Tractability of Constraint Satisfaction Problems and Projective Clone Homomorphisms

Antoine Mottet
École Normale Supérieure de Cachan

Mémoire réalisé durant l’Année de Recherche Prédoctorale à l’Étranger sous la direction de Manuel Bodirsky.
Contents

1 Introduction 2
  1.1 The Universal Algebraic Approach to Constraint Satisfaction 3
  1.2 From the Finite to the Infinite 5

2 Preliminaries 7
  2.1 Model Theory 7
  2.2 Universal Algebra 9
  2.3 Canonical Functions 11

3 The Complexity of First-Order Reducts of (\mathbb{N}; 0) 14
  3.1 The Role of Binary Injective Operations 15
  3.2 Polynomial-Time Tractable CSPs 16
    3.2.1 Syntactic Forms 16
    3.2.2 Algorithms 17
  3.3 The Classification 21
  3.4 Further work 23

4 Algebraic Criteria for the Tractability of Problems in MMSNP 25
  4.1 Introduction 25
    4.1.1 MMSNP 26
    4.1.2 Coloured Obstruction Sets 26
  4.2 Universal Structures for Connected Obstruction Sets 27
  4.3 A Dichotomy for Monochromatic Obstructions 29
    4.3.1 The Dichotomy for Structures of Type 1 32
    4.3.2 The Dichotomy for Structures of Type 2 33
  4.4 Conclusion 34
Chapter 1

Introduction

A constraint satisfaction problem is, informally, a decision problem where one is given a set of variables ranging over elements of some set $D$ together with constraints on these variables, and we are asked to determine whether there is an instantiation of the variables by values from $D$ that satisfies all the constraints. This general description captures a lot of decision problems. Consider e.g. the $k$-colourability of graphs, boolean satisfiability, scheduling problems, some variants of type checking, or feasibility of systems of equations.

Let us now see a formal definition of constraint satisfaction problems. Let $\tau$ be a finite relational signature, i.e. $\tau = \{R_1, \ldots, R_m\}$ where each $R_i$ is a relation symbol of arity $ar(R_i)$. Let $\Gamma$ be a $\tau$-structure, i.e. $\Gamma = (D; R^\Gamma_1, \ldots, R^\Gamma_m)$ where $R^\Gamma_i \subseteq D^{ar(R_i)}$. A first-order $\tau$-formula is a formula built from symbols from $\tau$, binary conjunctions, binary disjunctions, negations and universal and existential first-order quantifiers. A first-order $\tau$-formula is said to be primitive positive if it only contains symbols from $\tau$, conjunctions, and existential quantifiers. A sentence is a formula with no free variables. The constraint satisfaction problem associated with $\Gamma$, written CSP($\Gamma$), is defined as follows:

**Definition 1.1 (CSP($\Gamma$)).**

**Input:** a primitive positive $\tau$-sentence $\phi := \exists x_1, \ldots, x_n(\land_i \psi_i)$.

**Question:** does $\phi$ hold true in $\Gamma$?

This definition is easily seen to reflect the description of a CSP given in the first paragraph. Indeed, a set of constraints over a set of variables is akin to a conjunction of atoms over these variables, where each atom represents one of the constraints. If a primitive positive sentence is true in $\Gamma$, then there exists an instantiation of the variables that satisfies all the constraints, and vice versa. It is useful to note that when the domain of $\Gamma$ is finite, then CSP($\Gamma$) $\in$ NP, since if a sentence $\phi$ is true in $\Gamma$ then a satisfying assignment of the variables of $\phi$ can be represented with a polynomial size. When the domain of $\Gamma$ is infinite, however, the complexity of CSP($\Gamma$) ranges from constant-time solvable to undecidable. This motivates the following generic question:

**Question 1.** For which structures $\Gamma$ do we have that CSP($\Gamma$) is in P? For which structures $\Gamma$ is CSP($\Gamma$) NP-hard?

A celebrated result in the area of constraint satisfaction is Schaefer’s Theorem [Sch78], which answers Question 1 for structures whose domain has two elements:
Theorem 1.1. Let $\Gamma$ be a structure whose domain is $\{0, 1\}$. Then $\text{CSP}(\Gamma)$ is NP-complete, or at least one of the following holds:

- all the relations of $\Gamma$ contain the tuple $(0, \ldots, 0)$,
- all the relations of $\Gamma$ contain the tuple $(1, \ldots, 1)$,
- all the relations of $\Gamma$ are definable with Horn formulas,
- all the relations of $\Gamma$ are definable with dual-Horn formulas,
- all the relations of $\Gamma$ are definable with bijunctive formulas,
- all the relations of $\Gamma$ are definable with affine formulas.

In all these cases, $\text{CSP}(\Gamma)$ is in $\text{P}$.

This theorem has a logical flavour: it describes the complexity of $\text{CSP}(\Gamma)$ with conditions on the definability of the relations with certain restricted fragments of first-order logic. We will see that Schaefer’s dichotomy theorem can be formulated with vocabularies coming from different areas.

1.1 The Universal Algebraic Approach to Constraint Satisfaction

One of the central observation about CSPs is the following. Let $\Gamma$ be a structure of relational signature $\tau$, and let $\phi(x_1, \ldots, x_n)$ be a primitive positive $\tau$-formula with $n$ free variables. Let $R$ be the relation defined by $\phi$ in $\Gamma$, that is, $R$ is the subset of $\text{Dom}(\Gamma)^n$ defined by

$$\{(a_1, \ldots, a_n) \in \text{Dom}(\Gamma)^n \mid \Gamma \models \phi(a_1, \ldots, a_n)\}.$$  

Fact. The problems $\text{CSP}(\Gamma)$ and $\text{CSP}(\Gamma, R)$ are polynomial-time many-one equivalent.

Per this observation, the complexity of $\text{CSP}(\Gamma)$ is determined by the set of relations defined in $\Gamma$ by pp-formulas. This is where the so-called universal algebraic approach comes in. In this approach, the aim is to characterize the set of relations that are pp-definable in $\Gamma$ by a set of functions over $\text{Dom}(\Gamma)$, using a so-called Galois connection.

If $\Gamma$ has domain $D$ and $f: D^k \to D$ is a $k$-ary function on $D$, we say that $f$ is a polymorphism of $\Gamma$ if for each relation $R$ of $\Gamma$, and for all $a_1^1, \ldots, a_{ar(R)}^1, \ldots, a_1^k, \ldots, a_{ar(R)}^k$ tuples from $R$, we have that $f(a_1^1, \ldots, a_k^1)$ is a tuple in $R$, where we define $f(a_1^1, \ldots, a_k^1) = (f(a_1^1, \ldots, a_{ar(R)}^1), \ldots, f(a_1^k, \ldots, a_{ar(R)}^k))$. We also say that $f$ preserves $\Gamma$. We denote by $\text{Pol}(\Gamma)$ the set of polymorphisms of $\Gamma$. When $k = 1$, we say that $f$ is an endomorphism of $\Gamma$, and the set of endomorphisms of $\Gamma$ is denoted by $\text{End}(\Gamma)$. Moreover, when $f$ is unary and we have that $f(a_1, \ldots, a_{ar(R)}) \in R \Leftrightarrow (a_1, \ldots, a_{ar(R)}) \in R$, we say that $f$ is a self-embedding and we write $f \in \text{Emb}(\Gamma)$. Finally, a surjective embedding is called an automorphism, and $\text{Aut}(\Gamma)$ denotes the set of automorphisms of $\Gamma$. We can view $\text{Pol}, \text{Aut}, \text{Emb}$, and $\text{End}$ as operators that map a set of relations $\mathcal{R}$ over the same domain to a set of operations. Conversely, if $\mathcal{F}$ is a set of functions over some domain $D$, we define the operator $\text{Inv}$ by $\text{Inv}(\mathcal{F}) = \{R \subseteq D^k \mid f \text{ preserves } R\}$.

The following lemma is central:
Lemma 1.2. The operators Inv and Pol form a Galois connection. Moreover, if $\Gamma$ has a finite domain we have that $\text{Inv}(\text{Pol}(\Gamma))$ is the set of all the relations that are pp-definable in $\Gamma$.

Lemma 1.2 suggests that in order to find the complexity of CSP$(\Gamma)$ when Dom$(\Gamma)$ is finite, it is interesting to know what functions are in Pol$(\Gamma)$. Using polymorphisms, Schaefer’s Theorem can be rephrased as follows:

Theorem 1.3. Let $\Gamma$ be a structure whose domain is $\{0,1\}$. Then CSP$(\Gamma)$ is NP-complete, or at least one of the following holds:

- $\Gamma$ is preserved by the unary function $c_0 : x \mapsto 0$,
- $\Gamma$ is preserved by the unary function $c_1 : x \mapsto 1$,
- $\Gamma$ is preserved by the binary function $\land : (x,y) \mapsto x \land y$,
- $\Gamma$ is preserved by the binary function $\lor : (x,y) \mapsto x \lor y$,
- $\Gamma$ is preserved by the majority function $(x,y,z) \mapsto (x \land y) \lor (x \land z) \lor (y \land z)$,
- $\Gamma$ is preserved by the minority function $(x,y,z) \mapsto x \oplus y \oplus z$.

In all these cases, CSP$(\Gamma)$ is in P.

The motto in the theory of finite-domain CSPs is that whenever $\Gamma$ is a finite structure which has a “non-trivial” polymorphism, then CSP$(\Gamma)$ is tractable in polynomial-time. The tractability conjecture is the following statement (see [BJK05], in combination with [BK12]).

Conjecture. Let $\Gamma$ be a finite structure with a finite relational signature. Then either $\Gamma$ has a polymorphism $f$ that satisfies the identity

$$\forall x_1, \ldots, x_n \in \text{Dom}(\Gamma), f(x_1, \ldots, x_n) = f(x_2, \ldots, x_n, x_1)$$

and CSP$(\Gamma)$ is polynomial-time tractable, or CSP$(\Gamma)$ is NP-complete.

When a function satisfies the equation (1.1) we say that it is cyclic. Several equations are known to be equivalent to the cyclic equation, in the sense that there is a cyclic function in Pol$(\Gamma)$ if and only if Pol$(\Gamma)$ contains a function that satisfies another type of equation. For example, it is known that Pol$(\Gamma)$ contains a cyclic operation if and only if it also contains a 4-ary function $s$ that satisfies the Siggers equations [Sig10]:

$$\forall x, y, s(x,y,x,y) = s(x,y,y,y) = s(x,x,y,y).$$

(1.2)

A way to interpret this fact is that the equation itself is not what matters, but it is rather the fact that some non-trivial equation holds in Pol$(\Gamma)$ that is relevant. This leads us to the following considerations.

Definition 1.2. Let $C$ be a set of functions (of any arity) over some set $C$. We say that $C$ is a function clone if it satisfies the following two conditions:

- for all positive integer $n$ and all $1 \leq i \leq n$, the function
  $$\text{pr}^{(n)}_i : (x_1, \ldots, x_n) \mapsto x_i,$$
  called the $i$th projection of arity $n$, is in $C$;
• if $f \in \mathcal{C}$ has arity $m$ and $g_1, \ldots, g_m \in \mathcal{C}$ have arity $n$, then the function $f \circ (g_1, \ldots, g_m)$ that maps $(x_1, \ldots, x_n)$ to $f(g_1(x_1, \ldots, x_n), \ldots, g_m(x_1, \ldots, x_n))$ is in $\mathcal{C}$.

Note that for any structure $\Gamma$, $\text{Pol}(\Gamma)$ is a function clone. The smallest function clone on a set $D$ is the clone that only contain the projections. We denote by $1$ the clone that contains only the projections on the set $\{0,1\}$. A clone homomorphism between two clones $\mathcal{C}$ and $\mathcal{D}$ is a function $\xi : \mathcal{C} \rightarrow \mathcal{D}$ such that:

• $\xi(f)$ has the same arity as $f$ for all $f \in \mathcal{C}$,
• $\xi(p_{i,n}^C) = p_{i,n}^D$ for all $1 \leq i \leq n$,
• $\xi(f \circ (g_1, \ldots, g_m)) = \xi(f) \circ (\xi(g_1), \ldots, \xi(g_m))$.

Note that clone homomorphisms preserve equations, in the sense that any property of a function $f$ that can be expressed with universal quantifiers ranging over elements of the domain of $f$, equality, and conjunctions, are preserved by clone homomorphisms. One way of making the notion of having only “trivial polymorphisms” is by saying that there is a clone homomorphism from $\text{Pol}(\Gamma)$ to $1$, which is also called a projective homomorphism after [BPP14]. Since no equation holds in $1$, no equation holds in $\text{Pol}(\Gamma)$. The following theorem strengthen the idea that clone homomorphisms are relevant to the study of constraint satisfaction problems.

**Theorem 1.4.** Let $\Gamma, \Delta$ be two structures over a finite domain (not necessarily the same). If there exists a clone homomorphism $\xi : \text{Pol}(\Gamma) \rightarrow \text{Pol}(\Delta)$, then $\text{CSP}(\Delta)$ is polytime many-one reducible to $\text{CSP}(\Gamma)$.

Consider the following CSP interpretation of the boolean 1IN3-satisfiability problem. Let $\Delta$ be the structure over $\{0,1\}$ that contains a single relation $R = \{(1,0,0), (0,1,0), (0,0,1)\}$. It is known that $\text{Pol}(\Delta) = 1$ and (by Schaefer’s Theorem) that $\text{CSP}(\Delta)$ is NP-hard. By the previous theorem, we have that if there exists a clone homomorphism from $\text{Pol}(\Gamma)$ to $1$, then $\text{CSP}(\Gamma)$ is NP-hard. Conversely if there is no clone homomorphism from $\text{Pol}(\Gamma)$ to $1$ then $\text{Pol}(\Gamma)$ has a cyclic operation (see [BJK05]). The tractability conjecture can thus be phrased in terms of the existence of a clone homomorphism to $1$. For technical reasons, in this formulation one needs to look at a structure $\Delta$, which has the same CSP as $\Gamma$, and such that every function of $\text{Pol}(\Delta)$ satisfies $f(x, \ldots, x) = x$. This structure $\Delta$ is uniquely determined up to isomorphism, and is called the core of $\Gamma$. The notion of core will also appear in the context of infinite structures.

**Question 2.** Let $\Gamma$ be a finite structure with a finite signature, and let $\Delta$ be the core of $\Gamma$. Suppose that there is no clone homomorphism from $\text{Pol}(\Delta)$ to $1$. Does there exist a polynomial time algorithm that solves $\text{CSP}(\Gamma)$?

### 1.2 From the Finite to the Infinite

As stated above, the complexity of $\text{CSP}(\Gamma)$ is much wilder when $\Gamma$ is a structure over an infinite domain. We see now how the universal algebraic approach can be generalised in certain contexts, and how the study of the infinite case exposed new aspects that are blurred in the finite case.

Remember that a permutation group is a non-empty set of permutations of some set $D$ that is closed under composition and inverses. The orbit of an element $d \in D$ is the set $\{\alpha(d) \mid \alpha \in G\}$.
**Definition 1.3.** A permutation group $G$ over some set $D$ is called *oligomorphic* if for all positive integer $n$ there are only finitely many orbits of $n$-tuples of elements from $D$.

It is routine to check that $\text{Aut}(\Gamma)$ is a permutation group.

**Definition 1.4.** Let $\Gamma$ be a structure over a countably infinite domain. We say that $\Gamma$ is *$\omega$-categorical* if $\text{Aut}(\Gamma)$ is an oligomorphic permutation group.

