

On the Complexity of CSP-based Ideal Membership Problems

Andrei A. Bulatov*

Akbar Rafiey[†]

Abstract

In this paper we consider the Ideal Membership Problem (IMP for short), in which we are given real polynomials f_0, f_1, \dots, f_k and the question is to decide whether f_0 belongs to the ideal generated by f_1, \dots, f_k . In the more stringent version the task is also to find a proof of this fact. The IMP underlies many proof systems based on polynomials such as Nullstellensatz, Polynomial Calculus, and Sum-of-Squares. In the majority of such applications the IMP involves so called combinatorial ideals that arise from a variety of discrete combinatorial problems. This restriction makes the IMP significantly easier and in some cases allows for an efficient algorithm to solve it.

In 2019 Mastrolilli initiated a systematic study of IMPs arising from Constraint Satisfaction Problems (CSP) of the form $\text{CSP}(\Gamma)$, that is, CSPs in which the type of constraints is limited to relations from a set Γ . He described sets Γ on a 2-element set that give rise to polynomial time solvable IMPs and showed that for the remaining ones the problem is hard. We continue this line of research.

First, we show that many CSP techniques can be translated to IMPs thus allowing us to significantly improve the methods of studying the complexity of the IMP. We also develop universal algebraic techniques for the IMP that have been so useful in the study of the CSP. This allows us to prove a general necessary condition for the tractability of the IMP, and three sufficient ones. The sufficient conditions include IMPs arising from systems of linear equations over $\text{GF}(p)$, p prime, and also some conditions defined through special kinds of polymorphisms.

1 Introduction

The Ideal Membership Problem. The study of polynomial ideals and algorithmic problems related to them goes back to David Hilbert [24]. In spite of such a heritage, methods developed in this area till these days keep finding a wide range of applications in mathematics and computer science. In this paper we consider the Ideal Membership Problem (IMP for short), in which the goal is to decide whether a given polynomial belongs to a given ideal. It underlies such proof systems as Nullstellensatz and Polynomial Calculus.

To introduce the problem more formally, let \mathbb{F} be a field and $\mathbb{F}[x_1, x_2, \dots, x_n]$ denote the ring of polynomials over \mathbb{F} with indeterminates x_1, \dots, x_n . In this paper \mathbb{F} is always the field of

*abulatov@sfu.ca. Department of Computing Science, Simon Fraser University, Burnaby, BC, Canada. Research supported by an NSERC Discovery Grant.

[†]arafiey@sfu.ca. Department of Computing Science, Simon Fraser University, Burnaby, BC, Canada. Research supported by NSERC.

real or complex numbers. A set of polynomials $I \subseteq \mathbb{F}[x_1, x_2, \dots, x_n]$ is said to be an *ideal* if it is closed under addition and multiplications by elements from $\mathbb{F}[x_1, x_2, \dots, x_n]$. By the Hilbert Basis Theorem every ideal I has a finite generating set [23], that is, there exists $P = \{f_1, f_2, \dots, f_r\} \subseteq \mathbb{F}[x_1, x_2, \dots, x_n]$ such that for every $f_0 \in \mathbb{F}[x_1, x_2, \dots, x_n]$ the polynomial f_0 belongs to I if and only if there exists a *proof*, that is, polynomials h_1, \dots, h_r such that the identity $f_0 = h_1 f_1 + \dots + h_r f_r$ holds. Such proofs will also be referred to as *ideal membership* proofs. We then write $I = \langle P \rangle$. The Hilbert Basis Theorem allows one to state the IMP as follows: given polynomials f_0, f_1, \dots, f_r decide whether there exists a proof that $f_0 \in \langle f_1, \dots, f_r \rangle$.

In many cases combinatorial or optimization problems can be encoded as collections of polynomials, and the problem is then reduced to proving or refuting that some polynomial vanishes at specified points or is nonnegative at those points. Polynomial proof systems can then be applied to find a proof or a refutation of these facts. Polynomial Calculus and Nullstellensatz proof systems are some of the standard techniques to check for zeroes of a polynomial, and Sum-of-Squares (SOS) allows to prove or refute the nonnegativity of a polynomial. We may be interested in the length or degree of a proof in one of those systems. Sometimes such proofs can also be efficiently found — such proof systems are referred to as *automatizable* — and in those cases we are also concerned with the complexity of finding a proof.

The general IMP is a difficult problem and it is not even obvious whether or not it is decidable. The decidability was established in [22, 37, 39]. Then Mayr and Meyer [30] were the first to study the complexity of the IMP. They proved an exponential space lower bound for the membership problem for ideals generated by polynomials with integer and rational coefficients. Mayer [29] went on establishing an exponential space upper bound for the IMP for ideals over \mathbb{Q} , thus proving that such IMPs are **EXSPACE**-complete. The source of hardness here is that a proof that $f_0 \in \langle P \rangle$ may require polynomials of exponential degree. In the cases when the degree of a proof has a linear bound in the degree of f_0 , the IMP can be solved efficiently. (There is also the issue of exponentially long coefficients that we will mention later.)

Combinatorial ideals. By Hilbert’s Nullstellensatz polynomial ideals can often be characterized by the set of common zeroes of all the polynomials in the ideal. Such sets are known as *affine varieties* and provide a link between ideals and combinatorial problems, where an affine variety corresponds to the set of feasible solutions of a problem. Combinatorial problems give rise to a fairly narrow class of ideals known as *combinatorial* ideals. The corresponding varieties are finite, and therefore the ideals itself are zero-dimensional and radical. The former implies that the IMP can be decided in single-exponential time [18], while the latter will be important later for IMP algorithms. Indeed, if the IMP is restricted to radical ideals, it is equivalent to (negation of) the question: given f_0, f_1, \dots, f_r does there exists a zero of f_1, \dots, f_r that is not a zero of f_0 .

The special case of the IMP with $f_0 = 1$ has been studied for combinatorial problems in the context of lower bounds on Polynomial Calculus and Nullstellensatz proofs, see e.g. [5, 14, 21]. A broader approach of using polynomials to represent finite-domain constraints has been explored in [15, 26]. Clegg et al., [15], discuss a propositional proof system based on a bounded degree version of Buchberger’s algorithm [9] for finding proofs of unsatisfiability. Jefferson et al., [26] use a modified form of Buchberger’s algorithm that can be used to achieve the same benefits as the local-consistency algorithms which are widely used in constraint processing.

Complexity of the IMP and its applications in other proof systems. Whenever the degree of a proof h_1, \dots, h_r is bounded, that is, the degree of each h_i is bounded by a constant, there is an SDP program of polynomial size whose solutions are the coefficients of the proof. If in addition the solution of the SDP program can be represented by a polynomial number of bits (thus having low *bit complexity*), a proof can be efficiently found. This property also applies to SOS and provides one of the most powerful algorithmic methods for optimization problems.

It was recently observed by O'Donnell [32] that low degree of proofs does not necessarily implies its low bit complexity. More precisely, he presented a collection of polynomials and a polynomial such that there are low degree proofs of nonnegativity for these polynomials, that is, there also exists a polynomial size SDP whose solutions represent a SOS proof of that. However, the size of those solutions are always exponential (or the proof has high bit complexity), and therefore the Ellipsoid method will take exponential time to find them. It therefore is possible that every low degree SOS proof has high bit complexity. Raghavendra and Weitz [36] also demonstrated an example showing that this is the case even if all the constraints in the instance are Boolean, that is, on a 2-element set.

The examples of O'Donnell and Raghavendra-Weitz indicate that it is important to identify conditions under which low degree proofs (Nullstellensatz, Polynomial Calculus, or SOS) exist and also have low bit complexity. Raghavendra and Weitz [36] suggested some sufficient conditions of this kind for SOS proofs that are satisfied for a number of well studied problems such as Matching, TSP, and others. More precisely, they formulate three conditions that a polynomial system ought to satisfy to yield for a low bit complexity SOS proof. Two of these conditions hold for the majority of combinatorial ideals, while the third, the low bit complexity of Nullstellensatz, is the only nontrivial one. Noting that Nullstellensatz proofs are basically witnesses of ideal membership, the IMP is at the core of all the three proof systems.

IMP and CSP. The Constraint Satisfaction Problem (CSP) provides a general framework for a wide range of combinatorial problems. In a CSP we are given a set of variables and a collection of constraints, and we have to decide whether the variables can be assigned values so that all the constraints are satisfied. Following [26, 41, 28] every CSP can be associated with a polynomial ideal. Let CSP \mathcal{P} be given on variables x_1, \dots, x_n that can take values from a set $D = \{0, \dots, t-1\}$. The ideal $I(\mathcal{P})$ of $\mathbb{F}[x_1, \dots, x_n]$ whose corresponding variety equals the set of solutions of \mathcal{P} is constructed as follows. First, for every x_i the ideal $I(\mathcal{P})$ contains a *domain* polynomial $f_D(x_i)$ whose zeroes are precisely the elements of D . Then for every constraint $R(x_{i_1}, \dots, x_{i_k})$, where R is a predicate on D , the ideal $I(\mathcal{P})$ contains a polynomial $f_R(x_{i_1}, \dots, x_{i_k})$ that interpolates R , that is, for $(x_{i_1}, \dots, x_{i_k}) \in D^k$ it holds $f_R(x_{i_1}, \dots, x_{i_k}) = 0$ if and only if $R(x_{i_1}, \dots, x_{i_k})$ is true. This model generalizes a number of constructions used in the literature to apply Nullstellensatz or SOS proof systems to combinatorial problems, see, e.g., [5, 14, 21, 36].

The construction above also provides useful connections between the IMP and the CSP. For instance, the CSP \mathcal{P} is unsatisfiable if and only if the variety associated with $I(\mathcal{P})$ is empty, or equivalently, if and only if $1 \in I(\mathcal{P})$ (1 here denotes the polynomial of degree 0). In this sense it is related to the standard decision version of the CSP. However, since $I(\mathcal{P})$ is radical for any instance \mathcal{P} , the IMP reduces to verifying whether every point in the variety of $I(\mathcal{P})$ is a zero of f_0 . Thus, it is probably closer to the CSP Containment problem (given two CSP instances over the same variables, decide if every solution of the first one is also a solution of the second one), which has mainly been studied in the context of Database theory and Conjunctive Query

Containment, see, e.g., [27].

Studying CSPs in which the type of constraints is restricted has been a very fruitful research direction. For a fixed set of relations Γ over a finite set D , also called a *constraint language*, $\text{CSP}(\Gamma)$ denotes the CSP restricted to instances with constraints from Γ , [38, 20]. Feder and Vardi in [20] conjectured that every CSP of the form $\text{CSP}(\Gamma)$ is either solvable in polynomial time or is **NP**-complete. This conjecture was recently confirmed by Bulatov and Zhuk in [10, 42]. Mastrolilli in [28] initiated a similar study of the IMP. Let $\text{IMP}(\Gamma)$ be the IMP restricted to ideals produced by instances from $\text{CSP}(\Gamma)$. As was observed above, the complement of $\text{CSP}(\Gamma)$ reduces to $\text{IMP}(\Gamma)$, and so $\text{IMP}(\Gamma)$ is **coNP**-complete whenever $\text{CSP}(\Gamma)$ is **NP**-hard. Therefore the question posed in [28] is:

Problem 1.1. *For which constraint languages Γ is it possible to efficiently find a generating set for the ideal $\mathcal{I}(\mathcal{P})$, $\mathcal{P} \in \text{CSP}(\Gamma)$, that allows for low bit complexity proof of ideal membership?*

Mastrolilli [28] (along with [8]) resolved this question in the case when Γ is Boolean language, that is, over the set $\{0, 1\}$. He proved that in this case $\text{IMP}(\Gamma)$ is polynomial time solvable, and moreover ideal membership proofs can be efficiently found, too, for any Boolean Γ for which $\text{CSP}(\Gamma)$ is polynomial time solvable. However, $\text{IMP}(\Gamma)$ in [28] satisfies two restrictions. First, the result is obtained under the assumption that Γ contains the constant relations that allows one to fix a value of a variable. Mastrolilli called such languages *idempotent*. We will show that this suffices to obtain a more general result. Second, in the majority of cases a bound on the degree of the input polynomial f_0 has to be introduced. The IMP where the input polynomial has degree at most d will be denoted by $\text{IMP}_d, \text{IMP}_d(\Gamma)$. The exact result is that for any idempotent Boolean language Γ the problem $\text{IMP}_d(\Gamma)$ is polynomial time solvable for any d when $\text{CSP}(\Gamma)$ is polynomial time solvable, and an ideal membership proof can be efficiently found, and $\text{IMP}_0(\Gamma)$ is **coNP**-complete otherwise. We will reflect on the distinction between IMP and IMP_d later in the paper, but do not go deeply into that. The case when $\text{CSP}(\Gamma)$ is equivalent to solving systems of linear equations modulo 2 was also considered in [8] fixing a gap in [28]. There has been very little work done on $\text{IMP}(\Gamma)$ beyond 2-element domains. The only result we are aware of is [7] that proves that $\text{IMP}_d(\Gamma)$ is polynomial time when Γ is on a 3-element domain and is invariant under the so-called dual-discriminator operation, which imposes very strong restrictions on the relations from Γ ; we will discuss this case in greater details in Section 5.2.

The main tool for proving the tractability of $\text{IMP}(\Gamma)$ is constructing a Gröbner Basis of the corresponding ideal. It is not hard to see that the degree of polynomials in a Gröbner Basis of an ideal of $\mathbb{F}[x_1, \dots, x_n]$ that can occur in $\text{IMP}(\Gamma)$ is only bounded by $n|D|$, where Γ is over a set D . Therefore the basis and polynomials themselves can be exponentially large in general. In fact, this is the main reason why considering $\text{IMP}_d(\Gamma)$ instead makes the problem easier. For solving this problem it suffices to find a *d-truncated* Gröbner Basis, in which the degree of polynomials is bounded by d , and so such a basis always has polynomial size. Thus, the (possible) hardness of $\text{IMP}_d(\Gamma)$ is due to the hardness of constructing a Gröbner Basis.

Our contribution

In this paper we expand on [28] and [7, 8] in several ways. We consider $\text{IMP}(\Gamma)$ for languages Γ over arbitrary finite set and attempt to obtain general results about such problems. However, we mainly focus on a slightly different problem than Problem 1.1.

Problem 1.2. *For which constraint languages Γ the problem $\text{IMP}(\Gamma)$ [or $\text{IMP}_d(\Gamma)$] can be solved in polynomial time?*

Note that answering whether f_0 belongs to a certain ideal does not necessarily mean finding an ideal membership proof of that. However, we will argue that in many applications this is the problem we need to solve and therefore our results apply.

Expanding the constraint language. Firstly, in Section 3 we study reductions between IMP 's when the language Γ is enlarged in certain ways. Let Γ be a constraint language over a set D . By Γ^* we denote Γ with added *constant* relations, that is, relations of the form $\{(a)\}$, $a \in D$. Imposing such a constraint on a variable x essentially fixes the allowed values of x to be a . First, we prove that adding constant relations does not change the complexity of the IMP .

Theorem 1.3. *For any Γ over D the problem $\text{IMP}(\Gamma^*)$ is polynomial time reducible to $\text{IMP}(\Gamma)$, and for any d the problem $\text{IMP}_d(\Gamma^*)$ is polynomial time reducible to $\text{IMP}_{d+|D|(|D|-1)}(\Gamma)$.*

Theorem 1.3 has two immediate consequences. Since $\text{IMP}(\Gamma)$ is in **co-NP**, for any $\text{CSP}(\Gamma)$ instance \mathcal{P} there is always a proof that the input polynomial f_0 does not belong to the ideal $I(\mathcal{P})$. Any solution of \mathcal{P} that is not a zero of f_0 will do. Finding such a proof may be treated as a search version of $\text{IMP}(\Gamma)$. Through self-reducibility, Theorem 1.3 allows us to solve the search problem.

Theorem 1.4. *Let Γ be such that $\text{IMP}(\Gamma)$ [$\text{IMP}_{d+|D|(|D|-1)}(\Gamma)$] is solvable in polynomial time. Then for any instance (f_0, \mathcal{P}) of $\text{IMP}(\Gamma)$ [$\text{IMP}_d(\Gamma)$] such that $f_0 \notin I(\mathcal{P})$, a solution \mathbf{a} of \mathcal{P} such that $f_0(\mathbf{a}) \neq 0$ can also be found in polynomial time.*

Theorem 1.3 also provides a hint at a more plausible conjecture for which languages Γ the problem $\text{IMP}(\Gamma)$ or $\text{IMP}_d(\Gamma)$ is polynomial time. In particular, it allows to find an example of Γ such that $\text{CSP}(\Gamma)$ is tractable while $\text{IMP}_d(\Gamma)$ is not, even for a Γ on a 2-element set and $d = 1$. Later we state some results that might indicate that $\text{IMP}_d(\Gamma)$ is polynomial time for every Γ such that $\text{CSP}(\Gamma^*)$ is polynomial time. Note that the structure of such CSPs is now very well understood.

Another way of expanding a constraint language is by means of *primitive-positive (pp-) definitions* and *pp-interpretations*, and it is at the core of the so-called algebraic approach to the CSP. A relation R is said to be pp-definable in Γ if there is a first order formula Φ using only conjunctions, existential quantifiers, equality relation, and relations from Γ that is equivalent to R . Pp-interpretations are more complicated (see Section 3.3) and allow for certain encodings of R .

Theorem 1.5. (1) *Let Γ, Δ be constraint languages over the same set D , Δ is finite, and every relation from Δ is pp-definable in Γ . Then $\text{IMP}(\Delta)$ is polynomial time reducible to $\text{IMP}(\Gamma)$ and $\text{IMP}_d(\Delta)$ is polynomial time reducible to $\text{IMP}_d(\Gamma)$ for any d .*

(2) *Let Γ, Δ be constraint languages, Δ is finite, and Δ is pp-interpretable in Γ . Then $\text{IMP}(\Delta)$ is polynomial time reducible to $\text{IMP}(\Gamma)$, and there is a constant k such that $\text{IMP}_d(\Delta)$ is polynomial time reducible to $\text{IMP}_{kd}(\Gamma)$ for any d .*

The approach of Theorem 1.5 was first applied to various proof systems in [1], although that work is mostly concerned with proof complexity rather than computational complexity. Mastrolilli [28] ventured into pp-definability without proving any reductions. In particular, the

first part of Theorem 1.5 uses techniques from [28] for projections of ideals. It will later allow us to develop further universal algebra techniques for the IMP. The second part of that theorem will also work towards more powerful universal algebra methods.

Recall that according to [36] in order to find an SOS proof one needs to be able to find Nullstellensatz proofs efficiently. The reductions from Theorems 1.3, 1.5 do not always allow to recover such a proof efficiently only confirming such a proof exists, and therefore cannot be directly used in the conditions from [36]. In Section 3.4 we discuss how this issue can be avoided. To this end we consider the problem of checking if a polynomial is nonnegative at certain points, we call it *IsPOSITIVE*, without assuming any specific way of doing that. We show that the reductions from Theorems 1.3, 1.5 also work for *IsPOSITIVE*, and therefore can probably be useful in a number of applications.

