langle A; \mathscr{R}_{\mathrm{Con}}\rangle$, with $\underline{a}$ interpreted as $\{a\}$.

Theorem 20. *Let $\mathbb{A}$ be a core and $\mathscr{F}$ be a finite subset of $\mathrm{pp}(\mathscr{R}_{\mathrm{Con}})$. Then for any k, there exists a finite set $\mathscr{G}$ of pp-formulæ in the language of $\mathscr{R}$ such that the standard reduction from $\mathrm{CSP}(\mathbb{A}_{\mathrm{Con}})$ to $\mathrm{CSP}(\mathbb{A})$ takes $(k, \mathscr{F})$-robustly satisfiable instances of $\mathrm{CSP}(\mathbb{A}_{\mathrm{Con}})$ to the $(k, \mathscr{G})$-robustly satisfiable instances of $\mathrm{CSP}(\mathbb{A})$.*

Proof (Proof sketch). Let $\mathbb{B}$ be an instance of $\mathrm{CSP}(\mathbb{A}_{\mathrm{Con}})$. The standard reduction (first order by [33, Lemma 2.5]) involves adjoining a copy of $\mathbb{A}$ to the instance

$\mathbb{B}$, and replacing all hyperedges $b \in \underline{a}$ by identifying b with the adjoined copy of a; call this $\mathbb{B}^\sharp$. (A $(k, \mathscr{F})$-reflection, via the first order version of Lemma 16, can be used to circumvent some technical issues regarding identification of elements.) Our task is to show how to construct $\mathscr{G}$. Let diag($\mathbb{A}$) denote the *positive atomic diagram of* $\mathbb{A}$ on some set of variables $\{v_a \mid a \in A\}$; that is, the conjunction of all hyperedges of $\mathbb{A}$ (considered as atomic formulæ). We construct $\mathscr{G}$ by taking the conjunction of diag($\mathbb{A}$) with $\mathscr{F}$-types σ, and replacing each conjunct of the form $x \in \underline{a}$ in σ, by $x = v_a$.

Assume $\mathbb{B}$ is $(k, \mathscr{F})$-robustly satisfiable with respect to $\mathbb{A}_{\text{Con}}$ and consider a $\mathscr{G}$-compatible assignment ν from some k-set in $\mathbb{B}^\sharp$. Because $\mathbb{A}$ is a core, there is an automorphism α of $\mathbb{A}$ mapping witnesses to diag($\mathbb{A}$) to their named location (that is, taking v_a to a). Then $\alpha \circ \nu$ is $\mathscr{F}$-compatible into $\mathbb{A}_{\text{Con}}$, hence extends to a homomorphism ψ from $\mathbb{B}$. Then $\alpha^{-1} \circ \psi$ is a homomorphism from $\mathbb{B}^\sharp$ to $\mathbb{A}$ extending ν. □

12 Proof of ANT, Corollaries and Theorem 7

Proof (Proof of ANT). For part (1), we extend an idea from [24]. Our proof will use only the assumption that CSP($\mathbb{A}_{\text{Con}}$) is tractable. This is always true if $\mathbb{A}$ is a core with CSP($\mathbb{A}$) tractable. Now observe that an $\mathscr{F}$-compatible partial assignment $\nu : b_i \mapsto a_i$ from a subset $\{b_1, \dots, b_k\}$ of an instance $\mathbb{B}$ into $\mathbb{A}$ extends to a solution if and only if the structure obtained from $\mathbb{B}$ by adjoining the constraints $\{b_i \in \{a_i\} \mid i = 1, \dots, k\}$ is a YES instance of CSP($\mathbb{A}_{\text{Con}}$). Thus after polynomially many calls on the tractable problem CSP($\mathbb{A}_{\text{Con}}$), we can decide the $(k, \mathscr{F})$-robust satisfiability of $\mathbb{B}$. An almost identical argument will determine if $\mathbb{B}$ has no implied constraints, thus deciding $\text{CSP}_\infty(\mathbb{A})$.