The notion of $\omega$-categoricity allows us to extend some results that hold in the finite case to the infinite, as in the following:

**Lemma 1.5** (From [Bod10]). Let $\Gamma$ be an $\omega$-categorical structure. Then $\text{Inv}(\text{Pol}(\Gamma))$ is the set of relations that have a primitive positive definition in $\Gamma$.

The last notion that we define here is the topology of pointwise convergence. Let us fix a function clone $\mathcal{C}$ over a set $C$. We define a topology on $\mathcal{C}$ as follows: for any positive integers $n$ and $k$, any $n$-tuples $a^1, \ldots, a^k \in C^n$ and any elements $b_1, \ldots, b_k \in C$, let $S_{a^1, \ldots, a^k}^{b_1, \ldots, b_k} = \{f : C^n \to C \mid f \in \mathcal{C}, \forall 1 \leq i \leq n, f(a^i) = b_i\}$. The topology generated by these sets is called the topology of pointwise convergence. Under this topology, the composition of operations of $\mathcal{C}$ is continuous, and we say that $\mathcal{C}$ is a topological clone. Theorem 1.4 generalises to $\omega$-categorical structures using this topology:

**Theorem 1.6** (From [BP14]). Let $\Gamma, \Delta$ be two $\omega$-categorical structures with finite signatures. If there exists a continuous clone homomorphism $\xi : \text{Pol}(\Gamma) \to \text{Pol}(\Delta)$, then $\text{CSP}(\Delta)$ is polytime many-one reducible to $\text{CSP}(\Gamma)$.

**Remark 1.** Although Theorem 1.6 requires the existence of a continuous clone homomorphism, we do not know any example of $\omega$-categorical structures $\Gamma$ and $\Delta$ with finite signatures such that there exists a clone homomorphism $\xi : \text{Pol}(\Gamma) \to \text{Pol}(\Delta)$, but such that there does not exist any continuous one. This fact is a motivation for the study of reconstruction and automatic continuity, see e.g. [BPP13].

When the structures are not $\omega$-categorical, little is known on what general approach of constraint satisfaction problems one can have. In [BHM09], the Galois correspondence between polymorphisms and pp-definable relations is extended to certain types of structures, and it is shown that every CSP can be formulated with a structure in which a weak version of Lemma 1.5 holds. This new approach has been used successfully in [BMM15], where a complexity dichotomy is exhibited for the CSP of structures definable in $(\mathbb{Z}; \text{suc})$, some of which are not $\omega$-categorical.

**Structure of this document.** We provide several criteria for the tractability or NP-completeness for large classes of problems. In particular, we confirm some conjectures of topological and algebraic nature about the complexity of constraint satisfaction problems in the $\omega$-categorical setting. In Chapter 3, we classify the complexity of all the constraint satisfaction problems that are definable over the structure $(\mathbb{N}; 0)$, i.e., the structure over a countable domain with a single constant. In Chapter 4, we try to formulate a characterization of the tractability border for problems in MMSNP, which is a fragment of second-order existential logic. We attempt to use infinitary methods in order to obtain a dichotomy result for MMSNP, or fragments thereof. Chapter 2 is a collection of results from model theory and universal algebra that will be needed in the following chapters.

**Acknowledgements.** The third chapter is joint work with Manuel Bodirsky and François Bossière. The fourth chapter is extracted from a paper authored by Manuel Bodirsky and myself that will be presented during the Workshop on Qualitative Spatial and Temporal Reasoning held in Dresden, Germany.
Chapter 2

Preliminaries

In this chapter, we will present all the concepts and theorems that are needed in the following. The theorems are stated without proof, and we invite the interested reader to read [Bod10] in case she wants to learn more about the state-of-the-art techniques used in the area of infinite-domain constraint satisfaction. Some of the concepts presented here have already been encountered in the introduction above. We believe the repetition is not harmful, and the present chapter is meant to serve as a reference for the subsequent chapters.

2.1 Model Theory

Definition 2.1 (Relational signatures, structures). A relational signature $\tau$ is a set of symbols each of which is associated with a finite arity. A relational $\tau$-structure is a tuple $\Gamma = (D; \{ R^\Gamma : R \in \tau \})$ such that for each $R \in \tau$ of arity $n$, $R^\Gamma$ is a relation on $D$ of arity $n$. The set $D$ is called the domain, or universe of $\Gamma$ and denoted by $\text{Dom}(\Gamma)$.

In the following we always consider that the symbol $=$ is in the signature and that for every $\tau$-structure $\Gamma$ it is interpreted in $\Gamma$ by $\{(a, a) \mid a \in \text{Dom}(\Gamma)\}$.

Definition 2.2 (Homomorphisms). Let $\Gamma$ and $\Delta$ be $\tau$-structures. A homomorphism $h$ from $\Gamma$ to $\Delta$ is a function $\text{Dom}(\Gamma) \to \text{Dom}(\Delta)$ such that for each symbol $R$ from $\tau$ of arity $n$ and each $n$-tuple $(a_1, \ldots, a_n)$ from $\text{Dom}(\Gamma)$ we have

\[(a_1, \ldots, a_n) \in R^\Gamma \Rightarrow (h(a_1), \ldots, h(a_n)) \in R^\Delta. \quad (2.1)\]

If $\Gamma$ and $\Delta$ are the same structure, we also say that $h$ is an endomorphism. If the implication in (2.1) is in fact an equivalence, we say that $h$ is an embedding. A surjective embedding is also called an isomorphism. If there exists a homomorphism from $\Gamma$ to $\Delta$ we say that $\Gamma$ homomorphically maps to $\Delta$, and if $\Delta$ also homomorphically maps to $\Gamma$ we say that $\Gamma$ and $\Delta$ are homomorphically equivalent.

Definition 2.3 (Definable relations). Let $\Gamma$ be a relational structure and $R \subset \text{Dom}(\Gamma)^n$ be a relation. We say that $R$ is first-order definable (resp. ep-, pp-definable) in $\Gamma$ if there exists a first-order formula $\phi(x_1, \ldots, x_n)$ (resp. ep-, pp-formula) such that

\[(a_1, \ldots, a_n) \in R \Leftrightarrow \Gamma \models \phi(a_1, \ldots, a_n). \quad (2.2)\]
Example 1. The ternary relation over the rational number that contains the triples \((a, b, c)\) such that \(b\) is strictly between \(a\) and \(c\) is definable in \((\mathbb{Q}; <)\) by the formula \((x < y \land y < z) \lor (z < y \land y < x)\). The set of minimal elements in some partial order \((D; <)\) is definable by the first-order formula \(\forall y \neg (y < x)\).

**Definition 2.4** (First-order reduct). Let \(\Gamma\) and \(\Delta\) be relational structures over the same domain. We say that \(\Delta\) is a first-order reduct if all the relations of \(\Delta\) are first-order definable in \(\Gamma\).

**Definition 2.5** (Direct product). Let \(\Gamma_1, \Gamma_2\) be two \(\tau\)-structures. The direct product of \(\Gamma_1\) and \(\Gamma_2\), denoted by \(\Gamma_1 \times \Gamma_2\), is the \(\tau\)-structure whose domain is \(\text{Dom}(\Gamma_1) \times \text{Dom}(\Gamma_2)\) such that for every \(R \in \tau\), the interpretation of \(R\) in \(\Gamma_1 \times \Gamma_2\) is \(\{(a_1, b_1), \ldots, (a_k, b_k)\} \in \text{Dom}(\Gamma_1) \times \text{Dom}(\Gamma_2)^k | a \in R^{\Gamma_1}, b \in R^{\Gamma_2}\}\).

In other words, the direct product is the biggest structure over \(\text{Dom}(\Gamma_1) \times \text{Dom}(\Gamma_2)\) such that the projections \(p_i: \Gamma_1 \times \Gamma_2 \to \Gamma_i\) are homomorphisms.

A polymorphism of arity \(n\) of a structure \(\Gamma\) is a homomorphism from \(\Gamma^n\) to \(\Gamma\).

**Definition 2.6** (Theory of a structure). The first-order theory \(\text{Th}(\Gamma)\) of a structure \(\Gamma\) is the set of first-order sentences \(\phi\) such that \(\Gamma \models \phi\).

**Definition 2.7** (Type). Let \(\Gamma\) be a structure and \(a\) be a tuple of elements from \(\text{Dom}(\Gamma)\). The type of \(a\) in \(\Gamma\), denoted by \(\text{tp}^\Gamma(a)\) is the theory \(\text{Th}(\Gamma, a)\). Equivalently,

\[\text{tp}^\Gamma(a) = \{\phi(x) | \Gamma \models \phi(a)\}\].

**Definition 2.8** (Isolation). Let \(\Gamma\) be a structure, and \(\text{tp}(a)\) be a type in \(\Gamma\). We say that \(\text{tp}(a)\) is isolated by \(\phi(x)\) if for every first-order formula \(\psi()\), we have \(\Gamma \models \phi(a)\) iff \(\Gamma \models \psi(a)\). Informally, \(\text{tp}(a)\) is isolated if all that can be said about \(a\) in \(\Gamma\) can be said with a single formula.

**Definition 2.9** (\(\omega\)-categoricity). A countably infinite structure \(\Gamma\) is \(\omega\)-categorical if for every countable structure \(\Delta\) such that \(\text{Th}(\Gamma) = \text{Th}(\Delta)\), there exists an isomorphism between \(\Gamma\) and \(\Delta\).

Example 2. A standard back-and-forth argument shows that the structure \((\mathbb{Q}; <)\) is \(\omega\)-categorical. For a nonexample, consider the structure \((\mathbb{Z}; \text{succ})\) of the integers with successor. The structure that consists of two disjoint copies of \((\mathbb{Z}; \text{succ})\) has the same first-order theory, though the two structures are not isomorphic.

**Theorem 2.1** (Ryll-Nardzewski, Engelev, Svenonius). Let \(\Gamma\) be a countable structure with an at most countable signature. The following are equivalent:

- \(\Gamma\) is \(\omega\)-categorical,
- for each positive integer \(n\), there are only finitely many types of \(n\)-tuples from elements of \(\text{Dom}(\Gamma)\).

Many properties that hold for finite structures also hold for \(\omega\)-categorical structures. We close this section with the properties that we will use in the following.

**Theorem 2.2.** Let \(\Gamma\) be an \(\omega\)-categorical structure, and let \(R \subseteq \text{Dom}(\Gamma)^k\) be a relation over the domain of \(\Gamma\). The following hold:
• $R$ has a first-order definition in $\Gamma$ iff $R$ is preserved by all the automorphisms of $\Gamma$;

• $R$ has a primitive positive definition in $\Gamma$ iff $R$ is preserved by all the polymorphisms of $\Gamma$.

Moreover if $R$ is first-order definable in $\Gamma$ and consists of $n$ orbits of $k$-tuples under $\text{Aut}(\Gamma)$, then $R$ has a primitive positive definition in $\Gamma$ iff it is preserved by all the polymorphisms of arity $n$ of $\Gamma$.

**Definition 2.10** (Model-complete cores). A structure $\Gamma$ is called a model-complete core if all the endomorphisms of $\Gamma$ are self-embeddings, and moreover for every finite subset $A$ of $\text{Dom}(\Gamma)$ and every endomorphism $e$ of $\Gamma$, there exists an automorphism $\alpha$ of $\Gamma$ such that $e|_A = \alpha|_A$.

**Theorem 2.3.** Let $\Gamma$ be an $\omega$-categorical structure. There exists a structure $\Delta$ (finite or countably infinite), unique up to isomorphism, that is a model-complete core and that is homomorphically equivalent to $\Gamma$.

A structure $\Delta$ as in the theorem above is called the core of $\Gamma$, and up to isomorphism the core of $\Gamma$ is always an induced substructure of $\Gamma$.

It is useful to remark that a direct consequence of the definition of model-complete cores and Theorem 2.2 is the fact that the orbits of tuples under $\text{Aut}(\Gamma)$ are pp-definable in $\Gamma$ when $\Gamma$ is a model-complete core.

**Definition 2.11** (Homogeneous structures). A structure $\Gamma$ is called homogeneous (sometimes ultrahomogeneous in the literature) if every partial isomorphism, that is, every embedding from some induced substructure of $\Gamma$ to $\Gamma$, can be extended to an automorphism of $\Gamma$.

The notions of homogeneity of $\omega$-categoricity are intimately connected in our applications. A structure $\Gamma$ is said to have quantifier elimination if every first-order formula with free variables $\phi(x_1, \ldots, x_n)$ is equivalent in $\Gamma$ to some quantifier-free formula $\psi(x_1, \ldots, x_n)$.

**Theorem 2.4.** Let $\Gamma$ be a relational structure with a finite signature. Then $\Gamma$ is homogeneous if, and only if, it is $\omega$-categorical and has quantifier elimination.

**Definition 2.12** (Age of a structure). Let $\Gamma$ be a relational structure. The age of $\Gamma$, denoted by $\text{Age}(\Gamma)$, is the class of all the finite induced substructure of $\Gamma$.

### 2.2 Universal Algebra

We noted in the introduction that the complexity of $\text{CSP}(\Gamma)$ is completely determined by the set of relations that have a primitive positive definition in $\Gamma$. We also saw that when $\Gamma$ is finite or $\omega$-categorical, the complexity of $\text{CSP}(\Gamma)$ is determined by the set of polymorphisms of $\Gamma$. We can endow this set with the canonical algebraic structure given by (generalised) composition of functions, noting that the composition of polymorphisms of $\Gamma$ yields a polymorphism of $\Gamma$. The corresponding algebraic structure is called a function clone.

**Definition 2.13** (Function clone). A set $C$ of functions on some set $C$ is called a function clone if all the projections

$$\text{pr}^{(n)}_i : (x_1, \ldots, x_n) \mapsto x_i$$

are in $C$ and if for every $f \in C$ has arity $n$ and $g_1, \ldots, g_n \in C$ have arity $m$ then the function $f \circ (g_1, \ldots, g_n)$ is in $C$. 

If \( \mathcal{F} \) is a set of functions, we write \( \langle \mathcal{F} \rangle \) for the smallest function clone containing \( \mathcal{F} \).

We review in this section how the algebraic properties of this clone give us information about the complexity of CSP(\( \Gamma \)). We call an algebra a structure over a purely functional signature. We write \( A = (A, f_1, f_2, \ldots) \) for an algebra. If \( \tau \) is a functional signature and \( X \) is a set, we define inductively the set of \( \tau \)-terms with variables in \( X \) as follows:

- for every \( x \in X \), \( x \) is a \( \tau \)-term,

- if \( f \in \tau \) has arity \( n \) and \( t_1, \ldots, t_n \) are \( \tau \)-terms, then \( f(t_1, \ldots, t_n) \) is a \( \tau \)-term.

An element \( x \in X \) is free in a term \( t \) if \( x \) appears in \( t \).

Note that if \( A \) is a \( \tau \)-algebra and \( t \) is a term with \( n \) free variables, we can define the operation \( t^A : A^n \to A \) by structural induction on \( t \) in the obvious way. The set of all the operations \( t^A \) for \( t \) a \( \tau \)-term with finitely many free variables is called the term clone of \( A \), denoted by \( \text{Clo}(A) \). It follows directly from the definition of terms that \( \text{Clo}(A) \) is a function clone for every \( A \), whence the name.

