The Polynomial Hierarchy and ω-Categorical CSPs

Santiago Guzmán-Pro
Technische Universität Dresden, Germany

Jakub Rydval
Technische Universität Wien, Austria

Abstract

In 2008, Bodirsky and Grohe showed that for every Π_n^{P}-level of the Polynomial Hierarchy (PH) there are ω-categorical Constraint Satisfaction Problems (CSPs) complete for this level. We show that, in fact, there are ω-categorical CSPs complete for any level of the PH. To this end, we use a recent result of Bodirsky, Knäuer, and Rudolph for constructing ω-categorical CSPs from sentences of Monadic Second-Order logic (MSO) with certain preservation properties. As a secondary contribution, we develop a new tool for producing MSO sentences satisfying said preservation properties.

2012 ACM Subject Classification Theory of computation → Logic; Theory of computation → Computational complexity and cryptography

Keywords and phrases monadic second-order logic, constraint satisfaction, homomorphism-closed, ω-categoricity, polynomial hierarchy, conjunctive query, primitive positive formula, quantifiers

Digital Object Identifier 10.4230/LIPIcs.MFCS.2026.96

Related Version *Full Version*: https://arxiv.org/abs/2604.24539

Funding *Santiago Guzmán-Pro*: This research is funded by the European Union (ERC, POCOCOP, 101071674). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.
Jakub Rydval: This research was funded in whole or in part by the Austrian Science Fund (FWF) [ESP 1571225]. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

1 Introduction

The *Constraint Satisfaction Problem* (CSP) of a structure $\mathfrak{B}$ in a finite relational signature τ, denoted by CSP($\mathfrak{B}$), is the computational problem of deciding whether a given finite τ-structure $\mathfrak{A}$ homomorphically maps to $\mathfrak{B}$. The general CSP framework is remarkably expressive; in fact, it contains all decision problems up to polynomial-time Turing equivalence [7]. For this reason, CSPs are typically only studied under additional structural restrictions on $\mathfrak{B}$. One natural restriction is requiring the domain of the *template* structure $\mathfrak{B}$ to be finite. The class of finite-domain CSPs famously constitutes a large fragment of NP that admits a dichotomy between P and NP-completeness; this was proved in 2017 independently by Bulatov [13] and Zhuk [23, 24].

Besides its complexity-theoretic significance (in particular, of confirming the dichotomy conjecture of Feder and Vardi [17]), the result of Bulatov and Zhuk has the additional appeal that the border for tractability can be described by a neat algebraic condition for the *polymorphisms* of $\mathfrak{B}$ [3, 22]. A polymorphism of a relational structure $\mathfrak{B}$ is simply a homomorphism from a finite power of $\mathfrak{B}$ into $\mathfrak{B}$ itself.

The success of the *algebraic approach* in the finite-domain setting naturally raises the question of how far it extends to the infinite-domain case. Although infinite-domain CSPs in general can attain arbitrary complexities, there are interesting subclasses whose template cannot be captured by any finite-domain CSP, yet whose complexity is still determined by the polymorphisms of the template. One of the most celebrated examples is the family of *temporal* CSPs, i.e., CSPs of first-order reducts of $(\mathbb{Q}; <)$. Similar to finite-domain CSPs, temporal CSPs also enjoy a P vs. NP-complete dichotomy with a clean algebraic counterpart [8]. Based on strong empirical evidence, Bodirsky and Pinsker [11] conjectured that a P vs. NP-complete dichotomy characterized in terms of polymorphisms extends to all CSPs of *reducts of finitely bounded homogeneous* structures, forming a computationally well-behaved subclass of ω*-categorical* structures [2], and providing an infinite-domain analogue of the dichotomy theorem of Bulatov and Zhuk.

The notion of ω-categoricity was popularized by Bodirsky and Nešetřil [10] as a sufficient restriction ensuring that primitive positive definability of relations is determined by polymorphisms. A countable structure $\mathfrak{A}$ is ω-categorical if, for every positive integer k, the number of k-ary relations first-order definable in $\mathfrak{A}$ is finite [19]; with $(\mathbb{Q}; <)$ being a prototypical example. Bodirsky and Nešetřil showed that if two countable ω-categorical structures share the same set of polymorphisms, then they also share the same set of primitive positive definable relations [10]. Since expanding a CSP template by a primitive positive definable relation does not impact the complexity of the CSP (up to LOGSPACE-reductions), we say that the complexity of an ω-categorical CSP is determined by the polymorphisms of its template. Clearly, all finite structures are ω-categorical, but this class also contains numerous interesting infinite structures, and most of the research on the *algebraic methods* for infinite-domain CSPs has focused on this setting [1, 2, 3, 4, 12, 21].

We thus believe it is a central question to better understand which complexity classes appear in the complexity landscape of ω-categorical CSPs, and more specifically, which computational problems can be represented as ω-categorical CSPs.

Bodirsky and Grohe [7] in fact already considered the first question in 2008. From Theorem 2 in their paper it follows that, if $\mathcal{C}$ is any complexity class for which there exist $\mathrm{coNP}^{\mathcal{C}}$-complete problems, then there exists an ω-categorical structure $\mathfrak{B}$ with a finite relational signature such that $\mathrm{CSP}(\mathfrak{B})$ is $\mathrm{coNP}^{\mathcal{C}}$-complete. In particular, there are ω-categorical CSPs complete for the Π^{P}_n-levels of the *Polynomial Hierarchy* (PH). This result later received a universal-algebraic upgrade [18, Theorem 1.8], showing that the algebraic conditions for polymorphisms (*non-trivial identities*) that are expected to witness tractability of ω-categorical CSPs in NP can actually be satisfied by the polymorphisms of ω-categorical structures whose CSPs are complete for any complexity class lying above coNP, including coNP itself. Furthermore, the construction of $\mathrm{coNP}^{\mathcal{C}}$-complete ω-categorical CSPs from [7] also allows one to replicate Ladner's method of "blowing holes into SAT". This was used by Bodirsky and Grohe [7, Theorem 4] to show the existence of coNP-intermediate ω-categorical CSPs (unless P = coNP). In other words, there is no dichotomy between P and coNP-completeness for ω-categorical CSPs in coNP.

All results mentioned in the previous paragraph follow a similar strategy. First, given a formal language $\mathcal{L}$, its complement is turned into a set $\mathcal{F}$ of structures with complete Gaifman graphs.[1] Next, a class of finite structures $\mathcal{K}$ is defined by forbidding homomorphisms from all structures in $\mathcal{F}$. Since the Gaifman graph of each structure in $\mathcal{F}$ is complete, $\mathcal{K}$ possesses two important structural properties: preservation under unions (equivalently, the *free amalgamation property* [5]) and preservation under inverse homomorphisms. Via Fraïssé's theorem [19], this is enough to ensure that $\mathcal{K}$ is the CSP of an ω-categorical structure $\mathfrak{B}$. Solving the CSP of this structure $\mathfrak{B}$ amounts to verifying whether every complete substructure

[1] Two domain elements are adjacent in the Gaifman graph if they appear jointly in a relational tuple.

of an instance $\mathfrak{A}$ is not an encoding of a word from the complement of $\mathcal{L}$. This technique therefore naturally yields problems complete for any complexity class of the form $\mathrm{coNP}^{\mathcal{C}}$, while it is unlikely to produce completeness $\mathrm{CSP}(\mathfrak{B})$ for classes of the form $\mathrm{NP}^{\mathcal{C}}$. Indeed, to the best of our knowledge, no Σ_n^{P}-complete ω-categorical CSP has appeared in the literature for any $n \geq 2$ In the present article, we close this gap. To this end, we follow an alternative construction of ω-categorical CSPs which, in contrast to the Bodirsky–Grohe construction, is not inherently tied to the $\mathrm{coNP}^{\mathcal{C}}$ form, and we use it to exhibit examples of Σ_n^{P}-complete ω-categorical CSPs for $n \geq 2$.

Contribution

In 2020, Bodirsky, Knäuer, and Rudolph [9] identified a new general sufficient condition for the existence of ω-categorical CSPs. Intuitively, this condition relaxes the preservation under unions to disjoint unions, compensated by an additional logical definability requirement.

▶ **Theorem 1** (cf. Corollary 17 in [9]). *Let $\mathcal{K}$ be a class of finite structures that is preserved under inverse homomorphisms and disjoint unions and definable in monadic second-order logic. Then there exists a countable ω-categorical structure $\mathfrak{D}$ such that $\mathcal{K} = \mathrm{CSP}(\mathfrak{D})$.*

We employ Theorem 1 as a base construction method for showing that there exist ω-categorical CSPs complete for any level of the PH. Our key technical contribution concerns preservation properties of SO sentences (Section 3). Firstly, we pinpoint a syntactic condition, which we call $\forall$-*restriction*, that characterizes SO sentences preserved under substructures. Combining $\forall$-*restriction* with the standard notion of *negativity*, we characterize SO sentences preserved under inverse injective homomorphisms (Section 3.2). Secondly, we identify a syntactic transformation of SO sentences – which we call the "CSP hammer" – that turns $\forall$-restricted negative MSO sentences into MSO sentences that are further preserved under inverse homomorphisms and disjoint unions (Section 4). This yields a general method for constructing ω-categorical CSPs, which is fundamental for our applications in Sections 5 and 6. There, we construct specific MSO sentences Φ_n with such syntactic properties, and suitable for the CSP hammer. These sentences Φ_n are tailored so that their adjustment using the CSP hammer produces Σ_n^{P}-complete ω-categorical CSPs. On the way, we establish a previously unknown connection between *bounded alternation Quantified CSPs* (QCSPs) and ω-categorical CSPs – QCSPs will be shortly introduced.

▶ **Theorem 2.** *For every finite structure $\mathfrak{B}$ and every quantifier-prefix $\mathrm{Q}_1 \cdots \mathrm{Q}_n$ with at least one universal quantifier, there exists an ω-categorical structure $\mathfrak{D}$ such that $\mathrm{CSP}(\mathfrak{D})$ is expressible in $\mathrm{Q}_1 \cdots \mathrm{Q}_n$MSO and $\mathrm{Q}_1 \cdots \mathrm{Q}_n\mathrm{QCSP}(\mathfrak{B})$ reduces in polynomial-time to $\mathrm{CSP}(\mathfrak{D})$.*

Since there are finite-domain bounded alternation QCSPs complete for the $\Sigma_{2n+1}^{\mathrm{P}}$- and Π_{2n}^{P}-levels of the polynomial hierarchy, Theorem 2 has the following consequence.