Polymorphisms and sufficient conditions for tractability. Recall that a *polymorphism* of a constraint language Γ over a set D is a multi-ary operation on D that can be viewed as a multi-dimensional symmetry of relations from Γ . By $\text{Pol}(\Gamma)$ we denote the set of all polymorphisms of Γ . As in the case of the CSP, Theorem 1.5 implies that polymorphisms of Γ is what determines the complexity of $\text{IMP}(\Gamma)$. In Section 4.1 we show the following.

Corollary 1.6. *Let Γ, Δ be constraint languages over the same set D , Δ is finite. If $\text{Pol}(\Gamma) \subseteq \text{Pol}(\Delta)$, then $\text{IMP}(\Delta)$ is polynomial time reducible to $\text{IMP}(\Gamma)$ and $\text{IMP}_d(\Delta)$ is polynomial time reducible to $\text{IMP}_d(\Gamma)$ for any d .*

Corollary 1.6 allows us to represent IMPs through polymorphisms and classify the complexity of IMPs according to the corresponding polymorphisms. The method has been initiated by Mastrolilli and Bharathi [7, 8, 28], although mainly for 2- and one case of 3-element sets. We apply this approach to obtain three sufficient conditions for tractability of the IMP.

Theorem 1.7. *Let Γ be a constraint language over a set D . Then if one of the following conditions holds, $\text{IMP}_d(\Gamma)$ is polynomial time solvable for any d .*

1. Γ has the dual-discriminator polymorphism (i.e. a ternary operation g such that $g(x, y, z) = x$ unless $y = z$, in which case $g(x, y, z) = y$);
2. Γ has a semilattice polymorphism (i.e. a binary operation f such that $f(x, x) = x$, $f(x, y) = f(y, x)$, and $f(f(x, y), z) = f(x, f(y, z))$);
3. $|D| = p$, p prime, and Γ has an affine polymorphism modulo p (i.e. a ternary operation $h(x, y, z) = x \oplus y \oplus z$, where \oplus, \ominus are addition and subtraction modulo p , or, equivalently, of the field $\text{GF}(p)$). In this case every CSP can be represented as a system of linear equations over $\text{GF}(p)$.

The three polymorphisms occurring in Theorem 1.7 have played an important role in the CSP research. For one reason they completely cover the tractable cases when $|D| = 2$ and therefore the results of [28, 8], although we used some results (on semilattice polymorphisms) and approaches (on affine polymorphisms) from [28, 8].

For the first part of Theorem 1.7 we come up with a technique of preprocessing the input polynomial f_0 that allows us to get rid of permutation constraints and greatly simplify the proof for the dual-discriminator polymorphism, including the special case $|D| = 3$ considered in [7]. This, however, makes it difficult to recover a proof for the original polynomial. Note also that in

this case $\text{IMP}(\Gamma)$ — without restrictions on the degree of the input polynomials — is polynomial time.

In the second part of Theorem 1.7 we use the fact, see [33], that any language with a semilattice polymorphism is pp-interpretable in a language on a 2-element set also having a semilattice polymorphism. Then Theorem 1.7(2) follows from Theorem 1.5(2) and the results of [28].

As is mentioned, the third part of Theorem 1.7 is in fact about systems of linear equations over $\text{GF}(p)$, since every instance of $\text{CSP}(\Gamma)$, where Γ has an affine polymorphism, is equivalent to a system of linear equations over $\text{GF}(p)$. This part is technically most involved. We use the approach from [8] based on FGML algorithm [19] to construct a bounded degree Gröbner Basis. By this we therefore show that not only $\text{IMP}_d(\Gamma)$ is polynomial time, but also that an ideal membership proof can be found in polynomial time. This requires to develop a polynomial-like representation of functions on the set D^n , which may be of some independent interest. The role of monomials in this representation is played by linear polynomials over $\text{GF}(p)$. Let F_k denote the set of linear polynomials over $\text{GF}(p) = \mathbb{Z}_p$

$$F_k = \left\{ \bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus \beta \mid \alpha_i \in \mathbb{Z}_p, i \in \{1, \dots, k\}, \beta \in \{0, \dots, p-2\} \right\},$$

and

$$\mathcal{F}_k = F_k \cup \dots \cup F_0 \cup \{1\}.$$

Proposition 1.8. *Let $|D| = p$. For any real function $f : D^k \rightarrow \mathbb{R}$ there are reals $\gamma_1, \dots, \gamma_s$ and $g_1, \dots, g_s \in \mathcal{F}_{k-1}$ such that*

$$f(x_1, \dots, x_k) = \gamma_1 g_1 + \dots + \gamma_s g_s,$$

on D^k , where multiplication and addition are as real numbers.

Algebras. Finally, in Section 4.2 we prove that the standard features of the universal algebraic approach to the CSP work for IMP as well. These include reductions for standard algebraic constructions such as *subalgebras*, *direct powers*, and *homomorphic images*. They easily follow from Theorem 1.5(2). A more general construction of direct product requires a more general version of $\text{CSP}(\Gamma)$, and therefore of $\text{IMP}(\Gamma)$, the multi-sorted one, in which every variable can have its own domain of values. We do not venture into that direction in the paper. One implication of these results is a necessary condition for tractability of $\text{IMP}(\Gamma)$ that follows from a similar one for the CSP.

2 Preliminaries

2.1 Ideals, varieties and the Ideal Membership Problem

Let \mathbb{F} denote an arbitrary field. Let $\mathbb{F}[x_1, \dots, x_n]$ be the ring of polynomials over a field \mathbb{F} and indeterminates x_1, \dots, x_n . Sometimes it will be convenient not to assume any specific ordering or names of the indeterminates. In such cases we use $\mathbb{F}[X]$ instead, where X is a set of indeterminates, and treat points in \mathbb{F}^X as mappings $\varphi : X \rightarrow \mathbb{F}$. The value of a polynomial $f \in \mathbb{F}[X]$ is then written as $f(\varphi)$. Let $\mathbb{F}[x_1, \dots, x_n]_d$ denote the subset of polynomials of degree at most d . An

ideal of $\mathbb{F}[x_1, \dots, x_n]$ is a set of polynomials from $\mathbb{F}[x_1, \dots, x_n]$ closed under addition and multiplication by a polynomial from $\mathbb{F}[x_1, \dots, x_n]$. We will need ideals represented by a generating set.

Definition 2.1. The ideal (of $\mathbb{F}[x_1, \dots, x_n]$) generated by a finite set of polynomials $\{f_1, \dots, f_m\}$ in $\mathbb{F}[x_1, \dots, x_n]$ is defined as

$$\mathbf{I}(f_1, \dots, f_m) \stackrel{\text{def}}{=} \left\{ \sum_{i=1}^m t_i f_i \mid t_i \in \mathbb{F}[x_1, \dots, x_n] \right\}.$$

Definition 2.2. The set of polynomials that vanish in a given set $S \subset \mathbb{F}^n$ is called the vanishing ideal of S and denoted

$$\mathbf{I}(S) \stackrel{\text{def}}{=} \{f \in \mathbb{F}[x_1, \dots, x_n] : f(a_1, \dots, a_n) = 0 \ \forall (a_1, \dots, a_n) \in S\}.$$

Definition 2.3. An ideal \mathbf{I} is radical if $f^m \in \mathbf{I}$ for some integer $m \geq 1$ implies that $f \in \mathbf{I}$. For an arbitrary ideal \mathbf{I} the smallest radical ideal containing \mathbf{I} is denoted $\sqrt{\mathbf{I}}$. In other words $\sqrt{\mathbf{I}} = \{f \in \mathbb{F}[x_1, \dots, x_n] \mid f^m \in \mathbf{I} \text{ for some } m\}$.

Another common way to denote $\mathbf{I}(f_1, \dots, f_m)$ is by $\langle f_1, \dots, f_m \rangle$ and we will use both notations interchangeably.

Definition 2.4. Let $\{f_1, \dots, f_m\}$ be a finite set of polynomials in $\mathbb{F}[x_1, \dots, x_n]$. We call

$$\mathbf{V}(f_1, \dots, f_m) \stackrel{\text{def}}{=} \{(a_1, \dots, a_n) \in \mathbb{F}^n \mid f_i(a_1, \dots, a_n) = 0 \ 1 \leq i \leq m\}$$

the affine variety defined by f_1, \dots, f_m .

Similarly, for an ideal $\mathbf{I} \subseteq \mathbb{F}[x_1, \dots, x_n]$ we denote by $\mathbf{V}(\mathbf{I})$ the set $\mathbf{V}(\mathbf{I}) = \{(a_1, \dots, a_n) \in \mathbb{F}^n \mid f(a_1, \dots, a_n) = 0 \ \forall f \in \mathbf{I}\}$.

The Weak Nullstellensatz states that in any polynomial ring, algebraic closure is enough to guarantee that the only ideal which represents the empty variety is the entire polynomial ring itself. This is the basis of one of the most celebrated mathematical results, Hilbert's Nullstellensatz.

Theorem 2.5 (The Weak Nullstellensatz). Let \mathbb{F} be an algebraically closed field and let $\mathbf{I} \subseteq \mathbb{F}[x_1, \dots, x_n]$ be an ideal satisfying $\mathbf{V}(\mathbf{I}) = \emptyset$. Then $\mathbf{I} = \mathbb{F}[x_1, \dots, x_n]$.

One might hope that the correspondence between ideals and varieties is one-to-one provided only that one restricts to algebraically closed fields. Unfortunately, this is not the case¹. Indeed, the reason that the map \mathbf{V} fails to be one-to-one is that a power of a polynomial vanishes on the same set as the original polynomial.

Theorem 2.6 (Hilbert's Nullstellensatz). Let \mathbb{F} be an algebraically closed field. If $f_0, f_1, \dots, f_s \in \mathbb{F}[x_1, \dots, x_n]$, then $f_0 \in \mathbf{I}(\mathbf{V}(f_1, \dots, f_s))$ if and only if $f_0^m \in \langle f_1, \dots, f_s \rangle$ for some integer $m \geq 1$.

By definition, radical ideals consist of all polynomials which vanish on some variety V . This together with Theorem 2.6 suggests that there is a one-to-one correspondence between affine varieties and radical ideals.

¹For example $\mathbf{V}(x) = \mathbf{V}(x^2) = \{0\}$ works over any field.

Theorem 2.7 (The Strong Nullstellensatz). *Let \mathbb{F} be an algebraically closed field. If I is an ideal in $\mathbb{F}[x_1, \dots, x_n]$, then $I(V(I)) = \sqrt{I}$.*

The following theorem is a useful tool for finding unions and intersections of varieties. We will use it in the following subsections where we construct ideals corresponding to CSP instances.

Theorem 2.8 ([17]). *If I and J are ideals in $\mathbb{F}[x_1, \dots, x_n]$, then*

- i. $V(I \cap J) = V(I) \cup V(J)$,
- ii. $V(I + J) = V(I) \cap V(J)$.

2.2 The Constraint Satisfaction Problem

We use $[k]$ to denote $\{1, \dots, k\}$. Let D be a finite set, it will often be referred to as a domain. An n -ary relation on D is a set of n -tuples of elements from D ; we use \mathbf{R}_D to denote the set of all finitary relations on D . A constraint language is a subset of \mathbf{R}_D , and may be finite or infinite.

A constraint over a constraint language $\Gamma \subseteq \mathbf{R}_D$ is a pair $\langle \mathbf{s}, R \rangle$ with $\mathbf{s} = (x_1, \dots, x_k)$ a list of variables of length k (not necessarily distinct), called the *constraint scope*, and R a k -ary relation on D , belonging to Γ , called the *constraint relation*. Another common way to denote a constraint $\langle \mathbf{s}, R \rangle$ is by $R(\mathbf{s})$, that is, to treat R as a predicate, and we will use both notations interchangeably. A constraint is satisfied by a mapping $\varphi : \{x_1, \dots, x_k\} \rightarrow D$ if $(\varphi(x_1), \dots, \varphi(x_k)) \in R$.

Definition 2.9 (Constraint Satisfaction Problem). *The constraint satisfaction problem over a constraint language $\Gamma \subseteq \mathbf{R}_A$, denoted $\text{CSP}(\Gamma)$, is defined to be the decision problem with instance $\mathcal{P} = (X, D, C)$, where X is a finite set of variables, D is a domain, and C is a set of constraints over Γ with variables from X . The goal is to decide whether or not there exists a solution, i.e. a mapping $\varphi : X \rightarrow D$ satisfying all of the constraints. We will use $\text{Sol}(\mathcal{P})$ to denote the (possibly empty) set of solutions of \mathcal{P} .*

It is known [10, 42] that for any constraint language Γ (finite or infinite) on a finite set the problem $\text{CSP}(\Gamma)$ is either solvable in polynomial time or is **NP**-complete. We will use this fact to determine the complexity of the Ideal Membership Problem.

2.3 The ideal-CSP correspondence

Here, we explain how to construct an ideal corresponding to a given instance of CSP. Constraints are in essence varieties, see e.g. [26, 41]. Following [28, 41], we shall translate CSPs to polynomial ideals and back. Let $\mathcal{P} = (X, D, C)$ be an instance of $\text{CSP}(\Gamma)$. Without loss of generality, we assume that $D \subset \mathbb{F}^2$. Let $\text{Sol}(\mathcal{P})$ be the (possibly empty) set of all solutions of \mathcal{P} . We wish to map $\text{Sol}(\mathcal{P})$ to an ideal $I(\mathcal{P}) \subseteq \mathbb{F}[X]$ such that $\text{Sol}(\mathcal{P}) = V(I(\mathcal{P}))$.

First we show how to map a constraint to a generating system of an ideal. Consider a constraint $\langle \mathbf{s}, R \rangle$ from C where $\mathbf{s} = (x_{i_1}, \dots, x_{i_k})$ is a k -tuple of variables from X . Every $\mathbf{v} = (v_1, \dots, v_k) \in R$ corresponds to some point $\mathbf{v} \in \mathbb{F}^k$. The ideal $\langle \{\mathbf{v}\} \rangle \stackrel{\text{def}}{=} \langle x_{i_1} - v_1, \dots, x_{i_k} - v_k \rangle$ is an ideal in $\mathbb{F}[\mathbf{s}]$. Moreover, $\langle \{\mathbf{v}\} \rangle$ is radical [17] and $V(\langle \{\mathbf{v}\} \rangle) = \mathbf{v}$. Hence, we can rewrite the relation R as $R = \bigcup_{\mathbf{v} \in R} V(\langle \{\mathbf{v}\} \rangle)$. By Theorem 2.8, we have the following

$$R = \bigcup_{\mathbf{v} \in R} V(\langle \{\mathbf{v}\} \rangle) = V(I(R(\mathbf{s}))), \quad \text{where } I(R(\mathbf{s})) = \bigcap_{\mathbf{v} \in R} \langle \{\mathbf{v}\} \rangle. \quad (1)$$

²In fact, we will mainly assume $\mathbb{F} = \mathbb{R}$ and $D = \{0, 1, \dots, |D| - 1\}$

Note that $I(R(\mathbf{s})) \subseteq \mathbb{F}[\mathbf{s}]$ is zero-dimensional as its variety is finite. Moreover, $I(R(\mathbf{s}))$ is a radical ideal since it is the intersection of radical ideals (see [17], Proposition 16, p.197). Equation (1) means that relation $R(\mathbf{s})$ is a variety of \mathbb{F}^k . Similar to [28], the following is a generating system for $I(R(\mathbf{s}))$:

$$I(R(\mathbf{s})) = \left\langle \prod_{\mathbf{v} \in R} \left(1 - \prod_{j=1}^k \delta_{v_j}(x_{i_j})\right), \prod_{a \in D} (x_{i_1} - a), \dots, \prod_{a \in D} (x_{i_k} - a) \right\rangle, \quad (2)$$

where $\delta_{v_j}(x_{i_j})$ are indicator polynomials, i.e. equal to 1 when $x_{i_j} = v_j$ and 0 when $x_{i_j} \in D \setminus \{v_j\}$; polynomials $\prod_{a \in D} (x_{i_k} - a)$ force variables to take values in D and will be called the *domain polynomials*. Including a domain polynomial for each variable has the advantage that it ensures that the ideals generated by our systems of polynomials are radical (see Lemma 8.19 of [6]).

Recall that X is a set of variables and $\mathbf{s} \subseteq X$. Let $I^X(R(\mathbf{s}))$ be the smallest ideal (with respect to inclusion) of $\mathbb{F}[X]$ containing $I(R(\mathbf{s})) \subseteq \mathbb{F}[\mathbf{s}]$. This is called the $\mathbb{F}[X]$ -*module* of $I(R(\mathbf{s}))$. The set $\text{Sol}(\mathcal{P}) \subset \mathbb{F}^X$ of solutions of $\mathcal{P} = (X, D, C)$ is the intersection of the varieties of the constraints:

$$\text{Sol}(\mathcal{P}) = \bigcap_{\langle \mathbf{s}, R \rangle \in C} \mathbf{V} \left(I(R(\mathbf{s}))^{\mathbb{F}[X]} \right) = \mathbf{V}(I(\mathcal{P})), \quad \text{where} \quad (3)$$

$$I(\mathcal{P}) = \sum_{\langle \mathbf{s}, R \rangle \in C} I^X(R(\mathbf{s})). \quad (4)$$

The equality in (4) follows by Theorem 2.8. Finally, by Proposition 3.22 of [41], the ideal $I(\mathcal{P})$ is radical. Note that in the general version of Hilbert's Nullstellensatz, Theorems 2.6 and 2.7, it is necessary to work in an algebraically closed field. However, in our case it is not needed due to the presence of domain polynomials. Hence, by Theorem 2.5 and Theorem 2.7, we have the following properties.

Theorem 2.10. *Let \mathcal{P} be an instance of the CSP(Γ) and $I(\mathcal{P})$ defined as in (4). Then*

$$\mathbf{V}(I(\mathcal{P})) = \emptyset \Leftrightarrow 1 \in I(\mathcal{P}) \Leftrightarrow I(\mathcal{P}) = \mathbb{F}[X], \quad (\text{Weak Nullstellensatz})$$

$$I(\mathbf{V}(I(\mathcal{P}))) = \sqrt{I(\mathcal{P})}, \quad (\text{Strong Nullstellensatz})$$

$$\sqrt{I(\mathcal{P})} = I(\mathcal{P}). \quad (\text{Radical Ideal})$$

2.4 The Ideal Membership Problem

In the general Ideal Membership Problem we are given an ideal $I \subseteq \mathbb{F}[x_1, \dots, x_n]$, usually by some finite generating set, and a polynomial f_0 . The question then is to decide whether or not $f_0 \in I$. If I is given through a CSP instance, we can be more specific.

Definition 2.11. *The IDEAL MEMBERSHIP PROBLEM associated with a constraint language Γ over a set D is the problem $\text{IMP}(\Gamma)$ in which the input is a pair (f_0, \mathcal{P}) where $\mathcal{P} = (X, D, C)$ is a CSP(Γ) instance and f_0 is a polynomial from $\mathbb{F}[X]$. The goal is to decide whether f_0 lies in the combinatorial ideal $I(\mathcal{P})$. We use $\text{IMP}_d(\Gamma)$ to denote $\text{IMP}(\Gamma)$ when the input polynomial f_0 has degree at most d .*

As $I(\mathcal{P})$ is radical, by the Strong Nullstellensatz an equivalent way to solve the membership problem $f_0 \in I(\mathcal{P})$ is to answer the following question:

Does there exist an $\mathbf{a} \in \mathbf{V}(I(\mathcal{P}))$ such that $f_0(\mathbf{a}) \neq 0$?