For any relational structure \( \Gamma \), the set \( \text{Pol}(\Gamma) \) that contains the polymorphisms of \( \Gamma \) is a function clone. By enumerating these polymorphisms in some arbitrary way, we form an algebra \( C = (\text{Dom}(\Gamma), f_1, f_2, \ldots) \) such that \( \text{Clo}(C) = \text{Pol}(\Gamma) \). The algebra \( C \) is said to be a polymorphism algebra of \( \Gamma \). We have seen above that in some sense, the set of polymorphisms of an \( \omega \)-categorical structure \( \Gamma \) encodes a description of the relations that have a primitive positive definition in \( \Gamma \), and thus also encodes the complexity of CSP(\( \Gamma \)). By using the terminology of universal algebra, we shift the interest from the study of algebras to the study of varieties of algebras, which we present now.

**Definition 2.14.** Let \( A, B \) be two \( \tau \)-algebras. A homomorphism of algebras is a function \( h : A \to B \) such that for every \( f \in \tau \), we have that \( h(f^A(a_1, \ldots, a_n)) = f^B(h(a_1), \ldots, h(a_n)) \). We write \( H(A) \) for the class of algebras \( B \) for which there exists a surjective homomorphism \( A \to B \).

If \( K \) is a class of \( \tau \)-algebras, we write \( H(K) \) for the union of all the \( H(A) \) with \( A \in H \).

**Definition 2.15** (Subalgebra). Let \( A \) be a \( \tau \)-algebra and \( S \) be a subset of \( A \). If for all \( f \in \tau \) we have \( f(S^n) \subseteq S \), then we can define a \( \tau \)-algebra \( S \) on \( S \) by setting \( f^S(a_1, \ldots, a_n) = f^A(a_1, \ldots, a_n) \). We say that \( S \) is a subalgebra of \( A \), and we write \( S(A) \) for the class of subalgebras of \( A \).

As above, we write \( S(K) \) for \( \bigcup_{A \in K} S(A) \).

**Definition 2.16** (Product of algebras). Let \( (A_i)_{i \in I} \) be an indexed family of \( \tau \)-algebras. We define the product algebra of \( (A_i)_{i \in I} \) on the set \( \prod_{i \in I} A_i \) by setting

\[
(f \prod_{i \in I} A_i(s_1, \ldots, s_n))(i) = f^A_i(s_1(i), \ldots, s_n(i))
\]

and we let \( P(K) \) be the class of all the algebras obtained by taking products of algebras in \( K \). We also write \( P_{fin}(K) \) for the class of algebras obtained by taking finitary products of algebras in \( K \).

**Definition 2.17** (Variety). Let \( \tau \) be a functional signature. A variety \( \mathcal{V} \) is a class of \( \tau \)-algebras that is closed under products, subalgebras, and homomorphic images.

**Theorem 2.5.** Let \( A \) be an algebra. The smallest variety that contains \( A \) is the class \( HSP(A) \).
Theorem 2.6 (From [BP14]). Let \( \Gamma, \Delta \) be finite or \( \omega \)-categorical structures (not necessarily over the same signature or domain). Suppose that there exists a polymorphism algebra \( C \) of \( \Gamma \), and \( A \in \text{HSP}_{\text{fin}}(C) \) such that \( \text{Clo}(A) \subseteq \text{Pol}(\Delta) \). Then there exists a polynomial-time reduction from \( \text{CSP}(\Delta) \) to \( \text{CSP}(\Gamma) \).

This theorem is a very important tool, as it allows us to derive NP-hardness of \( \text{CSP}(\Gamma) \) from the NP-hardness of CSPs that were already classified before. For example, if \( \Gamma \) and \( \Delta \) have the same domain and \( \text{Pol}(\Delta) \subseteq \text{Pol}(\Gamma) \), we get (by letting \( A = C \) be any polymorphism algebra of \( \Gamma \)) that \( \text{CSP}(\Delta) \) reduces to \( \text{CSP}(\Gamma) \). We will see examples of this in the following chapters, where we will mainly be interested in the case where \( A \) is an algebra such that \( \text{Clo}(A) \) only has projections, in which case we write \( 1 \) for \( \text{Clo}(A) \). Every such algebra is a polymorphism algebra of an \( \text{NP} \)-hard problem, and applying Theorem 2.6 above gives us that \( \text{CSP}(\Gamma) \) is \( \text{NP} \)-hard.

2.3 Canonical Functions

The so-called canonical functions are such a powerful tool to understand the polymorphisms of some structures that this section is devoted to them. First, we have to recall some basic notions of point-set topology.

Definition 2.18 (Topology). Let \( X \) be a set. A topology \( \mathcal{T} \) on \( X \) is a set of subsets of \( X \) such that:

- \( \emptyset \) and \( X \) are in \( \mathcal{T} \),
- if \( U \) and \( V \) are in \( \mathcal{T} \) then so is \( U \cap V \),
- if \( (U_i)_{i \in I} \) is a family of members of \( \mathcal{T} \), then \( \bigcup_{i \in I} U_i \) is in \( \mathcal{T} \).

Let \( X \) be a set, and \( \mathcal{B} \) is a set of subsets of \( X \) that contains \( \emptyset, X \) and that is closed under finite intersections. Then we call the smallest topology that contains \( \mathcal{B} \) is called the topology generated by \( \mathcal{B} \), and \( \mathcal{B} \) is called a base for that topology. In the following, we take \( X \) to be the set of all the finitary operations on some domain \( D \). Formally, \( X = \bigcup_{n \geq 1} D^n \).

Let \( n \geq 1 \) be a positive integer, \( f \in X \) be an \( n \)-ary function and \( A \subset D \) be a finite subset of \( D \). Let

\[
S(f, A) := \{ g \in D^n \mid f|_A = g|_A \} \subseteq X
\]

be a subset of \( X \). We let \( \mathcal{B} \) be the union of all the \( S(f, A) \) for \( f \in X \) and every finite subset \( A \) of \( D \). The topology generated by \( \mathcal{B} \) is called the topology of pointwise convergence on \( X \). This topology gets its name from the fact that the convergence notion associated with the topology can be characterized by the property that a sequence of \( k \)-ary functions \( (f_n) \) of \( X \) converges to some \( k \)-ary function \( f \in X \) iff for every \((a_1, \ldots, a_k) \in D^k \), there exists some \( n \in \mathbb{N} \) such that \( f_m(a_1, \ldots, a_k) = f(a_1, \ldots, a_k) \) for every \( m \geq n \). In other words, on every \( k \)-tuple the sequence \( (f_n)_{n \in \mathbb{N}} \) stabilises from some point on.

While there are usually many topologies one can define on a set of functions, the topology of pointwise convergence is particularly well-suited for our purposes. One of its main features is that for every structure \( \Gamma \), the set \( \text{Pol}(\Gamma) \) is a closed subset of \( X \), that is, if a sequence of polymorphisms of \( \text{Pol}(\Gamma) \) converges to some \( f \in X \), then \( f \) is a polymorphism of \( \Gamma \), too. This is simply due to the fact that the property of being a polymorphism is local, i.e., can be verified only by looking at finite parts of \( D \).
Definition 2.19 (Canonical function). Let $\Gamma, \Delta$ be structures. A function $f: \text{Dom}(\Gamma)^n \to \text{Dom}(\Delta)$ is called canonical from $\Gamma$ to $\Delta$ if for every automorphisms $\alpha_1, \ldots, \alpha_n$ of $\Gamma$ and every $m$-tuples $s_1, \ldots, s_n \in \text{Dom}(\Gamma)^m$, there exists an automorphism $\beta$ of $\Delta$ such that

$$f(s_1, \ldots, s_n) = \beta(f(\alpha_1(s_1), \ldots, \alpha_n(s_n)))$$

(2.3)

i.e., perturbations by automorphisms of the domain can be “corrected” using an automorphism of the codomain.

Remark 2. This definition can also be stated using the notion of types from model theory. From this point of view, a function is canonical if the type of the tuple $f(s_1, \ldots, s_n)$ in $\Delta$ is determined by the types of $s_1, \ldots, s_n$ in $\Gamma$.

The main theorem that we will use in the following states that for some appropriate structures, the study of polymorphisms can be carried out by looking at canonical polymorphisms only. The next notations are borrowed from Ramsey theory. If $A, B$ are two structures, we write $\binom{B}{A}$ for the set of all the embeddings $e: A \to B$.

Definition 2.20 (Partition arrow). Let $A, B, C$ be structures over the same signature. Let $r$ be a positive integer. We write $C \rightarrow (B)^A_r$ as a shorthand for the following statement: for every colouring $\chi: \binom{C}{A} \to [r]$, there exists an embedding $e \in \binom{C}{B}$ such that $\chi$ is constant on $e \circ \binom{B}{A}$.

In other words, if $C \rightarrow (B)^A_r$ holds, whenever the copies of $A$ in $C$ are coloured with colours $\{1, \ldots, r\}$, there exists a copy $B'$ of $B$ in $C$ such that all the copies of $A$ that sit inside $B'$ have the same colour.

Definition 2.21 (Ramsey class). Let $\mathcal{C}$ be a class of finite structures. We say that $\mathcal{C}$ has the Ramsey property iff for all $A, B \in \mathcal{C}$ and $r \geq 1$, there exists $C \in \mathcal{C}$ such that $C \rightarrow (B)^A_r$.

Example 3. Let $\mathcal{C}$ be the class of all the finite linear orders. Then the classical Ramsey theorem states precisely that $\mathcal{C}$ has the Ramsey property.

Definition 2.22 (Ramsey structure). A homogeneous structure $\Gamma$ is called Ramsey if its age is Ramsey.

Example 4. The structure $\langle \mathbb{Q}; < \rangle$ is homogeneous, and its age is the set of all finite linear orders. Thus $\langle \mathbb{Q}; < \rangle$ is a Ramsey structure.

Theorem 2.7 (From [BPT13]). Let $\Delta$ be an $\omega$-categorical ordered Ramsey structure. Let $f: \text{Dom}(\Delta)^n \to \text{Dom}(\Delta)$ and $c_1, \ldots, c_k$ be elements of $\text{Dom}(\Delta)$. Then there exists a function $g: \text{Dom}(\Delta)^n \to \text{Dom}(\Delta)$ that is canonical from $\langle \Delta; c_1, \ldots, c_k \rangle$ to $\Delta$, such that $g$ and $f$ agree on $\{c_1, \ldots, c_k\}^n$, and such that

$$g \in \langle \{f\} \cup \text{Aut}(\Delta; c_1, \ldots, c_k) \rangle.$$
Theorem 2.8. Let $\Delta$ be an $\omega$-categorical ordered Ramsey structure. Let $f: \text{Dom}(\Delta)^n \to \text{Dom}(\Delta)$. There exists a canonical function $g: \text{Dom}(\Delta)^n \to \text{Dom}(\Delta)$ in

$\{\beta \circ f \circ (\alpha_1, \ldots , \alpha_n) \mid \beta, \alpha_1, \ldots , \alpha_n \in \text{Aut}(\Delta)\}$,

in which case we say that $g$ is interpolated by $f$ modulo automorphisms of $\Delta$.

Theorems 2.7 and 2.8 are not comparable in strength. Theorem 2.7 gives us a function for which we can control a finite number of values that function takes. The counterpart is that we have no control on how that function is obtained, i.e., we only know that it is a limit of functions in the smallest clone containing $f$ and automorphisms of $\Delta$. On the other hand, the function given by Theorem 2.8 is a limit of functions that have a simple expression, that allows us to keep some general properties of $f$. For example, since (left- or right-)composing an injective function with bijections yields an injective function, the canonical function we obtain from an injective function is again injective.
Chapter 3

The Complexity of First-Order Reducts of \((\mathbb{N}; 0)\)

We provide here criteria for the complexity of the CSP of first-order reducts of the structure \((\mathbb{N}; 0)\), the structure over \(\mathbb{N}\) with only one symbol naming a constant. All the reducts of this structure are \(\omega\)-categorical, and fall into the scope of Theorem 2.2. The border between polynomial-time tractability and \(\text{NP}\)-completeness can be described in terms of projective clone homomorphisms:

**Theorem 3.1.** Let \(\Gamma\) be a finite-signature first-order reduct of \((\mathbb{N}; 0)\), and let \(\Delta\) be the model-complete core of \(\Gamma\). Then \(\text{CSP}(\Delta)\) is \(\text{NP}\)-complete if there is a continuous clone homomorphism from \(\text{Pol}(\Delta)\) to \(1\), and is in \(\mathbb{P}\) otherwise.

Note that for any countable set \(D\) and \(d \in D\), the structure \((D; d)\) is isomorphic to \((\mathbb{N}; 0)\). In particular, \((\mathbb{Q}^+; 0)\) is isomorphic to \((\mathbb{N}; 0)\). Note that the age of \((\mathbb{Q}^+; 0, <)\) consists of all the finite linear orders where the minimal element is singled out, and is therefore a Ramsey class by Ramsey’s theorem. It follows that there exists a linear order \(<\) on \(\mathbb{N}\) such that \((\mathbb{N}; 0, <)\) is a Ramsey class (in fact, the order \(<\) is simply obtained by taking the image of the canonical order on \(\mathbb{Q}^+\) under any bijection between \(\mathbb{Q}^+\) and \(\mathbb{N}\) taking \(0\) to \(0\)).

In order to prove Theorem 3.1, we build on previous classification results. The first one is Schaefer’s Theorem, that we already mentioned, and that gives a complexity classification for the CSP of a structure over a two-element set. The second result we use is from [BK06] and gives a classification of all the CSPs of first-order reducts of \((\mathbb{N}; =)\), the structure over a countable set that contains only the equality in its language. The description of reducts of \((\mathbb{N}; =)\) whose CSP is polynomial-time tractable is as follows:

**Theorem 3.2.** (From [BK06]) Let \(\Gamma\) be a first-order reduct of \((\mathbb{N}; =)\). Then \(\text{CSP}(\Gamma)\) is in \(\mathbb{P}\) if \(\Gamma\) has a constant unary or an injective binary polymorphism. Otherwise it is \(\text{NP}\)-complete.

**Remark 3.** Although it is not done in [BK06], we can rephrase the previous theorem in terms of clone homomorphisms. In this case, Theorem 3.2 states that the CSP of a reduct of \((\mathbb{N}; 0)\) is in \(\mathbb{P}\) if there is no clone homomorphism from \(\text{Pol}(\Gamma)\) to \(1\), and is \(\text{NP}\)-complete otherwise. We will use Theorem 3.2 in both forms in the following. The same applies to Schaefer’s Theorem, which we will use both with the existence of polymorphisms satisfying some properties as in Theorem 1.3, or the equivalent version in terms of clone homomorphisms.
Figure 3.1: The binary operations that are canonical with respect to \((\mathbb{N}; \prec, 0)\) and injective on \((\mathbb{N} \setminus \{0\})^2\). If \(f\) is drawn, its dual \((x, y) \mapsto f(y, x)\) is not displayed. The symbols \(c\) and \(d\) denote constants different from 0 and \(\text{inj}\) means that the function is injective and not 0.

### 3.1 The Role of Binary Injective Operations

The machinery of canonical functions is very powerful in our case. Depending on their behaviour on elements, we can separate the binary functions that are canonical with respect to \((\mathbb{N}; 0, \prec)\) and injective on \((\mathbb{N} \setminus \{0\})^2\) into finitely many classes, which are pictured in Figure 3.1. We name three of these classes, which will be important in the following:

**Definition 3.1.** Let \(g: \mathbb{N}^2 \to \mathbb{N}\) be canonical with respect to \((\mathbb{N}; 0, \prec)\) and injective on \((\mathbb{N} \setminus \{0\})^2\).

- \(g\) is in \((\text{inj}, 0, 0)\) if \(g\) has type 1 in Figure 3.1.
- \(g\) is in \((\text{inj}, \text{inj}, 0)\) if \(g\) has type 2 in Figure 3.1.
- \(g\) is in \((\text{inj}, \text{inj}, \text{inj})\) if \(g\) has type 3 in Figure 3.1.