▶ **Theorem 3.** *For every $n \geq 1$, there are $\Sigma_{2n+1}^{\mathrm{P}}$- and Π_{2n}^{P}-complete ω-categorical CSPs.*

We cannot use Theorem 2 to obtain complete problems for the Σ_{2n}^{P}- or Π_{2n+1}^{P}-levels of the PH because finite-domain $(\exists\forall)^n$QCSPs cannot ever be Σ_{2n}^{P}- or Π_{2n+1}^{P}-complete. We then prove the following analogue of Theorem 3 for the remaining levels.

▶ **Theorem 4.** *For every $n \geq 1$, there are Σ_{2n}^{P}- and Π_{2n+1}^{P}-complete ω-categorical CSPs.*

The existence of Π_n^{P}-complete ω-categorical CSPs was originally proved in [7]. However, our methodology is robust enough to naturally yield Π_n^{P}-complete ω-categorical CSPs while constructing Σ_n^{P}-complete ω-categorical CSPs. These results, and the fact that there are NP-complete finite-domain CSPs (e.g., 3-Sat), together prove our main theorem.

▶ **Theorem 5.** *For every level Σ_n^{P} or Π_n^{P} $(n \geq 1)$ in the polynomial hierarchy, there exists an ω-categorical structure $\mathfrak{D}$ such that* $\mathrm{CSP}(\mathfrak{D})$ *is complete for that level.*

2 Preliminaries

We assume familiarity with basic computation complexity. In particular, we denote by Σ_n^{P} the complexity class $\mathrm{NP}^{\Sigma_{n-1}^{\mathrm{P}}}$ (i.e., languages recognizable in non-deterministic polynomial time with a $\Sigma_{n-1}^{\mathrm{P}}$ oracle), where $\Sigma_n^{\mathrm{P}} = \mathrm{P}$. Similarly, Π_n^{P} is the complexity class $\mathrm{coNP}^{\Pi_{n-1}^{\mathrm{P}}}$, and equivalently, Π_n^{P} is the collections of languages whose complement is in Σ_n^{P}. It is not hard to notice that $\Sigma_{n-1}^{\mathrm{P}} \subseteq \Pi_n^{\mathrm{P}} \subseteq \Sigma_{n+1}^{\mathrm{P}}$. The *polynomial hierarchy* (PH) is the union $\bigcup_{n\in\mathbb{N}} \Sigma_n^{\mathrm{P}}$.

We use the bar notation $\bar{\cdot}$ for tuples, and denote the set $\{1,\ldots,n\}$ by $[n]$. We extend the membership relation on sets to tuples by ignoring the ordering on the entries. For example, we might write $x \in \bar{x}$ to denote that x is an entry in $\bar{x}$.

Structures. A (*relational*) *signature* τ is a set of *relation symbols*, where each $R \in \tau$ is associated with a natural number called *arity*. A (*relational*) τ*-structure* $\mathfrak{A}$ consists of a set A (the *domain*) together with the relations $R^{\mathfrak{A}} \subseteq A^k$ for each $R \in \tau$ with arity k. An *expansion* of $\mathfrak{A}$ is a σ-structure $\mathfrak{A}^+$ with $A = A^+$ such that $\tau \subseteq \sigma$ and $R^{\mathfrak{A}^+} = R^{\mathfrak{A}}$ for each relation symbol $R \in \tau$. Conversely, we then call $\mathfrak{A}$ a *reduct* of $\mathfrak{A}^+$. We denote by $(\mathfrak{A}, R)$ the expansion of a structure $\mathfrak{A}$ by a relation R over its domain. The *substructure* of a τ-structure $\mathfrak{B}$ on a subset $A \subseteq B$ is the τ-structure $\mathfrak{A}$ with domain A and relations $R^{\mathfrak{A}} = R^{\mathfrak{B}} \cap A^k$ for every $R \in \tau$ of arity k. Conversely, we call $\mathfrak{B}$ a *superstructure* of $\mathfrak{A}$. The *union* of two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ is the τ-structure $\mathfrak{A} \cup \mathfrak{B}$ with domain $A \cup B$ and relations $R^{\mathfrak{A}\cup\mathfrak{B}} := R^{\mathfrak{A}} \cup R^{\mathfrak{B}}$ for every $R \in \tau$. If $A \cap B = \emptyset$, we call $\mathfrak{A} \cup \mathfrak{B}$ a *disjoint union* and write $\mathfrak{A} + \mathfrak{B}$. A *homomorphism* $h\colon \mathfrak{A} \to \mathfrak{B}$ for τ-structures $\mathfrak{A}, \mathfrak{B}$ is a mapping $h\colon A \to B$ that *preserves* each relation of τ, i.e., whenever $\bar{t} \in R^{\mathfrak{A}}$ for some relation symbol $R \in \tau$, then $h(\bar{t})$ (computed componentwise) is an element of $R^{\mathfrak{B}}$. We write $\mathfrak{A} \to \mathfrak{B}$ if $\mathfrak{A}$ maps homomorphically into $\mathfrak{B}$. A class $\mathcal{K}$ of finite structures is *preserved* or *closed* under:

- *homomorphisms* if, for every $\mathfrak{A} \in \mathcal{K}$, every finite structure $\mathfrak{B}$ with a homomorphism from $\mathfrak{A}$ is also contained in $\mathcal{K}$;
- *inverse homomorphisms* if, for every $\mathfrak{B} \in \mathcal{K}$, every finite structure $\mathfrak{A}$ with a homomorphism to $\mathfrak{B}$ is also contained in $\mathcal{K}$;
- *disjoint unions* if, for all $\mathfrak{A}, \mathfrak{B} \in \mathcal{K}$, the union of two isomorphic copies of $\mathfrak{A}$ and $\mathfrak{B}$ with disjoint domains is also contained in $\mathcal{K}$.

Logic. We assume that the reader is familiar with classical *First-Order* logic FO. A first-order τ-formula ϕ is *primitive positive* (pp) if it is of the form $\exists x_1,\ldots,x_m(\phi_1 \wedge \cdots \wedge \phi_n)$, where each ϕ_i is *atomic*, i.e., of the form $\bot$, $(x_j = x_k)$, or $R(x_{i_1},\ldots,x_{i_\ell})$ for some $R \in \tau$. *Quantified primitive positive* (qpp) formulas generalize pp-formulas by allowing both existential *and* universal quantification. We use the stylized letter Q for the quantifiers $\exists$ and $\forall$; given $\mathrm{Q} \in \{\exists,\forall\}$, we denote the complementary quantifier by $\overline{\mathrm{Q}}$.

Second-order logic SO is the extension of FO which additionally allows existential and universal quantification over relations; that is, if R is a relation symbol and ϕ is a SO $\tau \cup \{R\}$-formula, then $\exists R.\, \phi$ and $\forall R.\, \phi$ are SO τ-formulas. *Monadic second-order logic* MSO is the restriction of SO to unary SO variables. In $\exists R.\, \phi$ and $\forall R.\, \phi$, we then refer to R as a *SO variable*. The *model-checking problem* for a fixed τ-formula ϕ is the computational problem of deciding whether ϕ holds in a given finite τ-structure $\mathfrak{A}$.

We say that a formula ϕ is *k-ary* if it has k free FO variables; we use the notation $\phi(\bar{x})$ to indicate that the free FO variables of ϕ are among $\bar{x}$. This does not mean that the truth value of ϕ depends on each entry in $\bar{x}$. A *sentence* is a formula without free FO variables; we reserve uppercase letters for sentences and lowercase for formulas. We say that a formula ϕ is *normalized* if it is in prenex normal form. By this we mean that ϕ starts with an SO quantifier-prefix, followed by an FO quantifier-prefix, and ends with a quantifier-free formula. We say that ϕ is *CNF-* or *DNF-normalized* if it is normalized and its quantifier-free part is in *Conjunctive Normal Form* (CNF) or *Disjunctive Normal Form* (DNF), respectively. With a simple inductive argument over the definition of SO, one can verify that every SO formula is logically equivalent to a CNF- and a DNF-normalized one. We use $\mathrm{fm}(\Phi)$ to denote the class of all finite models of a sentence Φ; when we say that Φ is closed under inverse homomorphisms or disjoint unions, we mean that $\mathrm{fm}(\Phi)$ has the respective property.

We say that a class $\mathcal{L}$ of SO formulas is *closed under internal conjunctions and disjunctions* if the following holds: if ϕ_1 and ϕ_2 are $\tau \cup \{S_1, \dots, S_n\}$-formulas and $\mathrm{Q}_1, \dots, \mathrm{Q}_n$ quantifiers such that the τ-formulas $\mathrm{Q}_1 S_1 \cdots \mathrm{Q}_n S_n.\, \phi_1$ and $\mathrm{Q}_1 S_1 \cdots \mathrm{Q}_n S_n.\, \phi_2$ are contained in $\mathcal{L}$, then $\mathrm{Q}_1 S_1 \cdots \mathrm{Q}_n S_n \big(\phi_1 \wedge \phi_2\big)$ and $\mathrm{Q}_1 S_1 \cdots \mathrm{Q}_n S_n \big(\phi_1 \vee \phi_2\big)$ are contained in $\mathcal{L}$ as well. For a CNF- or DNF-normalized SO formula ϕ, we say that ϕ is *positive* (*negative*) if atomic τ-formulas of the form $R(\bar{x})$ for $R \in \tau$ may only appear positively (negatively) in clauses of ϕ, i.e., every negative (positive) atom contains an SO variable or the default equality predicate. Given a syntactic condition on CNF-normalized SO formulas, we *naturally* extend it to normalized SO formulas ϕ by instead requiring the syntactic condition to be satisfied by some formula ϕ' obtained from ϕ by converting the quantifier-free part of ϕ into CNF using the De Morgan's laws and the distributive law of propositional logic. In particular, we naturally extend the notions of positive and negative CNF-normalized SO formulas to normalized SO formulas. Note that we could have equivalently used DNF instead of CNF. Finally, we define the sets of *positive* and *negative SO formulas* (with SO prefix upfront) as the closure of positive and negative normalized SO formulas under internal conjunctions and disjunctions, respectively. Alternatively, we could define positivity and negativity in terms of τ-atoms being under the scope of evenly and oddly many negations, respectively. However, the current definition is more suitable for our applications.