In the **yes** case, such an \mathbf{a} exists if and only if $f_0 \notin \mathbf{I}(\mathbf{V}(I(\mathcal{P})))$ and therefore f_0 is **not** in the ideal $I(\mathcal{P})$. This observation also implies

Lemma 2.12. *For any constraint language Γ the problem $\text{IMP}_0(\Gamma)$ is equivalent to $\text{not-CSP}(\Gamma)$.*

As was observed in the Introduction, $\text{IMP}(\Gamma)$ belongs to **coNP** for any Γ over a finite domain. We say that $\text{IMP}(\Gamma)$ is *tractable* if it can be solved in polynomial time. We say that $\text{IMP}(\Gamma)$ is *d-tractable* if $\text{IMP}_d(\Gamma)$ can be solved in polynomial time for every d .

3 Expanding the constraint language

In this section we discuss constructions on relations that allow us to reduce one IMP with a fixed constraint language to another. First we show that adding so-called *constant* relations does not change the complexity of the problem. Second, we will consider languages on the same domain, and prove that *primitive positive (pp-, for short) definitions* between constraint languages provides a reduction between the corresponding IMPs. Third, we will turn our attention to the case where two languages are defined on different domains. In this case, we study a stronger notion called *primitive positive interpretability*. We prove that if a language Γ pp-interprets a language Δ , then $\text{IMP}(\Delta)$ is reducible to $\text{IMP}(\Gamma)$. Finally, we discuss how ideal membership proofs can (or cannot) be recovered under these reductions.

3.1 Constant relations and the search problem

We start with expansion of a constraint language Γ on a set D by constant relations. A *constant* relation R_a , $a \in D$, is the unary relation, that is, a subset of D , that contains just one element a . Using it in a CSP is equivalent to *pinning* a variable to a fixed value a . Expansion by constant relations is very important for CSPs. It preserves the complexity of the decision version of the problem when Γ is a *core*, see, [12]), and it preserves the complexity of the counting version of the problem for any Γ , see [11]. For $A \subseteq D$ let Γ^A denote $\Gamma \cup \{R_a \mid a \in A\}$. We also call constraints of the form $\langle x, R_a \rangle$ *pinning* constraints.

Proposition 3.1. *Let Γ be a constraint language on a set D and $A \subseteq D$. Then $\text{IMP}_d(\Gamma^A)$ is polynomial time reducible to $\text{IMP}_{d+|A|(|D|-1)}(\Gamma)$.*

Proof. Let $\mathcal{P} = (X, D, C)$ be an instance of $\text{CSP}(\Gamma^A)$ and let $I(\mathcal{P})$ be the ideal corresponding to \mathcal{P} . Suppose (f_0, \mathcal{P}) is an instance of $\text{IMP}_d(\Gamma^A)$ where we want to decide if $f_0 \in I(\mathcal{P})$. First, we perform some preprocessing of (f_0, \mathcal{P}) . Note that if \mathcal{P} contains constraints $\langle x, R_a \rangle, \langle x, R_b \rangle, a \neq b$, then \mathcal{P} has no solution and so $1 \in I(\mathcal{P})$ implying $f_0 \in I(\mathcal{P})$. Let X_a denote the set of variables x , for which there is a constraint $\langle x, R_a \rangle \in C$. Introduce new variable x_a for each $a \in A$ and replace every $x \in X_a$, with x_a in both f_0 and \mathcal{P} . In particular, let

$$X' = \left(X \setminus \bigcup_{a \in A} X_a \right) \cup \{x_a \mid a \in A\}.$$

The resulting instance (f'_0, \mathcal{P}') has the following properties:

- The solutions of \mathcal{P} and \mathcal{P}' are in one-to-one correspondence, since for every solution φ of \mathcal{P} we have $\varphi(x) = a$ for each $x \in X_a$, and so the mapping $\varphi' : X' \rightarrow D$ such that $\varphi(x_a) = a$ for $a \in A$ and $\varphi'(x) = \varphi(x)$ otherwise is a solution of \mathcal{P}' and vice versa.
- $f_0(\varphi) = 0$ if and only if $f'_0(\varphi') = 0$.

Now let $\mathcal{P}^* = (X', D, C^*)$ be an instance of $\text{CSP}(\Gamma)$ where C^* consists of all constraint from C' except the ones of the form $\langle x, R_a \rangle$, $a \in A$. We define a new polynomial f_0^* as follows.

$$f_0^* = \left(\prod_{a \in A} \prod_{b \in D \setminus \{a\}} (x_a - b) \right) \cdot f'_0.$$

Observe that, for any $a \in A$ and $\varphi^* : X' \rightarrow D$, if $\varphi^*(x_a) \neq a$ then $f_0^*(\varphi^*) = 0$. Suppose $\varphi' \in \mathbf{V}(I(\mathcal{P}'))$. As φ' satisfies all the pinning constraints in C' , we have $f'_0(\varphi') \neq 0$ if and only if $f_0^*(\varphi') \neq 0$. Moreover, suppose $\varphi^* \in \mathbf{V}(I(\mathcal{P}^*))$ and $f_0^*(\varphi^*) \neq 0$. This implies that

1. $\prod_{a \in A} \prod_{b \in D \setminus \{a\}} (\varphi^*(x_a) - b) \neq 0$, which means φ^* satisfies all the pinning constraints in C , and hence $\varphi^* \in \mathbf{V}(I(\mathcal{P}'))$, and
2. $f'_0(\varphi') \neq 0$.

Combining the preprocessing step with the second one there exists $\varphi \in \mathbf{V}(I(\mathcal{P}))$ such that $f_0(\varphi) \neq 0$ if and only if there exists $\varphi^* \in \mathbf{V}(I(\mathcal{P}^*))$ such that $f_0^*(\varphi^*) \neq 0$. This completes the proof of the proposition. \square

Proposition 3.1 together with the fact that $\text{IMP}(\Gamma)$ is a subproblem of $\text{IMP}(\Gamma^A)$ implies a close connection between the complexity of $\text{IMP}(\Gamma)$ and $\text{IMP}(\Gamma^A)$.

Corollary 3.2. *For any constraint language Γ on D and any $A \subseteq D$, the problem $\text{IMP}(\Gamma^A)$ is tractable (d -tractable) if and only if $\text{IMP}(\Gamma)$ is tractable (d -tractable), and $\text{IMP}(\Gamma^A)$ is **coNP**-complete if and only if $\text{IMP}(\Gamma)$ is **coNP**-complete.*

Recall that $\Gamma^* = \Gamma^D$. Then

Theorem 3.3. *For any Γ over D the problem $\text{IMP}(\Gamma^*)$ is polynomial time reducible to $\text{IMP}(\Gamma)$, and for any d the problem $\text{IMP}_d(\Gamma^*)$ is polynomial time reducible to $\text{IMP}_{d+|D|(|D|-1)}(\Gamma)$.*

However, Proposition 3.1 leaves some room for possible complexity of $\text{IMP}_d(\Gamma)$ for small d , less than $|D|(|D|-1)$.

Example 3.4. Fix D , $\ell \leq |D|$. Let NEQ_s , $s \leq |D|$ denote the s -ary disequality relation on D given by

$$\text{NEQ}_s = \{(a_1, \dots, a_s) \mid |\{a_1, \dots, a_s\}| = s\}.$$

In particular, $\text{CSP}(\text{NEQ}_2)$ is equivalent to $|D|$ -Coloring. Now let a $(\ell + 2)$ -ary relation R be defined as follows

$$R = (\text{NEQ}_2 \times \text{NEQ}_\ell) \cup \{(a_1, \dots, a_{2+\ell}) \mid |\{a_1, \dots, a_{2+\ell}\}| < \ell\},$$

and let $\Gamma = \{R\}$. It is easy to see, $\text{CSP}(\Gamma)$ is polynomial time, as assigning the same value to all variables always provides a solution. As we observed in Lemma 2.12 this implies that IMP_0 is also easy. Actually, f_0 of degree 0 never belongs to the ideal except $f_0 = 0$.

It can also be shown that for any $A \subseteq D$ with $|A| < \ell$ assigning a constant $a \in A$ to all variables except those bound by the pinning constraints is also a solution of $\text{CSP}(\Gamma^A)$. Therefore, $\text{IMP}_0(\Gamma^A)$ is easy for any such set. On the other hand, if $|A| = \ell$, say, $A = \{a_1, \dots, a_\ell\}$, then $\text{CSP}(\Gamma^A)$ can simulate $|D|$ -Coloring by using $R(x, y, a_1, \dots, a_\ell)$. (This will be made more precise in Section 3.2.) Therefore, $\text{IMP}_0(\Gamma^A)$ is **coNP**-complete in this case. Clearly that playing with the exact definition of R one can construct a language Γ such that $\text{IMP}_0(\Gamma^A)$ becomes **coNP**-complete for any specified collection of subsets A while remains easy for the rest of the subsets.

In the case of a 2-element D we can show a more definitive result.

Proposition 3.5 (see also [28]). *Let Γ be a constraint language on the set $\{0, 1\}$. Then*

- (1) $\text{IMP}_d(\Gamma^*)$ is polynomial time equivalent to $\text{IMP}_{d+2}(\Gamma^*)$.
- (2) $\text{IMP}_0(\Gamma)$ is polynomial time [**coNP**-complete] if and only if $\text{CSP}(\Gamma)$ is polynomial time [**coNP**-complete].
- (3) If $\text{CSP}(\Gamma^{\{0\}})$ or $\text{CSP}(\Gamma^{\{1\}})$ is **NP**-complete then $\text{IMP}_1(\Gamma)$ is **coNP**-complete.

Items (1),(3) follow from Proposition 3.1 and item (2) follows from Lemma 2.12. Moreover, replacing the relation NEQ_2 in Example 3.4 with the NOT-ALL-EQUAL relation, one can construct constraint languages Γ such that the borderline between easiness and hardness in the sequence $\text{IMP}_0(\Gamma), \text{IMP}_1(\Gamma), \text{IMP}_2(\Gamma)$ lies in any desirable place.

Proposition 3.1 also provide a connection between the decision version of the IMP and its search version. Since $\text{IMP}(\Gamma)$ is in **coNP**, here by the search IMP we understand the following problem. Let (f_0, \mathcal{P}) be an instance of $\text{IMP}(\Gamma)$ such that $f_0 \notin I(\mathcal{P})$, the problem is to find an assignment $\varphi \in \mathbf{V}(I(\mathcal{P}))$ such that $f_0(\varphi) \neq 0$.

Corollary 3.6. *A decision problem $\text{IMP}(\Gamma)$ is tractable [d -tractable] if and only if the corresponding search problem is tractable [d -tractable].*

Proof. One direction is trivial as the tractability of the search problem implies the tractability of the corresponding decision problem.

For the converse, let Γ be a constraint language over a finite set D such that $\text{IMP}(\Gamma)$ is (d -) tractable. Consider (f_0, \mathcal{P}) , an instance of $\text{IMP}(\Gamma)$, where $\mathcal{P} = (X, D, C)$ is an instance of $\text{CSP}(\Gamma)$. By the choice of Γ , we can decide in polynomial time whether there exists φ such that $\varphi \in \mathbf{V}(I(\mathcal{P}))$ but $f_0(\varphi) \neq 0$. Suppose such φ exists and hence $f_0 \notin I(\mathcal{P})$. Then for each $x \in X$ there must be some $a \in D$, for which in the following instance (f'_0, \mathcal{P}') of the IMP we have $f'_0 \notin I(\mathcal{P}')$:

1. define f'_0 to be the polynomial obtained from f_0 by substituting a for x .
2. define \mathcal{P}' with $\mathcal{P}' = (X, D, C' = C \cup \{\langle x, \{a\} \rangle\})$.

Checking whether $f'_0 \in I(\mathcal{P}')$ is an instance of $\text{IMP}(\Gamma^*)$ and therefore can be done in polynomial time. Hence, by considering each possible value $a \in D$ we can find a value for x that is a part of $\varphi \in \mathbf{V}(I(\mathcal{P}))$ such that $f_0(\varphi) \neq 0$. Repeating the process for each variable in turn we can find a required φ . The algorithm requires solving at most $|X| \cdot |D|$ instances of $\text{IMP}(\Gamma^*)$, each of which can be solved in polynomial time. \square

3.2 Primitive positive definability

One of the most useful reductions between CSPs is by means of primitive-positive definitions.

Definition 3.7 (pp-definability). *Let Γ, Δ be constraint languages on the same set D . We say that Γ pp-defines Δ (or Δ is pp-definable from Γ) if for each relation (predicate) $R \subseteq D^k$ in Δ there exists a first order formula L over variables $\{x_1, \dots, x_m, x_{m+1}, \dots, x_{m+k}\}$ that uses predicates from Γ , equality relations, and conjunctions such that*

$$R(x_{m+1}, \dots, x_{m+k}) = \exists x_1 \dots \exists x_m L$$

Such an expression is often called a primitive positive (pp-) formula.

Mastrolilli showed that there is an analogue of existential quantification on the IMP side.

Definition 3.8. *Given $I = \langle f_1, \dots, f_s \rangle \subseteq \mathbb{F}[X]$, for $Y \subseteq X$, the Y -elimination ideal $I_{X \setminus Y}$ is the ideal of $\mathbb{F}[X \setminus Y]$ defined by*

$$I_{X \setminus Y} = I \cap \mathbb{F}[X \setminus Y]$$

In other words, $I_{X \setminus Y}$ consists of all consequences of $f_1 = \dots = f_s = 0$ that do not depend on variables from Y .

Theorem 3.9 ([28]). *Let $\mathcal{P} = (X, D, C)$ be an instance of the $\text{CSP}(\Gamma)$, and let $I(\mathcal{P})$ be the corresponding ideal. For any $Y \subseteq X$ let I_Y be the $(X \setminus Y)$ -elimination ideal. Then, for any partial solution $\varphi_Y \in \mathbf{V}(I_Y)$ there exists an extension $\psi : X \setminus Y$ such that $(\varphi, \psi) \in \mathbf{V}(I(\mathcal{P}))$.*

Let Γ, Δ be constraint languages on the same domain D such that Γ pp-defines Δ . This means that for every relation R from Δ there is a pp-definition in Γ

$$R(x_{m_R+1}, \dots, x_{m_R+k_R}) = \exists x_1 \dots \exists x_{m_R} L_R.$$

Suppose $\mathcal{P}_\Delta = (X, D, C)$ is an instance of $\text{CSP}(\Delta)$. This instance can be converted into an instance $\mathcal{P}_\Gamma = (X', D, C')$ of $\text{CSP}(\Gamma)$, see Theorem 2.16 in [12], in such a way that $X \subseteq X'$ and the instance \mathcal{P}_Δ has a solution if and only if \mathcal{P}_Γ does. Moreover, it can be shown that $\mathcal{P}_\Gamma, \mathcal{P}_\Delta$ satisfy the following condition.

The Extension Condition. Every solution of \mathcal{P}_Δ can be extended to a solution of \mathcal{P}_Γ , and, vice versa, the restriction of every solution of \mathcal{P}_Γ onto variables from X is a solution of \mathcal{P}_Δ .

As usual, let $I(\mathcal{P}_\Delta)$ be the ideal of $\mathbb{F}[X]$ corresponding to \mathcal{P}_Δ and $I(\mathcal{P}_\Gamma)$ the ideal of $\mathbb{F}[X']$ corresponding to \mathcal{P}_Γ . We would like to relate the set of solutions of \mathcal{P}_Δ to the variety of the $X' \setminus X$ -elimination ideal of $I(\mathcal{P}_\Gamma)$. The next lemma states that the variety of the $X' \setminus X$ -elimination ideal of $I(\mathcal{P}_\Gamma)$ is equal to the the variety of $I(\mathcal{P}_\Delta)$.

Lemma 3.10 ([28], Lemma 6.1, paraphrased). *Let $I_X = I(\mathcal{P}_\Gamma) \cap \mathbb{F}[X]$ be the $X' \setminus X$ -elimination ideal of $I(\mathcal{P}_\Gamma)$. Then*

$$\mathbf{V}(I(\mathcal{P}_\Delta)) = \mathbf{V}(I_X).$$

We can now prove a reduction for pp-definable constraint languages.

Theorem 3.11. *If Γ pp-defines Δ , then $\text{IMP}(\Delta)$ is polynomial time reducible to $\text{IMP}(\Gamma)$.*

Proof. Let $(f_0, \mathcal{P}_\Delta)$, $\mathcal{P}_\Delta = (X, D, C_\Delta)$, be an instance of $\text{IMP}(\Delta)$ where $X = \{x_{m+1}, \dots, x_{m+k}\}$, $f_0 \in \mathbb{F}[x_{m+1}, \dots, x_{m+k}]$, $k = |X|$, and m will be defined later, and $I(\mathcal{P}_\Delta) \subseteq \mathbb{F}[x_{m+1}, \dots, x_{m+k}]$. From this we construct an instance $(f'_0, \mathcal{P}_\Gamma)$ of $\text{IMP}(\Gamma)$ where $f'_0 \in \mathbb{F}[x_1, \dots, x_{m+k}]$ and $I(\mathcal{P}_\Gamma) \subseteq \mathbb{F}[x_1, \dots, x_{m+k}]$ such that $f_0 \in I(\mathcal{P}_\Delta)$ if and only if $f'_0 \in I(\mathcal{P}_\Gamma)$.

Using pp-definitions of relations from Δ we convert the instance \mathcal{P}_Δ into an instance $\mathcal{P}_\Gamma = (\{x_1, \dots, x_{m+k}\}, D, C_\Gamma)$ of $\text{CSP}(\Gamma)$ such that every solution of $\mathcal{P}_\Delta, \mathcal{P}_\Gamma$ satisfy the Extension Condition above. Such an instance \mathcal{P}_Γ can be constructed in polynomial time as follows.

By the assumption each $S \in \Delta$, say, t_S -ary, is pp-definable in Γ by a pp-formula involving relations from Γ and the equality relation, $=_D$. Thus,

$$S(y_{q_S+1}, \dots, y_{q_S+t_S}) = \exists y_1, \dots, y_{q_S} (R_1(w_1^1, \dots, w_{l_1}^1) \wedge \dots \wedge R_r(w_1^r, \dots, w_{l_r}^r)),$$

where $w_1^1, \dots, w_{l_1}^1, \dots, w_1^k, \dots, w_{l_k}^k \in \{y_1, \dots, y_{m_S+t_S}\}$ and $R_1, \dots, R_r \subseteq \Gamma \cup \{=_D\}$.