**Proposition 3.3.** Let \(f: \mathbb{N}^2 \to \mathbb{N}\) be canonical with respect to \((\mathbb{N}; 0, \prec)\) and be such that \(f(0) = 0\) and \(f\) is injective on \((\mathbb{N} \setminus \{0\})^2\). Then \(f\) is of one of the forms that appear in Figure 3.1.

**Proof.** We look at the types of pairs in \((\mathbb{N}; \prec)\). By quantifier elimination in \((\mathbb{N}; 0, \prec)\), and since all the types are isolated in \((\mathbb{N}; 0, \prec)\), each of these types is determined by a single quantifier-free formula. These are the only possibilities (up to permutation of \(x\) and \(y\)):

- \(x = 0 \land y = 0\),
- \(x = 0 \land y \neq 0\),
- \(x \neq 0 \land y \neq 0 \land x \neq y\).

We show that \(f\) has one of the requested forms on \((\mathbb{N} \setminus \{0\}) \times 0\), the dual case being similar. Let \(a, b, c\) be in \(\mathbb{N} \setminus \{0\}\) such that \(a, b,\) and \(c\) are all different. The types of \((a, b)\) and \((a, c)\) are the
same. By canonicity of \( f \), the types of \((f(a,0), f(b,0)), (f(a,0), f(c,0)), \) and \((f(b,0), f(c,0))\) are all the same. If \( f(a,0) = 0 \), then \( f(b,0) = 0 \) and then \( f(c,0) = 0 \), too. Hence, \( f \) is zero on \( \mathbb{N} \times 0 \). If \( f(a,0) \neq 0 \), there are two cases. Either \( f(a,0) = f(b,0) \), then \( f(a,0) = f(c,0) \) and \( f \) is constant on \((\mathbb{N} \setminus \{0\}) \times 0\). Finally, if \( f(a,0) \neq f(b,0) \) and \( f(b,0) \neq f(c,0) \) then \( f \) is injective (and nonzero) on \((\mathbb{N} \setminus \{0\}) \times 0\).

**Proposition 3.4.** Let \( g \) be a binary operation over \( \mathbb{N} \) such that \( g(0) = 0 \) and \( g \) injective on \((\mathbb{N} \setminus \{0\}) \times (\mathbb{N} \setminus \{0\})\). Then \( \langle \{g\} \cup \text{Aut}(\mathbb{N}; \prec, 0) \rangle \) contains an operation \( h \) which belongs to one of the following classes: \((\text{inj, inj, inj}), (\text{inj, inj, 0}), \) or \((\text{inj, 0, 0})\).

**Proof.** Let \( h \) be given by Theorem 2.8. If \( h \) is in case 1, 2, or 3, we are done. Suppose that \( h \) falls into case 5, that is, \( h \) has the following behaviour:

<table>
<thead>
<tr>
<th></th>
<th>inj</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>inj</td>
</tr>
</tbody>
</table>

Consider the function \( h' : (x, y) \mapsto h(h(x, y), h(y, x)) \). It is easy to check that this function is in \((\text{inj, inj, inj})\) and in \( \langle \{g\} \cup \text{Aut}(\mathbb{N}; \prec, 0) \rangle \). If \( h \) is in case 6, the function \( (x, y) \mapsto h(h(x, y), y) \) is in case 2, i.e. it is of type \((\text{inj, inj, 0})\). Finally, if \( h \) is in case 4 then \( (x, y) \mapsto h(h(y, x), x) \) is in case 5, and we are done.

3.2 Polynomial-Time Tractable CSPs

3.2.1 Syntactic Forms

Let \( \Gamma \) be a first-order reduct of \( (\mathbb{N}; 0) \). Rather than working with primitive positive sentences in the language of \( \Gamma \), our algorithms for CSP(\( \Gamma \)) will take as input a set of clauses that are obtained by replacing every relation symbol from \( \Gamma \) by a corresponding definition of the relation in \((\mathbb{N}; 0)\).

Since \((\mathbb{N}; 0)\) has the elimination of quantifiers, we obtain this way a set of quantifier-free clauses. When \( \Gamma \) is preserved by some function in \((\text{inj, inj, inj}), (\text{inj, inj, 0}), \) or \((\text{inj, 0, 0})\), we see in this section that we can assume that the input of our algorithms have a specific syntactic form.

When working with quantifier-free formulas, we will always want to get rid of unnecessary clauses or literals in order to only work with formulas that contain precisely the information that is relevant. This idea is formalised in the following definition:

**Definition 3.2** (Reduced formula). A quantifier-free first-order formula \( \Phi \) in CNF (conjunctive normal form) is reduced over \( \Gamma \) if every formula obtained from \( \Phi \) by removing a literal or a clause is not equivalent to \( \Phi \) over \( \Gamma \).

**Remark 4.** Let \( \Gamma \) be a structure. Every quantifier-free formula is logically equivalent to a reduced formula over \( \Gamma \).

**Proposition 3.5.** Let \( \Gamma \) such that \( \text{Pol}(\Gamma) \cap \{\text{inj, inj, inj}\} \neq \emptyset \). Then every relation \( R \) of \( \Gamma \) has a Horn definition in \((\mathbb{N}; 0)\), i.e., there exist \((\phi_{i,1}(\bar{x}), \ldots, \phi_{i,k_i}(\bar{\bar{x}}))_{1 \leq i \leq k}\) such that for all \( \bar{v}, \Gamma \models R(v_1, \ldots, v_n) \) if

\begin{align*}
\gamma_i(\bar{v}) &:= \phi_{i,1}(\bar{x}) \land \ldots \land \phi_{i,k_i}(\bar{x}) \\
\delta_i(\bar{v}) &:= \neg \gamma_i(\bar{v}) \land \phi_{i,1}(\bar{x}) \land \ldots \land \phi_{i,k_i}(\bar{x}) \\
\end{align*}
and only if \((\mathbb{N}, 0) \models \bigwedge_{i \leq k} \bigvee_{j \leq k} \phi_{i,j}(v)\), with \(\phi_{i,j}\) atomic formulas or negations of atomic formulas of \((\mathbb{N}; 0)\), such that for all \(i \leq k\), at most one of the \(\phi_{i,j}\) is positive.

**Proof.** Let \(f \in \text{Pol}(\Gamma) \cap (\text{inj}, \text{inj}, \text{inj})\). Let \(\Phi\) be a reduced quantifier-free definition of \(R\) in \((\mathbb{N}; 0)\). Such a formula exists by Remark 4. We assume for contradiction that \(\Phi\) is not Horn, that is, has a clause \(C : \bigvee_{i \leq n} \phi_i(x)\) where the \(\phi_i\) are atomic formulas or negations of atomic formulas, with \(\phi_1\) and \(\phi_2\) both positive. Let \(x\) be the variables of \(C\). Since \(\Phi\) is reduced, there exists a truth assignment \(\sigma_1 : x \rightarrow \mathbb{N}\) such that \(\phi_i(\sigma_1(x))\) is false for all \(i \neq 1\) but true for \(i = 1\) (otherwise, we could remove \(\phi_1\) from \(C\) to obtain an equivalent formula, contradicting the assumption that \(\Phi\) is reduced). Similarly, there exists a truth assignment \(\sigma_2 : x \rightarrow \mathbb{N}\) such that \(\phi_i(\sigma_2(x))\) is false for all \(i \neq 2\) but true for \(i = 2\). One of the following applies:

- either \(\phi_1 \models x = y\), and \(\phi_2 \models z = t\),
- either \(\phi_1 \models x = y\), and \(\phi_2 \models t = 0\),
- or \(\phi_1 \models x = 0\), and \(\phi_2 \models y = 0\).

In all these three cases, it is straightforward to check that \(f(\sigma_1(x), \sigma_2(x))\) does not satisfy \(C\), contradicting the fact that \(f\) preserves \(R\). Consequently, \(R\) has a Horn definition on \((\mathbb{N}, 0)\). \(\square\)

**Proposition 3.6.** Let \(\Gamma\) be such that \(\text{Pol}(\Gamma) \cap (\text{inj}, 0, 0) \neq \emptyset\). Then every relation \(R\) of \(\Gamma\) is definable by a quantifier-free formula in conjunctive normal form and such that in each clause, at most one literal is \(x \neq 0\) or \(x = y\), and all the other literals are \(x \neq y\) or \(x = 0\).

**Proof.** The proof is similar to the proof of Proposition 3.5. \(\square\)

A clause which has the form described in Proposition 3.6 is called quasi-Horn.

### 3.2.2 Algorithms

**Proposition 3.7.** Let \(\Gamma\) be a reduct of \((\mathbb{N}; 0)\) such that \(\text{Pol}(\Gamma) \cap (\text{inj}, \text{inj}, \text{inj}) \neq \emptyset\). Then Algorithm 7 solves \(\text{CSP}(\Gamma)\) in polynomial time.

**Proof.** By Proposition 3.5 every relation \(R\) of \(\Gamma\) has a Horn definition on the language \((\mathbb{N}; 0, =)\), i.e., there exists \((\phi_{i,1}(x), \ldots, \phi_{i,k}(x))\) such that for all \(v, \Gamma \models R(v_1, \ldots, v_n)\) if and only if \((\mathbb{N}, 0) \models \bigwedge_{i \leq k} \bigvee_{j \leq k} \phi_{i,j}(v)\), such that for all \(i \leq k\), at most one of the \(\phi_{i,j}\) is positive. Let \(\exists v_1, \ldots, v_r \bigwedge R_i(v)\) be an instance of \(\text{CSP}(\Gamma)\). We can replace in linear time all the symbols \(R_i\) by one of their Horn definition in \((\mathbb{N}; 0)\). The set of clauses \(\Phi\) obtained by such a substitution is such that \((\mathbb{N}; 0) \models \exists v_1, \ldots, v_r \bigwedge \Phi\) iff \(\Gamma \models \exists v_1, \ldots, v_r \bigwedge R_i(v)\). Consider a run of Algorithm 7 on \(\Phi\).

The main observation we make about this algorithm is that the sets of clauses \(\Phi = \Phi_0, \ldots, \Phi_m\) before each iteration of the main loop are such that \(\bigwedge \Phi_i \Leftrightarrow \bigwedge \Phi\). This simply comes from the fact that if \(\Psi_i \models \ell\), i.e., if every satisfying assignment to the variables of \(\Psi_i\) also satisfies \(\ell\), then every satisfying assignment of \(\Phi_i\) also satisfies \(\ell\), and therefore the literal \(\neg \ell\), which appears somewhere in a clause of \(\Phi_i\), cannot be satisfied. It is therefore sound to remove it entirely from \(\Phi_i\).

This observation is enough to prove that if \(\Phi\) is rejected by the algorithm, then \((\mathbb{N}; 0) \not\models \exists v_1, \ldots, v_r \bigwedge \Phi\), and therefore \(\Gamma \not\models \exists v_1, \ldots, v_r \bigwedge R_i\).

Suppose now that the algorithm accepts \(\Phi\). Let \(\Psi\) be the set of clauses of \(\Phi\) that consist of only one positive literal. Since \(\Psi\) does not contain a literal and its negation, we see that \(\bigwedge \Psi\) has
a satisfying assignment. Enumerate all the negative literals of Φ by \( \ell_1, \ldots, \ell_q \). By construction, for each such literal there exists a satisfying assignment \( s_i \) of Ψ that also satisfies \( \ell_i \). Let \( f \) be a polymorphism of Γ that is of type \( \mathrm{inj}, \mathrm{inj}, \mathrm{inj} \). Then we see that the assignment of the variables of Φ that maps \( v_i \) to \( f(s_1(v_i), f(s_2(v_i), \ldots, f(s_{q-1}(v_i), s_q(v_i)) \ldots) \) satisfies Φ, by injectivity of \( f \).

It remains to prove that Algorithm 1 runs in polynomial time. But this follows easily by noting that to test whether \( \Psi \models \neg \ell \), where \( \ell \) is a negative literal, can be done in a trivial way. Indeed, if we consider the graph whose vertices are the variables \( 0, v_1, \ldots, v_r \) and whose edges are \( \{v_i, v_j\} \) where \( v_i = v_j \) is in Ψ, then \( \Psi \models x = y \) iff there is a path from \( x \) to \( y \) in the graph, and \( \Psi \models x = 0 \) iff there is a path from \( x \) to \( 0 \). This test can be carried out in polynomial time, which concludes the proof.

Data: a set \( \Phi \) of Horn clauses with variables \( v \)
Result: accepts if \( (\mathbb{N}; 0) \models \exists v \bigwedge \Phi \), rejects otherwise
Ψ := ϕ;
repeat
| Ψ := \{C ∈ Φ | C is a clause without negative literals\};
| if Ψ contains a literal and its negation then
| reject;
| forall the negative literals \( \ell \) in a clause of \( \Phi \) do
| if \( \Psi \models \neg \ell \) then
| Remove the literal \( \ell \) from every clause in which it appears;
| end
| end
| if one of the clauses of \( \Phi \) is empty then
| reject;
| end
until Φ doesn’t change;
accept;

Algorithm 1: a polynomial-time algorithm for sets of Horn clauses.

Proposition 3.8. Let \( \Gamma \) be a reduct of \((\mathbb{N}; 0)\) such that \( \mathrm{Pol}(\Gamma) \cap (\mathrm{inj}, 0, 0) \neq \emptyset \). Then Algorithm 2 solves \( \mathrm{CSP}(\Gamma) \) in polynomial time.

Proof. By Proposition 3.6, for every relation \( R \) of \( \Gamma \), there exists a conjunctive normal form of \( R \) such that every clause is quasi-Horn.

First note that the quasi-Horn clauses which contain at most one literal \( x = 0 \) are preserved under \( (\mathrm{inj}, \mathrm{inj}, \mathrm{inj}) \), and it is therefore possible to test the satisfiability of these clauses in polynomial time using Algorithm 1.

Let \( \exists v \wedge R_i(v) \) be an instance of \( \mathrm{CSP}(\Gamma) \), and let \( \Phi \) be initialized to the corresponding set of quasi-Horn clauses. Consider a run of Algorithm 2 on \( \Phi \).