▶ **Example 6.** The following MSO $\{E\}$-sentence Φ, testing whether a binary relation E contains a directed cycle, is positive:

$$\begin{aligned}\exists S\, \forall C \big(\exists x.\, E(x,x)\big) \vee \Big(&\big(\exists x, y.\, S(x) \wedge S(y) \wedge \neg(x = y)\big)\\ \wedge \Big(&\big(\exists a, b.\, S(a) \wedge S(b) \wedge C(a) \wedge E(a,b) \wedge \neg C(b)\big)\\ &\vee \big(\forall x, y.\, \neg S(x) \vee \neg S(y) \vee \neg C(x) \vee C(y)\big)\Big)\Big).\end{aligned}$$

Negating Φ and applying De Morgan's laws instead yields a negative sentence.

It is useful to take the following game-theoretic perspective at SO (see, e.g., [16]). The evaluation of an SO τ-formula ϕ on a structure $\mathfrak{A}$ corresponds to a game between two players: an *existential player* (EP) and a *universal player* (UP) who assign values to the existentially and universally quantified variables, respectively. To every moment of the game we associate a partial function $[\![\cdot]\!]$ describing values assigned to the variables of ϕ by either of the players:

- if x is a FO variable, then $[\![x]\!]$ is either undefined or an element of A;
- if S is an SO variable of arity k, then $[\![S]\!]$ is either undefined or a subset of A^k.

The goal of the EP is to satisfy the quantifier-free part of ϕ evaluated at $\bar{a}$; the goal of the UP is the opposite. Clearly, $\mathfrak{A} \models \phi(\bar{a})$ if the EP has a winning strategy in this game, i.e., can respond to all moves of the UP while keeping the quantifier-free part of ϕ evaluated at $\bar{a}$ satisfied. Otherwise, $\mathfrak{A} \not\models \phi(\bar{a})$ and the UP has a winning strategy, i.e., can violate the quantifier-free part of ϕ evaluated at $\bar{a}$ regardless of the moves of the EP.

Constraint satisfaction problems. The *(Quantified) Constraint Satisfaction Problem* for a structure $\mathfrak{B}$, denoted by (Q)CSP($\mathfrak{B}$), is the computational problem of deciding whether a given (q)pp τ-sentence holds in $\mathfrak{B}$. By *constraints*, we refer to the conjuncts in the quantifier-free part of a given instance of (Q)CSP($\mathfrak{B}$). In the CSP framework, we also consider the homomorphism perspective, where an instance is viewed as a finite τ-structure. Then CSP($\mathfrak{B}$) is (the membership problem for) the class of all finite structures $\mathfrak{A}$ which homomorphically map into $\mathfrak{B}$. The two definitions are equivalent [14]. *Bounded alternation* QCSPs are QCSPs restricted to a fixed quantifier prefix $\mathrm{Q}_1 \dots \mathrm{Q}_n$. In particular, CSPs are bounded alternation QCSPs restricted to the quantifier prefix $\exists$. Finite-domain *bounded alternation* QCSPs have been investigated in [15].

We usually think of an instance of a QCSP as a game between two players: an *existential player* (EP) and a *universal player* (UP) who assign values to the existentially and universally quantified variables, respectively. To every moment of the game we associate a partial function $[\![\cdot]\!]$ from the variables into the domain of the parametrizing structure $\mathfrak{B}$ describing values assigned to the variables by either of the players. The instance is true if and only if the EP has a winning strategy in this game, i.e., can respond to all moves of the UP while keeping all constraints satisfied. Otherwise, the instance is false and the UP has a winning strategy, i.e., can violate a constraint regardless of the moves of the EP.

3-CNF. The *satisfiability problem* for 3-CNF asks whether a given propositional formula in conjunctive normal form with at most three variables per clause has a satisfying assignment. This problem can naturally be represented as a finite-domain CSP by interpreting each clause as a ternary relational constraint over $\{0, 1\}$, e.g., $(\neg x_i \vee x_j \vee x_k)$ as $R_{100}(x_i, x_j, x_k)$, where $R_{100} \coloneqq \{0, 1\}^3 \setminus \{(1, 0, 0)\}$. Allowing existential *and* universal quantification over the input variables yields the computational decision problem Quantified 3-CNF. Naturally, we denote by $(\forall\exists)^n$3-CNF and $\exists(\forall\exists)^n$3-CNF the fragment of Quantified 3-CNF where we only consider quantifier prefixes of the form $(\forall\exists)^n$ and $\exists(\forall\exists)^n$, respectively.

3 Preservation lemma

We begin this section by giving a syntactic characterization of the fragment of SO closed under injective homomorphisms (Lemma 10). We then dualize the said syntactic characterization to obtain one that characterizes the fragment of SO closed under inverse injective homomorphisms (Lemma 14). In Section 4, we convert the dualized syntactic characterization into a tool for producing SO sentences closed under inverse homomorphisms and disjoint unions. This tool will be used in Sections 5 and 6 to prove our main results.

3.1 Injective homomorphisms

In Lemma 10, we give a syntactic condition for SO sentences which implies closure under injective homomorphisms. Bodirsky, Feller, Knäuer, and Rudolph [6, Theorem 59] recently identified a syntactic transformation $\Phi \mapsto (\Phi^{\mathrm{shom}})^{\mathrm{sup}}$ (shortly introduced) for SO sentences with the property that $\mathrm{fm}(\Phi) = \mathrm{fm}((\Phi^{\mathrm{shom}})^{\mathrm{sup}})$ if and only if Φ is preserved under homomorphisms. Here, we present a syntactic condition for SO sentences which implies closure under injective homomorphisms (Lemma 10). In particular, for every normalized SO sentence Φ, the sentence $(\Phi^{\mathrm{shom}})^{\mathrm{sup}}$ satisfies our syntactic condition for preservation under injective homomorphisms, and hence our condition captures the fragment of SO closed under injective homomorphisms up to logical equivalence.

We begin by defining the transformations $\Phi \mapsto \Phi^{\mathrm{shom}}$ and $\Phi \mapsto \Phi^{\mathrm{sup}}$ from [6]. They define Φ^{shom} as follows

$$\exists_{R\in\tau} R_p \big(\Phi'_p \wedge \bigwedge\nolimits_{R\in\tau} \forall \bar{x}.\, \neg R_p(\bar{x}) \vee R(\bar{x})\big),$$

where Φ'_p is obtained from Φ by replacing every atom of the form $R(\bar{x})$ for $R \in \tau$ with $R_p(\bar{x})$. The SO τ-sentence Φ^{sup} is obtained from Φ by adding:

- a fresh SO variable S existentially quantified at the beginning of the SO quantifier prefix,
- a fresh FO variable z existentially quantified at the beginning of the FO quantifier prefix, together with the clause $S(z)$,
- for every existential FO variable y, a clause ψ that contains a disjunct $\neg S(x)$ for every universal FO variable x, and a disjunct $S(y)$,
- for every universal FO variable x, the disjunct $\neg S(x)$ to every clause in the quantifier-free part of Φ.

Now, we proceed to introduce our syntactic conditions.

▶ **Definition 7.** *A CNF-normalized SO sentence Φ is* $\exists$-guarded *if, for every (disjunctive) clause ϕ in Φ and every universally quantified FO variable x in ϕ, there is a disjunct $\neg S(\bar{x})$ in ϕ for some existentially quantified SO variable S such that $x \in \bar{x}$. We* naturally[2] *extend this definition to normalized SO sentences. The set of* $\exists$-guarded SO sentences *with SO quantifiers up front is the closure of $\exists$-guarded normalized SO sentences under internal conjunctions and disjunctions.*

It is easy to verify that the sentence in Example 6 is $\exists$-guarded; we only need to break it down to its internal CNF-normalized conjuncts and disjuncts, and apply Definition 7 inductively. It is also not hard to observe that Φ^{sup} is $\exists$-guarded.

Intuitively, the condition of $\exists$-guarding allows the EP to restrict the game of evaluating an SO sentence in a structure $\mathfrak{A}$ to any particular substructure $\mathfrak{B}$; all the EP has to do is to evaluate every existentially quantified SO variable within B. This reasoning can be used to show that every $\exists$-guarded SO sentence is preserved under superstructures, because if $\mathfrak{B}$ satisfies the SO sentence, then the EP can choose to restrict the game on $\mathfrak{B}$. The following lemma extends [6, Corollary 55] to a syntactic characterization of the SO-fragment closed under superstructures.

▶ **Lemma 8.** *For an SO sentence Φ, the following are equivalent:*

1. *Φ is preserved under superstructures.*
2. *Φ is logically equivalent to an $\exists$-guarded SO sentence.*

Proof. We only prove the statement for normalized SO sentences; it will become clear that the argument extends to the full inductive definition. Let τ be the signature of Φ.

"(2)⇒(1)" We may assume that Φ itself is CNF-normalized; otherwise we replace its quantifier-free part with the CNF version of it witnessing $\exists$-guarding. Suppose that $\mathfrak{A} \models \Phi$ and $\mathfrak{B}$ is a superstructure of $\mathfrak{A}$. We claim that the EP has the following winning strategy on $\mathfrak{B}$ based on the winning strategy on $\mathfrak{A}$. For a SO variable S:

[2] See definition of positive normalized SO formulas.

- if S is existentially quantified and the EP chooses $X \subseteq A^k$ for $[\![S]\!]$ on $\mathfrak{A}$, then the EP chooses $X \subseteq A^k \subseteq B^k$ for $[\![S]\!]$ on $\mathfrak{B}$;
- if S is universally quantified and the UP reacts by choosing $Y \subseteq B^k$ for $[\![S]\!]$ on $\mathfrak{B}$, then we consider the situation where the UP chooses $Y \cap A^k \subseteq A^k$ for $[\![S]\!]$ on $\mathfrak{A}$;

For FO variables, the EP follows the winning strategy on $\mathfrak{A}$ and interprets any responses by the UP with values outside of A as arbitrary responses inside A.

Suppose that, when playing on $\mathfrak{B}$, the UP chooses $[\![x]\!] \in B \setminus A$. Since Φ is $\exists$-guarded, every clause ϕ contains a disjunct $\neg S(\bar{x})$ in ϕ for some existentially quantified SO variable S. By the strategy of the EP, we have that $[\![S]\!] \subseteq A^k$ for some k. Hence, this choice of the UP trivially satisfies ϕ. Alternatively, the strategy of the UP on $\mathfrak{B}$ simulates a strategy on $\mathfrak{A}$. Hence, the whole game on $\mathfrak{B}$ simulates a game on $\mathfrak{A}$, and by the choice of strategy for the EP, we conclude that the EP wins.