Now, for every constraint $B = \langle \mathbf{s}, S \rangle \in C_\Delta$, where $\mathbf{s} = (x_{i_1}, \dots, x_{i_t})$ create a fresh copy of $\{y_1, \dots, y_{q_S}\}$ denoted by Y_B , and add the following constraints to C_Γ

$$\langle (w_1^1, \dots, w_{l_1}^1), R_1 \rangle, \dots, \langle (w_1^r, \dots, w_{l_r}^r), R_r \rangle.$$

We then set $m = \sum_{B \in C} |Y_B|$ and assume that $\cup_{B \in C} Y_B = \{x_1, \dots, x_m\}$. Note that the problem instance obtained by this procedure belongs to $\text{CSP}(\Gamma \cup \{=_D\})$. All constraints of the form $\langle (x_i, x_j), =_D \rangle$ can be eliminated by replacing all occurrences of the variable x_i with x_j . Moreover, it can be checked (see also Theorem 2.16 in [12]) that $\mathcal{P}_\Delta, \mathcal{P}_\Gamma$ satisfy the Extension Condition.

Let $I(\mathcal{P}_\Gamma) \subseteq \mathbb{F}[x_1, \dots, x_{m+k}]$ be the ideal corresponding to \mathcal{P}_Γ and set $f'_0 = f_0$. Since $f_0 \in \mathbb{F}[x_{m+1}, \dots, x_{m+k}]$ we also have $f_0 \in \mathbb{F}[x_1, \dots, x_{m+k}]$. Hence, $(f_0, \mathcal{P}_\Gamma)$ is an instance of $\text{IMP}(\Gamma)$. We prove that $f_0 \in I(\mathcal{P}_\Delta)$ if and only if $f_0 \in I(\mathcal{P}_\Gamma)$.

Suppose $f_0 \notin I(\mathcal{P}_\Delta)$, this means there exists $\varphi \in \mathbf{V}(I(\mathcal{P}_\Delta))$ such that $f(\varphi) \neq 0$. By Theorem 3.9, φ can be extended to a point $\varphi' \in \mathbf{V}(I(\mathcal{P}_\Gamma))$. This in turn implies that $f_0 \notin I(\mathcal{P}_\Gamma)$. Conversely, suppose $f_0 \notin I(\mathcal{P}_\Gamma)$. Hence, there exists $\varphi' \in \mathbf{V}(I(\mathcal{P}_\Gamma))$ such that $f_0(\varphi') \neq 0$. Projection of φ' to its last k coordinates gives a point $\varphi \in \mathbf{V}(I_X)$. By Lemma 3.10, $\varphi \in \mathbf{V}(I(\mathcal{P}_\Delta))$ which implies $f_0 \notin I(\mathcal{P}_\Delta)$. \square

Remark 3.12. The smallest set of all relations pp-defined from a constraint language $\Gamma \subseteq \mathbf{R}_A$ is called the relational clone of Γ , denoted by $\langle \Gamma \rangle$. Hence, as a corollary to Theorem 3.11, for a finite set of relations Γ , $\text{IMP}(\Gamma)$ is tractable [d -tractable] if and only if $\text{IMP}(\Delta)$ is tractable [d -tractable] for any finite $\Delta \subseteq \langle \Gamma \rangle$. Similarly, $\text{IMP}(\Gamma)$ is coNP-complete if and only if $\text{IMP}(\Delta)$ is coNP-complete for some finite $\Delta \subseteq \langle \Gamma \rangle$.

3.3 Primitive positive interpretability

Pp-definability is a useful technique that tells us what additional relations can be added to a constraint language without changing the complexity of the corresponding problem class, and provides a tool for comparing different languages on the same domain. Next, we discuss a more powerful tool that can be used to compare the complexity of the IMP for languages over different domains.

Definition 3.13 (pp-interpretability). Let Γ, Δ be constraint languages over finite domains D, E , respectively, and Δ is finite. We say that Γ pp-interprets Δ if there exists a natural number ℓ , a set $F \subseteq D^\ell$, and an onto mapping $\pi : F \rightarrow E$ such that Γ pp-defines the following relations

1. the relation F ,
2. the π -preimage of the equality relation on E , and
3. the π -preimage of every relation in Δ ,

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

$$\pi^{-1}(S)(x_{11}, \dots, x_{1k}, x_{21}, \dots, x_{2k}, \dots, x_{\ell 1}, \dots, x_{\ell k}) \quad \text{is true}$$

if and only if

$$S(\pi(x_{11}, \dots, x_{\ell 1}), \dots, \pi(x_{1k}, \dots, x_{\ell k})) \quad \text{is true.}$$

Example 3.14. Suppose $D = \{0, 1\}$ and $E = \{0, 1, 2\}$ and define relations $R_D = \{(0, 0), (0, 1), (1, 1)\}$ and $R_E = \{(0, 0), (0, 1), (0, 2), (1, 1), (1, 2), (2, 2)\}$. Note that relations R_D, R_E are orders $0 \leq 1$ and $0 \leq 1 \leq 2$ on D, E , respectively. Set $\Gamma = \{R_D\}$ and $\Delta = \{R_E\}$.

Let $n = 2$ and define $F = \{(0, 0), (0, 1), (1, 1)\} \subseteq D^2$. The language Γ pp-defines F i.e. $F = \{(x, y) \mid x \leq y \text{ and } x, y \in \{0, 1\}\}$. Now define mapping $\pi : F \rightarrow E$ as follows $\pi((0, 0)) = 0, \pi((0, 1)) = 1, \pi((1, 1)) = 2$. The π -preimage of the relation R_E is the relation

$$R_F = \{(0, 0, 0, 0), (0, 0, 0, 1), (0, 0, 1, 1), (0, 1, 0, 1), (0, 1, 1, 1), (1, 1, 1, 1)\}.$$

The language Γ pp-defines $\Gamma' = \{R_F\}$ through the following pp-formula

$$R_F = \{(x_1, x_2, y_1, y_2) \mid (x_1 \leq y_1) \wedge (x_2 \leq y_2) \wedge (x_1 \leq x_2) \wedge (y_1 \leq y_2), \text{ and } (x_1, x_2, y_1, y_2) \in \{0, 1\}^4\}.$$

Consider instance $(\{x, y, z\}, E, C)$ of $\text{CSP}(\Delta)$ where the set of constraint is $C = \{ \langle (x, y), R_E \rangle, \langle (y, z), R_E \rangle \}$. This basically means the requirements $(x \leq y) \wedge (y \leq z)$. This instance is equivalent to the following instance of $\text{CSP}(\Gamma')$:

$$\langle (x_1, x_2, y_1, y_2), R_F \rangle \wedge \langle (y_1, y_2, z_1, z_2), R_F \rangle \quad (5)$$

As was pointed out, Γ pp-defines F as well as R_F . Hence, we define $(\{x_1, x_2, y_1, y_2, z_1, z_2\}, D, C')$, an instance of $\text{CSP}(\Gamma)$, with the constraints

$$\langle (x_1, x_2), R_D \rangle \wedge \langle (y_1, y_2), R_D \rangle \wedge \langle (z_1, z_2), R_D \rangle \quad (6)$$

$$\wedge \langle (x_1, y_1), R_D \rangle \wedge \langle (x_2, y_2), R_D \rangle \quad (7)$$

$$\wedge \langle (y_1, z_1), R_D \rangle \wedge \langle (y_2, z_2), R_D \rangle \quad (8)$$

Note that (6) is a pp-definition of relation F and forces $(x_1 x_2), (y_1 y_2), (z_1 z_2) \in F$. Moreover, equations (7) and (8) are equivalent to

$$(x_1 \leq y_1) \wedge (x_2 \leq y_2) \wedge (y_1 \leq z_1) \wedge (y_2 \leq z_2) \quad (9)$$

Applying the mapping π , every solution of the instance $(\{x_1, x_2, y_1, y_2, z_1, z_2\}, D, C')$ can be transformed to a solution of instance $(\{x, y, z\}, E, C)$ and back.

One can interpolate mapping π in Definition 3.13 by a polynomial of low degree. It is a known fact that given $N + 1$ distinct $\mathbf{x}_0, \dots, \mathbf{x}_N \in \mathbb{R}^\ell$ and corresponding values y_0, \dots, y_N , there exists a polynomial p of degree at most ℓN that interpolates the data i.e. $p(\mathbf{x}_j) = y_j$ for each $j \in \{0, \dots, N\}$ (such a polynomial can be obtained by a straightforward generalization of the Lagrange interpolating polynomial, see, e.g., [34]). Hence, we can interpolate the mapping π by a polynomial of total degree at most $\ell|E|$.

Theorem 3.15. *Let Γ, Δ be constraint languages on sets D, E , respectively, and let Γ pp-interprets Δ . Then $\text{IMP}(\Delta)$ [$\text{IMP}_d(\Delta)$] is polynomial time reducible to $\text{IMP}(\Gamma)$ [respectively, $\text{IMP}_{\ell|E|}(\Gamma)$].*

Proof. Let $(f_0, \mathcal{P}_\Delta)$ be an instance of $\text{IMP}(\Delta)$ where $f_0 \in \mathbb{F}[x_1, \dots, x_n]$, $\mathcal{P}_\Delta = (\{x_1, \dots, x_n\}, E, C_\Delta)$, an instance of $\text{CSP}(\Delta)$, and $I(\mathcal{P}_\Delta) \subseteq \mathbb{F}[x_1, \dots, x_n]$.

The properties of the mapping π from Definition 3.13 allow us to rewrite an instance of $\text{CSP}(\Delta)$ to an instance of $\text{CSP}(\Gamma')$ over the constraint language Γ' . Recall that, by Definition 3.13, Γ' contains all the ℓk -ary relations $\pi^{-1}(S)$ on D where $S \in \Delta$ is k -ary relation.

Note that Γ' is pp-definable from Γ . By Theorem 3.11, $\text{IMP}(\Gamma')$ is reducible to $\text{IMP}(\Gamma)$. It remains to show $\text{IMP}(\Delta)$ is reducible to $\text{IMP}(\Gamma')$. To do so, from instance $(f_0, \mathcal{P}_\Delta)$ of $\text{IMP}(\Delta)$ we construct an instance $(f'_0, \mathcal{P}_{\Gamma'})$ of $\text{IMP}(\Gamma')$ such that $f_0 \in I(\mathcal{P}_\Delta)$ if and only if $f'_0 \in I(\mathcal{P}_{\Gamma'})$.

Let p be a polynomial of total degree at most $\ell|E|$ that interpolates mapping π . For every $f_0 \in \mathbb{F}[x_1, \dots, x_n]$, let $f'_0 \in \mathbb{F}[x_{11}, \dots, x_{\ell 1}, \dots, x_{1n}, \dots, x_{\ell n}]$ be the polynomial that is obtained from f_0 by replacing each indeterminate x_i with $p(x_{1i}, \dots, x_{\ell i})$. Clearly, for any assignment $\varphi : \{x_1, \dots, x_n\} \rightarrow E$, $f_0(\varphi) = 0$ if and only if $f'_0(\psi) = 0$ for every $\psi : \{x_{11}, \dots, x_{\ell n}\} \rightarrow D$ such that

$$\varphi(x_i) = \pi(\psi(x_{1i}), \dots, \psi(x_{\ell i}))$$

for every $i \leq n$. Moreover, for any such φ, ψ it holds $\varphi \in \mathbf{V}(I(\mathcal{P}_\Delta))$ if and only if $\psi \in \mathbf{V}(I(\mathcal{P}_{\Gamma'}))$. This yields that

$$(\exists \varphi \in \mathbf{V}(I(\mathcal{P}_\Delta)) \wedge f_0(\varphi) \neq 0) \iff (\exists \psi \in \mathbf{V}(I(\mathcal{P}_{\Gamma'})) \wedge f'_0(\psi) \neq 0)$$

This completes the proof of the theorem. □

3.4 Recovering proofs, SOS, and proofs of nonnegativity

Proposition 3.1 and Theorems 3.11, 3.15 only allow us to reduce $\text{IMP}(\Delta)$ to $\text{IMP}(\Gamma)$ for some constraint languages as decision problems telling nothing about finding an ideal membership proof for an instance of $\text{IMP}(\Delta)$, even when such a proof can be found for $\text{IMP}(\Gamma)$. In this section we discuss issues around this phenomenon, prove some partial results, and show that for some applications the reductions in Proposition 3.1 and Theorems 3.11, 3.15 suffice.

Recall that an ideal membership proof of low bit complexity is one of bounded degree and with polynomially sized coefficients. We start with an observation that a proof for an $\text{IMP}(\Delta)$ instance can be transformed to a proof for an $\text{IMP}(\Gamma)$ instance with only a modest increase in degree and sizes of coefficients.

Proposition 3.16. *Let Γ be a constraint language.*

- (1) *For any $\text{IMP}(\Gamma^*)$ instance (f_0, \mathcal{P}) , where \mathcal{P} has a solution, if there is a low bit complexity [low degree] ideal membership proof for $f_0 \in I(\mathcal{P})$, then there is a low bit complexity [low degree] ideal membership proof of $f_0^* \in I(\mathcal{P}^*)$, where \mathcal{P}^* is the instance constructed in Proposition 3.1.*
- (2) *Let Δ be pp-definable in Γ . Then for any $\text{IMP}(\Delta)$ instance (f_0, \mathcal{P}) , if there is a low bit complexity [low degree] ideal membership proof for $f_0 \in I(\mathcal{P})$, then there is a low bit complexity [low degree] ideal membership proof of $f_0 \in I(\mathcal{P}_\Gamma)$, where \mathcal{P}_Γ is the instance constructed in Theorem 3.11.*
- (3) *Let Δ be pp-interpretable in Γ . Then for any $\text{IMP}(\Delta)$ instance (f_0, \mathcal{P}) , if there is a low bit complexity [low degree] ideal membership proof for $f_0 \in I(\mathcal{P})$, then there is a low bit complexity [low degree] ideal membership proof of $f_0 \in I(\mathcal{P}_\Gamma)$, where \mathcal{P}_Γ is the instance constructed in Theorem 3.15.*

Proof. (1) Given (f_0, \mathcal{P}) , an instance of $\text{IMP}(\Gamma^*)$, $\mathcal{P} = (X, D, C)$, suppose that

$$f_0 = g_1 f_1 + \cdots + g_k f_k,$$

where $f_1, \dots, f_k \in I(\mathcal{P})$ are polynomials that are encodings of relations R_1, \dots, R_k from Γ^* . Also suppose that $R_1, \dots, R_\ell \in \Gamma$ and R_i is a constant relation for $i > \ell$. In other words, $f_i = x - a_i$ for some $x \in X$ and $a_i \in D$. Then let f'_0 and $f'_i, g'_i, i \in [k]$, denote the polynomials obtained by substituting x_a for any x such that \mathcal{P} contains a constraint $\langle (x), R_a \rangle$, or, equivalently, one of $f_{\ell+1}, \dots, f_k$ equals $x - a$. We then obviously have

$$f'_0 = g'_1 f'_1 + \cdots + g'_k f'_k,$$

where every $f'_j, \ell < j \leq k$, is of the form $x_{a_j} - a_j$, and every $f'_j, j \in [\ell]$ is the encoding of a constraint from Γ .

Now, by the construction in Proposition 3.1,

$$\begin{aligned} f_0^* &= \left(\prod_{a \in D} \prod_{b \in D \setminus \{a\}} (x_a - b) \right) \cdot f'_0 \\ &= \sum_{i=1}^k \left(\prod_{a \in D} \prod_{b \in D \setminus \{a\}} (x_a - b) \right) \cdot g'_i f'_i \\ &= \sum_{i=1}^{\ell} g_i^* f'_i + \sum_{i=\ell+1}^k \left(\prod_{a \in D \setminus \{a_i\}} \prod_{b \in D \setminus \{a\}} (x_a - b) \right) g'_i \prod_{c \in D} (x_{a_i} - c) \\ &= \sum_{i=1}^{\ell} g_i^* f'_i + \sum_{i=\ell+1}^k g_i^* f_i^*, \end{aligned}$$

where

$$\begin{aligned} g_i^* &= \left(\prod_{a \in D} \prod_{b \in D \setminus \{a\}} (x_a - b) \right) \cdot g'_i \quad \text{for } i \in [\ell], \quad \text{and} \\ g_i^* &= \left(\prod_{a \in D \setminus \{a_i\}} \prod_{b \in D \setminus \{a\}} (x_a - b) \right) g'_i, \quad f_i^* = \prod_{c \in D} (x_{a_i} - c) \quad \text{for } \ell < i \leq k. \end{aligned}$$

Observing that $\prod_{c \in D} (x_{a_j} - c)$ is a domain polynomial we conclude the result.

(2) First, we convert pp-definitions of relations from Δ into ideal membership proofs. Suppose that for $R \in \Delta$ we have

$$R(x_{m_R+1}, \dots, x_{m_R+k_R}) = \exists x_1 \dots \exists x_{m_R} L_R,$$

where L_R is a conjunction of relations from Γ (since equality relations can be replaced with identifying variables, we assume they are a part of Γ). Consider L as an instance of $\text{CSP}(\Gamma)$ and let $I(L)$ be the corresponding ideal in $\mathbb{F}[x_1, \dots, x_{m_R+k_R}]$. Let also $f_R \in \mathbb{F}[x_{m_R+1}, \dots, x_{m_R+k_R}]$ be the encoding of R . Then f_R vanishes in every point from $\mathbf{V}(I(L))$ by Theorem 3.9, implying $f_R \in I(L)$ and there is an ideal membership proof of this fact:

$$f_R = h_1 f_1 + \cdots + h_\ell f_\ell, \tag{10}$$

where f_1, \dots, f_ℓ are encodings of relations from Γ .

Now, if for an $\text{IMP}(\Delta)$ instance (f_0, \mathcal{P}) there is a proof $f_0 = g_1 f_{R_1} + \dots + g_s f_{R_s}$ where f_{R_i} is the encoding of $R_i \in \Gamma$, $i \in [s]$, then a proof that $f_0 \in I(\mathcal{P}_\Gamma)$ can be obtained by plugging in representation of the form (10). Clearly, it only increases the degree and the size of the coefficients of the proof by a constant factor, as the proof (10) is constant degree, because Γ, Δ are fixed.

(3) We use the notation from the proof of Theorem 3.15. Let π be the mapping from F to E and p a polynomial interpolating π . If (f_0, \mathcal{P}) is an instance of $\text{IMP}(\Delta)$ and $f_0 = g_1 f_1 + \dots + g_k f_k$ an ideal membership proof that $f_0 \in I(\mathcal{P})$. In order to obtain an ideal membership proof that $f'_0 \in I(\mathcal{P}')$ it suffices to substitute p for every variable in $f_0, f_i, g_i, i \in [k]$. Then we can complete the proof using item (2) of Proposition 3.16. \square

In the special case of pp-interpretability when the mapping π is injective, recovery of ideal membership proofs is to some extent possible.