As it was the case in Algorithm 1, the sequence of sets \( \Phi_0, \Phi_1, \ldots \) we obtain at the beginning of each loop is such that \( \wedge \Phi_i \iff \wedge \Phi \). This is enough to conclude that if the algorithm rejects \( \Phi \), then \((\mathbb{N}; 0) \not\models \exists v \wedge \Phi \).
Suppose now that Algorithm 2 accepts \( \phi \). As in the proof of Proposition 3.7, we know that for each literal \( x = 0 \) that remains at the end of the algorithm there exists an assignment that satisfies \( \Psi \land x = 0 \). Using a function from \((\text{inj},0,0)\), we can then build an assignment that satisfies \( \Phi \), thus \( (N;0) |\exists v \land \Phi \). \( \square \)

**Data:** a set \( \Phi \) of quasi-Horn clauses with variables \( v \)

**Result:** accepts if \( (N;0) |\exists v \land \Phi \), rejects otherwise

\[
\Psi := \{ C \in \Phi | C \text{ has no literal of the form } x = 0 \}; \\
\text{if } (N;0) \not|\exists v \land \Psi \text{ then reject;} \\
\text{end}
\]

**end**

**until \( \Phi \) doesn’t change;**

**accept;**

**Algorithm 2:** An algorithm for sets of quasi-Horn clauses.

**Definition 3.3.** We denote by q-minority the set of all ternary operations \( h \) such that for all \( x, y, z, x', y', z' \) in \( N \), we have the following:

- \( h \) is injective on \((N \setminus \{0\})^3\)
- \( h(x, y, z) = 0 \) if and only if \( x = y = z = 0 \) or exactly one of \( x, y, z \) equals 0
- \( h(x, y, z) = h(x', y', z') \) if and only if: \( (x = x' \land y = y' \land z = z') \) or \( h(x, y, z) = h(x', y', z') = 0 \)

**Definition 3.4.** We denote by q-majority the set of all ternary operations \( h \) such that for all \( x, y, z, x', y', z' \) in \( N \), we have the following:

- \( h \) is injective on \((N \setminus \{0\})^3\)
- \( h(x, y, z) = 0 \) if and only if at least two of \( x, y, z \) are set to 0
- \( h(x, y, z) = h(x', y', z') \) if and only if: \( (x = x' \land y = y' \land z = z') \) or \( h(x, y, z) = h(x', y', z') = 0 \)

**Definition 3.5.** Let \( R \) be a \( n \)-ary relation over \( N \). We define the equality-horn characterization of \( R(\bar{x}) \) the set \( \tau(R(\bar{x})) \) of formulas over the language of \((N,=)\) such that:

- for every \( \phi(\bar{x}) \in \tau(R(\bar{x})) \), the free variables of \( \phi \) belong to \( \bar{x} \)
- for every \( \phi(\bar{x}) \in \tau(R(\bar{x})) \), \( \phi \) is Horn, i.e., of the form \( (x_1 = x'_1 \land \cdots \land x_k = x'_k) \Rightarrow x_{k+1} = x'_{k+1} \)
• for every $\phi(\tau) \in \tau(R(\tau))$, we have: $(\mathbb{N},=) \models \forall \tau.(R(\tau) \Rightarrow \phi(\tau))$

• for every Horn formula $\phi(\tau)$ over $(\mathbb{N},=)$, if $(\mathbb{N},=) \models \forall \tau.(R(\tau) \Rightarrow \phi(\tau))$, then $\phi(\tau) \in \tau(R)$.

**Notation 1.** Let $\zeta: \mathbb{N} \rightarrow \{0,1\}$ be such that: $\zeta(x) = 0$ if $x = 0$, and $\zeta(x) = 1$ if $x \neq 0$. For any relation $R \subseteq \mathbb{N}^n$, we denote by $\rho(R)$ the relation over $\{0,1\}$ defined as follows:

$$\rho(R) := \{(\zeta(x_1),\ldots,\zeta(x_n)) \mid (x_1,\ldots,x_n) \in R\}$$

**Proposition 3.9.** Let $\Gamma$ be a reduct of $(\mathbb{N},0)$ such that $\text{Pol}(\Gamma) \cap (q\text{-minority} \cup q\text{-majority}) \neq \emptyset$. Then the following algorithm solves $\text{CSP}(\Gamma)$ in polynomial time.

**Algorithm 3:** An algorithm for sets of clauses preserved under an operation in $q\text{-minority}$ or $q\text{-majority}$. 

**Data:** a pp-sentence $\Phi = \exists x.\varphi$ in the language of $\Gamma$

**Result:** accepts if $\Gamma \models \Phi$, rejects otherwise 

$\Psi_H := \{\tau(R_i(\tau)) \mid R_i(\tau) \text{ appears in } \Phi\}$

repeat

| forall the pairs of variables $(x,y)$ of $\Phi$ do |
| if Solve CSP$((\mathbb{N},=)(\exists x. \bigwedge \Psi_H \land x \neq y))$ rejects then |
| Replace every occurrence of $y$ by $x$ in $\Psi_H$ |
| Replace every occurrence of $x$ by $y$ in $\Phi$ |
| end |
| until $\Psi_H$ doesn’t change; |
| $\Psi_{0,1} := \{S_i(\tau) \mid S_i = \rho(R_i) \text{ for } R_i(\tau) \text{ appearing in } \Phi\}$ |
| if Solve CSP$((\{0,1\};0,=)(\exists x'. \bigwedge \Psi_{0,1}'))$ rejects then |
| reject; |
| accept; |
| Algorithm 3 accepts $\Phi$ |

**Proof.** First note that $\Phi$ is equi-satisfiable at each step of the algorithm. Hence, we suppose now that $\Phi$ is the set of clauses obtained after the first step of the algorithm.

Suppose that $\Phi$ is satisfiable and let $(a_1,\ldots,a_n)$ be such that $\Gamma \models \Phi(a_1,\ldots,a_n)$. Then $(\zeta(a_1),\ldots,\zeta(a_n))$ is a solution of $\Psi_{0,1}$. Indeed, let $R(\tau)$ be any clause of $\Phi$. Then $\Gamma \models R(\tau)$. Hence, the tuple $(\zeta(a_1),\ldots,\zeta(a_n))$ is in $\tau(R(\tau))$. So Algorithm 3 accepts $\Phi$.

Conversely, suppose that Algorithm 3 accepts $\Phi$. We show that there exists an assignment of the variables which satisfies $\Phi$. Since Algorithm 3 accepts $\Phi$, Solve CSP$((\{0,1\};0)(\exists x'. \bigwedge \Psi_{0,1}'))$ accepts. So there exists $(b_1,\ldots,b_n) \in \{0,1\}^n$ such that $(\{0,1\};0) \models \Psi_{0,1}'(\overline{b})$. Let $I := \{i \mid b_i \neq 0\}$ and let $g$ be any injection from $I$ to $\mathbb{N} \setminus \{0\}$. We now define $\pi \in \mathbb{N}^n$ as follows: if $i \in I$, $a_i = g(i)$, and if $i \notin I$, $a_i = 0$. We now show that $\pi$ satisfies $\Phi$. Let $R(x_1,\ldots,x_k)$ be a clause of $\Phi$. We know that $(\{0,1\},0,=) \models \rho(R(b_1,\ldots,b_k))$, so there exists $(c_1,\ldots,c_k) \in R$ such that $(b_1,\ldots,b_k) = (\zeta(c_1),\ldots,\zeta(c_k))$. For notational simplicity, we suppose that $c_{j+1} = \cdots = c_k = 0$ and $c_1,\ldots,c_j$ are all distinct from 0. Assume that $c_1 = c_2$. Thanks to the first process of the algorithm, we know that there exists a tuple $(c_1',c_2',\ldots,c_k') \in R$ such that $c_1' \neq c_2'$. We now make a case distinction:
Lemma 3.11. Let \( \Gamma \) be a reduct of \((\mathbb{N}; 0)\) such that \( \text{Neq}_1 \) has no primitive positive definition in \( \Gamma \). Then the model-complete core of \( \Gamma \) is a finite structure with at most two elements.

Proof. Since \( \text{Neq}_1 \) is not in pp-definable in \( \Gamma \), and since \( \Gamma \) is \( \omega \)-categorical, it follows from Theorem 2.2 that there exists a polymorphism of \( \Gamma \) that does not preserve \( \text{Neq}_1 \). Moreover, \( \text{Neq}_1 \) consists of only one orbit of pairs under \( \text{Aut}(\mathbb{N}; 0) \), so that it must be violated by a unary polymorphism, i.e., an endomorphism of \( \Gamma \). Let \( g \) be such an endomorphism. That \( \text{Neq}_1 \) is violated by \( g \) means that there exists two distinct elements \( x_0, x_1 \) different than 0 such that \( g(x_0) = g(x_1) \).

Let \( \Delta \) be the model-complete core of \( \Gamma \), and let \( h \) be a homomorphism from \( \Gamma \) to \( \Delta \). Suppose that \( \Delta \) has at least three distinct elements, \( x, y, \) and \( z \). At least two of these are different than 0, and suppose without loss of generality that \( x \) and \( y \) are different than 0. There is an automorphism \( \alpha \) of \( \Gamma \) that maps \( x \) to \( x_0 \) and \( y \) to \( x_1 \). Then the function \( g \circ \alpha|_{\text{Dom}(\Delta)} \) is an endomorphism of \( \Delta \) that is not injective, contradicting the fact that \( \Delta \) is a model-complete core.

Lemma 3.10. Let \( \Gamma \) be a reduct of \((\mathbb{N}; 0)\) such that \( \text{Neq}_2 \) is preserved by \( \text{Neq}_2 \) in \( \Gamma \). Then there exists a binary polymorphism of \( \Gamma \) which is injective on \((\mathbb{N} \setminus \{0\})^2\), or there exists a continuous clone homomorphism from \( \text{Pol}(\Gamma) \) to \( 1 \).

Proof. Suppose that \( \Gamma \) has no binary polymorphism which is injective on \((\mathbb{N} \setminus \{0\})^2\). Since \( \text{Neq}_2 \) is preserved by all the polymorphisms of \( \Gamma \), \( \Gamma \) does not have an endomorphism that is constant on \( \mathbb{N} \setminus \{0\} \) either. Let \( C \) be a polymorphism algebra of \( \Gamma \). Since \( \mathbb{N} \setminus \{0\} \) is preserved by the polymorphisms of \( \Gamma \), \( C \) induces on \( \mathbb{N} \setminus \{0\} \) a subalgebra \( A \). The clone \( \text{Clo}(A) \) is included in the

3.3 The Classification

In this final section, we give a proof of Theorem 3.1.

Definition 3.6. The following three relations play an important role in our classification:

- Let \( \text{Neq}_1 \) be the following unary relation: \( x \in \text{Neq}_1 \) if and only if \( x \neq 0 \).
- Let \( \text{Neq}_2 \) be the following binary relation: \( (x, y) \in \text{Neq}_2 \) if and only if \( 0 \neq x \neq y \neq 0 \).
- Let \( N \) be the following binary relation: \( (x, y) \in N \) if and only if \( x = 0 \Leftrightarrow y \neq 0 \).

Note that both \( \text{Neq}_1 \) and \( \text{Neq}_2 \) consist in only one orbit, and \( N \) consists in at most two orbits under the action of \( \text{Aut}(\Gamma) \) on \( \mathbb{N}^2 \).

Lemma 3.11. Let \( \Gamma \) be a reduct of \((\mathbb{N}; 0)\) such that \( \text{Neq}_1, \text{Neq}_2 \in \langle \Gamma \rangle_{\text{pp}} \). Then there exists a binary polymorphism of \( \Gamma \) which is injective on \((\mathbb{N} \setminus \{0\})^2\), or there exists a continuous clone homomorphism from \( \text{Pol}(\Gamma) \) to \( 1 \).

Proof. Suppose that \( \Gamma \) has no binary polymorphism which is injective on \((\mathbb{N} \setminus \{0\})^2\). Since \( \text{Neq}_2 \) is preserved by all the polymorphisms of \( \Gamma \), \( \Gamma \) does not have an endomorphism that is constant on \( \mathbb{N} \setminus \{0\} \) either. Let \( C \) be a polymorphism algebra of \( \Gamma \). Since \( \mathbb{N} \setminus \{0\} \) is preserved by the polymorphisms of \( \Gamma \), \( C \) induces on \( \mathbb{N} \setminus \{0\} \) a subalgebra \( A \). The clone \( \text{Clo}(A) \) is included in the
polymorphism clone of a structure $\Delta$, which is a reduct of $(\mathbb{N};=)$ since $\text{Clo}(A)$ contains all the permutations. It follows from Theorem 3.2 that there exists a continuous clone homomorphism from $\text{Pol}(\Delta)$ to $1$. Moreover, by Theorem 3.7, since $A$ is in $\text{HSP}_{\text{in}}(C)$ there exists a continuous clone homomorphism from $\text{Pol}(\Gamma)$ to $\text{Clo}(A)$. By composing the two homomorphisms, we obtain a continuous clone homomorphism from $\text{Pol}(\Gamma)$ to $1$. \hfill $\Box$

**Proposition 3.12.** Let $\Gamma$ be a first-order reduct of $(\mathbb{N};0)$ such that $0, \text{Neq}_1, \neq$, and $E$ are pp-definable over $\Gamma$. If there is no continuous clone homomorphism from $\text{Pol}(\Gamma)$ to $1$, $\text{CSP}(\Gamma)$ is in $P$.

**Proof.** Since $E$ is pp-definable in $\Gamma$, $E$ is preserved by every polymorphism of $\Gamma$. Hence, we can associate with every $f \in \text{Pol}(\Gamma)$ the mapping $\overline{f}$ such that $\overline{f}(cl(x_1), \ldots, cl(x_n)) = cl(f(x_1, \ldots, x_n))$ for all $x_1, \ldots, x_n \in \mathbb{N}$. Note that we can consider the mapping $\overline{f}$ as an operation of $\{0,1\}$, $0$ being the class of $0$, and $1$ being the class of all the nonzero elements. Let $C$ be the set $\{\overline{f} \mid f \in \text{Pol}(\Gamma)\}$. One can easily see that since $\text{Pol}(\Gamma)$ is a clone over $\mathbb{N}$, $C$ is a clone over $\{0,1\}$.

If there were a continuous clone homomorphism from $C$ to $1$, we would obtain a continuous clone homomorphism from $\text{Pol}(\Gamma)$ to $1$, contradicting our hypothesis on $\Gamma$. Therefore, by Theorem 1.3 there exists a ternary polymorphism $f$ of $\Gamma$ such that $\overline{f}$ is cyclic. Moreover, since there is no continuous clone homomorphism from $\text{Pol}(\Gamma)$ to $1$, we have by Lemma 3.11 that $\Gamma$ has a binary polymorphism which is injective on $(\mathbb{N}\setminus\{0\})^2$. Through Proposition 3.4 we may even assume that this polymorphism $g$ is in $\langle \text{inj}, \text{inj}, \text{inj} \rangle, \langle \text{inj}, 0, 0 \rangle$, or in $\langle \text{inj}, \text{inj}, 0 \rangle$. In the first two cases, we can already conclude that $\text{CSP}(\Gamma)$ is in $P$, by Propositions 3.8 and 3.7. So assume in the following that $g$ is in $\langle \text{inj}, \text{inj}, 0 \rangle$.

Let $h$ be the ternary function defined by $h(x,y,z) = g(g(x,y),g(x,z))$. We can easily check that $h$ is injective on $(\mathbb{N}\setminus\{0\}) \times \mathbb{N}^2$, and is identically $0$ on $0 \times \mathbb{N}^2$. Define now $f'(x,y,z) = f(h(x,y,z),h(y,z,x),h(z,x,y))$. Again, we can check that $f'$ is injective on $(\mathbb{N}\setminus\{0\})^3$, and that $\overline{f'}$ is still cyclic. Note that since $f'$ preserves $E$, whether $f'(x,y,z)$ is $0$ or nonzero depends solely on the number of zeros in $(x,y,z)$. Finally, for $X,Y,Z \in \{\{0\},(\mathbb{N}\setminus\{0\})\}$, $f'$ is injective on $X \times Y \times Z$ iff $f$ is nonzero on $X \times Y \times Z$, and is identically $0$ otherwise.