"(1)⇒(2)" This direction follows from [6, Corollary 55] using $\Phi \mapsto \Phi^{\text{sup}}$ because Φ^{sup} is clearly $\exists$-guarded. ◀

Next, Corollary 58 in [6] shows that $\text{fm}(\Phi) = \text{fm}(\Phi^{\text{shom}})$ if and only if Φ is preserved under surjective homomorphisms.

The following lemma provides a syntactic characterisation of the SO-fragment closed under bijective homomorphisms. It is compatible with $\Phi \mapsto \Phi^{\text{shom}}$ in the sense that Φ^{shom} is positive whenever Φ is a normalized SO sentence, but it does not extend [6, Corollary 58] in its full generality. Recall the definition of positivity for SO formulas from Section 2.

▶ **Lemma 9.** *For an SO sentence Φ, the following are equivalent:*

1. *Φ is preserved under bijective homomorphisms.*
2. *Φ is logically equivalent to a positive SO sentence.*

Proof. Again, we only prove the statement for normalized SO sentences; it will become clear that the argument extends to the full inductive definition. Let τ be the signature of Φ.

"(2)⇒(1)" We may assume that Φ itself is CNF-normalized; otherwise we replace its quantifier-free part with the CNF version of it witnessing positivity. Let $\mathfrak{A}$ and $\mathfrak{B}$ be two finite τ-structures over the signature τ of Φ such that $\mathfrak{A} \models \Phi$ and there exists a bijective homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$. Without loss of generality we assume that $A = B$, and that $R^{\mathfrak{A}} \subseteq R^{\mathfrak{B}}$ for all $R \in \tau$. We claim that the EP has the following winning strategy on $\mathfrak{B}$ based on the winning strategy on $\mathfrak{A}$: For a SO variable S:

- if S is existentially quantified and the EP chooses $X \subseteq A^k$ for $[\![S]\!]$ on $\mathfrak{A}$, then the EP chooses $X \subseteq B^k = A^k$ for $[\![S]\!]$ on $\mathfrak{B}$;
- if S is universally quantified and the UP reacts by choosing $Y \subseteq B^k$ for $[\![S]\!]$ on $\mathfrak{B}$, then we consider the situation where the UP chooses $Y \subseteq A^k = B^k$ for $[\![S]\!]$ on $\mathfrak{A}$.

The FO variables are treated analogously.

Denote by $\mathfrak{A}^+$ and $\mathfrak{B}^+$ the expansions of $\mathfrak{A}$ and $\mathfrak{B}$ resulting from the above evaluations of the SO variables in Φ. Let $\phi_1 \vee \cdots \vee \phi_k$ be a clause in the quantifier-free part of Φ. Since the strategy of the EP on $\mathfrak{A}$ is a winning strategy, there exists $i \in [k]$ such that $\mathfrak{A}^+ \models \phi_i([\![\bar{x}]\!])$. For every SO variable S, we have that $S^{\mathfrak{A}^+} = S^{\mathfrak{B}^+}$. Hence, if $\phi_i(\bar{x})$ contains an SO variable S, i.e., $\phi_i(\bar{x}) \in \{S(\bar{x}), \neg S(\bar{x})\}$, then $\mathfrak{B} \models \phi_i(h([\![\bar{x}]\!]))$ is immediate. Also, if ϕ_i is of the form $(x_1 = x_2)$ or $\neg(x_1 = x_2)$, then $\mathfrak{B} \models \phi_i(h([\![\bar{x}]\!]))$ is immediate because h is bijective. Otherwise, since Φ is positive, $\phi_i(\bar{x})$ must be of the form $R(\bar{x})$ for $R \in \tau$. Then $\mathfrak{B} \models \phi_i(h([\![\bar{x}]\!]))$ follows because $R^{\mathfrak{A}} \subseteq R^{\mathfrak{B}}$.

"(1)⇒(2)" This direction follows from [6, Corollary 58] using $\Phi \mapsto \Phi^{\text{shom}}$. ◀

The transformations $\Phi \mapsto \Phi^{\text{sup}}$ and $\Phi \mapsto \Phi^{\text{shom}}$ used to prove "(1)⇒(2)" in Lemmas 8 and 9 can be combined, yielding an $\exists$-guarded positive sentence. This observation shows "(1)⇒(2)" from the next theorem. Similarly, the implications "(2)⇒(1)" from Lemmas 8 and 9 show that "(3)⇒(1)" from the next lemma also holds. With these arguments, and the fact that (2) clearly implies (3), we conclude that the following lemma holds.

▶ **Lemma 10.** *For an SO sentence Φ, the following are equivalent:*
1. *Φ is preserved under injective homomorphisms.*
2. *Φ is logically equivalent to an $\exists$-guarded positive SO sentence.*
3. *Φ is logically equivalent to an $\exists$-guarded SO sentence and a positive SO sentence.*

3.2 Inverse injective homomorphisms

Note that $\mathcal{K}$ is closed under inverse injective homomorphisms if and only if the *complement* of $\mathcal{K}$, i.e., the class of all finite τ-structures not in $\mathcal{K}$, is closed under injective homomorphisms. Hence, by Lemma 10, Φ is logically equivalent to an $\exists$-guarded positive SO sentence if and only if $\neg\Phi$ is closed under inverse injective homomorphisms. The following fact can be shown using the De Morgan's laws and the distributive law of propositional logic.

▶ **Lemma 11.** *For every SO sentence $\Phi := \mathrm{Q}_1 S_1 \cdots \mathrm{Q}_m S_m.\, \Phi_*$, the following are equivalent:*
1. *there exists a positive SO sentence $\mathrm{Q}_1 S_1 \cdots \mathrm{Q}_m S_m.\, \Psi_\exists$ logically equivalent to Φ.*
2. *there exists a negative SO sentence $\overline{\mathrm{Q}}_1 S_1 \cdots \overline{\mathrm{Q}}_m S_m.\, \Psi_\forall$ logically equivalent to $\neg\Phi$.* ⌟

We can therefore say that negativity constitutes the *dual* notion to positivity. We shall now dualize the notion of $\exists$-guarding.

▶ **Definition 12.** *A DNF-normalized SO sentence Φ is* $\forall$-restricted *if, for every (conjunctive) clause ϕ in Φ and every existentially quantified FO variable x in ϕ, there is a conjunct $S(\bar{x})$ in ϕ for some universally quantified SO variable S such that $x \in \bar{x}$. We* naturally[3] *extend this definition to normalized SO sentences. The set of* $\forall$-restricted SO sentences *with SO prefix up front is the closure of $\forall$-restricted normalized SO sentences under internal conjunctions and disjunctions.*

Analogously to $\exists$-guarding, the condition of $\forall$-restriction allows the UP to restrict the game of evaluating an SO sentence in a structure to any particular substructure by evaluating every universally quantified SO variable within this substructure. Hence, the UP has the power to verify that the $\forall$-restricted SO sentence is satisfied on all substructures, and this reasoning can be used to show that every $\forall$-restricted SO sentence is preserved under substructures. Instead of giving the obvious counterpart to Lemma 9, we give a lemma that captures the duality between $\exists$-guarding and $\forall$-restriction.

▶ **Lemma 13.** *For every SO sentence $\Phi := \mathrm{Q}_1 S_1 \cdots \mathrm{Q}_m S_m.\, \Phi_*$, the following are equivalent:*
1. *there exists an $\exists$-guarded SO sentence $\mathrm{Q}_1 S_1 \cdots \mathrm{Q}_m S_m.\, \Psi_\exists$ logically equivalent to Φ.*
2. *there exists an $\forall$-restricted SO sentence $\overline{\mathrm{Q}}_1 S_1 \cdots \overline{\mathrm{Q}}_m S_m.\, \Psi_\forall$ logically equivalent to $\neg\Phi$.*

Note that Lemmas 11 and 13 are formulated in such a way that allows us to exploit the inductive nature of the involved definitions. For example, to check whether a given sentence is $\exists$-guarded, we can first decompose it into its normalized components with the same SO quantifier prefix, and then, for each component we check $\exists$-guarding, or verify $\forall$-restriction

[3] See definition of positive normalized SO formulas.

for its negation. The important part is that, after the application of Lemma 13, we keep the original SO quantifier prefix. Finally, combining Lemma 11 and Lemma 13, we obtain the following dual statement as a corollary to Lemma 10.

▶ **Lemma 14.** *For an SO sentence Φ, the following are equivalent:*

1. *Φ is preserved under inverse injective homomorphisms.*
2. *Φ is logically equivalent to a $\forall$-restricted negative SO sentence.*
3. *Φ is logically equivalent to a $\forall$-restricted SO sentence and a negative SO sentence.* ◀

4 The CSP hammer

Recall from the paragraph above Lemma 9 that the notion of positivity characterizing the preservation by bijective homomorphisms is compatible with the transformation $\Phi \mapsto \Phi^{\mathrm{shom}}$ from [6, Corollary 58] characterizing preservation by surjective homomorphisms. More specifically, Φ^{shom} is positive whenever Φ is a normalized SO sentence. However, not every positive SO sentence is preserved by surjective homomorphisms; dually, not every negative SO sentence is preserved under inverse surjective homomorphisms.

▶ **Example 15.** For a binary symbol N, consider the following MSO $\{N\}$-sentence:

$$\Phi := \forall A\, \exists E\, \forall x, y \Big(\big(\neg N(x,y) \vee \neg A(x) \vee E(y)\big) \wedge \big(\neg N(x,y) \vee \neg E(x) \vee A(y)\big) \Big).$$

It says that the EP has to copy the A-colouring of vertices by the UP with an E-colouring across N-edges. Clearly, Φ is CNF-normalized, negative because $\{N\}$-atoms only appear negatively, and $\forall$-restricted because it does not contain any existentially quantified FO variables. Hence, it is preserved under inverse injective homomorphisms (Lemma 14). However, it is not preserved under inverse surjective homomorphisms. Indeed, $\mathfrak{G} := ([3]; \{(1,2),(2,1),(3,2),(2,3)\})$ has a surjective homomorphism to $\mathfrak{K}_2 = ([2]; \{(1,2),(2,1)\})$. On $\mathfrak{K}_2$, the EP wins because every vertex only has a single neighbour. On $\mathfrak{G}$, however, the UP has a winning strategy by choosing to A-colour 1 while not A-colouring 3.

It is not clear to us how to generalize the notion of positivity (negativity) so that it would capture the full fragment of SO closed under (inverse) surjective homomorphisms. By extension, we face the same obstacle in identifying a fragment of SO that characterizes preservation under inverse homomorphisms; it is also worth noting that testing whether a given SO sentence is preserved by homomorphisms is in fact undecidable [6].