Proposition 3.17. *Let a constraint language Δ over a set E be pp-interpretable in a constraint language Γ over a set D in such a way that the mapping π in the definition of pp-interpretability is injective. Then for any $\text{IMP}(\Delta)$ instance (f_0, \mathcal{P}) , if there is a low bit complexity [low degree] ideal membership proof for $f'_0 \in I(\mathcal{P}')$ of $\text{IMP}(\pi^{-1}(\Delta))$, then there is a low bit complexity [low degree] ideal membership proof of $f_0 \in I(\mathcal{P})$, where (f'_0, \mathcal{P}') is the instance constructed in Theorem 3.15.*

Proof. We use the ‘inverse’ functions $\pi_i^{-1}, i \in [\ell]$, given by $\pi_i^{-1}(b) = a_i$ whenever $\pi(a_1, \dots, a_\ell) = b$ for $b \in E$ and some $a_1, \dots, a_\ell \in D$. Since π is injective, the definition is consistent. Now given an instance (f_0, \mathcal{P}) of $\text{IMP}(\Delta)$, the corresponding instance (f'_0, \mathcal{P}') of $\text{IMP}(\pi^{-1}(\Delta))$ and an ideal membership proof $f'_0 = g'_1 f'_1 + \dots + g'_k f'_k$ that $f'_0 \in I(\mathcal{P}')$, we substitute $\pi_j^{-1}(x_{ij})$ for each x_{ij} of $f'_0, f'_s, g'_s, s \in [k]$. As is easily seen, the result is an ideal membership proof that $f_0 \in I(\mathcal{P})$. \square

While ideal membership proofs and therefore the IMP allow one to prove that a polynomial has zeroes wherever a collection of other polynomials has, proving *nonnegativity* requires a different approach. Following, e.g., [36] Sum-of-Squares proof of nonnegativity are defined as follows.

Definition 3.18. *Let $P = \{p_1, \dots, p_n\}$ and $Q = \{q_1, \dots, q_n\}$ be two sets of polynomials, and let $S = \{\mathbf{x} \in \mathbb{R}^n \mid \forall p \in P : p(\mathbf{x}) = 0, \forall q \in Q : q(\mathbf{x}) \geq 0\}$. A proof of nonnegativity is defined as follows: Polynomial $f_0(\mathbf{x})$ has a Sum-of-Squares (SOS) proof of nonnegativity from P and Q if there is a polynomial identity of the form*

$$f_0(\mathbf{x}) = \sum_{i=1}^{t_0} h_i^2(\mathbf{x}) + \sum_{i=1}^m \left(\sum_{j=1}^{y_i} s_j^2(\mathbf{x}) \right) q_i(\mathbf{x}) + \sum_{i=1}^n \lambda_i(\mathbf{x}) p_i(\mathbf{x}).$$

We say the proof has degree d if $\max\{\deg h_i^2, \deg s_j^2 q_i, \deg \lambda_i p_i\} = d$.

We show that SOS proofs allow for a similar approach as studied above in this section.

Let Γ be a constraint language on a finite set D . By $\text{SOS}(\Gamma)$ we denote the class of SOS problems as in Definition 3.18 where the set P are from the set of encodings of relations from Γ . Note that this sets can also be thought of as an instance of $\text{CSP}(\Gamma)$. Thus, an instance of $\text{SOS}(\Gamma)$ is a triple (f_0, \mathcal{P}, Q) , where f_0 is a polynomial, \mathcal{P} is an instance of $\text{CSP}(\Gamma)$, and Q a set of polynomials, all over the same set of variables. The question is, whether there exists an SOS

proof of nonnegativity of f_0 from the encodings of constraints from \mathcal{P} and the set Q . By $\text{SOS}_d(\Gamma)$ we denote the problem $\text{SOS}(\Gamma)$ restricted to instances in which the input polynomial has degree at most d .

An analog of Proposition 3.16 holds for SOS proofs, as we will prove later. However, we can also prove analogs of Proposition 3.1, Theorem 3.11, 3.15, although to this end we actually need to consider the problem that SOS proofs are sought to solve.

Definition 3.19. Define the problem *IsPOSITIVE* (IPOS) as follows. Let $P = \{p_1, \dots, p_n\}$ and $Q = \{q_1, \dots, q_n\}$ be two sets of polynomials, and let $S = \{\mathbf{x} \in \mathbb{R}^n \mid \forall p \in P : p(\mathbf{x}) = 0, \forall q \in Q : q(\mathbf{x}) \geq 0\}$. Given a polynomial $f_0(\mathbf{x})$ decide whether $f_0(\mathbf{x})$ is nonnegative in every point of S .

For a constraint language Γ by $\text{IPOS}(\Gamma), \text{IPOS}_d(\Gamma)$ we denote the class of IPOS problems restricted the same way as for SOS. Then reductions from Proposition 3.1, Theorem 3.11, 3.15 hold for $\text{IPOS}(\Gamma)$.

Theorem 3.20. Let Γ be a constraint language on a set D .

- (1) Let $A \subseteq D$. Then $\text{IPOS}_d(\Gamma^A)$ is polynomial time reducible to $\text{IPOS}_{d+2|A|(|D|-1)}(\Gamma)$.
- (2) If Γ pp-defines Δ , then $\text{IPOS}(\Delta)$ [$\text{IPOS}_d(\Delta)$] is polynomial time reducible to $\text{IPOS}(\Gamma)$ [$\text{IPOS}_d(\Gamma)$].
- (3) Let Δ be a constraint language on sets E , and let Γ pp-interprets Δ . Then $\text{IPOS}(\Delta)$ [$\text{IPOS}_d(\Delta)$] is polynomial time reducible to $\text{IPOS}(\Gamma)$ [$\text{IPOS}_{kd}(\Gamma)$ for a constant k].

Proof. The proof of Theorem 3.20 repeats the proofs of Proposition 3.1, Theorems 3.11, 3.15 almost verbatim. So, we mainly highlight the differences caused by the need to take care of the polynomials from Q .

(1) Let (f_0, \mathcal{P}, Q) be an instance of $\text{IPOS}(\Gamma^A)$. First, by substituting the pinned variables with variables of the form x_a we obtain a polynomial f'_0 and an instance \mathcal{P}' in the same way as in the proof of Proposition 3.1. We also set Q' to be the set of polynomials obtained from polynomials of Q by the same substitution. Let $p = \prod_{a \in D} \prod_{b \in D \setminus \{a\}} (x_a - b)$ and set:

- $f_0^\dagger = p^2 \cdot f'_0$;
- \mathcal{P}^\dagger contains the same constraints as \mathcal{P}' except the constraints of the form $\langle (x_a), R_a \rangle$;
- $Q^\dagger = Q'$.

Let X be the set of variables in (f_0, \mathcal{P}, Q) and X^\dagger that in $(f_0^\dagger, \mathcal{P}^\dagger, Q^\dagger)$. If $S^\dagger = \{\mathbf{x} \in \mathbb{R}^{|X^\dagger|} \mid \forall f \in I(\mathcal{P}^\dagger) : f(\mathbf{x}) = 0, \forall q \in Q^\dagger : q(\mathbf{x}) \geq 0\}$, then it is straightforward that f_0 is nonnegative on $S = \{\mathbf{x} \in \mathbb{R}^{|X|} \mid \forall f \in I(\mathcal{P}) : f(\mathbf{x}) = 0, \forall q \in Q : q(\mathbf{x}) \geq 0\}$ if and only if f_0^\dagger is nonnegative on S^\dagger .

(2) Let (f_0, \mathcal{P}, Q) be an instance of $\text{IPOS}(\Delta)$. In this case we perform the same transformation of the instance \mathcal{P} as in the proof of Theorem 3.11 leaving f_0 and Q unchanged. Since f_0 does not change, its nonnegativity follows.

(3) Let (f_0, \mathcal{P}, Q) be an instance of $\text{IPOS}(\Delta)$. Let $F \subseteq D^\ell$ and $\pi : F \rightarrow E$ be from the definition of pp-interpretability, and p an interpolation of π . Then f_0 and \mathcal{P} are transformed into f'_0 and \mathcal{P}' by substituting p for every variable. In this case we also have to perform a similar transformation of polynomials from Q . Let (f'_0, \mathcal{P}', Q') be the resulting instance of $\text{IPOS}(\Gamma')$. A proof that f_0 and f'_0 are nonnegative simultaneously is now straightforward. \square

Using reductions from Theorem 3.20 an analog of Proposition 3.16 is immediate.

Proposition 3.21. *Let Γ be a constraint language.*

- (1) *For any $\text{SOS}(\Gamma^*)$ instance (f_0, \mathcal{P}, Q) , if there is a low bit complexity [low degree] SOS proof of nonnegativity of f_0 from $\mathcal{I}(\mathcal{P}), Q$, then there is a low bit complexity [low degree] SOS proof of nonnegativity of f_0^* from \mathcal{P}^+ and Q^+ , where \mathcal{P}^+, Q^+ are as in the proof of Theorem 3.20 (1).*
- (2) *Let Δ be pp-definable in Γ . Then for any $\text{SOS}(\Delta)$ instance (f_0, \mathcal{P}, Q) , if there is a low bit complexity [low degree] SOS proof of nonnegativity of f_0 from $\mathcal{I}(\mathcal{P}), Q$, then there is a low bit complexity [low degree] SOS proof of nonnegativity of f_0 from $\mathcal{I}(\mathcal{P}_\Gamma), Q$, where \mathcal{P}_Γ is the instance constructed in Theorem 3.20 (2).*
- (3) *Let Δ be pp-interpretable in Γ . Then for any $\text{SOS}(\Delta)$ instance (f_0, \mathcal{P}, Q) , if there is a low bit complexity [low degree] SOS proof of nonnegativity of f_0 from $\mathcal{I}(\mathcal{P}), Q$, then there is a low bit complexity [low degree] SOS proof of nonnegativity of f'_0 from $\mathcal{I}(\mathcal{P}'), Q'$, where (f'_0, \mathcal{P}', Q') is the instance constructed in Theorem 3.20 (3).*
- (4) *Let a constraint language Δ be pp-interpretable in Γ in such a way that the mapping π in the definition of pp-interpretable is injective. Then for any $\text{SOS}(\Delta)$ instance (f_0, \mathcal{P}, Q) , if there is a low bit complexity [low degree] SOS proof of nonnegativity of f'_0 from $\mathcal{I}(\mathcal{P}'), Q'$, then there is a low bit complexity [low degree] SOS proof of nonnegativity of f_0 from $\mathcal{I}(\mathcal{P}), Q$, where (f'_0, \mathcal{P}', Q') is the instance of $\text{SOS}(\pi^{-1}(\Delta))$ constructed in Theorem 3.20 (3).*

4 Polymorphisms and algebras

4.1 Polymorphisms and necessary condition for tractability

Sets of relations closed under pp-definitions allow for a succinct representation through polymorphisms. Let R be a k -ary relation on a set D and ψ an n -ary operation on the same set. Operation ψ is said to be a *polymorphism* of R if for any $\mathbf{a}^1, \dots, \mathbf{a}^n \in R$ the tuple $\psi(\mathbf{a}^1, \dots, \mathbf{a}^n)$ belongs to R . Here by $\psi(\mathbf{a}^1, \dots, \mathbf{a}^n)$ we understand the component-wise action of ψ , that is, if $\mathbf{a}^i = (a_1^i, \dots, a_k^i)$ then

$$\psi(\mathbf{a}^1, \dots, \mathbf{a}^n) = (\psi(a_1^1, \dots, a_1^n), \dots, \psi(a_k^1, \dots, a_k^n)).$$

For more background on polymorphisms, their properties, and links to the CSP the reader is referred to a relatively recent survey [4]. Most of the standard results we use below can be found in this survey.

A polymorphism of a constraint language Γ is an operation that is a polymorphism of every relation in Γ . The set of all polymorphisms of the language Γ is denoted by $\text{Pol}(\Gamma)$. For a set Ψ of operations by $\text{Inv}(\Psi)$ we denote the set of relations R such that every operation from Ψ is a polymorphism of R . The operators Pol and Inv induces so called *Galois correspondence* between sets of operations and constraint languages. There is a rich theory of this correspondence, however, for the sake of this paper we only need one fact.

Theorem 4.1. *Let Γ, Δ be constraint languages on a finite set D . Then $\text{Pol}(\Gamma) \subseteq \text{Pol}(\Delta)$ if and only if Γ pp-defines Δ . In particular, $\text{Inv}(\text{Pol}(\Gamma))$ is the set of all relations pp-definable in Γ .*

Combining Theorem 4.1 and Theorem 3.11, polymorphisms of constraint languages provide reductions between IMPs.

Corollary 4.2. *Let Γ, Δ be constraint languages on a finite set D and Δ finite. If $\text{Pol}(\Gamma) \subseteq \text{Pol}(\Delta)$ then $\text{IMP}(\Delta)$ [$\text{IMP}_d(\Delta)$] is polynomial time reducible to $\text{IMP}(\Gamma)$ [$\text{IMP}_d(\Gamma)$].*

Corollary 4.2 amounts to saying that similar to $\text{CSP}(\Gamma)$ the complexity of $\text{IMP}(\Gamma)$ is determined by the polymorphisms of Γ .

Next we use the known necessary condition for CSP tractability [12] to obtain some necessary conditions for tractability of $\text{IMP}(\Gamma)$.

A *projection* is an operation $\psi : D^k \rightarrow D$ such that there is $i \in [k]$ with $\psi(x_1, \dots, x_k) = x_i$ for any $x_1, \dots, x_k \in D$. If the only polymorphisms of a constraint language are projections, every relation is pp-definable in Γ implying that $\text{IMP}(\Gamma)$ is **coNP**-complete.

Theorem 3.3 is another ingredient for our necessary condition. Recall that for a language Γ by Γ^* we denote the language with added *constant relations* R_a for all $a \in D$. It is known that every polymorphism ψ of all the constant relations is *idempotent*, that is, satisfies the condition $\psi(x, \dots, x) = x$. Therefore, by Theorem 3.3 it suffices to focus on idempotent polymorphisms.

Proposition 4.3. *Let Γ be a constraint language over a finite set D . If the only idempotent polymorphisms of Γ are projections then $\text{IMP}_{|D|(|D|-1)}(\Gamma)$ is **coNP**-complete.*

Example 4.4. Consider the relation $N = \{0, 1\}^3 \setminus \{(0, 0, 0), (1, 1, 1)\}$. This relation corresponds to **NOT-ALL-EQUAL SATISFIABILITY** problem and it is known that the idempotent operations from $\text{Pol}(\{N\})$ are projections [35]. $\text{CSP}(\{N\})$ was shown to be **NP**-complete by Schaefer [38] and $\text{IMP}(\{N\})$ is shown to be **coNP**-complete in [28].

4.2 Algebras and a better necessary condition

In this section we briefly review the basics of structural properties of (universal) algebra in application to the IMP. Universal algebras have been instrumental in the study of CSPs, and, although we do not go deeper into this theory in this paper, we expect they should be useful for IMPs as well. We follow textbooks [13, 31] and texts on the algebraic theory of the CSP, see, e.g., [2, 3, 12, 11].

Algebras. An *algebra* is a pair $\mathcal{D} = (D, \Psi)$ where D is a set (always finite in this paper) and Ψ is a set of operations on D (perhaps multi-ary). The operations from Ψ are called *basic*, and any operation that can be obtained from operations in Ψ by means of composition is called a *term* operation. The set of all term operations will be denoted by $\text{Term}(\mathcal{D})$. For example, Ψ can be the set $\text{Pol}(\Gamma)$ for some constraint language Γ on D , in which case \mathcal{D} is called the *algebra of polymorphisms* of Γ and will be denoted $\text{Alg}(\Gamma)$. Thus, $\text{Alg}(\Gamma) = (D, \text{Pol}(\Gamma))$.

By Corollary 4.2 the algebra $\text{Alg}(\Gamma)$ determines the complexity of $\text{IMP}(\Gamma)$ and $\text{IMP}_d(\Gamma)$ for sufficiently large d . The advantage of using algebras rather than just polymorphisms is that it unlocks a variety of structural methods that cannot be easily applied if we only use polymorphisms. Algebra $\mathcal{D} = (D, \Psi)$ is said to be *tractable* [*d-tractable*] if for any finite constraint language Γ such that $\Psi \subseteq \text{Pol}(\Gamma)$ the problem $\text{IMP}(\Gamma)$ is tractable [*d-tractable*]. Algebra \mathcal{D} is said to be **coNP**-complete if for some finite language Γ with $\Psi \subseteq \text{Pol}(\Gamma)$ the problem $\text{IMP}(\Gamma)$ is

coNP-complete. In the rest of this section apart from another necessary condition of tractability we prove several results that deduce the tractability [d -tractability, **coNP-completeness**] of a certain algebra derivative from \mathcal{D} from a similar property of \mathcal{D} .

Idempotent algebras. The first step will be to reduce the kind of algebras we have to study. By Theorem 3.3 idempotent polymorphisms determine the complexity of $\text{IMP}(\Gamma)$. On the algebraic side, an algebra \mathcal{D} is said to be *idempotent* if each of its basic operations (and therefore each of its term operations) is idempotent. Every algebra $\mathcal{D} = (D, \Psi)$ can be converted into an idempotent algebra simply by throwing out all the non-idempotent term operations. Let $\text{Term}_{id}(\mathcal{D})$ denote the set of all idempotent operations from $\text{Term}(\mathcal{D})$. Then the *full idempotent reduct* of $\mathcal{D} = (D, \Psi)$ is the algebra $\text{Id}(\mathcal{D}) = (D, \text{Term}_{id}(\mathcal{D}))$.

Proposition 4.5. *For any finite algebra \mathcal{D} , \mathcal{D} is tractable [d -tractable] if and only if $\text{Id}(\mathcal{D})$ is tractable [d -tractable]. Also $\text{Id}(\mathcal{D})$ is **coNP-complete** if and only if \mathcal{D} is **coNP-complete**.*

Proof. Note that an operation ψ on a set D is idempotent if and only if it preserves all the relations in the set $\Gamma_{\text{CON}} = \{R_a \mid a \in D\}$, consisting of all constant relations R_a on D . Hence, $\text{Inv}(\text{Term}_{id}(\mathcal{D}))$ is the relational clone generated by $\text{Inv}(\Psi) \cup \Gamma_{\text{CON}}$, or, in other words, every relation R such that $\text{Term}_{id}(\mathcal{D}) \subseteq \text{Pol}(R)$ is pp-definable in $\text{Inv}(\Psi) \cup \Gamma_{\text{CON}}$.

Let Δ be a finite set from $\text{Inv}(\text{Term}_{id}(\mathcal{D}))$. By the observation above there is a finite $\Gamma \subseteq \text{Inv}(\Psi) \cup \Gamma_{\text{CON}}$ such that Γ pp-defines Δ . By Theorem 3.3 for any d the problem $\text{IMP}_d(\Delta)$ can be reduced in polynomial time to $\text{IMP}_{d+|D|(|D|-1)}(\Gamma)$, and the result follows. \square

Subalgebras, homomorphisms, and direct powers. The following standard algebraic constructions have been very useful in the study of the CSP.