We now have enough pieces of information to conclude that $\text{CSP}(\Gamma)$ is tractable. There are only $4$ types of functions that satisfy the properties of $f'$: on triples with two zero entries (tuples of type $A$), $f'$ is either zero or injective, and triples with one zero entry (tuples of type $B$), $f'$ is either zero or injective. For triples with $0$ and $3$ zero entries, $f'$ is known to be respectively $0$ and injective. If $f'$ is zero on the tuples of type $A$ and $B$, $f'(x,y,z)$ is in $\langle \text{inj}, 0, 0 \rangle$, and we are done by Proposition 3.8. If $f'$ is injective on the triples of type $A$ and $B$, then $f'(x,y,z)$ is in $\langle \text{inj}, \text{inj}, \text{inj} \rangle$, and we are done by Proposition 3.7. If $f'$ is injective on triples of type $A$ and zero on triples of type $B$, $f'$ is in q-minority. If $f'$ is injective on triples of type $B$ and zero on triples of type $A$, $f'$ is in q-majority. In the last two cases, we conclude with Proposition 3.9. \hfill $\Box$

**Proposition 3.13.** Let $\Gamma$ be a first-order reduct of $(\mathbb{N};0)$ such that $0, \text{Neq}_1, \text{Neq}_2$ are pp-definable over $\Gamma$ but $N$ is not. If there is no continuous clone homomorphism from $\text{Pol}(\Gamma)$ to $1$, then $\text{CSP}(\Gamma)$ is in $P$.

**Proof.** Since there is no continuous clone homomorphism from $\text{Pol}(\Gamma)$ to $1$, we obtain by Lemma 3.11 a polymorphism $g$ of $\Gamma$ that is injective on $(\mathbb{N}\setminus\{0\})^2$. Since $N$ is not pp-definable in $\Gamma$, and $\Gamma$ is $\omega$-categorical, there exists a polymorphism $f$ of $\Gamma$ that violates $N$. Furthermore, since $N$ consists of at most two orbits of pairs under $\text{Aut}(\Gamma)$, we may assume that $f$ is binary by Theorem 2.2. The fact...
that $f$ violates $N$ means that there exist $x_0, y_0$, both not 0, such that $f(x_0, 0)$ and $f(0, y_0)$ are in the same orbit. By letting $\alpha$ be an automorphism of $\Gamma$ that maps $x_0$ to 1, and $\beta$ be an automorphism that maps $y_0$ to 1, we may replace $f$ by $(x, y) \mapsto f(\alpha x, \alpha y)$ and assume that $x_0 = y_0 = 1$.

Assume first that $f(x_0, 0) = f(0, y_0) = 0$. We prove by induction that for every $n \in \mathbb{N}$, there exists a binary polymorphism $f_n$ of $\Gamma$ such that $f_n(x_0, 0) = f_n(0, x) = 0$ for every $x \in [n]$. There is nothing to prove in the case $n = 1$, by letting $f_1$ be $f$. Suppose now that $f_n$ exists. If $f_n(0, n+1) \neq 0$, let $f$ be an automorphism of $\Gamma$ that maps $f_n(0, n+1)$ to $n$. Define $h(x, y) = f_n(x, \alpha f_n(x, y))$. For $y \in [n]$, we have $h(0, y) = f_n(0, \alpha f_n(0, y)) = f_n(0, 0) = 0$. For $y = n + 1$, we have $h(0, n + 1) = f_n(0, \alpha f_n(0, n + 1)) = f_n(0, n) = 0$. Finally, for $x \in [n]$, we have $h(x, 0) = f_n(x, \alpha f_n(x, 0)) = f_n(x, 0) = 0$. If $h(n + 1, 0) = 0$, we are done by letting $f_{n+1}$ be $h$. Otherwise, let $\beta$ be an automorphism of $\Gamma$ that maps $h(n + 1, 0)$ to $n$. Define $f_{n+1}(x, y) = h(\beta h(x, y), y)$. We can check that $f_{n+1}$ has the desired properties, and this finishes the proof of the claim.

A standard compactness argument now shows that there exists a polymorphism of $\Gamma$ that is 0 on $\mathbb{N} \times 0$ and $0 \times \mathbb{N}$. For $n \in \mathbb{N}$, define an equivalence relation $\sim_n$ on the partial functions $[n]^2 \to \mathbb{N}$ by setting $f \sim_n g$ iff there exists an automorphism $\alpha$ of $\Gamma$ such that $\alpha \circ f(x, y) = g(x, y)$ for all $(x, y) \in [n]^2$. The $\omega$-categoricity of $\Gamma$ entails that $\sim_n$ has finite index for all $n \in \mathbb{N}$. Let $T$ be the tree whose vertices on level $n$ are the $\sim_n$-equivalence classes of functions that are 0 on $[n] \times 0$ and $0 \times [n]$, and where there is an edge between $[f_n]$ and $[f_{n+1}]$ if there are $g_n \sim_n f_n, g_{n+1} \sim_{n+1} f_{n+1}$ such that $g_n = g_{n+1}|[n]^2$. This tree has vertices on each level by the previous claim, and has finitely many vertices on each level since $\sim_n$ has finite index. By König’s tree lemma, there exists an infinite branch $([k_n])_n$ in $T$. We build the desired polymorphism $h$ of $\Gamma$ by induction on $n$. Define $h|_{[0]^2} = h_0$. Suppose now that $h$ is defined on $[n]^2$ and that $h \sim_n h_n$. By definition of the edge relation, there exists $\alpha$ such that $\alpha \circ h_{n+1} = h_n$. We extend $h$ to $[n+1]^2$ simply by setting $h(x, y) = \alpha h_{n+1}(x, y)$ for $x, y \in [n+1]$. To finish the proof, note that $h(g(x, y), g(y, x))$ is in $\text{CSP}(\Gamma)$, and $h$ is a polymorphism of $\Gamma$ by Proposition 3.8.

When $f(1, 0) \neq 0$ and $f(0, 1) \neq 0$, it suffices to show that there also exists for all $n$ a function $f_n$ which is nonzero on $[n] \times 0$ and $0 \times [n]$, and then the previous compactness argument works the same. We can assume that $f$ is nonzero on $0 \times \mathbb{N}$ and that $f$ has some 0 on $([\mathbb{N}] \setminus \{0\}) \times 0$, otherwise we conclude with the previous case. Then we can simply define $f_{n+1}(x, y)$ as $f_n(\alpha_{n+1} x, f_n(x, y))$, where $\alpha_{n+1}$ is an automorphism of $\Gamma$ that maps $n + 1$ to 1. The verifications are left to the reader.

Proof of Theorem 3.4. Let $\Delta$ be the model-complete core of $\Gamma$. If $\Gamma$ has an endomorphism $g$ and two points $x_0, x_1$ different than 0 such that $g(x_0) = g(x_1)$, then $\Delta$ is a structure with at most two elements by Lemma 3.10. In this case, the complexity of CSP($\Delta$) is described by Schaefer’s Theorem, and we are done.

Otherwise, all the endomorphisms of $\Gamma$ are injective on $\mathbb{N} \setminus \{0\}$. In particular, $\Gamma$ has no finite-range endomorphism and $\Delta$ is an infinite substructure of $\Gamma$. If $\Delta$ does not contain 0, it is isomorphic to a reduct of $(\mathbb{N}; =)$, and we are done by Theorem 3.2. Otherwise $\Delta$ contains 0 and is isomorphic to a reduct of $(\mathbb{N}; 0)$. Since $\Delta$ is a model-complete core and the relations $0, \text{Neq}_1$, and $\text{Neq}_2$ are orbits, they are pp-definable in $\Delta$. Depending on whether $E$ is pp-definable in $\Gamma$, Propositions 3.12 and 3.13 allow us to finish the proof.

3.4 Further work

It is interesting to note that if we allow an arbitrarily large number of constants $c_1, \ldots, c_n$, the CSPs whose templates are first-order reducts of $(\mathbb{N}; c_1, \ldots, c_n)$ are a proper superclass of the finite-
domain CSPs. Here, the reducts we studied were powerful enough to interpret all the boolean
constraint satisfaction problems, for which we already had a dichotomy result. Moreover, our result
is a natural extension to Schaefer’s dichotomy theorem, since here again our result can be rephrased
as follows:

**Theorem 3.14.** Let $\Gamma$ be a reduct of $(\mathbb{N}; 0)$ with a finite signature. Then if $\Gamma$ has a ternary
polymorphism that is cyclic modulo automorphisms, $\text{CSP}(\Gamma)$ is in $\text{P}$. Otherwise $\text{CSP}(\Gamma)$ is $\text{NP}$-
complete.

Most of the proofs given here seem to generalize well to the case where we throw in more
constants. We therefore pose the following question:

**Question 3.** Assume that the Tractability Conjecture is true, and let $\Gamma$ be a reduct of $(\mathbb{N}; 0, \ldots, n)$
with a finite signature. Is it true that $\text{CSP}(\Gamma)$ is in $\text{P}$ if the model-complete core of $\Gamma$ has a cyclic
polymorphism modulo automorphisms, and is $\text{NP}$-complete otherwise?

A possible proof strategy for this question would be as follows:

1. Let $\Delta$ be the model-complete core of $\Gamma$. If $\Delta$ is already a reduct of $(\mathbb{N}; 0, \ldots, n - 1)$, conclude
   by induction. If $\Delta$ is a finite structure, conclude using the tractability conjecture.
2. Otherwise $\Delta$ is a reduct of $(\mathbb{N}; 0, \ldots, n)$ where all the constants are pp-definable.
3. Use Theorem 3.2 to obtain a polymorphism which is injective on $\mathbb{N} \setminus \{0, \ldots, n\}$.
4. Combine the polymorphism of $\Delta$ that is cyclic modulo automorphisms and the injective
   polymorphism to obtain a polymorphism $\omega$ that is cyclic modulo automorphisms and canonical
   with respect to $(\mathbb{N}; 0, \ldots, n)$.
5. Reduce $\text{CSP}(\Delta)$ to the CSP of a finite structure, in which $\omega$ “becomes” a cyclic polymorphism.
   Conclude with the tractability conjecture.

In this strategy, the only point that remains unclear is how to adapt our current proofs to fulfill
the fourth step, though we believe that achieving this goal is within the reach of our current proof
techniques.
Chapter 4

Algebraic Criteria for the Tractability of Problems in MMSNP

4.1 Introduction

Temporal and spatial constraint formalisms that have been called qualitative in the literature typically turn out to have an $\omega$-categorical constraint language (Hir96, BC09, BW11). The class of $\omega$-categorical constraint languages includes constraint languages over finite domains, but does not include languages over the integers, rationals, or reals that can express addition. A relatively restricted subclass of $\omega$-categorical constraint languages that still contains all finite-domain constraint languages is the class of constraint languages whose CSP is expressible in MMSNP. We believe that this class is also of interest to temporal and spatial reasoning, since the problems in MMSNP can be viewed as the simplest (least expressive) qualitative reasoning problems. And indeed, by fascinating results of Feder and Vardi (FV99, and Kun Kun13), this class probably exhibits a complexity dichotomy into P and NP-complete (unless the dichotomy conjecture for finite-domain CSPs fails, too).

The result of Feder, Vardi, and Kun is deep and relies on a sophisticated deterministic construction of expander hypergraphs. This chapter is an initial study in the long-term goal to obtain an alternative universal-algebraic proof of the theorem of Feder-Vardi and Kun. The methods for tackling this project are most probably also useful for complexity analysis of much wider classes of qualitative reasoning problems in temporal and spatial reasoning.

The result of this chapter is a new and purely universal-algebraic proof of a complete complexity classification for a subclass of MMSNP due to Bodirsky-Chen-Feder (BCF12) (which does not rely on the finite-domain dichotomy conjecture), namely for MMSNP problems with monochromatic obstruction sets (for definitions, see the following subsections).
4.1.1 MMSNP

In descriptive complexity, one measures the complexity of a problem in terms of the expressive power that a logic needs to be able to describe this problem, in contrast to measuring, e.g., the time that a deterministic Turing machine needs to take to decide the problem. A class $C$ of relational structures is said to be described by a formula $\phi$ from some logic $L$ if

$$A \in C \iff A \models \phi$$

holds, for some relation $\models$ that depends on $L$. We write $\text{Models}(\phi)$ for the class of finite structures $A$ such that $A \models \phi$. Keeping in mind that we consider the logic itself as a resource, just as time or space, a logic gives rise to a complexity class, namely the class of all the problems that can be described by a formula in $L$. For brevity, this complexity class is just denoted by $L$. In [Fag74], Fagin proved that existential second-order logic ESO, that consists of the formulas of the form

$$\exists R_1, \ldots, \exists R_k \phi$$

where $\phi$ is a first-order formula, equals NP.

We will study here a fragment of ESO obtained by imposing several restrictions on the formulas. The logic SNP (strict NP) was introduced by Kolaitis and Vardi in [KV87]. It consists of existential second-order formulas in which the first-order part is itself universal. That is, SNP consists of formulas of the form

$$\exists R_1, \ldots, \exists R_k \forall x \phi_0$$

where $\phi_0$ is a quantifier-free formula. The logic SNP is rich, in the sense that every NP problem is polynomial-time equivalent to a problem of the form $\text{Models}(\phi)$ with $\phi$ in SNP. In [FV93], Feder and Vardi presented the following three conditions over formulas of SNP:

**Monadicity** all the relation symbols $R_1, \ldots, R_k$ are monadic, i.e., they are unary relation symbols;

**Monotony** all the relation symbols from $\tau$ (the input signature) must be negative, i.e. they appear under an odd number of negation symbols;

**No disequalities** there are no disequalities in $\phi_0$.

The resulting logic is called MMSNP, and the relation symbols that are not existentially quantified in an MMSNP formula $\phi$ are called the input signature of $\phi$.

4.1.2 Coloured Obstruction Sets

It will be convenient to take a combinatorial perspective when working with problems in MMSNP. Let $\tau$ and $\rho$ be disjoint, finite, relational signatures, where $\rho$ consists of unary relational symbols. Let $G$ be a $\tau$-structure. A proper $\rho$-colouring of $G$ is an expansion $G^*$ of $G$ by the relations of $\rho$, such that each element of $\text{Dom}(G)$ is contained in exactly one $R^{G^*}$ for $R \in \rho$. If $\mathcal{F}$ is a set of $(\tau \cup \rho)$-structures, we say that $G^*$ is $\mathcal{F}$-free if for every structure $F$ of $\mathcal{F}$ there is no homomorphism $F \to G^*$. The class of all the finite or countably infinite $\mathcal{F}$-free structures is denoted by $\text{Forb}_b(\mathcal{F})$.

Let now $\mathcal{C}$ be a class of finite $\tau$-structures, and $\mathcal{F}$ be a set of finite $(\tau \cup \rho)$-structures. We say that $\mathcal{F}$ is a coloured obstruction set for $\mathcal{C}$ if $\mathcal{C}$ is precisely the class of $\tau$-structures which have an $\mathcal{F}$-free proper $\rho$-colouring. The connection between coloured obstruction sets and MMSNP is given by the following theorem.
Theorem 4.1 (from [FV99]). Let $C$ be a class of finite $\tau$-structures. There exists an MMSNP formula that defines $C$ if and only if there exists a finite unary signature $\rho$ and a finite set of $(\tau \cup \rho)$-structures that is a coloured obstruction set for $C$.

Example 5. Let $\tau = \{E\}$ consist of one binary relation symbol. Let $\phi$ be the $\tau$-formula
\[
\exists R \exists B \forall x \forall y \forall z (-(R(x) \land R(z) \land R(y) \land E(x, y) \land E(x, z) \land E(y, z)) \\
\land \neg (B(x) \land B(z) \land B(y) \land E(x, y) \land E(x, z) \land E(y, z))),
\]
which is in MMSNP. A finite graph $G$ satisfies $\phi$ if and only if $G$ can be 2-coloured in such a way that no triangle in $G$ is monochromatic. Let $\rho = \{R, B\}$. An obstruction set for $\phi$ simply consists of two monochromatic triangles of colour $R$ and $B$.