In the present section, we introduce the CSP hammer, a certain syntactic transformation $\Phi \mapsto \Phi^{\mathrm{CSP}}$ for $\forall$-restricted negative SO sentences. The purpose of the CSP hammer is to turn every $\forall$-restricted negative SO sentence into one that defines a CSP and whose model-checking is polynomial-time equivalent to the original when restricted to a particular subset of instances. To this end, we first introduce the *restriction* transformation on FO sentences. Let Φ be a FO τ-sentence, and U a fresh unary relation symbol not in τ. We set

$$\Phi_{|U} := \big(\forall z.\, \neg U(z)\big) \vee \Phi',$$

where Φ' is obtained from Φ by the following syntactic replacement. We replace every atomic or negated atomic formula $\psi(x_1,\dots,x_n)$ in Φ with the formula

$$\big(\bigvee\nolimits_{i\in I_\forall} \neg U(x_i)\big) \vee \big(\bigwedge\nolimits_{i\in I_\exists} U(x_i) \wedge \psi(x_1,\dots,x_n)\big), \tag{1}$$

where $I_\exists$ and $I_\forall$ consist of the indices $i \in [n]$ such that x_i is existentially and universally quantified in Φ, respectively. The purpose of $\Phi_{|U}$ is captured by the next observation.

▶ **Observation 16.** *Let $\mathfrak{A}$ be an arbitrary $(\tau \cup \{U\})$-structure with $U^{\mathfrak{A}} \neq \emptyset$. Then $\mathfrak{A}$ satisfies $\Phi_{|U}$ if and only if the substructure of $\mathfrak{A}$ on $U^{\mathfrak{A}}$ satisfies Φ.* ⌟

Now, let $\Phi := \mathrm{Q}_1 S_1 \dots \mathrm{Q}_n S_n.\, \Psi$ be a $\forall$-restricted negative SO τ-sentence such that Ψ is a FO $(\tau \cup \sigma)$-sentence for $\sigma := \{S_1, \dots, S_n\}$. Moreover, let U and N be fresh symbols of arities 1 and 2, respectively. Fix k to be the first index $i \in [n]$ such that $\mathrm{Q}_i = \forall$; if no such index exists, set $k = n + 1$. We set

$$\Phi^{\mathrm{CSP}} := \mathrm{Q}_1 S_1 \dots \mathrm{Q}_{k-1} S_{k-1} \forall U \mathrm{Q}_k S_k \dots \mathrm{Q}_n S_n.\, \big(\Psi^N_{\neg\mathrm{edge}} \vee (\Psi^N_{\neg\mathrm{loop}} \wedge \Psi)\big)_{|U} \tag{2}$$

where $\Psi^N_{\neg\mathrm{edge}}$ and $\Psi^N_{\neg\mathrm{loop}}$ are FO $(\tau \cup \sigma \cup \{N\})$-sentences defined below:

$$\Psi^N_{\neg\mathrm{edge}} := \exists x, y\big((x \neq y) \wedge \neg N(x, y)\big) \tag{3}$$

$$\Psi^N_{\neg\mathrm{loop}} := \forall x.\, \neg N(x, x). \tag{4}$$

It is not hard to see that $\Phi \mapsto \Phi^{\mathrm{CSP}}$ preserves negativity, and $\forall$-restriction – via the second disjunct in (1). It also preserves the alternation-depth for the second-order quantifier prefix of Φ whenever Φ contains at least one universally quantified SO variable. If there is no universal SO variable, then model-checking of Φ is in NP, which for us is an uninteresting case. We introduce some nomenclature which will be useful in the proof of the next lemma. We say that a $(\tau \cup \{N\})$-structure $\mathfrak{A}$ is:

- *N-surjective* if $\mathfrak{A}$ does not satisfy $\Psi^N_{\neg\mathrm{edge}}$;
- *N-injective* if $\mathfrak{A}$ satisfies $\Psi^N_{\neg\mathrm{loop}}$.

▶ **Observation 17.** *If a $\tau \cup \{N\}$-structure $\mathfrak{A}$ satisfies Φ^{CSP}, then it satisfies $\Psi^N_{\neg\mathrm{loop}}$.*

The next lemma contains the core properties of the CSP hammer.

▶ **Lemma 18.** *Let Φ be a $\forall$-restricted negative SO τ-sentence. Then all of the following hold:*

1. *Φ^{CSP} is preserved under disjoint unions.*
2. *Φ^{CSP} is preserved under inverse homomorphisms.*
3. *Model-checking for Φ reduces in polynomial time to model-checking for Φ^{CSP}.*

Proof. We assume that Φ is of the form $\mathsf{Q}_1 S_1 \dots \mathsf{Q}_n S_n.\, \Psi$, where Ψ a FO $(\tau \cup \sigma)$-sentence for $\sigma := \{S_1, \dots, S_n\}$. We verify the three items in the lemma. For convenience, we set

$$\Psi' := \Psi^N_{\neg\mathrm{edge}} \vee (\Psi^N_{\neg\mathrm{loop}} \wedge \Psi).$$

For (1), let $\mathfrak{A}, \mathfrak{B} \in \mathrm{fm}(\Phi^{\mathrm{CSP}})$ be arbitrary. Consider the following strategy of the EP for evaluating Φ^{CSP} on $\mathfrak{A} + \mathfrak{B}$ based on the winning strategies of the EP for evaluating Φ^{CSP} on $\mathfrak{A}$ and on $\mathfrak{B}$. For a SO variable S (of arity k):

- if S is existential and the EP chooses $[\![S]\!] := X_A \subseteq A^k$ on $\mathfrak{A}$ and $[\![S]\!] := X_B \subseteq B^k$ on $\mathfrak{B}$, then the EP chooses $[\![S]\!] := X_A \cup X_B \subseteq (A \cup B)^k$ on $\mathfrak{A} + \mathfrak{B}$;
- if S is universal and the UP chooses $[\![S]\!] := Y \subseteq (A \cup B)^k$ on $\mathfrak{A} + \mathfrak{B}$, then we consider the situation where the UP chooses $[\![S]\!] := Y \cap A^k$ on $\mathfrak{A}$ and $[\![S]\!] := Y \cap B^k$ on $\mathfrak{B}$.

Denote by $\mathfrak{A}^+$, $\mathfrak{B}^+$, and $(\mathfrak{A} + \mathfrak{B})^+$ the expansions of $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{A} + \mathfrak{B}$ resulting from the above evaluations of the SO variables in Φ, respectively. We now consider two possible cases for the evaluation of $[\![U]\!]$ and show that, in both of them, $\mathfrak{A} + \mathfrak{B} \models (\Psi')_{|U}$.

Case 1: $[\![U]\!] \subseteq A$ or $[\![U]\!] \subseteq B$. Without loss of generality, we assume that $[\![U]\!] \subseteq A$. Since $\mathfrak{A}^+ \models (\Psi')_{|U}$, it follows from Observation 16 that $(\mathfrak{A} + \mathfrak{B})^+ \models (\Psi')_{|U}$.

Case 2: $[\![U]\!] \cap A \neq \varnothing$ and $[\![U]\!] \cap B \neq \varnothing$. In this case, the substructure of $(\mathfrak{A} + \mathfrak{B})^+$ with domain $[\![U]\!]$ satisfies $\Psi^N_{\neg\mathrm{edge}}$ and hence also $(\Psi')_{|U}$.

For (2), observe first that Φ^{CSP} is logically equivalent to a $\forall$-restricted negative SO sentence. Indeed, the addition of $\Psi^N_{\neg\mathrm{edge}}$ and $\Psi^N_{\neg\mathrm{loop}}$ preserves negativity, and the application of $(\cdot)_{|U}$ ensures $\forall$-restriction. Hence, we conclude via Lemma 14 that Φ^{CSP} is preserved under inverse injective homomorphisms. Note that every homomorphism $\mathfrak{A} \to \mathfrak{B}$ decomposes to a surjective homomorphsism $\mathfrak{A} \to \mathfrak{A}_q$ and an injective homomorphism $\mathfrak{A}_q \to \mathfrak{B}$. It thus suffices to verify that Φ is preserved under inverse surjective homomorphisms.

To this end, let $\mathfrak{A}$ and $\mathfrak{B}$ be two finite τ-structures over the signature of Φ^{CSP} such that $\mathfrak{B} \models \Phi$ and there exists a surjective homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$. We claim that the EP has the following winning strategy on $\mathfrak{A}$ based on the winning strategy of the EP on $\mathfrak{B}$. For a SO variable S:

- if S is existentially quantified and the EP chooses $X \subseteq B^k$ for $[\![S]\!]$ on $\mathfrak{B}$, then the EP chooses $h^{-1}(X) \subseteq A^k$ for $[\![S]\!]$ on $\mathfrak{A}$;
- if $S = U$ and the UP reacts by choosing $Y_U \subseteq A$ for $[\![U]\!]$, then we consider the situation where UP chooses $h(Y_U) \subseteq B$ for $[\![U]\!]$ in $\mathfrak{B}$;
- if S is universally quantified with $S \neq U$ and UP reacts by choosing $Y \subseteq A^k$ for $[\![S]\!]$ on $\mathfrak{A}$, then we consider the situation where UP chooses $h(Y \cap Y_U^k) \subseteq B^k$ for $[\![S]\!]$ on $\mathfrak{B}$.

Denote by $\mathfrak{A}^+$ and $\mathfrak{B}^+$ the expansions of $\mathfrak{A}$ and $\mathfrak{B}$ resulting from the above evaluations of the SO variables in Φ. We show that $\mathfrak{A}^+ \models (\Psi')_{|U}$.