Definition 4.6. *Let $\mathcal{D} = (D, \Psi)$ be an algebra.*

- **(Subalgebra)** *Let $E \subseteq D$ such that, for any $\psi \in \Psi$ and for any $b_1, \dots, b_k \in E$, where k is the arity of ψ , we have $\psi(b_1, \dots, b_k) \in E$. In other words, ψ is a polymorphism of E or $E \in \text{Inv}(\Psi)$. The algebra $\mathcal{E} = (E, \Psi|_E)$, where $\Psi|_E$ consists of the restrictions of all operations in Ψ to E , is called a subalgebra of \mathcal{D} .*
- **(Direct power)** *For a natural number k the k -th direct power \mathcal{D}^k of \mathcal{D} is the algebra $\mathcal{D}^k = (D^k, \Psi^k)$, where Ψ^k consists of all the operations from Ψ acting on D^k component-wise (see the definition of polymorphism).*
- **(Homomorphic image)** *Let E be a set and $\chi : D \rightarrow E$ a mapping such that for any (say, k -ary) $\psi \in \Psi$ and any $a_1, \dots, a_k, b_1, \dots, b_k \in D$, if $\chi(a_i) = \chi(b_i)$, $i \in [k]$, then $\chi(\psi(a_1, \dots, a_k)) = \chi(\psi(b_1, \dots, b_k))$. The algebra $\mathcal{E} = (E, \Psi_\chi)$ is called a homomorphic image of \mathcal{D} , where for every $\psi \in \Psi$ the set Ψ_χ contains ψ/χ given by $\psi/\chi(c_1, \dots, c_k) = \chi(\psi(a_1, \dots, a_k))$ and $a_1, \dots, a_k \in D$ are such that $c_i = \chi(a_i)$, $i \in [k]$.*

If an algebra is the algebra of polymorphisms of some constraint language, the concepts above are related to pp-definitions and pp-interpretations. We will use the following easy observation.

Lemma 4.7. *Let $\mathcal{D} = (D, \Psi) = \text{Alg}(\Gamma)$ for a constraint language Γ over D .*

- *If $\mathcal{E} = (E, \Psi|_E)$ is a subalgebra of \mathcal{D} then E is pp-definable in Γ .*

- Every relation from $\text{Inv}(\Psi^k)$ is pp-interpretable in Γ .
- Let $\mathcal{E} = (E, \Psi_\chi)$ be a homomorphic image of \mathcal{D} . Then every relation from $\text{Inv}(\Psi_\chi)$ is pp-interpretable in Γ .

The standard algebraic constructions also include direct products of different algebras. Direct products also have a strong connection to the CSP and therefore IMP. However, they require a more general framework, multi-sorted CSPs and IMPs. These are beyond the scope of this paper.

We are now ready to prove the reductions induced by subalgebras, direct powers, and homomorphic images.

Theorem 4.8. *Let \mathcal{D} be an algebra and \mathcal{E} its subalgebra [direct power, homomorphic image]. If \mathcal{D} is tractable [d-tractable], then so is \mathcal{E} . If \mathcal{E} is **coNP**-complete, then \mathcal{D} is also **coNP**-complete.*

Proof. The theorem is almost straightforward from Lemma 4.7 and Theorems 3.11, 3.15. Let $\mathcal{D} = (D, \Psi)$, $\mathcal{E} = (E, \Psi')$ and $\Delta \subseteq \text{Inv}(\Psi')$, a finite set. Note that $\mathcal{D} = \text{Alg}(\Gamma)$ for $\Gamma = \text{Inv}(\Psi)$.

If \mathcal{E} is a subalgebra of \mathcal{D} , that is, $\Psi' = \Psi|_E$ then by Lemma 4.7 E is pp-definable in Γ and therefore $\Delta \subseteq \text{Inv}(\Psi|_E) \subseteq \text{Inv}(\Psi)$. The result follows.

If \mathcal{E} is a direct power, say, $\mathcal{E} = (D^k, \Psi^k)$, then since every relation from $\text{Inv}(\Psi^k)$ is pp-interpretable in Γ , there is a finite set $\Gamma' \subseteq \Gamma$ that pp-interprets Δ . Then by Theorem 3.15 $\text{IMP}(\Delta)$ can be reduced to $\text{IMP}(\Gamma')$ in polynomial time. The result follows.

In the case when \mathcal{E} is a homomorphic image of Γ , the proof is identical to the previous case due to Lemma 4.7. \square

Stronger necessary condition for tractability. Subalgebras, direct powers, and homomorphic images allow us to state a stronger condition for tractability of constraint languages. In the case of CSP, when a constraint language Γ contains all the constant relations, $\text{CSP}(\Gamma)$ is **NP**-complete if and only if $\text{Alg}(\Gamma)$ has a homomorphic image \mathcal{D} of a subalgebra such that all the term operations of \mathcal{D} are projections. Using Theorem 3.3 we can make this condition even stronger (although only necessary).

Theorem 4.9. *Let Γ be a constraint language with the property that there exists a homomorphic image \mathcal{E} of a subalgebra of a direct power of $\text{Id}(\text{Alg}(\Gamma))$ such that all the term operations of \mathcal{E} are projections. Then $\text{IMP}_{|D|(|D|-1)}(\Gamma)$ is **coNP**-complete.*

Proof. Let $\mathcal{E} = (E, \Psi)$. Since the term operations of \mathcal{E} are only projections, by Proposition 4.3 there is a finite set Δ such that $\text{IMP}_0(\Delta)$ is **coNP**-complete. Then, by Lemma 4.7, Γ^* pp-interprets Δ , and we obtain the result by Theorems 3.3 and 3.15. \square

5 Sufficient conditions for tractability

5.1 The Ideal Membership Problem and Gröbner Bases

A possible way to answer to the IMP is via polynomial division. Informally, if a remainder of division of f_0 by generating polynomials of $I(\mathcal{P})$ is zero then $f_0 \in I(\mathcal{P})$. Let us recall some standard notations from algebraic geometry that are needed to present a division algorithm and the notion of Gröbner Bases. We follow notation in [17].

A monomial ordering \succ on $\mathbb{F}[x_1, \dots, x_n]$ is a relation \succ on $\mathbb{Z}_{\geq 0}^n$, or equivalently, a relation on the set of monomials x^α , $\alpha \in \mathbb{Z}_{\geq 0}^n$ (see [17], Definition 1, p.55). Each monomial $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ corresponds to an n -tuple of exponents $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_{\geq 0}^n$. This establishes a one-to-one correspondence between the monomials in $\mathbb{F}[x_1, \dots, x_n]$ and $\mathbb{Z}_{\geq 0}^n$. Any ordering \succ we establish on the space $\mathbb{Z}_{\geq 0}^n$ will give us an ordering on monomials: if $\alpha \succ \beta$ according to this ordering, we will also say that $x^\alpha \succ x^\beta$.

Definition 5.1. Let $\alpha = (\alpha_1, \dots, \alpha_n), \beta = (\beta_1, \dots, \beta_n) \in \mathbb{Z}_{\geq 0}^n$ and $|\alpha| = \sum_{i=1}^n \alpha_i$, $|\beta| = \sum_{i=1}^n \beta_i$. Lexicographic order and graded lexicographic order are defined as follows.

1. We say $\alpha \succ_{\text{lex}} \beta$ if the leftmost nonzero entry of the vector difference $\alpha - \beta \in \mathbb{Z}^n$ is positive. We will write $x^\alpha \succ_{\text{lex}} x^\beta$ if $\alpha \succ_{\text{lex}} \beta$.
2. We say $\alpha \succ_{\text{grlex}} \beta$ if $|\alpha| > |\beta|$, or $|\alpha| = |\beta|$ and $\alpha \succ_{\text{lex}} \beta$.

Definition 5.2. For any $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_{\geq 0}^n$ let $x^\alpha \stackrel{\text{def}}{=} \prod_{i=1}^n x_i^{\alpha_i}$. Let $f = \sum_{\alpha} a_{\alpha} x^{\alpha}$ be a nonzero polynomial in $\mathbb{F}[x_1, \dots, x_n]$ and let \succ be a monomial order.

1. The multidegree of f is $\text{multideg}(f) \stackrel{\text{def}}{=} \max(\alpha \in \mathbb{Z}_{\geq 0}^n : a_{\alpha} \neq 0)$.
2. The degree of f is $\deg(f) = |\text{multideg}(f)|$ where $|\alpha| = \sum_{i=1}^n \alpha_i$. In this paper, this is always according to grlex order.
3. The leading coefficient of f is $\text{LC}(f) \stackrel{\text{def}}{=} a_{\text{multideg}(f)} \in \mathbb{F}$.
4. The leading monomial of f is $\text{LM}(f) \stackrel{\text{def}}{=} x^{\text{multideg}(f)}$ (with coefficient 1).
5. The leading term of f is $\text{LT}(f) \stackrel{\text{def}}{=} \text{LC}(f) \cdot \text{LM}(f)$.

Definition 5.3 (A division algorithm). Let \succ be a monomial order on $\mathbb{Z}_{\geq 0}^n$, and let $F = \{f_1, \dots, f_s\} \subset \mathbb{F}[x_1, \dots, x_n]$. Then every $f \in \mathbb{F}[x_1, \dots, x_n]$ can be written as $f = q_1 f_1 + \dots + q_s f_s + r$, where $q_i, r \in \mathbb{F}[x_1, \dots, x_n]$, and either $r = 0$ or r is a linear combination, with coefficients in \mathbb{F} , of monomials, none of which is divisible by any of $\text{LT}(f_1), \dots, \text{LT}(f_s)$. Furthermore, if $q_i f_i \neq 0$, then $\text{multideg}(f) \succeq \text{multideg}(q_i f_i)$. We call r a remainder of f on division by F . Also, we say that f reduces to r modulo F , written $f \rightarrow_F r$.

The above definition suggests the following procedure to compute a remainder. Repeatedly, choose an $f_i \in F$ such that $\text{LT}(f_i)$ divides some term t of f and replace f with $f - \frac{t}{\text{LT}(f_i)} f_i$, until it cannot be further applied. Note that the order we choose the polynomial f_i is not specified. Unfortunately, depending on the generating polynomials of the ideal a remainder of division may not be unique. Moreover, such a remainder depends on the order we do division.

Example 5.4. Let $f = x^2 y - xy^2 + y$ and $I = \langle f_1, f_2 \rangle$ with $f_1 = x^2$ and $f_2 = xy - 1$. Consider the grlex order with $x \succ_{\text{lex}} y$. On one hand, $f = 0 \cdot f_1 + (x - y) \cdot f_2 - x$. On the other hand, $f = y \cdot f_1 - y \cdot f_2 + 0$.

The Hilbert Basis Theorem states that every ideal has a finite generating set (see, e.g., Theorem 4 on page 77 [17]). Fortunately, for every ideal there is a finite generating set that suits our purposes. That is, there is a generating set so that the remainder of division by that set is uniquely defined, no matter in which order we do the division.

Definition 5.5. Fix a monomial order on the polynomial ring $\mathbb{F}[x_1, \dots, x_n]$. A finite subset $G = \{g_1, \dots, g_t\}$ of an ideal $I \subseteq \mathbb{F}[x_1, \dots, x_n]$ different from $\{0\}$ is said to be a Gröbner Basis (or standard basis) if

$$\langle \text{LT}(g_1), \dots, \text{LT}(g_t) \rangle = \langle \text{LT}(I) \rangle$$

where $\langle \text{LT}(I) \rangle$ denotes the ideal generated by the leading terms of elements of I .

Definition 5.6 (*d-truncated Gröbner Basis*). If G is a Gröbner Basis of an ideal, the *d-truncated Gröbner Basis* G' of G is defined as

$$G' = G \cap \mathbb{F}[x_1, \dots, x_n]_d,$$

where $\mathbb{F}[x_1, \dots, x_n]_d$ is the set of polynomials of degree less than or equal to d .

Proposition 5.7 ([17], Proposition 1, p.83). Let $I \subseteq \mathbb{F}[x_1, \dots, x_n]$ be an ideal and let $G = \{g_1, \dots, g_t\}$ be a Gröbner Basis for I . Then given $f \in \mathbb{F}[x_1, \dots, x_n]$, there is a unique $r \in \mathbb{F}[x_1, \dots, x_n]$ with the following two properties:

1. No term of r is divisible by any of $\text{LT}(g_1), \dots, \text{LT}(g_t)$,
2. There is $g \in I$ such that $f = g + r$.

In particular, r is the remainder on division of f by G no matter how the elements of G are listed when using the division algorithm.

The remainder r is called the *normal form* of f by G , denoted by $f|_G$. Note that, although the remainder r is unique, even for a Gröbner Basis, the "quotients" q_i produced by the division algorithm $f = q_1g_1 + \dots + q_tg_t + r$ can change if we list the generators in a different order. As a corollary of Proposition 5.7, we get the following criterion for when a given polynomial lies in an ideal.

Corollary 5.8 ([17], Corollary 2, p.84). Let $G = \{g_1, \dots, g_t\}$ be a Gröbner Basis for an ideal $I \subseteq \mathbb{F}[x_1, \dots, x_n]$ and let $f \in \mathbb{F}[x_1, \dots, x_n]$. Then $f \in I$ if and only if the remainder on division of f by G is zero.

There is a criterion, known as Buchberger's criterion, that tells us whether a given generating set of an ideal is a Gröbner Basis. In order to formally express this criterion, we need to define the notion of S-polynomials.

Definition 5.9 (S-polynomial). Let $f, g \in \mathbb{F}[x_1, \dots, x_n]$ be nonzero polynomials. If $\text{multideg}(f) = \alpha$ and $\text{multideg}(g) = \beta$, then let $\gamma = (\gamma_1, \dots, \gamma_n)$, where $\gamma_i = \max(\alpha_i, \beta_i)$ for each i . We call x^γ the least common multiple of $\text{LM}(f)$ and $\text{LM}(g)$, written $x^\gamma = \text{lcm}(\text{LM}(f), \text{LM}(g))$. The S-polynomial of f and g is the combination

$$S(f, g) = \frac{x^\gamma}{\text{LT}(f)} \cdot f - \frac{x^\gamma}{\text{LT}(g)} \cdot g.$$

Theorem 5.10 (Buchberger's Criterion [17], Theorem 3, p.105). Let I be a polynomial ideal. Then a basis $G = \{g_1, \dots, g_t\}$ of I is a Gröbner Basis of I if and only if for all pairs $i \neq j$, the remainder on division of $S(g_i, g_j)$ by G (listed in some order) is zero.

Proposition 5.11 ([17], Proposition 4, p.106). We say the leading monomials of two polynomials f, g are relatively prime if $\text{lcm}(\text{LM}(f), \text{LM}(g)) = \text{LM}(f) \cdot \text{LM}(g)$. Given a finite set $G \subseteq \mathbb{F}[x_1, \dots, x_n]$, suppose that we have $f, g \in G$ such that the leading monomials of f and g are relatively prime. Then $S(f, g) \rightarrow_G 0$.

5.2 The dual-discriminator

Here we deal with a *majority* operation. Over the Boolean domain there is only one majority operation, called the *dual-discriminator*. In the Boolean case, Mastrolilli [28] proved that the $\text{IMP}(\Gamma)$ is tractable when the constraint language Γ is closed under the dual-discriminator operation. Later, Bharathi and Mastrolilli [7] expand this tractability result to constraint languages over the ternary domain. We establish tractability result for any finite domain. We start off with the definition of a majority polymorphism and explain an appealing structure of majority closed relations.

Definition 5.12. Let μ be a 3-ary operation from D^3 to D . If for all $x, y \in D$ we have $\mu(x, x, y) = \mu(x, y, x) = \mu(y, x, x) = x$, then μ is called a majority operation.

For a (n -ary) relation R and $T \subseteq [n]$ by $\text{pr}_T R$ we denote the *projection* of R onto T , that is, the set of tuples $(a_i)_{i \in T}$ such that there is $(b_1, \dots, b_n) \in R$ with $b_i = a_i$ for each $i \in T$.

Proposition 5.13 ([25]). Let R be a relation of arity n with a majority polymorphism, and let $C = \langle S, R \rangle$ constraining the variables in S with relation R . For any problem \mathcal{P} with constraint C , the problem \mathcal{P}' which is obtained by replacing C by the set of constraints

$$\{((S[i], S[j]), \text{pr}_{i,j}(R)) \mid 1 \leq i \leq j \leq n\}$$

has exactly the same solutions as \mathcal{P} .

The above proposition suggests that, without loss of generality, we may assume that all the constraints are binary when a constrain language has a majority polymorphism μ . Let Γ be a language over a set D such that $\mu \in \text{Pol}(\Gamma)$, and let $\mathcal{P} = (X, D, C)$ be an instance of $\text{CSP}(\Gamma)$. We assume each constraint in C is binary. That is

$$C = \{C_{ij} = \langle (x_i, x_j), R_{ij} \rangle \mid R_{ij} \subseteq D_i \times D_j \text{ where } D_i, D_j \subseteq D\}.$$

Throughout this section we assume all the constraint are binary. Relations closed under the dual-discriminator operation admit a much nicer structure that has been characterized, see [40]. Indeed, such a characterization states that constraints can only be of three types. Let us first define the dual-discriminator operation before formulating the characterization. The dual-discriminator operation is defined as follows.

$$\nabla(x, y, z) = \begin{cases} y & \text{if } y = z, \\ x & \text{otherwise.} \end{cases}$$

Lemma 5.14 ([16, 40]). Suppose $\nabla \in \text{Pol}(\Gamma)$. Then each constraint $C_{ij} = \langle (x_i, x_j), R_{ij} \rangle$ is one of the following three types.

1. A complete constraint: $R_{ij} = D_i \times D_j$ for some $D_i, D_j \subseteq D$,
2. A permutation constraint: $R_{ij} = \{(a, \pi(a)) \mid a \in D_i\}$ for some $D_i \subseteq D$ and some bijection $\pi : D_i \rightarrow D_j$, where $D_j \subseteq D$,
3. A two-fan constraint: $R_{ij} = \{(\{a\} \times D_j) \cup (D_i \times \{b\})\}$ for some $D_i, D_j \subseteq D$ and $a \in D_i, b \in D_j$.

Here, we discuss a preprocessing step in order to handle permutation constraints and transform the instance into an instance without any permutation constraints. Such a transformation simplifies the problem and yields to an efficient computation of a Gröbner Basis. Let $\mathcal{P} = (X, D, C)$ be an instance of $\text{CSP}(\Gamma)$ such that $\nabla \in \text{Pol}(\Gamma)$. Suppose the set of constraints C contains a permutation constraint $C_{ij} = \langle (x_i, x_j), R_{ij} \rangle$ with $R_{ij} = \{(a, \pi(a)) \mid a \in D_i, \pi(a) \in D_j\}$. Define instance $\mathcal{P}' = (X \setminus \{x_j\}, D, C')$ as follows.

1. $C'_{st} = C_{st}$ if $s \neq j$ and $t \neq j$,
2. replace each constraint $C_{sj} = \langle (x_s, x_j), R_{sj} \rangle$ by $C'_{si} = \langle (x_s, x_i), R'_{si} \rangle$ where $R'_{si} = \{(a, \pi^{-1}(b)) \mid (a, b) \in R_{sj}\}$,
3. replace each constraint $C_{js} = \langle (x_j, x_s), R_{js} \rangle$ by $C'_{is} = \langle (x_i, x_s), R'_{is} \rangle$ where $R'_{is} = \{(\pi^{-1}(a), b) \mid (a, b) \in R_{js}\}$.

Note that instance \mathcal{P} has a solution if and only if instance \mathcal{P}' has a solution. The next lemma states the above preprocessing step does not change the complexity of the IMP. In the next lemma we use a polynomial interpolation of permutation π . The permutation π can be interpolated by a polynomial $p : \mathbb{R} \rightarrow \mathbb{R}$ such that $p(a) = \pi(a)$ for all $a \in D_i$.