Note that there is no canonical coloured obstruction set for a given formula $\phi$: two different families $\mathcal{F}$ and $\mathcal{F}'$ of structures can be a coloured obstruction set for the same class of structures. For example, consider the family $\mathcal{F}'$ that we obtain by adding to the example above any monochromatic graph $H$ that contains a triangle. If a structure $G$ satisfies $\phi$, we know that $G$ has some $\{R, B\}$-colouring $G'$ such that no monochromatic triangle homomorphically maps to $G'$, which implies that $H$ does not homomorphically map to $G'$ either. Conversely, if $G'$ is $\mathcal{F}$-free it is in particular $\mathcal{F}'$-free, and so $G$ satisfies $\phi$. This shows that each of $\mathcal{F}$ and $\mathcal{F}'$ is an obstruction set for the same class of structures. More generally, we see that if $\mathcal{F}$ contains two structures $A, B$ such that $A$ homomorphically maps to $B$, then $\mathcal{F} \setminus \{A\}$ and $\mathcal{F}$ are obstruction sets for the class of structures. We will thus assume in the following that whenever $A \in \mathcal{F}$ and $A \rightarrow B$, then $B \not\in \mathcal{F}$, and say in this case that $\mathcal{F}$ is reduced.

A problem in MMSNP with obstruction set $\mathcal{F}$ is polynomially equivalent to a problem with obstruction set $\mathcal{F}'$ where $\mathcal{F}'$ consists of connected structures only (we say that a structure is connected if and only if it is not the disjoint union of two of its substructures), see for example [MS07]. We will henceforth only consider obstruction sets that consist of finitely many finite connected structures. We call the associated logic connected MMSNP, that is, connected MMSNP consists of those MMSNP sentences which have an associated obstruction set consisting only of connected structures. We will see in Section 4.2 that we will even consider only coloured obstruction sets on which we impose more conditions, by a result in [BM].

4.2 Universal Structures for Connected Obstruction Sets

We want to study problems in MMSNP using constraint satisfaction problems with infinite templates. For a formula $\phi$ in connected MMSNP, we will see below that there exists an $\omega$-categorical structure $\Gamma$, called the template associated with $\phi$, such that CSP($\Gamma$) = MODELS($\phi$). Moreover, $\Gamma$ enjoys good combinatorial properties that will help us in the analysis of its polymorphisms. First, we present a restricted notion of canonical functions:

Definition 4.1. Let $D$ be a set and $\Delta$ be a structure whose domain is $D$. A function $f: D^n \rightarrow D$ is said to be 1-canonical with respect to $\Delta$ if for every $x_1, \ldots, x_n$, the type of $f(x_1, \ldots, x_n)$ in $\Delta$ is determined by the types of $x_1, \ldots, x_n$ in $\Delta$.

A structure $U$ whose age is some class $C$ is called universal for $C$.

Theorem 4.2 (Theorem 3.1 in [HN13]). Let $\mathcal{F}$ be a regular family of finite connected structures over a finite signature $\tau$. There exists a finite signature $\sigma$ and a homogeneous $(\tau \cup \sigma)$-structure $\mathcal{V}'$ such that the $\tau$-reduct $\mathcal{V}$ of $\mathcal{V}'$ is universal for Forb$_h(\mathcal{F})$.  

27
The regularity hypothesis in Theorem 4.2 is a condition that is true of every finite family of finite connected structures. We now proceed with the definition of the desired structure $\Gamma$. Fix a coloured obstruction set $F$ of $(\tau \cup \rho)$-structures, where $\rho$ is unary, and let $V'$ be as above. Consider the substructure $W'$ induced by $V'$ on the set of elements that have exactly one colour in $\rho$. Let $U'$ be the model-complete core of $W'$, that exists by Theorem 2.3. The relation between the different structures is depicted in Figure 4.1. We write $U, V, W$ (without the prime) to denote the $(\tau \cup \rho)$-reducts of the respective primed structures.

Note that if $G$ is a $\tau$-structure and $G^*$ is a proper $\rho$-colouring of $G$, then $G^*$ homomorphically maps to $U$ if and only if $G^*$ is $F$-free. This is because if $G^*$ is $F$-free, it embeds in $V$ by Theorem 4.2, and since $G^*$ is properly coloured it embeds in the part of $V$ that is properly coloured, i.e., it embeds in $W$. Since $W$ and $U$ are homomorphically equivalent, we obtain a homomorphism from $G^*$ to $U$.

**Proposition 4.3.** Let $\Gamma$ be the $\tau$-reduct of $U$, and let $G$ be a finite $\tau$-structure. Then $G$ homomorphically maps to $\Gamma$ (i.e., $G$ is a yes-instance for $\text{CSP}(\Gamma)$) if and only if $G \models \phi$.

**Proof.** Suppose that $G \models \phi$, so there exists a proper $\rho$-colouring $G^*$ of $G$ that is $F$-free. By the previous remark $G^*$ homomorphically maps to $U$, so the $\tau$-reduct $G$ of $G^*$ maps to $\Gamma$.

Conversely, suppose that $h: G \to \Gamma$ is a $\tau$-homomorphism. Define a proper $\rho$-colouring of $G$ as follows. For $C_i \in \rho$, set $x \in C_i^{G^*}$ if and only if $h(x) \in C_i^U$, so that $h$ is a $(\tau \cup \rho)$-homomorphism from $G^*$ to $U$. It follows that $G^*$ is $F$-free, and that $G \models \phi$. \hfill $\square$

A structure is said to be 2-connected when the removal of any element leaves the structure connected. If $F$ is a coloured obstruction set that consists only of 2-connected structures, the structure $U$ constructed above satisfies some additional property:

**Theorem 4.4** (From [BM]). Let $\phi$ be a connected MMSNP formula. There exists a 2-connected coloured obstruction set $F$ for $\phi$ and an $\omega$-categorical structure $U$ which is universal for $\text{Forb}_h(F)$ and whose 1-types are isolated by a quantifier-free formula.
This theorem implies the following lemma, which will be used in almost every proof that follows.

**Lemma 4.5.** Whenever two elements \( x \) and \( y \) of \( U \) satisfy the same atomic formulas in \( U \), there exists an automorphism of \( U \) that maps \( x \) to \( y \).

**Proof.** If \( x \) and \( y \) satisfy the same quantifier-free formulas, they have the same 1-type in \( U \) by Theorem 4.3. Since \( U \) is \( \omega \)-categorical, it follows that there exists an automorphism of \( U \) that maps \( x \) to \( y \). \( \square \)

Finally, the following theorem is similar to Theorem 2.7 and states that we can 1-canonize functions with respect to \( U \).

**Theorem 4.6 (From [BF].** Let \( \phi \) be a connected MMSNP sentence, and let \( \Gamma \) be the template associated with \( \phi \). Let \( f \) be a function over \( \text{Dom}(\Gamma) \) and \( x_1, \ldots, x_n \) be elements of \( \text{Dom}(\Gamma) \). Then there exists a function \( g \) that is 1-canonical with respect to \( (U, x_1, \ldots, x_n) \), that coincides with \( f \) on \( \{x_1, \ldots, x_n\} \), and that is locally generated by \( \text{Aut}(U, x_1, \ldots, x_n) \cup \{f\} \).

### 4.3 A Dichotomy for Monochromatic Obstructions

We prove in this section the following theorem.

**Theorem 4.7.** Let \( \phi \) be a connected MMSNP sentence with input signature \( \tau \). Suppose that \( \phi \) has a coloured obstruction set \( F \) that is both monochromatic and 2-connected. On loopless instances, the problem \textsc{Models}(\( \phi \)) is in \( \mathsf{P} \) or \( \mathsf{NP} \)-complete.

Since we consider the problem \textsc{Models}(\( \phi \)) on loopless input structures only, we can assume that \( F \) contains all the coloured loops. That is, for each \( C \in \rho \) and each \( R \in \tau \), \( F \) contains the \((\tau \cup \rho)\)-structure \( F \) on one element \( x \) such that \( C^F = \{x\} \), \( R^F = \{(x, \ldots, x)\} \), and all the other relations are interpreted as the empty set.

Moreover, a very simple case to analyse is when for some \( C \in \rho \), the structures in \( F \) of colour \( C \) are precisely the loops. In this case the problem \textsc{Models}(\( \phi \)) is trivially in \( \mathsf{P} \): given a loopless input structure \( G \), the proper \( \rho \)-colouring of \( G \) that is obtained by assigning the colour \( C \) to every element in \( G \) is \( F \)-free. Therefore, we assume in the following that for each \( C \in \rho \), there exists some structure \( F \in F \) of colour \( C \) that contains more than one element.

**Lemma 4.8.** Suppose that \( C^n \in \mathsf{pp-definable} \) in \( \Gamma \) for all \( i \in I \). Then the relation \( N \) that contains precisely the pairs \((x, y)\) \( \in \text{Dom}(\Gamma)^2 \) such that \( x \) and \( y \) are not in the same colour has a \( \mathsf{pp} \)-definition in \( \Gamma \).

**Proof.** Let \( \{C_1, \ldots, C_n\} \) be the symbols of \( \rho \). Suppose that \( N \) is violated by some polymorphism \( f: \Gamma^n \rightarrow \Gamma \), which means that there are tuples \((x_1, \ldots, x_n), (y_1, \ldots, y_n)\) such that \( x_i \) and \( y_i \) are not in the same colour for all \( i \in [n] \) and \( f(x_1, \ldots, x_n) \) and \( f(y_1, \ldots, y_n) \) are in the same colour, say \( C_k \) for \( k \in [c] \). Let \( G \) be a structure in \( F \) of colour \( C_k \) and of size greater than 1. Let \( a_1, a_2 \) be two points of \( \text{Dom}(G) \). Since \( F \) is reduced, the structure \( G_i \) \((1 \leq i \leq n)\) obtained from \( G \) by changing the colour of \( a_1 \) to the colour of \( x_i \) in \( U \) and the colour of \( a_2 \) to the colour of \( y_i \) in \( U \) is \( F \)-free. This means that we can find for \( 1 \leq i \leq n \) homomorphisms \( h_i: G_i \rightarrow \Gamma \) such that \((h_1(a_1), \ldots, h_n(a_1)) = (x_1, \ldots, x_n) \) and \((h_1(a_2), \ldots, h_n(a_2)) = (y_1, \ldots, y_n) \). Then the function \( f \circ (h_1, \ldots, h_n) \) is a homomorphism from \( G \) to \( \Gamma \), which is a contradiction to the fact that \( G \) is in \( \mathcal{F} \). \( \square \)
We now turn ourselves to the case where at least one colour is not pp-definable in $\Gamma$.

**Definition 4.2.** Let $\Delta$ be a structure, and let $x, y$ be two elements of $\Delta$. We say that $y$ is in the weak orbit of $x$ if there exists an endomorphism of $\Delta$ that maps $x$ to $y$.

We write $x \prec y$ when $y$ is in the weak orbit of $x$, without specifying $\Delta$ when it is clear from the context. It is easy to see that $\prec$ is a reflexive and transitive relation. It turns out that in $\Gamma$, $\prec$ is also symmetric. In the proofs below, we will use 1-canonical functions and one 1-type will be important to us in particular.

**Proposition 4.9.** Let $x, y$ be two elements of $\Gamma$, and suppose that there exists an endomorphism $e$ of $\Gamma$ that maps $x$ to $y$. Then there exists an endomorphism of $\Gamma$ that maps $y$ to $x$. Moreover, whenever $x$ and $y$ are in the same weak orbit, the obstructions in $F$ with the same colour as $x$ or $y$ are the same.

**Proof.** Let $C^U_a$ be the colour of $x$ in $U$, and $C^U_{a_1}$ be that of $y$. Since two elements of the same colour in $U$ are in the same orbit, it suffices to prove that some element coloured $C^U_{a_1}$ can be mapped by an endomorphism of $\Gamma$ to an element coloured $C^U_a$. Let $f$ be locally generated by $\{e\} \cup \text{Aut}(U, x)$ and be 1-canonical as a map from $(U, x)$ to $U$.

Let $z$ be an element in $C^U_{a_2}$. If $f(z)$ is in $C^U_{a_1}$, we are done. Otherwise, let $f(z) \in C^U_{a_3}$ with $a_3 \neq a_1$, and moreover $a_3 \neq a_2$: otherwise, let $G$ be a structure in $F$ of colour $C_{a_2}$. For any element $b$ in $G$, the structure $G'$ obtained by changing the colour of $b$ to $C_{a_1}$ is $F$-free. Thus, $G'$ homomorphically maps to $U$ and by Lemma 4.3, we may even assume that the image of $b$ under this homomorphism is $x$. By applying $f$, we obtain a homomorphic image of $G$ in $U$, a contradiction.

Iterating this argument, we build a sequence $\{C^U_{a_2}, C^U_{a_3}, \ldots, C^U_{a_r}\}$ such that for all $i \in \{2, \ldots, r-1\}$ all the elements in $C^U_{a_i}$ are mapped under $f$ to $C^U_{a_{i+1}}$, and where $a_r \in \{a_1, \ldots, a_{r-1}\}$. We argue now that $a_r$ must be $a_1$, whereupon we obtain that $y$ can be mapped to $x$ using $f$ and automorphisms of $U$.

Let $i \in \{1, \ldots, r-1\}$ be such that $a_i = a_r$. We claim that the obstructions in $F$ whose colour are among $a_1, \ldots, a_{r-1}$ are the same. Suppose that some structure $G \in F$ of colour $C_{a_j}$ is such that the structure $H$ over the same domain obtained by recolouring the elements and giving them the colour $C_{a_j}$ is $F$-free. Hence, there exists a homomorphism from $H$ to $U$. When we apply $f$ we obtain a homomorphic image of $H$, now coloured $C_{a_{k+1}}$. Iterating this argument, we finally obtain a homomorphic image of $G$ in $U$, a contradiction.

We are now ready to prove that $a_1 = a_r$. If that was not the case, $C^U_{a_1}$ would have in-degree 2, in the sense that elements from both $C^U_{a_{r-1}}$ and $C^U_{a_{r-1}}$ map to $C^U_{a_1}$ under $f$. By the observation in the previous paragraph, $C^U_{a_{r-1}}$ has the same obstructions as $C^U_{a_r}$. Let $G$ be such an obstruction, of colour $C_{a_{r-1}}$, and distinguish an element $a$ in $G$. Change the colour of $a$ to $C_{a_1}$, and call $H$ this structure. We have that $H$ is $F$-free, so we have a homomorphism from $H$ to $U$. By applying $f$, we then obtain a homomorphic image of $G$ in $U$, a contradiction. Therefore, it must be the case that $a_1 = a_r$, and by the previous paragraph, all the obstructions in $F$ of colour $C_{a_1}, \ldots, C_{a_r}$ are the same.

We say that a weak orbit $O$ spans $C^U_i$ if $O \cap C^U_i \neq \emptyset$. Since weak orbits are preserved by automorphisms, and since $\text{Aut}(\Gamma)$ acts transitively on each colour, it follows that $C^U_i \subseteq O$. A weak orbit $O$ is trivial if there exists a unique $i \in \{1, \ldots, c\}$ such that $O$ spans $C^U_i$. As in the case where the colours were pp-definable in $\Gamma$, our aim is to prove that the relation $N$ is pp-definable in $\Gamma$. Here, however, one needs to restrict $N$ to a non-trivial weak orbit of $\text{Dom}(\Gamma)$.