Now to prove ANT part (2). Let $\mathbf{A}$ denote the polymorphism algebra of $\mathbb{A}_{\text{Con}}$. One of the fundamental consequences of the algebraic method is that if $\mathbb{A}$ has no WNU polymorphism, then the polymorphism algebra of 3SAT is a homomorphic image of a subalgebra of $\mathbf{A}$ (direct powers are not required; see [41, Prop 3.1]). For CSPs, these facts will give a first order reduction from 3SAT to some finite set of relations $\mathscr{S}^A$ in pp($\mathbb{A}_{\text{Con}}$): see [33]. The first step of this reduction is to reduce through homomorphic preimages and subalgebras. Ham [21, Sect. 8] showed that these initial reductions also preserve the $(Y_{(\ell,\mathscr{F})}, N_{\text{CSP}})$ promise, with only minor modification to $\mathscr{F}$. Combining this with Theorem 18 then Lemma 13 we find that for all ℓ there exists an $\mathscr{F}_2$ such that $(Y_{(\leq \ell,\mathscr{F}_2)}, N_{\text{CSP}})$ is NP-complete for $\langle A, \mathscr{S}^A \rangle$. Then (using $\ell = arity(\mathscr{S}) \times k$) we can use Theorem 17 then Lemma 13 to find that for every k there exists $\mathscr{F}_3$ such that $(Y_{(\leq k,\mathscr{F}_3)}, N_{\text{CSP}})$ is NP-complete for $\mathbb{A}_{\text{Con}}$ with respect to first order reductions. By Theorem 20 the same is true for $\mathbb{A}$, with an amended compatibility condition $\mathscr{F}$ depending on k. Lemma 16 then extends the promise to $(Y_{(k,\mathscr{F}),\mathsf{Q}}, N_{\text{CSP}})$, as required. □

Proof (Proof of Corollary 8). Let $\mathbb{A}$ be a finite relational structure without a quasi WNU polymorphism. By Chen and Larose [13, Lemma 6.4] the core retract $\mathbb{A}^\flat$ of $\mathbb{A}$ has no WNU. Hence the ANT applies to $\mathbb{A}^\flat$. Now $\mathsf{Q}^+(\mathbb{A})$ contains $\mathsf{Q}^+(\mathbb{A}^\flat)$, which contains the YES promise in the ANT and is disjoint from the NO promise. Hence membership in $\mathtt{Q}^+(\mathbb{A})$ is NP-complete with respect to first order reductions, and hence is also not finitely axiomatisable in first order logic, even at the finite level. The same argument using Theorem 7(2) implies non-finite axiomatisability in the case that $\mathbb{A}^\flat$ does not have bounded width. □

Proof (Proof of the Gap Dichotomy for Simple Graphs 9). If $\mathbb{G}$ is bipartite, then CSP($\mathbb{G}$) is tractable and so is deciding membership in $\mathtt{Q}^+(\mathbb{G})$: there are only five distinct quasivarieties [12,38]. Otherwise, $\mathbb{G}$ is not bipartite and so neither is its core retract. Hence $\mathbb{G}$ has no quasi WNU; see [6]. Then apply Corollary 8. □

Proof (Proof of Theorem 7). Due to space constraints we give only a very brief overview of the method. A CSP has *bounded width* provided that there exists j such that the existence of a homomorphism from $\mathbb{B}$ to $\mathbb{A}$ is equivalent to a family of partial homomorphisms on all subsets of size at most $j+1$, with the family satisfying a compatibility condition, known as a $(j, j+1)$*-strategy*; see [5]. When $k > j$ and input $\mathbb{B}$ satisfies the $Y_{(k,\mathscr{F}),\mathtt{Q}}$ promise, there is an obvious choice for a $(j, j+1)$-strategy: the family of all maps that can extend to $\mathscr{F}$-compatible assignments on k points. The property that this family forms a $(j, j+1)$-strategy can be expressed as a first order sentence ξ. When CSP($\mathbb{A}$) has bounded width (so that NO instances do not have $(j, j+1)$-strategies) the sentence ξ must fail on instances satisfying the N_{CSP} promise, and must hold on those satisfying the $Y_{(k,\mathscr{F}),\mathtt{Q}}$ promise.

Now assume that $\mathbb{A}$ does not have bounded width. In this case, a direct analogue of the arguments of Sect. 12 lead back to a structure $\mathbb{C}$ (encoding ternary linear equations over an abelian group) whose CSP is $\mathtt{Mod}_p(\mathtt{L})$-complete; see proof of [33, Theorem 4.1]. The rest of the proof parallels that of the ANT 5, except that Sects. 9 and 10 are replaced by constructions concerning linear systems of equations.

13 Discussion and Extensions

We have shown in the ANT that the fundamental intractability result of [11] can be replaced by an unbounded hierarchy of intractable promise problems, and demonstrated in Theorem 7 a collapse in several intermediate complexity classes for these problems. We feel these results are just the beginning of new applications to ideas relating to the detection of more general implied constraints (as in [7]), minimal networks (as in [18]), as well as to other areas of mathematics and computer science, such as the quantum-theoretic applications in [3] and the semigroup-theoretic applications of [24]. Some further consequences of the ANT omitted from the present work include a substantial extension of the Ham's "Gap Trichotomy Theorem" [21] to the $(Y_{(k,\mathscr{F}),\mathtt{Q}}, N_{\mathrm{CSP}})$ promise.