Note that, if $[\![U]\!]$ does not induce a N-surjective structure in $\mathfrak{A}$, then $\mathfrak{A}^+ \models (\Psi^N_{\neg\mathrm{edge}})_{|U}$ and hence also $\mathfrak{A}^+ \models (\Psi')_{|U}$. Hence, we assume that $[\![U]\!]$ induces a N-surjective structure in $\mathfrak{A}$. Note that, every model of Φ^{CSP} is a model $\Psi^N_{\neg\mathrm{loop}}$ (Observation 17), and so, every model of Φ^{CSP} is N-injective. Since the substructure of $\mathfrak{A}$ on $[\![U]\!]$ is N-surjective and h is a homomorphism, we have that h must be injective on $U^{\mathfrak{A}^+}$. In sum, we have that h is a bijection from $U^{\mathfrak{A}^+}$ to $U^{\mathfrak{B}^+}$. In what follows, it will be convenient to identify the sets $U^{\mathfrak{A}^+}$ and $U^{\mathfrak{B}^+}$, and simply denote them by $[\![U]\!]$. Since $\mathfrak{B}^+ \models (\Psi')_{|U}$, the EP has a winning strategy to show that the substructure of $\mathfrak{B}^+$ with domain $[\![U]\!]$ satisfies Ψ'. Now, the EP can mimic this strategy when evaluating Ψ' on the substructure of $\mathfrak{A}^+$ induced by $[\![U]\!]$. The fact that this is a winning strategy follows from the following two facts:

- atomic $(\tau \cup \{N\})$-formulas cannot appear positively in Ψ' and, since h is a homomorphism, for every $R \in \tau \cup \{N\}$, $R^{\mathfrak{A}^+} \cap [\![U]\!]^k \subseteq R^{\mathfrak{B}^+} \cap [\![U]\!]^k$
- by the strategy of EP for SO variables and the identification $U^{\mathfrak{A}^+} = [\![U]\!] = U^{\mathfrak{B}^+}$, for every SO variable S the equality $S^{\mathfrak{A}^+} \cap [\![U]\!]^k = S^{\mathfrak{B}^+} \cap [\![U]\!]^k$ holds.

Since the EP has a winning strategy for evaluating Ψ' on the substructure of $\mathfrak{A}^+$ induced by $[\![U]\!]$, we conclude that $\mathfrak{A}^+ \models (\Psi')_{|U}$.

For (3), we show that, for every τ-structure $\mathfrak{A}$, we have that

$$\mathfrak{A} \models \Phi \quad \text{if and only if} \quad (\mathfrak{A}, \neq) \models \Phi^{\mathrm{CSP}}.$$

First, observe that one direction is trivial. Indeed, if $(\mathfrak{A}, \neq) \models \Phi^{\mathrm{CSP}}$, then the EP simply projects the winning strategy onto evaluating Φ^{CSP} on $\mathfrak{A}$ for the case where $[\![U]\!] = A$. Now suppose that $\mathfrak{A} \models \Phi$. Since Φ is $\forall$-restricted, it is preserved under substructures due to Lemma 14. This means that the EP has a winning strategy for evaluating Φ on any substructure of $\mathfrak{A}$. But note that the notion of $\forall$-restriction implies an even stronger property. Namely, EP must have $\mathsf{Q}_1, \dots, \mathsf{Q}_k$-uniform winning strategies for evaluating Φ on the substructures of $\mathfrak{A}$ in the following sense. EP can evaluate the first block of existential SO variables $S_1, \dots, S_k$ to $S_1^{\mathfrak{A}}, \dots, A_k^{\mathfrak{A}}$ in $\mathfrak{A}$ so that: for each substructure $\mathfrak{A}'$, EP has a winning strategy for evaluating Φ in $\mathfrak{A}'$ where the evaluation of each S_i for $i \in [k]$ is the intersection of $S_i^{\mathfrak{A}}$ with $(A')^{s_i}$ where s_i is the arity of S_i. The high level reason is that Φ is universally restricted and EP cannot possibly know which substructure the UP will restrict the game to using the universal variables until the first universal SO is played (which is after the existential SO variables $S_1, \ldots, S_k$). From this observation, it follows immediately that the EP can extend the $\mathsf{Q}_1, \ldots, \mathsf{Q}_k$-uniform winning strategy for evaluating Φ on the substructures of $\mathfrak{A}$ to a winning strategy for evaluating Φ^{CSP} on $(\mathfrak{A}, \neq)$; the choice of $[\![U]\!]$ done by the UP restricts the substructures of $\mathfrak{A}$ the UP might want to play on. Formally, we have the following two cases for a SO variable S.

If $S \in \{S_1, \ldots, S_k\}$ and the EP chooses $X \subseteq A^k$ for $[\![S]\!]$ on $\mathfrak{A}$, then the EP chooses the same set $X \subseteq A^k$ for $[\![S]\!]$ on $(\mathfrak{A}, \neq)$. Otherwise U precedes S. Then:

- if S is existentially quantified and the EP chooses $[\![S]\!] = X \subseteq A^k$ for evaluating Φ on $\mathfrak{A}$, then EP chooses $[\![S]\!] = X \subseteq A^k$ for evaluating Φ^{CSP} on $(\mathfrak{A}, \neq)$;
- if $S \neq U$ is universally quantified and the UP reacts by choosing $[\![S]\!] = X \subseteq A^k$ for evaluating Φ^{CSP} on $(\mathfrak{A}, \neq)$, then we consider the case when UP chooses $[\![S]\!] = X \cap [\![U]\!]^k \subseteq A^k$ for evaluating Φ on $\mathfrak{A}$. ◀

5 From finite-domain QCSPs to ω-categorical CSPs

In the present section, we give a general construction of ω-categorical CSPs from bounded alternation finite-domain QCSPs. Let $\mathfrak{B}$ be a finite structure with the relational signature τ and domain $B = \{b_1, \ldots, b_\ell\}$. Also, fix an FO quantifier-prefix $\mathsf{Q}_1\bar{x}_1 \cdots \mathsf{Q}_n\bar{x}_n$ for some $n \geq 1$, where $\mathsf{Q}_i \in \{\exists, \forall\}$ for every $i \in [n]$. We define a new signature τ' consisting of the symbols from τ, and, for every $i \in [n]$, the unary symbol E_i or A_i, depending on whether Q_i equals $\exists$ or $\forall$. We shall refer to the ith symbol E_i or A_i by Q_i. Consider the MSO τ'-sentence

$$\Phi_{\mathfrak{B}} := \mathsf{Q}_1 Q_{1,b_1} \cdots \mathsf{Q}_1 Q_{1,b_\ell} \cdots \mathsf{Q}_n Q_{n,b_1} \cdots \mathsf{Q}_n Q_{n,b_\ell} \big(\Psi_\forall \vee (\Psi_\exists \wedge \Psi_{\mathfrak{B}})\big),$$

where $\Psi_\forall$, $\Psi_\exists$, and $\Psi_{\mathfrak{B}}$ are FO formulas defined below. Intuitively, $Q_{k,b}$ for $k \in [n]$ and $b \in B$ represents the set of all variables v in an instance of $\mathsf{Q}_1 \cdots \mathsf{Q}_n\mathrm{QCSP}(\mathfrak{B})$ for which the player controlling Q_k has chosen $[\![x]\!] := b$. We now give definitions of $\Psi_\forall$, $\Psi_\exists$, and $\Psi_{\mathfrak{B}}$.

First, $\Psi_\forall$ states that the UP did not play by the rules:

$$\Psi_\forall := \bigvee\nolimits_{i \in [n] \colon Q_i = A_i} \bigvee\nolimits_{b \in B} \exists x \big(\neg A_i(x) \wedge A_{i,b}(x)\big) \vee \bigvee\nolimits_{c \in B \setminus \{b\}} \exists x \big(A_{i,b}(x) \wedge A_{i,c}(x)\big). \tag{5}$$

The sentence $\Psi_\exists$ states that the EP plays by the rules:

$$\Psi_\exists := \bigwedge\nolimits_{i \in [n] \colon Q_i = E_i} \forall x \big(\neg E_i(x) \vee \bigvee\nolimits_{b \in B} E_{i,b}(x)\big). \tag{6}$$

Finally, $\Psi_{\mathfrak{B}}$ states that either the UP has cheated – by violating the converse of eq. (5) – or the Q_is and $Q_{i,b}$s are compatible with the relations of $\mathfrak{B}$:

$$\Psi_{\mathfrak{B}} = \bigwedge\nolimits_{R \in \tau \text{ with arity } m} \bigwedge\nolimits_{(i_1, \ldots, i_m) \in [n]^m} \forall \bar{x} \big(\psi_{\tau'} \vee \psi_\forall \vee \psi_\exists\big), \tag{7}$$

where $\psi_{\tau'}$, $\psi_\forall$, and $\psi_\exists$ are shortcuts for the following formulas (that depend on R and the tuple $(i_1, \ldots, i_m)$):

$$\psi_{\tau'} := \neg R(x_1, \ldots, x_m) \vee \bigvee\nolimits_{k \in [m]} \neg Q_{i_k}(x_k),$$
$$\psi_\forall := \bigvee\nolimits_{k \in [m] \colon Q_{i_k} = A_{i_k}} \bigwedge\nolimits_{b \in B} \neg A_{i_k, b}(x_k),$$
$$\psi_\exists := \bigvee\nolimits_{(t_1, \ldots, t_m) \in R^{\mathfrak{B}}} \bigwedge\nolimits_{k \in [m]} Q_{i_k, t_k}(x_k).$$

▶ **Lemma 19.**

1. $\Phi_{\mathfrak{B}}$ *is* $\forall$*-restricted and negative.*
2. $\mathrm{Q}_1 \cdots \mathrm{Q}_n\mathrm{QCSP}(\mathfrak{B})$ *reduces in polynomial-time to model-checking for* $\Phi_{\mathfrak{B}}$*.*

Proof. Showing (1) is straightforward; recall that $\forall$-restriction and negativity are closed under internal conjunctions and disjunctions.

For (2), we use the following reduction. Given an instance $\Phi := \mathsf{Q}_1\bar{x}_1 \cdots \mathsf{Q}_n\bar{x}_n.\,\phi$ of $\mathrm{QCSP}(\mathfrak{B})$, we obtain the instance $\mathfrak{A}$ of the model-checking problem for $\Phi_{\mathfrak{B}}$, where:

- $\mathfrak{A} \models Q_i(v)$ if and only if v appears in $\bar{x}_i$, and
- $\mathfrak{A} \models R(v_1, \ldots, v_m)$ for $R \in \tau$ of arity m if and only if this atom appears in ϕ.