30
Lemma 4.10. Weak orbits are pp-definable in $\Gamma$.

Proof. Let $O$ be a weak orbit, and let $f: \Gamma^n \to \Gamma$ be a polymorphism of $\Gamma$. Let $(x_1, \ldots, x_n)$ be a tuple such that $x_i \in O$ for all $i \in \{1, \ldots, n\}$. Since $x_2, \ldots, x_n$ are in the weak orbit of $x_1$, there are endomorphisms $e_2, \ldots, e_n$ of $\Gamma$ such that $e_i(x_1) = x_i$ for all $i \in \{2, \ldots, n\}$. Let us consider the endomorphism $g: x \mapsto f(x, e_2(x), \ldots, e_n(x))$ of $\Gamma$. We have $g(x_1) = f(x_1, x_2, \ldots, x_n)$, so that $f(x_1, \ldots, x_n)$ is in the same weak orbit as $x_1$, i.e., $f(x_1, \ldots, x_n) \in O$. Therefore, $O$ is preserved by all the polymorphisms of $\Gamma$, and is pp-definable in $\Gamma$. \qed

The next lemma plays a role similar to the role of Lemma 4.8 above, with a few technicalities added.

Lemma 4.11. Let $O$ be a weak orbit of $\Gamma$, and let $N$ be the binary relation that contains precisely the pairs $(x, y) \in \text{Dom}(\Gamma)^2$ such that $x$ and $y$ are not in the same orbit. The relation $N \cap O^2$ has a primitive positive definition in $\Gamma$.

Proof. Let $C_{a_1}, \ldots, C_{a_k}$ be the colours spanned by the weak orbit $O$. By Proposition 4.9, there exists a $\tau$-structure $G$ that is in $\mathcal{F}$ for all possible monochromatic colourings of its elements with the colours $C_{a_1}, \ldots, C_{a_k}$. Let $f: \Gamma^n \to \Gamma$ be a polymorphism of $\Gamma$, and suppose that $f$ violates $N \cap O^2$. As in the proof of Lemma 4.8, there exist tuples $(x_1, \ldots, x_n), (y_1, \ldots, y_n)$ of elements from $O$ such that $x_i$ and $y_i$ are not in the same $C_{a_i}$ for all $i \in [n]$ and $f(x_1, \ldots, x_n)$ and $f(y_1, \ldots, y_n)$ are in the same colour, say $C_{a_r}$ for $r \in [c]$. Let $g$ be locally generated by $\{f\} \cup \text{Aut}(U, x_1, y_1, \ldots, x_n, y_n)$ such that $g$ agrees with $f$ on $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ and such that $g$ is 1-canonical.

Let $b$ be an element in $G$. For $i \in \{1, \ldots, n\}$, let $H_i$ be the following structure. The domain is the disjoint union of $k$ copies of $G \setminus \{b\}$, to which we add two distinguished elements $\Box$ and $\Diamond$. The copies of $G \setminus \{b\}$ are given the colours $C_{a_1}, \ldots, C_{a_k}$, $\Box$ is given the colour of $x_i$ and $\Diamond$ is given the colour of $y_i$. Within each copy of $G \setminus \{b\}$, the relations of $\tau$ are those induced by $G$ on $G \setminus \{b\}$. For every $i$, the structure $H_i$ is $\mathcal{F}$-free. Indeed, since the structures of $\mathcal{F}$ are monochromatic and connected, the image of a homomorphism from a structure of $\mathcal{F}$ to $H_i$ is a connected substructure included in one of $C_{a_1}^U, \ldots, C_{a_k}^U$. Each colour in $H_i$ contains at most two connected components: one isomorphic to $G \setminus \{b\}$, and one that consists of only one point (either $\Diamond$ or $\Box$, but not both by assumption that $x_i$ and $y_i$ have different colours). But $G \setminus \{b\}$ and structures of size 1 are $\mathcal{F}$-free, so $H_i$ is $\mathcal{F}$-free. Therefore, for each $i \in \{1, \ldots, n\}$ there exists a homomorphism $h_i: H_i \to U$. By Lemma 4.5, we can suppose that $h_i(\Box) = x_i$ and $h_i(\Diamond) = y_i$. Note that the $\tau$-reducts of all the $H_i$ are equal, and let $H$ be this $\tau$-structure. We thus have a $\tau$-homomorphism $h: H \to \Gamma^n$ given by $h(x) = (h_1(x), \ldots, h_n(x))$. We claim that the image of $g \circ h$ in $\text{Dom}(\Gamma)$ contains a copy of $G$ which is included in a single colour in $U$, which contradicts the fact that $U$ is $\mathcal{F}$-free. Note that $g(\Box, \ldots, \Box)$ and $g(\Diamond, \ldots, \Diamond)$ are both in $C_{a_r}^U$. Moreover, if $x$ and $y$ are taken from the same copy of $G \setminus \{b\}$, then $g(h(x))$ and $g(h(y))$ have the same colour in $U$ by 1-canonicity of $g$. This implies that if $x$ and $y$ are taken in different copies of $G \setminus \{b\}$, then $g(h(x))$ and $g(h(y))$ have different colours. Indeed, if they had the same colours we would obtain a homomorphic copy of $G$ in $U$, a contradiction. By the same reasoning, the colour of a copy of $G \setminus \{b\}$ under $g \circ h$ cannot be $C_{a_r}^U$.

Finally, the colour of a copy of $G \setminus \{b\}$ under $g \circ h$ is one of $C_{a_1}^U, \ldots, C_{a_k}^U$, since $O$ is preserved by $g \circ h$ according to Lemma 4.10. We have a contradiction: there are $k$ copies of $G \setminus \{b\}$, which ought to have different colours under $g \circ h$, and since there are $k$ colours to pick from one of the copies must have the copy $C_{a_r}^U$.

We conclude that all the polymorphisms of $\Gamma$ preserve $N \cap O^2$, so that $N \cap O^2$ is pp-definable in $\Gamma$. \qed
Theorem 4.12. Let $E$ be the binary relation that contains precisely the pairs $(x, y) \in \text{Dom}(\Gamma)^2$ such that $x$ and $y$ are in the same orbit. Then exactly one of the following statements is true:

1. all the colours are pp-definable in $\Gamma$, and $E$ is pp-definable in $\Gamma$,

2. there exists a non-trivial weak orbit $O$ in $\Gamma$, and the relation $E \cap O^2$ is pp-definable in $\Gamma$.

Proof. Suppose that all the colours are pp-definable in $\Gamma$. By Lemma 4.8, the relation $E$ that

Proposition 4.13. Let $f$ be a polymorphism of $\Gamma$ such that the set of terms of $A$ is conservative clearly follows from Proposition 4.13. We now describe a $A/E$ whose domain is Dom($\Gamma$) and such that the element $G[i_1, \ldots, i_k]$ in $U$ is one of $C_{b_1}, \ldots, C_{b_n}$.

Proof. This proof uses only ideas that we already used before and is thus left for the interested reader.

The fact that $A/E$ is conservative clearly follows from Proposition 4.13. We now describe a structure $\Gamma_{fin}$ over $\{1, \ldots, c\}$ such that $\text{Clo}(A/E) \subseteq \text{Pol}(\Gamma_{fin})$ and such that $\text{CSP}(\Gamma)$ is polynomial-time reducible to $\text{CSP}(\Gamma_{fin})$. The first property gives that $\Gamma_{fin}$ is pp-interpretive in $\Gamma$, and thus that $\text{CSP}(\Gamma)$ and $\text{CSP}(\Gamma_{fin})$ are polynomial-time equivalent. Since the tractability conjecture has been verified for conservative algebras (three times, Bulatov first $[Bul11]$, then Barto $[Bar11]$, then Bulatov again $[Bul14]$), we finally obtain the dichotomy in this case.

Let $G$ be a $\tau$-structure of size $k$, and assume that $\text{Dom}(G) = \{1, \ldots, k\}$. For each $k$-tuple $(i_1, \ldots, i_k)$ of integers in $\{1, \ldots, c\}$, we define $G[i_1, \ldots, i_k]$ to be the expansion of $G$ by the relations $C_{i_1}, \ldots, C_{i_k}$ and such that the element $j$ of $\text{Dom}(G)$ has the colour $C_{i_j}$ in $G[i_1, \ldots, i_k]$. Define

\[
R^\tau_G = \{(i_1, \ldots, i_k) \in \{1, \ldots, c\}^k \mid G[i_1, \ldots, i_k] \text{ is } F\text{-free}\}.
\]

The induced constraint language of $\Gamma$, denoted by $\Gamma_{fin}$, is the structure whose domain is $\{1, \ldots, c\}$, that contains every unary relation and all the relations of the form $R^\tau_G$ where $G$ is the $\tau$-reduct of some structure in $F$. 32
Proposition 4.14 (See Lemma 1 in [BCF12]). The constraint satisfaction problem of $\Gamma$ reduces to the constraint satisfaction problem of $\Gamma_{\text{fin}}$, in polynomial-time.

Proposition 4.15. The polymorphism clone of $\Gamma_{\text{fin}}$ contains all the operations of $A/E$. Therefore, $\Gamma_{\text{fin}}$ has a primitive positive interpretation in $\Gamma$.

**Proof.** For a reduct $G$ of some structure $G^*$ in $F$, let $\phi_G(x_1, \ldots, x_{|G|})$ be the canonical query of $G$. The relation defined by $\phi_G$ in $\Gamma$ is preserved by $\text{Pol}(\Gamma)$, of course, and the relation $R_G^F$ is simply the image of this relation under the factor map. Therefore $R_G^F$ is preserved by all the operations in $A/E$. The fact that all the unary relations are preserved by $A/E$ is due to Proposition 4.13.

Theorem 4.16. Let $\mathcal{F}$ be a finite set of finite connected monochromatic structures that satisfies the conclusion of Theorem 4.4. Suppose that $\mathcal{F}$ contains a structure of colour $C_i$ for every $i \in \{1, \ldots, c\}$, and that $\mathcal{F}$ contains all the loops. Let $U$ be universal for the class $\text{Forb}_0(\mathcal{F})$ and $\Gamma$ be the $\tau$-reduct of $U$. Suppose moreover that the relations $C_i^U$ are pp-definable in $\Gamma$. Then the constraint satisfaction problem of $\Gamma$ is in P if and only if $\Gamma$ has a polymorphism that is cyclic modulo automorphisms, and it is NP-complete otherwise.

**Proof.** If $\text{Pol}(\Gamma)$ has no cyclic polymorphism modulo automorphisms, $A/E$ has no cyclic term, and $\text{CSP}(\Gamma)$ is NP-hard. If $\Gamma$ has a cyclic polymorphism modulo automorphisms, $A/E$ has a cyclic term, so $\text{CSP}(\Gamma_{\text{fin}})$ is tractable by Bulatov [Bul11], and $\text{CSP}(\Gamma)$ is in P by Proposition 4.14.

4.3.2 The Dichotomy for Structures of Type 2

We prove here that every structure of type 2 in Theorem 4.12 has an NP-hard CSP. Let $O$ be a non-trivial weak orbit of $\Gamma$, and let $\{C_{a_i}^U, \ldots, C_{a_j}^U\}$ be the colours spanned by $O$. Since $O$ is pp-definable in $\Gamma$, $O$ is a subuniverse of the polymorphism algebra $A$ of $\Gamma$. Moreover, the relation $E$ is a congruence of the algebra $O$ generated by $O$, and $O/E$ can be thought as an algebra over $\{1, \ldots, r\}$ by the identification $x/E = i$ iff $x \in C_{a_i}^U$.

As in the previous section, there exists a finite structure $\Gamma_{\text{fin}}$ over $\{1, \ldots, r\}$ such that $\text{Clo}(O/E) \subseteq \text{Pol}(\Gamma_{\text{fin}})$. The structure $\Gamma_{\text{fin}}$ is defined as above, where we only have the relations of the type $R_G^F \cap \{1, \ldots, r\}^{[G]}$ with $G$ being a reduct of some structure $G^*$ of $F$ whose colour is among $\{C_{a_1}, \ldots, C_{a_j}\}$ (note that we do not add all the unary relations, as this time $O/E$ might not be a conservative algebra). We show that $\text{CSP}(\Gamma_{\text{fin}})$ cannot have a cyclic polymorphism, or equivalently that it cannot have a Siggers polymorphism [Sig10], i.e. a function of arity 4 that satisfies the equations

$$\forall x, y \in \{1, \ldots, r\} \ f(y, x, x, x) = f(x, x, y, y) = f(x, y, y, y).$$

Let $G$ be a $\tau$-structure of size at least 2 such that all the monochromatic colourings of $G$ with a colour in $C_{a_1}, \ldots, C_{a_j}$ are in $\mathcal{F}$, that exists by Proposition 4.9. Suppose that $\text{Dom}(G) = \{1, \ldots, k\}$. We consider the four following colourings of $G$:

1. $C_{a_1}^G = \{2, 3, \ldots, k\}$, $C_{a_2}^G = \{1\}$, $C_{a_j}^G = \emptyset$ for $j > 2$;
2. $C_{a_1}^G = \{1, 2\}$, $C_{a_2}^G = \{3, \ldots, k\}$, $C_{a_j}^G = \emptyset$ for $j > 2$;
3. $C_{a_1}^G = \{1, 3, \ldots, k\}$, $C_{a_2}^G = \{2\}$, $C_{a_j}^G = \emptyset$ for $j > 2$;
4. $C_{a_1}^G = \{1\}$, $C_{a_2}^G = \{2, \ldots, k\}$, $C_{a_j}^G = \emptyset$ for $j > 2$. 

33
These four colourings give rise to four $\mathcal{F}$-free $(\tau \cup \rho)$-structures, which embed in $\mathbf{U}$ by embeddings $h_1, h_2, h_3, h_4$. Suppose that $\mathbf{O}/E$ has a Siggers term. This means that $\text{Pol}(\Gamma)$ contains a function of arity four whose behaviour satisfies the Siggers equations over $O$, if we only look at the colour of its arguments. It follows from the Siggers equations that the function $g: G \to \Gamma$ defined by $x \mapsto f(h_1(x), h_2(x), h_3(x), h_4(x))$ is a homomorphism whose image is contained in a single $C_{a_i}^\mathbf{U}$, contradicting the fact that $\mathbf{U}$ is $G$-free.

**Theorem 4.17.** Let $\mathcal{F}$ be a finite set of finite connected monochromatic structures that satisfies the conclusion of Theorem 4.4. Suppose that $\mathcal{F}$ contains a structure of colour $C_i$ of size at least 2 for every $i \in \{1, \ldots, c\}$, and that $\mathcal{F}$ contains all the loops. Let $\mathbf{U}$ be universal for the class $\text{Forb}_h(\mathcal{F})$ and $\Gamma$ be the $\tau$-reduct of $\mathbf{U}$. Suppose moreover that some relation $C_i^\mathbf{U}$ is not pp-definable in $\Gamma$. Then $\text{CSP}(\Gamma)$ is $\text{NP}$-hard and $\Gamma$ does not have a cyclic polymorphism modulo automorphisms.

**Proof.** The idea of the proof of $\text{NP}$-hardness is described above. The fact that $\Gamma$ does not have a cyclic polymorphism modulo automorphisms comes from the fact that $\Gamma$ interprets a finite-structure whose polymorphisms are essentially unary. It is well-known that this implies that $\text{Pol}(\Gamma)$ does not satisfy any non-trivial equation. \qed

### 4.4 Conclusion

In our result we only study the complexity of the problems of the form $\text{MODELS}(\phi)$ on loopless input structures only. This restriction also appeared very naturally in [BCF12], where their argument is based on the notion of chromatic number of a structure, a generalization of the eponymous idea in graph theory. Although our methods are very different and model-theoretic in nature, they also fail when we allow structures with loops. It seems important to find a way to overcome this difficulty algebraically, as it is a necessary step in the pursuit of a universal-algebraic proof of the Feder-Vardi-Kun result that MMSNP and constraint satisfaction problems over finite domains are intimately related.
Bibliography