Some specific new directions this work should be taken include the extension of ANT to noncore templates and to infinite templates, where a much wider array of important computational problems can be found. Another difficult question: can the promise supplied by the PCP Theorem be added as a restriction to N_{CSP} in the ANT? (We write $N_{\varepsilon\,\mathrm{CSP}}$ for this condition: ε proportion of the constraints must fail.) The answer is nearly yes, but not quite. It is quite routine to carry through the failure of a positive fraction of constraints through steps 1–5 of the proof of the ANT part (2), and through step 6 with more difficulty, thereby achieving the NP-completeness of $(Y_{(k,\mathscr{F})}, N_{\varepsilon\,\mathrm{CSP}})$ for core templates without a WNU. Surprisingly though, $N_{\varepsilon\,\mathrm{CSP}}$ does not in general survive reflection, as the following example demonstrates. Let $\mathbb{2}^+$ denote the template on $\{0,1\}$ with the fundamental ternary relation r of +1-in-3SAT and the 4-ary total relation $s := \{0,1\}^4$. This has no WNU, as +1-in-3SAT has no WNU, so the ANT part (2) and claims just made imply that both $(Y_{(k,\mathscr{F})}, N_{\varepsilon\,\mathrm{CSP}})$ and $(Y_{(k,\mathscr{F}),\mathtt{Q}}, N_{\mathrm{CSP}})$ are NP-complete. Yet $(Y_{(k,\mathscr{F}),\mathtt{Q}}, N_{\varepsilon\,\mathrm{CSP}})$ for $\mathbb{2}^+$ falls into $\mathtt{AC}^0$! Indeed the first order property τ stating that s is total must hold on instances without implied constraints, and fail on any large enough instance $\mathbb{B}$ satisfying $N_{\varepsilon\,\mathrm{CSP}}$: the number of r-constraints is at most $|B|^3$ compared to the $|B|^4$-many s-constraints required by τ, and no s-constraint can fail into $\mathbb{2}^+$. For +1-in-3SAT itself we can show that $(Y_{(k,\mathscr{F}),\mathtt{Q}}, N_{\varepsilon\,\mathrm{CSP}})$ remains NP-complete.

References

1. Allender, E., Bauland, M., Immerman, N., Schnoor, H., Vollmer, H.: The complexity of satisfiability problems: refining Schaefer's Theorem. J. Comput. System Sci. **75**, 245–254 (2009)
2. Almeida, J.: Finite Semigroups and Universal Algebra. World Scientific, Singapore (1994)
3. Abramsky, S., Gottlob, G., Kolaitis, P.G.: Robust constraint satisfaction and local hidden variables in quantum mechanics. In: IJCAI 2013, pp. 440–446 (2013)
4. Arora, S., Safra, S.: Probabilistic checking of proofs: a new characterization of NP. J. ACM **45**(1), 70–122 (1998)
5. Barto, L., Kozik, M.: Constraint satisfaction problems of bounded width. In: Proceedings of FOCS 2009 (2009)
6. Barto, L., Kozik, M., Niven, T.: 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**, 1782–1802 (2008/2009)
7. Beacham, A., Culberson, J.: On the complexity of unfrozen problems. Disc. Appl. Math. **153**, 3–24 (2005)
8. Bulatov, A.A.: A dichotomy theorem for constraint satisfaction problems on a 3-element set. J. ACM **53**, 66–120 (2006)
9. Bulatov, A.A.: Complexity of conservative constraint satisfaction problems. ACM Trans. Comput. Log. **12**(4), 24 (2011)
10. Bulatov, A.A.: A dichotomy theorem for nonuniform CSPs. arXiv:1703.03021v2
11. Bulatov, A.A., Jeavons, P.G., Krokhin, A.: Classifying the complexity of constraints using finite algebras. SIAM J. Comput. **34**(3), 720–742 (2005)
12. Caicedo, X.: Finitely axiomatizable quasivarieties of graphs. Algebra Univers. **34**, 314–321 (1995)
13. Chen, H., Larose, B.: Asking the metaquestions in constraint tractability. arxiv:1604.00932
14. Feder, T., Vardi, M.Y.: The computational structure of monotone monadic SNP and constraint satisfaction: a study through datalog and group theory. SIAM J. Comput. **28**(1), 57–104 (1998)

15. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman & Co., New York (1979)
16. Geiger, D.: Closed systems of functions and predicates. Pacific J. Math. **27**, 95–100 (1968)
17. Gorbunov, V.A.: Algebraic Theory of Quasivarieties. Consultants Bureau, New York (1998)
18. Gottlob, G.: On minimal constraint networks. In: Lee, J. (ed.) CP 2011. LNCS, vol. 6876, pp. 325–339. Springer, Heidelberg (2011). doi:10.1007/978-3-642-23786-7_26
19. Goldreich, O.: On promise problems: a survey. In: Goldreich, O., Rosenberg, A.L., Selman, A.L. (eds.) Theoretical Computer Science. LNCS, vol. 3895, pp. 254–290. Springer, Heidelberg (2006). doi:10.1007/11685654_12
20. Ham, L.: A gap trichotomy theorem for Boolean constraint problems: extending Schaefer's theorem. In: ISAAC 2016, pp. 36: 1–36: 12 (2016)
21. Ham, L.: Gap theorems for robust satisfiability of constraint problems: Boolean CSPs and beyond. Theoret. Comp. Sci. **676**, 69–91 (2017)
22. Hell, P., Nešetřil, J.: On the complexity of H-colouring. J. Combin. Theory Ser. B **48**(1), 92–110 (1990)
23. Idziak, P., Markovic, P., McKenzie, R., Valeriote, M., Willard, R.: Tractability and learnability arising from algebras with few subpowers. SIAM J. Comput. **39**, 3023–3037 (2010)
24. Jackson, M.: Flexible constraint satisfiability and a problem in semigroup theory. arXiv:1512.03127
25. Jackson, M., Kowalski, T., Niven, T.: Digraph related constructions and the complexity of digraph homomorphism problems. Int. J. Algebra Comput. **26**, 1395–1433 (2016)
26. Jackson, M., Trotta, B.: Constraint satisfaction, irredundant axiomatisability and continuous colouring. Stud. Logica. **101**, 65–94 (2013)
27. Jeavons, P.: On the algebraic structure of combinatorial problems. Theor. Comput. Sci. **200**(1–2), 185–204 (1998)
28. Jeavons, P., Cohen, D.A., Martin, C.: Cooper.: constraints, consistency and closure. Artif. Intell. **101**(1–2), 251–265 (1998)
29. Jeavons, P., Cohen, D., Gyssens, M.: A unifying framework for tractable constraints. In: Montanari, U., Rossi, F. (eds.) CP 1995. LNCS, vol. 976, pp. 276–291. Springer, Heidelberg (1995). doi:10.1007/3-540-60299-2_17
30. Jeavons, P., Cohen, D.A., Gyssens, M.: Closure properties of constraints. J. ACM **44**, 527–548 (1997)
31. Jeavons, P., Cohen, D.A., Pearson, J.: Constraints and universal algebra. Ann. Math. Artif. Intell. **24**(1–4), 51–67 (1998)
32. Jonsson, P., Krokhin, A.: Recognizing frozen variables in constraint satisfaction problems. Theoret. Comp. Sci. **329**, 93–113 (2004)
33. Larose, B., Tesson, P.: Universal algebra and hardness results for constraint satisfaction problems. Theoret. Comput. Sci. **410**, 1629–1647 (2009)
34. Maltsev, A.I.: Algebraic Systems. Springer, Heidelberg (1973)
35. Maróti, M., McKenzie, R.: Existence theorems for weakly symmetric operations. Algebra Univers. **59**, 463–489 (2008)
36. Marques-Silva, J.P., Sakallah, K.A.: GRASP: a search algorithm for propositional satisfiability. IEEE Trans. Comput. **48**, 506–521 (1999)
37. Monasson, R., Zecchina, R., Kirkpatrick, S., Selman, B., Troyansky, L.: Determining computational complexity from characteristic phase transitions. Nature **400**, 133–137 (1998)

38. Nešetřil, J., Pultr, A.: On classes of relations and graphs determined by subobjects and factorobjects. Disc. Math. **22**, 287–300 (1978)
39. Rafiey, A., Kinne, J., Feder, T.: Dichotomy for digraph homomorphism problems, arXiv:1701.02409v2
40. Schaefer, T.J.: The complexity of satisfiability problems. In: STOC , pp. 216–226 (1978)
41. Valeriote, M.A.: A subalgebra intersection property for congruence distributive varieties. Canad. J. Math. **61**, 451–464 (2009)
42. Zhuk, D.: The proof of the CSP dichotomy conjecture, arXiv:1704.01914