Lemma 5.15. *Let $I(\mathcal{P})$ and $I(\mathcal{P}')$ be the corresponding ideals to instances \mathcal{P} and \mathcal{P}' , respectively. Given a polynomial f_0 , define polynomial f'_0 to be the polynomial where we replace every occurrence of x_j with $p(x_i)$. Then $f_0 \in I(\mathcal{P})$ if and only if $f'_0 \in I(\mathcal{P}')$.*

Proof. Note that, by our construction, there is a one-to-one correspondence between the points in $\mathbf{V}(I(\mathcal{P}))$ and the points in $\mathbf{V}(I(\mathcal{P}'))$. That is, each $\mathbf{a} = (a_1, \dots, a_i, \dots, a_j = p(a_i), \dots, a_n) \in \mathbf{V}(I(\mathcal{P}))$ corresponds to $\mathbf{a}' = (a_1, \dots, a_{j-1}, a_{j+1}, \dots, a_n) \in \mathbf{V}(I(\mathcal{P}'))$. Furthermore, for all $\mathbf{a} \in \mathbf{V}(I(\mathcal{P}))$ and the corresponding $\mathbf{a}' \in \mathbf{V}(I(\mathcal{P}'))$ we have $f_0(\mathbf{a}) = f'_0(\mathbf{a}')$. This yields that

$$(\exists \mathbf{a} \in \mathbf{V}(I(\mathcal{P})) \wedge f_0(\mathbf{a}) \neq 0) \iff (\exists \mathbf{a}' \in \mathbf{V}(I(\mathcal{P}')) \wedge f'_0(\mathbf{a}') \neq 0)$$

This finishes the proof of the lemma. \square

The preprocessing step and Lemma 5.15 suggest that we can massage any instance of $\text{IMP}(\Gamma)$ and obtain an instance $\text{IMP}(\Gamma)$ without permutation constraints. This can be carried out in polynomial time by processing permutation constraints one by one in turn.

Lemma 5.16. *Let $\mathcal{P} = (X, D, C)$ be an instance of $\text{CSP}(\Gamma)$ such that $\nabla \in \text{Pol}(\Gamma)$ and C contains no permutation constraint. Then a Gröbner Basis of the corresponding ideal $I(\mathcal{P})$ can be computed in polynomial time.*

Proof. Satisfiability of instance \mathcal{P} can be decided in polynomial time [16] so we may assume that $1 \notin I(\mathcal{P})$, else $G = \{1\}$ is a Gröbner Basis. Moreover, we assume for every $x_i, x_j, x_k \in X$ if $(a, b) \in R_{ij}$ then there exists c such that $(a, c) \in R_{ik}$ and $(c, b) \in R_{kj}$. This corresponds to the so-called *arc consistency* notion which is used in solving CSPs extensively. Note that such an assumption does not change $\text{Sol}(\mathcal{P})$. Equivalently, it does not change $\mathbf{V}(I(\mathcal{P}))$ which, by Theorem 2.10, means the ideal $I(\mathcal{P})$ does not change. This arc consistency assumption has a consequence in terms of polynomials which enables us to prove our set of polynomials is indeed a Gröbner Basis.

First, let us give a set G of polynomials that represent binary constraints. Initially, G contains all the domain polynomials i.e. $G = \{ \prod_{a \in D} (x_i - a) \mid x_i \in X \}$. We proceed as follows.

- i) To each complete constraint $C_{ij} = \langle (x_i, x_j), R_{ij} = D_i \times D_j \rangle$ we associate two polynomials $g_i = \prod_{a \in D_i} (x_i - a)$ and $g_j = \prod_{b \in D_j} (x_j - b)$ and replace $\prod_{a \in D} (x_i - a)$ by g_i , and replace $\prod_{a \in D} (x_j - a)$ by g_j .
- ii) To each two-fan constraint $C_{ij} = \langle (x_i, x_j), R_{ij} = \{(\{a\} \times D_j) \cup (D_i \times \{b\})\} \rangle$ we associate polynomial $g_{ij} = (x_i - a)(x_j - b)$ and set $G = G \cup \{g_{ij}\}$. Furthermore, we replace domain polynomial $\prod_{a \in D} (x_i - a)$ by $\prod_{a \in D_i} (x_i - a)$, and replace domain polynomial $\prod_{a \in D} (x_j - a)$ by $\prod_{a \in D_j} (x_j - a)$.

Observe that, as a consequence of arc consistency, for every two polynomials $f = (x_i - a)(x_j - b)$ and $g = (x_i - c)(x_k - d)$ in G with $a \neq c$ polynomial $h = (x_j - b)(x_k - d)$ is also in G .

Consider grlex order with $x_1 \succ_{\text{lex}} \dots \succ_{\text{lex}} x_n$. Now, we prove that for any two polynomials $f, g \in G$ we have $S(f, g) \rightarrow_G 0$. We dismiss the cases where $\text{LM}(f)$ and $\text{LM}(g)$ are relatively prime. In these cases, by Proposition 5.11, we have $S(f, g) \rightarrow_G 0$. Hence, we focus on the cases where $\text{LM}(f)$ and $\text{LM}(g)$ are not relatively prime.

1. Suppose $f = (x_i - a)(x_j - b)$ and $g = (x_i - a)(x_k - d)$. Then

$$S(f, g) = x_k \cdot f - x_j \cdot g = d \cdot x_i \cdot x_j - b \cdot x_i \cdot x_k + a \cdot b \cdot x_k - a \cdot d \cdot x_j = d \cdot f - b \cdot g.$$

Observe that $\text{multideg}(S(f, g)) \succeq \text{multideg}(d \cdot f)$ and $\text{multideg}(S(f, g)) \succeq \text{multideg}(b \cdot g)$. Hence, by Definition 5.3, we have $S(f, g) \rightarrow_{\{f, g\}} 0$.

2. Suppose $f = (x_i - a)(x_j - b)$ and $g = (x_i - c)(x_k - d)$ where $a \neq c$. Then $S(f, g) = x_k \cdot f - x_j \cdot g$ and $S(f, g) \rightarrow_{\{f, g\}} (c - a)(x_j - b)(x_k - d)$. However, $h = (x_j - b)(x_k - d)$ is in G and hence $S(f, g) \rightarrow_G 0$.

3. Suppose $f = \prod_{a \in D_i} (x_i - a)$ and $g = (x_i - c)(x_j - b)$ where $c \in D_i \subseteq D$. Define $f_1 = \prod_{a \in D_i \setminus \{c\}} (x_i - a)$ and $g_1 = (x_j - b)$. Hence, $f = (x_i - c) \cdot f_1$ and $g = (x_i - c) \cdot g_1$.

$$\begin{aligned} S(f, g) &= x_j \cdot f - (x_i^{|D_i|-1}) \cdot g \\ &= [(x_j - b) + b] \cdot f - \left[\prod_{a \in D_i \setminus \{c\}} (x_i - a) - \left(\prod_{a \in D_i \setminus \{c\}} (x_i - a) - x_i^{|A_i|-1} \right) \right] \cdot g \\ &= (x_j - b) \cdot f - (b) \cdot f - \prod_{a \in D_i \setminus \{c\}} (x_i - a) \cdot g + \left(\prod_{a \in D_i \setminus \{c\}} (x_i - a) - x_i^{|D_i|-1} \right) \cdot g \\ &= (-b) \cdot f + \left(\prod_{a \in D_i \setminus \{c\}} (x_i - a) - x_i^{|D_i|-1} \right) \cdot g = (g_1 - \text{LT}(g_1)) \cdot f + (f_1 - \text{LT}(f_1)) \cdot g \end{aligned}$$

We show that $\text{multideg}(S(f, g)) \succeq \text{multideg}((\text{LT}(g_1) - g_1) \cdot f)$ and $\text{multideg}(S(f, g)) \succeq \text{multideg}((f_1 - \text{LT}(f_1)) \cdot g)$. This follows by showing $\text{LM}((\text{LT}(g_1) - g_1) \cdot f) \neq \text{LM}((f_1 - \text{LT}(f_1)) \cdot g)$. By contradiction, if $\text{LM}((\text{LT}(g_1) - g_1) \cdot f) = \text{LM}((f_1 - \text{LT}(f_1)) \cdot g)$ then

$$\text{LM}(\text{LT}(g_1) - g_1) \cdot \text{LM}(f) = \text{LM}(f_1 - \text{LT}(f_1)) \cdot \text{LM}(g) \implies x_i^{|A_i|} = x_i^{|A_i|-2} \cdot x_i x_j$$

The latter is impossible, hence $\text{multideg}(S(f, g)) \succeq \text{multideg}((\text{LT}(g_1) - g_1) \cdot f)$ and $\text{multideg}(S(f, g)) \succeq \text{multideg}((f_1 - \text{LT}(f_1)) \cdot g)$. Therefore, we have $S(f, g) \rightarrow_{\{f, g\}} 0$ (recall Definition 5.3).

We have shown for any two polynomials $f, g \in G$ we have $S(f, g) \rightarrow_G 0$. Hence, by Buchberger's criterion (Theorem 5.10), G is a Gröbner Basis for $I(\mathcal{P})$. \square

Theorem 5.17. *Let Γ be a constraint language. If Γ has the dual-discriminator polymorphism, then $\text{IMP}(\Gamma)$ can be solved in polynomial time.*

Remark 5.18. *Note that unlike in the results of [28, 7], due to the preprocessing step, Theorem 5.17 does not always allow one to find a proof that a polynomial belongs to the ideal.*

5.3 Semilattice polymorphisms

In this section we study the $\text{IMP}(\Gamma)$ for languages Γ where $\text{Pol}(\Gamma)$ contains a *semilattice* operation. A binary operation $\psi(x, y)$ satisfying the following three conditions is said to be a semilattice operation:

1. Associativity: $\psi(x, \psi(y, z)) = \psi(\psi(x, y), z)$
2. Commutativity: $\psi(x, y) = \psi(y, x)$
3. Idempotency: $\psi(x, x) = x$

Mastrolilli [28] considered this problem for languages over the Boolean domain i.e., $D = \{0, 1\}$, and proved the following. We remark that a Boolean relation is closed under a semilattice operation if and only if it can be defined by a conjunction of *dual-Horn* clauses or can be defined by a conjunction of *Horn* clauses [25].

Theorem 5.19 ([28]). *Let Γ be a finite Boolean constraint language. If Γ has a semilattice polymorphism, then $\text{IMP}_d(\Gamma)$ can be solved in $n^{O(d)}$ time for $d \geq 1$.*

We extend this tractability result to languages over any finite domain D . That is, we prove that $\text{IMP}_d(\Gamma)$ is polynomial time solvable when Γ is a language over D and it has a semilattice polymorphism. To do so, we use a well-known result in semilattice theory. A *semilattice* is an algebra $\mathcal{D} = (D, \{\psi\})$, where ψ is a semilattice operation. Informally speaking, every semilattice is a subalgebra of a direct power of a 2-element semilattice.

Theorem 5.20 ([33]). *Let $\mathcal{D} = (D, \psi)$ be a finite semilattice where ψ is a semilattice operation. Then there is k such that \mathcal{D} is a subalgebra of the direct power \mathcal{B}^k of $\mathcal{B} = (\{0, 1\}, \varphi)$, where φ is a semilattice operation on $\{0, 1\}$.*

Armed with Theorem 5.20 a proof of tractability of semilattice IMPs is straightforward.

Theorem 5.21. *Let Γ be a finite constraint language over domain D . If Γ has a semilattice polymorphism, then $\text{IMP}_d(\Gamma)$ can be solved in polynomial time.*

Proof. Let ψ be a semilattice polymorphism of Γ . Then $\mathcal{D} = (D, \psi)$ is a semilattice, and therefore is a subalgebra of \mathcal{B}^k , where $\mathcal{B} = (\{0, 1\}, \varphi)$ is a 2-element semilattice. By Lemma 4.7, there is a finite constraint language Δ over $\{0, 1\}$ such that φ is a polymorphism of Δ . By Theorem 3.15, $\text{IMP}_d(\Gamma)$ reduces to $\text{IMP}_{ud}(\Delta)$ in polynomial time for a constant u . By Theorem 5.19, we get the result. \square

Example 5.22 (Totally Ordered Domain). Let $D = \{1, \dots, t\}$ be a finite domain and Γ be a language defined on D . A semilattice polymorphism ψ is conservative if $\psi(x, y) \in \{x, y\}$. Suppose Γ has a conservative semilattice polymorphism $\psi : D^2 \rightarrow D$. Note that ψ defines a total ordering on $\{1, \dots, t\}$ so that $u \leq v$ if and only if $\psi(u, v) = u$. Define $\pi : D \rightarrow \{0, 1\}^t$ to be the following mapping

$$\pi(i) = (0, \dots, 0, \overbrace{1, \dots, 1}^i).$$

Let $\mathcal{P} = (X, D, C)$ denote an instance of $\text{CSP}(\Gamma)$ where $X = \{x_1, \dots, x_n\}$. Construct CSP instance $\mathcal{P}' = (X', \{0, 1\}, C')$ with $X' = \{x_{11}, \dots, x_{1t}, \dots, x_{n1}, \dots, x_{nt}\}$ with the following set of constraints

1. $x_{ij} \leq x_{ik}$ for all $1 \leq i \leq n$ and $1 \leq k \leq j \leq t$,
2. if $R(x_{i_1}, \dots, x_{i_k}) \in C$ then $\pi(R)(x_{i_1 1}, \dots, x_{i_1 t}, \dots, x_{i_k 1}, \dots, x_{i_k t}) \in C'$.

Observe that $\mathbf{a} \in \mathbf{V}(\mathbf{I}(\mathcal{P}))$ if and only if $\pi(\mathbf{a}) \in \mathbf{V}(\mathbf{I}(\mathcal{P}'))$. Given $f_0 \in \mathbb{F}[x_1, \dots, x_n]$, define $f'_0 \in \mathbb{F}[x_{11}, \dots, x_{1t}, \dots, x_{n1}, \dots, x_{nt}]$ to be the polynomial obtained from f_0 where we replace each indeterminate x_i with $x_{i1} + \dots + x_{it}$. It is easy to check that, for $\mathbf{a} \in \mathbf{V}(\mathbf{I}(\mathcal{P}))$, we have $f_0(\mathbf{a}) = 0$ if and only if $\pi(\mathbf{a}) \in \mathbf{V}(\mathbf{I}(\mathcal{P}'))$ and $f'_0(\pi(\mathbf{a})) = 0$. Therefore, deciding if $f_0 \in \mathbf{I}(\mathcal{P})$ is equivalent to deciding if $f'_0 \in \mathbf{I}(\mathcal{P}')$ where the later one is polynomial time solvable by Theorem 5.19.

5.4 Linear equations mod p

In this section we focus on constraint languages that are expressible as a system of linear equations modulo a prime number. Let Γ be a constraint language over a set D with $|D| = p$, and p a prime number. Suppose Γ has an affine polymorphism modulo p (i.e. a ternary operation $\psi(x, y, z) = x \oplus y \oplus z$, where \oplus, \ominus are addition and subtraction modulo p , or, equivalently, of the field $\text{GF}(p)$). In this case every CSP can be represented as a system of linear equations over $\text{GF}(p)$. Without loss of generality, we may assume that the system of linear equations at hand is already in the *reduced row echelon* form. Transforming system of linear equations mod p in its reduced row echelon form to a system of polynomials in $\mathbb{R}[X]$ that are a Gröbner Basis is not immediate and requires substantial work. This is the case even if we restrict ourselves to lexicographic order. Let us elaborate on this by an example considering linear equations over $\text{GF}(2)$.

Example 5.23. We assume a lexicographic order \succ with $x_1 \succ_{\text{lex}} \dots \succ_{\text{lex}} x_n$. We also assume that the linear system has $r \leq n$ equations and is already in its reduced row echelon form with x_i as the leading monomial of the i -th equation. Let $\text{Supp}_i \subset [n]$ such that $\{x_j : j \in \text{Supp}_i\}$ is the set of variables appearing in the i -th equation of the linear system except for x_i . Let the i -th equation be $g_i = 0 \pmod{2}$ where $g_i = x_i \oplus f_i$, with $i \in [r]$ and f_i is the Boolean function $(\bigoplus_{j \in \text{Supp}_i} x_j) \oplus \alpha_i$ and $\alpha_i \in \{0, 1\}$. Define a polynomial $M(f_i) \in \mathbb{R}[x_1, \dots, x_n]$ interpolating f_i , that is, such that, for every $\mathbf{a} \in \{0, 1\}^n$, $f_i(\mathbf{a}) = 0$ if and only if $M(f_i)(\mathbf{a}) = 0$, and $f_i(\mathbf{a}) = 1$ if and only if $M(f_i)(\mathbf{a}) = 1$. Now, consider the following set of polynomials.

$$G = \{x_1 - M(f_1), \dots, x_r - M(f_r), (x_{r+1}^2 - x_{r+1}), \dots, (x_n^2 - x_n)\}.$$

Set $G \subset \mathbb{R}[x_1, \dots, x_n]$ is a Gröbner Basis with respect to lex order. This is because for every pair of polynomials in G the reduced S -polynomial is zero as the leading monomials of any two polynomials in

G are relatively prime. By Buchberger's Criterion (see Theorem 5.10) it follows that G is a Gröbner Basis with respect to the lex ordering.

However, this construction may not be computationally efficient as the polynomials $M(f_i)$ may have exponentially many monomials. This case was overlooked in [28]. Bharathi and Mastrolilli [8] resolved this issue in an elegant way. Having $G' = \{x_1 - f_1, \dots, x_r - f_r, (x_{r+1}^2 - x_{r+1}), \dots, (x_n^2 - x_n)\}$, they convert G' to set of polynomials in $\mathbb{R}[x_1, \dots, x_n]$ which is a d -truncated Gröbner Basis. Their conversion algorithm has time complexity $n^{O(d)}$ where $d = O(1)$. Their algorithm is a modification of the conversion algorithm by Faugère, Gianni, Lazard and Mora [19]. See [8] for more details.

We consider this problem for any fixed prime p and prove that a d -truncated Gröbner Basis can be computed in time $n^{O(d)}$. First, we give a very brief introduction to the FGLM conversion algorithm [19]. Next, we present our conversion algorithm that, given a system of linear equations mod p , produces a d -truncated Gröbner Basis in graded lexicographic order. The heart of our algorithm is finding linearly independent expressions mod p that helps us carry the conversion.

5.4.1 Gröbner Basis conversion

We say a Gröbner Basis $G = \{g_1, \dots, g_t\}$ is *reduced* if $\text{LC}(g_i) = 1$ for all $g_i \in G$, and if for all $g_i \in G$ no monomial of g_i lies in $\langle \text{LT}(G \setminus \{g_i\}) \rangle$. We note that for an ideal and a given monomial ordering the reduced Gröbner Basis of I is unique (see, e.g., [17], Theorem 5, p.93). Given the reduced Gröbner Basis of a zero-dimensional ideal $I \subset \mathbb{F}[X]$ with respect to a monomial order \succ_1 , where \mathbb{F} is a computable field, the FGLM algorithm computes a Gröbner Basis of I with respect to another monomial order \succ_2 . The complexity of the FGLM algorithm depends on the dimension of \mathbb{F} -vector space $\mathbb{F}[X]/I$. More precisely, let $\mathcal{D}(I)$ denote the dimension of \mathbb{F} -vector space $\mathbb{F}[X]/I$, then we have the following proposition.