"$\Rightarrow$" Suppose that Φ is a YES-instance of $\mathrm{QCSP}(\mathfrak{B})$. Before defining a strategy for the EP, notice that $\Psi_\forall$ says that if the evaluations of universal SO variables do not satisfy the inclusion $[\![A_{i,b}]\!] \subseteq A_i$, then EP can trivially win the FO part. So we assume that $[\![A_{i,b}]\!] \subseteq A_i$, and consider the following strategy of the EP evaluating the SO variables of $\Phi_{\mathfrak{B}}$ on $\mathfrak{A}$ based on the winning strategy of the EP on Φ:

- the EP evaluates SO variables on $\mathfrak{A}$: $[\![E_{i,b}]\!] := \{v \in \bar{x}_i \mid [\![v]\!] = b \text{ on } \Phi\}$;
- if the UP reacts by evaluating $[\![A_{i,b}]\!] = X \subseteq A$, then the EP considers the move of the UP on the QCSP instance where: $[\![v]\!] := b$ if and only if $v \in [\![A_{i,b}]\!]$ on $\mathfrak{A}$ – if this is not a valid move for a variable v, i.e., v that gets assigned no value $b \in B$ or two different values $b \neq b'$, then EP chooses an arbitrary value $[\![v]\!] \in B$.

Let $\mathfrak{A}^+$ be an expansion of $\mathfrak{A}$ resulting from the above evaluations of the SO variables in $\Phi_{\mathfrak{B}}$. We show that $\mathfrak{A}^+$ satisfies the FO part of $\Phi_{\mathfrak{B}}$, and it suffices to show that it satisfies $\Psi_\exists \wedge \Psi_{\mathfrak{B}}$. Clearly, $\mathfrak{A}^+ \models \Psi_\exists$ because the EP assigns a value to every existential variable on Φ. We must show that $\mathfrak{A}^+ \models \Psi_{\mathfrak{B}}$. Whenever $\psi_{\tau'}$ is not satisfied, we have the following two cases. Either the UP has not assigned values to the $A_{i,b}$s in a way that correctly represents an assignment of values to the universal variables in Φ, in which case $\psi_\forall$ holds, or otherwise $\psi_\exists$ holds because the strategy of the EP on $\mathfrak{A}$ reflects a winning strategy of the EP on Φ.

"$\Leftarrow$" Suppose that Φ is a NO-instance of $\mathrm{QCSP}(\mathfrak{B})$. Consider the following strategy of the UP for evaluating $\Phi_{\mathfrak{B}}$ on $\mathfrak{A}$ based on the winning strategy of the UP on Φ:

- the UP evaluates SO variables on $\mathfrak{A}$: $[\![A_{i,b}]\!] := \{v \in \bar{x}_i \mid [\![v]\!] = b \text{ on } \Phi\}$;
- when EP reacts by evaluating SO variables by $[\![E_{i,b}]\!]$ on $\mathfrak{A}$, the UP consideres the case when EP evaluates the QCSP variables from Φ: $[\![v]\!] := b$ if and only if $v \in [\![E_{i,b}]\!]$ on $\mathfrak{A}$.

Let $\mathfrak{A}^+$ be an expansion of $\mathfrak{A}$ resulting from the above evaluations of the SO variables in $\Phi_{\mathfrak{B}}$. We show that $\mathfrak{A}^+$ does not satisfy the FO part of $\Phi_{\mathfrak{B}}$.

First, either eq. (6) does not hold, in which case the EP loses immediately, or the assignment of values to the existentially quantified SO variables in $\Phi_{\mathfrak{B}}$ on $\mathfrak{A}$ correctly represents an assignment of values to the existentially quantified variables in Φ; we assume the latter. Regarding $\Psi_\forall$, eq. (5) holds by the strategy of the UP on $\mathfrak{A}$; hence, $\mathfrak{A}^+ \not\models \Psi_\forall$. We show that $\mathfrak{A}^+ \not\models \Psi_{\mathfrak{B}}$. By the strategy of the UP on $\mathfrak{A}$, if $\psi_{\tau'}$ does not hold, then $\psi_\forall$ does not hold either.

But then the last part of eq. (7) cannot be satisfied uniformly across all relations because the strategy of the UP on $\mathfrak{A}$ reflects a winning strategy of the UP on Φ. ◀

Proof of Theorem 2. Consider the MSO sentence $\Phi_{\mathfrak{B}}$. By Lemma 19(1), $\Phi_{\mathfrak{B}}$ is $\forall$-restricted and negative. Hence, the prerequisites of Lemma 18 are satisfied for $\Phi_{\mathfrak{B}}$. By items 1 and 2 of Lemma 18, $\Phi_{\mathfrak{B}}^{\mathrm{CSP}}$ is preserved by disjoint unions and inverse homomorphisms. By construction, $\Phi_{\mathfrak{B}}^{\mathrm{CSP}}$ is a $\mathrm{Q}_1 \cdots \mathrm{Q}_n\mathrm{MSO}$ sentence. It follows from Theorem 1 that there exists an ω-categorical structure $\mathfrak{D}$ with $\mathrm{CSP}(\mathfrak{D}) = \mathrm{fm}(\Phi_{\mathfrak{B}}^{\mathrm{CSP}})$.

By Lemma 19(2), $\mathrm{Q}_1 \cdots \mathrm{Q}_n\mathrm{QCSP}(\mathfrak{B})$ reduces in polynomial-time to model-checking for $\Phi_{\mathfrak{B}}$. It follows from item 3 of Lemma 18 that model-checking for $\Phi_{\mathfrak{B}}$ reduces in polynomial-time to model-checking for $\Phi_{\mathfrak{B}}^{\mathrm{CSP}}$. ◀

Proof of Theorem 3. As already mentioned, Theorem 3 follows from Theorem 2, and from the fact that there are $(\exists\forall)^n\exists$- and $\forall(\exists\forall)^n\exists$-QCSPs complete for the $\Sigma_{2n+1}^{\mathrm{P}}$-level and the Π_{2n}^{P}-level of the polynomial hierarchy ($n \geq 1$); e.g., $(\exists\forall)^n\exists$- and $\forall(\exists\forall)^n\exists$-3CNF. ◀

6 ω-categorical CSPs complete for Σ_{2n}^{P} and Π_{2n+1}^{P}

In the present section, we give a construction of Σ_{2n}^{P}-complete ω-categorical CSPs from the complement of $(\forall\exists)^n$3-CNF. There is an analogous construction of Π_{2n+1}^{P}-complete ω-categorical CSPs from the complement of $\exists(\forall\exists)^n$3-CNF; for the sake of simplicity, we choose to only provide details for the case of $(\forall\exists)^n$3-CNF. Fix $n \geq 1$. The strategy is to construct an $\exists$-guarded positive MSO τ-sentence Φ_* such that the SO quantifier prefix of Φ_* is $(\forall\exists)^n$ and $(\forall\exists)^n$3-CNF reduces in polynomial-time to model-checking for Φ_*. Then we hit $\neg\Phi_*$ with the CSP hammer.

We define the signature τ consisting of the binary symbols R, $\overline{R}$, and $\prec$, and the unary symbols S, T, $E_1, \dots, E_n$, and $A_1, \dots, A_n$. Now, we set

$$\Phi_* := \forall A_{1,0}\exists E_{1,1} \dots \forall A_{n,0}\exists E_{n,1}\exists V \left(\exists x.\, V(x)\right) \wedge \left(\neg\Psi_\forall \vee (\Psi_\exists \wedge \Psi_{|V})\right),$$

where V is a fresh unary symbol, and $\Psi_\forall$, $\Psi_\exists$, Ψ are FO formulas defined below. At a high level, $\Psi_\forall$ and $\Psi_\exists$ are sets of FO constraints that UP and the EP must follow in their SO evaluation strategies. Following this intuition, the FO part of Φ_* says that, whenever the UP plays by the rules $\Psi_\forall$, the EP must play by the rules $\Psi_\exists$ and satisfy Ψ in the substructure with domain $[\![V]\!]$. Formally, $\Psi_\forall$ is the conjunction of the following sentences:

$$\forall x\big(\neg A_{k,0}(x) \vee A_{\ell,0}(x)\big) \text{ for all } k,\ell \in [n] \text{ with } k < \ell, \tag{8}$$
$$\forall x\big(\neg A_{\ell,0}(x) \vee \neg A_k(x) \vee A_{k,0}(x)\big) \text{ for all } k,\ell \in [n] \text{ with } k < \ell, \tag{9}$$
$$\forall x\big(\neg A_{k,0}(x) \vee \neg A_\ell(x)\big) \text{ for all } k,\ell \in [n] \text{ with } k < \ell, \tag{10}$$
$$\forall x\big(\neg A_{n,0}(x) \vee \neg E_k(x)\big) \text{ for each } k \in [n], \tag{11}$$
$$\forall x,y\big(\neg A_{n,0}(x) \vee \neg A_{n,0}(y) \vee \neg\overline{R}(x,y)\big). \tag{12}$$

Now, $\Psi_\exists$ is the conjunction of the following formulas:

$$\forall x \bigwedge\nolimits_{1\leq k<\ell\leq n} \left(\neg E_{k,1}(x) \vee E_{\ell,1}(x)\right), \tag{13}$$
$$\forall y \bigwedge\nolimits_{k\in[n]} \Big(\neg E_{k,1}(y) \vee \bigvee\nolimits_{j\in[k]} E_j(y)\Big), \tag{14}$$
$$\forall w,z \left(\neg E_{n,1}(w) \vee \neg E_{n,1}(z) \vee R(w,z)\right). \tag{15}$$

Finally, Ψ is the conjunction of the following formulas:

$$\forall x,y.\, R(x,y) \tag{16}$$
$$\exists x,y\big(S(x) \wedge T(y)\big), \tag{17}$$
$$\forall x\Big(\neg A_{n,0}(x) \vee \bigvee\nolimits_{k\in[n]} E_{k,1}(x)\Big), \tag{18}$$
$$\forall x,y\,\exists a,b\Big((x=y) \vee \big((x \prec a) \wedge (b \prec y)\big) \vee \big((y \prec b) \wedge (a \prec x)\big)\Big). \tag{19}$$

▶ **Lemma 20.**
1. Φ_* *is logically equivalent to an* $\exists$*-guarded positive sentence.*
2. $(\forall\exists)^n$ *3-CNF reduces in polynomial-time to model-checking for* Φ_**.*

Proof of Theorem 4. Consider the MSO sentence Φ_*. By Lemma 20(1), Φ_* is logically equivalent to an $\exists$-guarded positive sentence. By Lemma 11, Lemma 13, and Lemma 14, $\neg\Phi_*$ is logically equivalent to an $\forall$-restricted negative sentence. Hence, the prerequisites of Lemma 18 are satisfied for $\neg\Phi_*$. By items 1 and 2 of Lemma 18, $(\neg\Phi_*)^{\mathrm{CSP}}$ is preserved by disjoint unions and inverse homomorphisms. Note that, by construction, $(\neg\Phi_*)^{\mathrm{CSP}}$ is an $(\exists\forall)^n$-MSO sentence. It follows from Theorem 1 that there exists an ω-categorical structure $\mathfrak{D}$ with $\mathrm{CSP}(\mathfrak{D}) = \mathrm{fm}((\neg\Phi_*)^{\mathrm{CSP}})$. By Lemma 20(2), $(\forall\exists)^n$3-CNF has a polynomial-time reduction to model-checking for Φ_*.

All our reductions are many-one, and hence it follows immediately that $(\exists\forall)^n$3-DNF has a polynomial-time many-one reduction to model-checking for $\neg\Phi_*$. It follows from item 3 of Lemma 18 that model-checking for $\neg\Phi_*$ reduces in polynomial-time to model-checking for $(\neg\Phi_*)^{\mathrm{CSP}}$. We conclude that $\mathrm{CSP}(\mathfrak{D})$ is Σ^{P}_{2n}-complete. The Π^{P}_{2n+1}-complete case can be treated analogously by adding an existential SO variable $\exists E_{0,1}$ to Φ^* in the outermost position, and then modifying $\Psi_\exists$ and Ψ, accordingly. ◀

7 Conclusion and outlook

We showed that there are ω-categorical CSPs complete for any level of the polynomial hierarchy above NP. To this end, we developed a new tool, the CSP hammer, for producing MSO sentences closed under inverse homomorphisms and disjoint unions. By a recent result of Bodirsky, Knäuer, and Rudolph [9], every such sentence defines an ω-categorical CSP. Next, in Section 5, we gave a general construction of ω-categorical CSPs from bounded alternation finite-domain QCSPs. When adjusted with the CSP hammer, this construction is able to produce $\Sigma^{\mathrm{P}}_{2n+1}$- and Π^{P}_{2n}-complete ω-categorical CSPs for every $n \geq 1$. One drawback of this method is that it does not always preserve the complexity of the original QCSP; achieving a fine-grained correspondence between ω-categorical CSPs and bounded alternation finite-domain QCSPs would require turning the CSP hammer into a more robust toolset. We ask the following question, which seems to be surprisingly non-trivial.

Question 1. Is it true that for every finite structure $\mathfrak{B}$ and every quantifier prefix $\mathrm{Q}_1 \cdots \mathrm{Q}_n$ there exists an ω-categorical structure $\mathfrak{D}$ such that $\mathrm{Q}_1 \cdots \mathrm{Q}_n\mathrm{QCSP}(\mathfrak{B})$ and $\mathrm{CSP}(\mathfrak{D})$ are polynomial-time equivalent?

We remark that a positive answer would imply that there are ω-categorical CSPs complete for some exotic complexity classes in PH such as DP or Θ^{P}_2 [25]. The obstacle seems to be getting around inverse surjective homomorphisms (see Example 15), which our CSP hammer resolves in an ad hoc manner. We ask the same question for the unbounded case.

Question 2. Is it true that for every finite structure $\mathfrak{B}$ there is an ω-categorical structure $\mathfrak{D}$ such that $\mathrm{QCSP}(\mathfrak{B})$ and $\mathrm{CSP}(\mathfrak{D})$ are polynomial-time equivalent?

In Section 6, we gave a polynomial-time reduction from the complement $(\forall\exists)^n$3-CNF to model checking for MSO sentences which, when combined with the CSP hammer, produces Σ^{P}_{2n}- and Π^{P}_{2n+1}-complete ω-categorical CSPs for every $n \geq 1$. In some aspects, our construction is quite similar to the one used in [7] for obtaining Π^{P}_n-complete and ultimately coNP-intermediate ω-categorical CSPs. One notable difference is that [7, Theorem 2] uses the

Fraïssé amalgamation method for constructing ω-categorical CSP templates while our result uses [9, Corollary 17], which itself goes back to [20, Theorem 4.27]. This observation raises the question whether one can replicate the construction of an intermediate ω-categorical CSP for NP instead of coNP. Below, we restate a question of Bodirsky [5, Question 49].

Question 3. Assuming P $\neq$ NP, does there exist an NP-intermediate ω-categorical CSP?

Regarding the universal-algebraic upgrade [18, Theorem 1.8] for [7, Theorem 2] in the Π_n^{P}-levels of PH, it would be interesting to know if a similar upgrade for Theorem 4 and 3 can be obtained for the Σ_n^{P}-levels of PH. For $n = 1$, this would confirm that the assumption of being a reduct of a finitely bounded homogeneous structure is essential in the algebraic dichotomy conjecture for infinite-domain CSPs [4].

Question 4. Does there exist an ω-categorical structure $\mathfrak{D}$ whose CSP is Σ_n^{P}-complete for some $n \geq 1$ and such that the polymorphisms of $\mathfrak{D}$ satisfy some non-trivial identities?

References

1 Libor Barto, Bertalan Bodor, Marcin Kozik, Antoine Mottet, and Michael Pinsker. Symmetries of Graphs and Structures that Fail to Interpret a Finite Thing. In *2023 38th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–13, 2023. doi:10.1109/LICS56636.2023.10175732.

2 Libor Barto, Michael Kompatscher, Miroslav Olšák, Van Pham Trung, and Michael Pinsker. Equations in oligomorphic clones and the constraint satisfaction problem for ω-categorical structures. *J. Math. Log.*, 19(02):Article 1950010, 2019. doi:10.1142/S0219061319500107.

3 Libor Barto, Jakub Opršal, and Michael Pinsker. The wonderland of reflections. *Isr. J. Math.*, 223(1):363–398, 2018. doi:10.1007/s11856-017-1621-9.

4 Libor Barto and Michael Pinsker. Topology is irrelevant (in a dichotomy conjecture for infinite domain constraint satisfaction problems). *SIAM J. Comput.*, 49(2):365–393, 2020. doi:10.1137/18M1216213.

5 Manuel Bodirsky. *Complexity of infinite-domain constraint satisfaction*, volume 52. Cambridge University Press, 2021. Postprint available online. doi:10.1017/9781107337534.

6 Manuel Bodirsky, Thomas Feller, Simon Knäuer, and Sebastian Rudolph. On logics and homomorphism closure. In *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–13. IEEE, 2021. doi:10.1109/LICS52264.2021.9470511.

7 Manuel Bodirsky and Martin Grohe. Non-dichotomies in constraint satisfaction complexity. In *Proceedings of the 35th International Colloquium on Automata, Languages, and Programming - ICALP'08*, pages 184–196. Springer, 2008. doi:10.1007/978-3-540-70583-3_16.

8 Manuel Bodirsky and Jan Kára. The complexity of temporal constraint satisfaction problems. *J. ACM.*, 57(2):9:1–9:41, 2010. A conference version appeared in the proceedings of the 40th ACM Symposium on Theory of Computing - STOC'08. doi:10.1145/1667053.1667058.

9 Manuel Bodirsky, Simon Knäuer, and Sebastian Rudolph. Datalog-Expressibility for Monadic and Guarded Second-Order Logic. In *Proceedings of the 48th International Colloquium on Automata, Languages, and Programming - ICALP'21*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 120:1–120:17, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ICALP.2021.120.

10 Manuel Bodirsky and Jaroslav Nešetřil. Constraint satisfaction with countable homogeneous templates. *J. Log. Comput*, 16(3):359–373, 2006. A conference version appeared in the proceedings of Computer Science Logic - CSL'03. doi:10.1093/logcom/exi083.

11 Manuel Bodirsky, Michael Pinsker, and András Pongrácz. Projective clone homomorphisms. *J. Symb. Log.*, 86(1):148–161, 2021. doi:10.1017/jsl.2019.23.

12 Johanna Brunar, Marcin Kozik, Tomáš Nagy, and Michael Pinsker. The sorrows of a smooth digraph: the first hardness criterion for infinite directed graph-colouring problems. In *2025 40th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 403–416, 2025. doi:10.1109/LICS65433.2025.00037.

13 Andrei A. Bulatov. A dichotomy theorem for nonuniform CSPs. In *Proceedings of the 58th IEEE Annual Symposium on Foundations of Computer Science - FOCS'17*, pages 319–330. IEEE, 2017. doi:10.1109/FOCS.2017.37.

14 A. K. Chandra and P. M. Merlin. Optimal implementation of conjunctive queries in relational data bases. In *Proceedings of the Ninth Annual ACM Symposium on Theory of Computing*, STOC '77, pages 77–90, New York, NY, USA, 1977. Association for Computing Machinery. doi:10.1145/800105.803397.

15 Hubie Chen. Existentially restricted quantified constraint satisfaction. *Inf. Comput.*, 207(3):369–388, 2009. doi:10.1016/j.ic.2008.11.001.

16 Grädel E. *Back and Forth Between Logic and Games*, pages 99–145. Cambridge University Press, 2011.

17 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, 1998. doi:10.1137/S0097539794266766.

18 Pierre Gillibert, Julius Jonusas, Michael Kompatscher, Antoine Mottet, and Michael Pinsker. When symmetries are not enough: A hierarchy of hard constraint satisfaction problems. *SIAM J. Comput.*, 51(2):175–213, 2022. doi:10.1137/20M1383471.

19 Wilfrid Hodges. *A Shorter Model Theory.* Cambridge University Press, 1997.

20 Barnaby Martin, Manuel Bodirsky, and Martin Hils. On the scope of the universal-algebraic approach to constraint satisfaction. *Logical Methods in Computer Science*, 8, 2012. doi:10.2168/LMCS-8(3:13)2012.

21 Antoine Mottet and Michael Pinsker. Smooth approximations: An algebraic approach to CSPs over finitely bounded homogeneous structures. *J. ACM.*, 71(5):36:1–36:47, 2024. doi:10.1145/3689207.

22 Mark H. Siggers. A strong Mal'cev condition for locally finite varieties omitting the unary type. *Algebra Univers.*, 64:15–20, 2010. doi:10.1007/s00012-010-0082-3.

23 Dmitriy Zhuk. A proof of CSP dichotomy conjecture. In *Proceedings of the 58th IEEE Annual Symposium on Foundations of Computer Science - FOCS'17*, pages 331–342. IEEE, 2017. doi:10.1109/FOCS.2017.38.

24 Dmitriy Zhuk. A proof of the CSP dichotomy conjecture. *J. ACM.*, 67(5):30:1–30:78, 2020. doi:10.1145/3402029.

25 Dmitriy Zhuk and Barnaby Martin. QCSP monsters and the demise of the Chen Conjecture. *J. ACM*, 69(5):35:1–35:44, 2022. doi:10.1145/3563820.