Proposition 5.24 (Proposition 4.1 in [19]). *Let I be a zero-dimensional ideal and G_1 be the reduced Gröbner Basis with respect to an ordering \succ_1 . Given a different ordering \succ_2 , there is an algorithm that constructs a Gröbner Basis G_2 with respect to ordering \succ_2 in time $O(n\mathcal{D}(I)^3)$.*

We cannot apply the FGLM algorithm directly as $\mathcal{D}(I)$ could be exponentially large in our setting. Note that $\mathcal{D}(I)$ is equal to the number of common zeros of the polynomials from $\langle G_1 \rangle$, which in the case of linear equations is $\mathcal{D}(I) = O(p^{n-r})$ where r is the number of equations in the reduced row echelon form. Furthermore, as we discussed, we are not given the explicit reduced Gröbner Basis G_1 with respect to lex ordering (the Gröbner Basis is presented to us as a system of linear equation mod p rather than polynomials in $\mathbb{R}[X]$). We shall present an algorithm that resolves these issues.

Let \mathcal{P} be an instance of $\text{CSP}(\Gamma)$ that is expressed as a system of linear equations \mathcal{S} over \mathbb{Z}_p with variables x_1, \dots, x_n . A system of linear equations over \mathbb{Z}_p can be solved by Gaussian elimination (this immediately tells us if $1 \in I(\mathcal{P})$ or not, and we proceed only if $1 \notin I(\mathcal{P})$). We assume a lexicographic order \succ_{lex} with $x_1 \succ_{\text{lex}} \dots \succ_{\text{lex}} x_n$. We also assume that the linear system has $r \leq n$ equations and it is already in its reduced row echelon form with x_i as the leading monomial of the i -th equation. Let $\text{Supp}_i \subset [n]$ such that $\{x_j : j \in \text{Supp}_i\}$ be the set of variables appearing in the i -th equation of the linear system except for x_i .

Fix a prime p and \oplus, \ominus, \odot denote addition, subtraction, and multiplication modulo p , respectively. We will call a linear polynomial over \mathbb{Z}_p a p -expression. Let the i -th equation be $g_i = 0 \pmod{p}$ where $g_i := x_i \ominus f_i$, with $i \in [r]$ and f_i is the p -expression $(\bigoplus_{j \in \text{Supp}_i} \alpha_j x_j) \oplus \alpha_i$ and

Algorithm 1 Conversion algorithm

Require: G_1 as in (11) that corresponds to $I(\mathcal{P})$, degree d .

- 1: Let Q be the list of all monomials of degree at most d arranged in increasing order with respect to grlex .
 - 2: $G_2 = \emptyset, B(G_2) = \{1\}$ (we assume $1 \notin I(\mathcal{P})$). Let b_i (arranged in increasing grlex order) be the elements of $B(G_2)$.
 - 3: **for** $q \in Q$ **do**
 - 4: **if** q is divisible by some LM in G_2 **then**
 - 5: Discard it and go to Step 3,
 - 6: **if** $q|_{G_1} = \sum_j k_j b_j|_{G_1}$ **then** \triangleright where $k_j \in \mathbb{R}, b_j \in B(G_2)$
 - 7: $G_2 = G_2 \cup \{q - \sum_j k_j b_j\}$
 - 8: **else**
 - 9: $B(G_2) = B(G_2) \cup \{q\}$
 - 10: **return** G_2
-

$\alpha_j, \alpha_i \in \mathbb{Z}_p$. We will assume that each variable x_i is associated with its p -expression f_i which comes from the mod p equations. This is clear for $i \leq r$; for $i > r$ the p -expression $f_i = x_i$ itself. Hence, we can write down the reduced Gröbner Basis in the lex order in an implicit form as follows.

$$G_1 = \{x_1 - f_1, \dots, x_r - f_r, \prod_{i \in \mathbb{Z}_p} (x_{r+1} - i), \dots, \prod_{i \in \mathbb{Z}_p} (x_n - i)\} \quad (11)$$

Given G_1 , our conversion algorithm, Algorithm 1, constructs a d -truncated Gröbner Basis over $\mathbb{R}[x_1, \dots, x_n]$ with respect to the grlex order. At the beginning of the algorithm, there will be two sets: G_2 , which is initially empty but will become the new Gröbner Basis with respect to the grlex order, and $B(G_2)$, which initially contains 1 and will grow to be the grlex monomial basis of the quotient ring $\mathbb{R}[x_1, \dots, x_n] / I(\mathcal{P})$ as a \mathbb{R} -vector space. In fact, $B(G_2)$ contains the reduced monomials (of degree at most d) with respect to G_2 . Every $f \in \mathbb{R}[x_1, \dots, x_n]$ is congruent modulo $I(\mathcal{P})$ to a unique polynomial r which is a \mathbb{R} -linear combination of the monomials in the *complement* of $\langle \text{LT}(I(\mathcal{P})) \rangle$. Furthermore, the elements of $\{x^\alpha \mid x^\alpha \notin \langle \text{LT}(I(\mathcal{P})) \rangle\}$ are "linearly independent modulo $I(\mathcal{P})$ " (see, e.g., Proposition 1 on page 248 of [17]). This suggests the following. In Algorithm 1, Q is the list of all monomials of degree at most d arranged in increasing order with respect to grlex ordering. The algorithm iterates over monomials in Q in increasing grlex order and at each iteration decides exactly one of the followings given the current sets G_2 and $B(G_2)$.

1. q should be discarded (if q is divisible by some LM in G_2), or
2. a polynomial with q as its leading monomial should be added to G_2 (if $q|_{G_1} = \sum_j k_j b_j|_{G_1}$; $b_j \in B(G_2)$), or
3. q should be added to $B(G_2)$.

The trickiest part is to decide if the current monomial $q|_{G_1}$ is a \mathbb{R} -linear combination of $b_j|_{G_1}$ with b_j being the current elements in $B(G_2)$. Provided this can be done correctly and in polynomial time, the correctness of Algorithm 1 follows by the analyses in [19] and it runs in polynomial

time as there are at most $O(n^d)$ monomials in Q . The rest of this section is devoted to provide a polynomial time procedure that correctly decides if $q|_{G_1} = \sum_j k_j b_j|_{G_1}$ holds for the current monomial q and the current b_j s in $B(G_2)$.

5.4.2 Expansion in a basis of p -expressions

For a monomial q , the normal form of q by G_1 , $q|_{G_1}$, is the remainder on division of q by G_1 in the lex order. $q|_{G_1}$ is unique and it does not matter how the elements of G_1 are listed when using the division algorithm. Here, it suffices for us to write $q|_{G_1}$ in terms of product of p -expressions. We start with a simple observation. Recall that r denotes the number of linear equations in the reduced row echelon form.

Observation 5.25. *Let $q = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ be a monomial such that for all $r < i$ we have $\alpha_i \leq p - 1$. Then, $q|_{G_1} = f_1^{\alpha_1} \cdots f_n^{\alpha_n}$ where each f_i is the p -expression associated to x_i .*

A keystone of our conversion algorithm is a relation between a product of p -expressions and a sum of p -expressions. Intuitively speaking, we will prove that a product of p -expressions can be written as a \mathbb{R} -linear combination of (linearly independent) p -expressions. Indeed, we provide a set of p -expressions and prove the p -expressions in this set are linearly independent and span the space of functions from \mathbb{Z}_p^d to \mathbb{C} . Consider the set $\mathcal{V}_{n,p}$ of functions from \mathbb{Z}_p^n to \mathbb{C} as a p^n -dimensional vector space, whose components are values of the function at the corresponding point of \mathbb{Z}_p^n . Let

$$F_n = \left\{ \bigoplus_{i=1}^n \alpha_i x_i \oplus x_{n+1} \oplus \beta \mid \alpha_i \in \{0, \dots, p-1\}, i \in [n], \beta \in \{0, \dots, p-2\} \right\}$$

be a collection of linear functions over \mathbb{Z}_p , and let

$$\mathcal{F}_n = F_n \cup \dots \cup F_0 \cup \{1\}.$$

Theorem 5.26. *For any n , the collection \mathcal{F}_n of p -expressions is linearly independent as a set of vectors from $\mathcal{V}_{n+1,p}$ and forms a basis of $\mathcal{V}_{n+1,p}$.*

As a first application of Theorem 5.26 we show that any p -expression can be written as a \mathbb{R} -linear combination of p -expressions in our basis \mathcal{F}_n .

Lemma 5.27. *Any p -expression $\alpha_1 x_1 \oplus \alpha_2 x_2 \oplus \dots \oplus \alpha_n x_n \oplus \beta$ with $\alpha_n \neq 0$ can be represented by a \mathbb{R} -linear combination of the p -expression basis from \mathcal{F}_{n-1} .*

Proof. By Theorem 5.26, for any x and y , the expression $y \oplus \alpha x \oplus \beta$ can be written as a \mathbb{R} -linear combination of p -expressions from \mathcal{F}_1 . Such a \mathbb{R} -linear combination can be found in constant time $p^{O(1)}$ as the number of functions in \mathcal{F}_1 is p^2 . We continue by assuming such a \mathbb{R} -linear combination of any $y \oplus \alpha x \oplus \beta$ is provided to us.

Now consider $\alpha_1 x_1 \oplus \alpha_2 x_2 \oplus \dots \oplus \alpha_n x_n \oplus \beta$. Introduce a new variable y and set $y = \alpha_1 x_1 \oplus \alpha_2 x_2 \oplus \dots \oplus \alpha_{n-1} x_{n-1}$. Hence, $\alpha_1 x_1 \oplus \alpha_2 x_2 \oplus \dots \oplus \alpha_n x_n \oplus \beta = y \oplus \alpha_n x_n \oplus \beta$. By the above discussion, $y \oplus \alpha_n x_n \oplus \beta$ can be written as a \mathbb{R} -linear combination of p -expressions from \mathcal{F}_1 . Therefore, there exist $c_\gamma, c_{\alpha\gamma}, \kappa \in \mathbb{R}$ so that

$$y \oplus \alpha_n x_n \oplus \beta = \sum_{\gamma=0}^{p-2} c_\gamma (y \oplus \gamma) + \sum_{\alpha \in [p-1], \gamma \in [p-2]} c_{\alpha\gamma} (\alpha y \oplus x_n \oplus \gamma) + \kappa \quad (12)$$

Note that p -expressions $y \oplus \gamma$ are in F_0 , p -expressions $\alpha y \oplus x_n \oplus \gamma$ are in F_1 , and κ is a constant. Substituting back for y , we observe that the second sum on the right hand side is already a \mathbb{R} -linear combination of p -expressions from F_{n-1} . Consider the first sum.

$$\sum_{\gamma=0}^{p-2} c_\gamma(y \oplus \gamma) = \sum_{\gamma=0}^{p-2} c_\gamma(\alpha_1 x_1 \oplus \alpha_2 x_2 \oplus \cdots \oplus \alpha_{n-1} x_{n-1} \oplus \gamma) \quad (13)$$

Now set $y = \alpha_1 x_1 \oplus \alpha_2 x_2 \oplus \cdots \oplus \alpha_{n-2} x_{n-2}$. This gives

$$\sum_{\gamma=0}^{p-2} c_\gamma(\alpha_1 x_1 \oplus \alpha_2 x_2 \oplus \cdots \oplus \alpha_{n-1} x_{n-1} \oplus \gamma) = \sum_{\gamma=0}^{p-2} c_\gamma(y \oplus \alpha_{n-1} x_{n-1} \oplus \gamma) \quad (14)$$

Note that each term $y \oplus \alpha_{n-1} x_{n-1} \oplus \gamma$ is expressible as a \mathbb{R} -linear combination of p -expressions from \mathcal{F}_1 . This leads us to the following.

$$\begin{aligned} \sum_{\gamma=0}^{p-2} c_\gamma(y \oplus \alpha_{n-1} x_{n-1} \oplus \gamma) &= \sum_{\gamma=0}^{p-2} c_\gamma \left(\sum_{\delta=0}^{p-2} c_\delta(y \oplus \delta) + \sum_{\alpha \in [p-1], \delta \in [p-2]} c_{\alpha\delta}(\alpha y \oplus x_{n-1} \oplus \delta) + \kappa' \right) \\ &= \sum_{\gamma=0}^{p-2} c_\gamma \left(\sum_{\delta=0}^{p-2} c_\delta(y \oplus \delta) \right) + \sum_{\gamma=0}^{p-2} c_\gamma \left(\sum_{\substack{\alpha \in [p-1], \\ \delta \in [p-2]}} c_{\alpha\delta}(\alpha y \oplus x_{n-1} \oplus \delta) \right) + \sum_{\gamma=0}^{p-2} c_\gamma \kappa' \\ &= \sum_{\gamma=0}^{p-2} c'_\gamma(y \oplus \gamma) + \sum_{\substack{\alpha \in [p-1], \\ \gamma \in [p-2]}} c'_{\alpha\gamma}(\alpha y \oplus x_{n-1} \oplus \gamma) + \kappa'' \end{aligned} \quad (15)$$

Note that p -expressions $y \oplus \gamma$ are in F_0 , p -expressions $\alpha y \oplus x_{n-1} \oplus \gamma$ are in F_1 , and κ'' is a constant. Substituting back for y , we observe that the second sum on the right hand side is already a \mathbb{R} -linear combination of p -expressions from F_{n-2} . Hence, it suffices to continue with the first term of the sum (15) which we can handle similar to the above procedure.

All in all, it requires to repeat the above procedure n times where at the i -th iteration we deal with a sum of $p-2$ p -expressions of form $\bigoplus_{i=1}^{n-i} \alpha_i x_i \oplus \beta$. This results in $O(np^{O(1)})$ running time. \square

Another application of Theorem 5.26 is transforming a multiplication of p -expressions to an equivalent \mathbb{R} -linear combination of the basis in \mathcal{F}_n . Suppose x_1, \dots, x_d are (not necessary distinct) variables that take values $0, \dots, p-1$. Let $x_1 \cdot x_2 \cdots x_d$ be their multiplication. In general, we are interested in a multiplication of p -expressions however, let us first discuss the simpler case of $x_1 \cdot x_2 \cdots x_d$. Unfortunately, the trick we used in the proof of Lemma 5.27 is no longer effective here. However, assuming d is a constant makes the situation easier. $x_1 \cdot x_2 \cdots x_d$ is a p^d -dimensional vector. By Theorem 5.26, the set \mathcal{F}_{d-1} of p -expressions spans the set $\mathcal{V}_{d,p}$ of functions from \mathbb{Z}_p^d to \mathbb{C} as a p^d -dimensional vector space. Hence, in constant time (depending on p and d), we can have a \mathbb{R} -linear combination of the basis in \mathcal{F}_{d-1} that represents $x_1 \cdot x_2 \cdots x_d$. We continue by assuming such a \mathbb{R} -linear combination of any $x_1 \cdot x_2 \cdots x_d$ is provided to us. The

next lemma states that we can have a \mathbb{R} -linear combination of the basis in \mathcal{F}_n for any $h_1 \cdot h_2 \cdots h_d$ where each h_i is a p -expression over variables x_1, \dots, x_n .

Lemma 5.28. *Let h_1, h_2, \dots, h_d be (not necessary distinct) p -expressions over variables x_1, \dots, x_n . The product $\mathcal{M} = h_1 \cdot h_2 \cdots h_d$ viewed as a function from \mathbb{Z}_p^n to \mathbb{C} can be represented as a \mathbb{R} -linear combination of the basis in \mathcal{F}_{n-1} .*

Proof. Let us treat h_i s as indeterminates. Define

$$H_t = \left\{ \bigoplus_{i=1}^t \alpha_i h_i \oplus h_{t+1} \oplus \beta \mid \alpha_i \in \{0, \dots, p-1\}, i \in [t], \beta \in \{0, \dots, p-2\} \right\}$$

to be a collection of linear functions over \mathbb{Z}_p , and let

$$\mathcal{H}_{d-1} = H_{d-1} \cup \dots \cup H_0 \cup \{1\}.$$

By Theorem 5.26 and the above discussion, \mathcal{M} can be written as a \mathbb{R} -linear combination of functions in \mathcal{H}_{d-1} . Therefore, there are coefficients $c_{\alpha_1 \dots \alpha_t \beta} \in \mathbb{R}$ and constant $\kappa \in \mathbb{R}$ so that

$$\mathcal{M} = \sum_{t=0}^{d-1} \sum_{\substack{\alpha_i \in [p-1], \\ \beta \in [p-2]}} c_{\alpha_1 \dots \alpha_t \beta} \left(\bigoplus_{i=1}^t \alpha_i h_i \oplus h_{t+1} \oplus \beta \right) + \kappa \quad (16)$$

Recall that each h_i is a p -expressions over variables x_1, \dots, x_n . By substituting back for each h_i and rearranging terms, each p -expression $\bigoplus_{i=1}^t \alpha_i h_i \oplus h_{t+1} \oplus \beta$ is equivalent to $\alpha'_1 x_1 \oplus \alpha'_2 x_2 \oplus \dots \oplus \alpha'_n x_n \oplus \beta'$ for some $\alpha'_1, \dots, \alpha'_n, \beta' \in [p-1]$. By Lemma 5.27, such expression can be written as a \mathbb{R} -linear combination of basis in \mathcal{F}_n .

Note that number of p -expressions in (16) is at most p^{d+1} . By Lemma 5.27, each p -expression can be written as a \mathbb{R} -linear combination of basis in \mathcal{F}_n in time $O(np^{O(1)})$. Therefore, in time $O(np^{O(d)})$ we can write $\mathcal{M} = h_1 \cdot h_2 \cdots h_d$ as a \mathbb{R} -linear combination of basis in \mathcal{F}_n which is polynomial in n for fixed p and d . \square

5.4.3 The correctness of the conversion algorithm

Now we have enough ingredients to prove Algorithm 1 runs in polynomial time and correctly decides if $q|_{G_1} = \sum_j k_j b_j|_{G_1}$ for every q . In the following theorem, suppose q is the current monomial considered by the algorithm. Furthermore, suppose sets G_2 and $B(G_2)$ have been constructed correctly so far.

Theorem 5.29. *Let $q = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ be a monomial of degree at most d . Suppose q is not divisible by any leading monomial of polynomials in the current set G_2 . Then, there exists a polynomial time algorithm that can decide whether $q|_{G_1} = \sum_j k_j b_j|_{G_1}$ where b_j are in the current set $B(G_2)$ in Algorithm 1.*

Proof. First, we discuss the case where for some $r < i$ we have $p-1 < \alpha_i$. Set $q = q' \cdot x_i^{\alpha_i}$ where $q' = x_1^{\alpha_1} \cdots x_{i-1}^{\alpha_{i-1}} \cdot x_{i+1}^{\alpha_{i+1}} \cdots x_n^{\alpha_n}$. Then $q|_{G_1} = q'|_{G_1} \cdot x_i^{\alpha_i}|_{G_1}$. Note that $x_i^{\alpha_i}|_{G_1}$ is a linear combination

of $x_i^{p-1}, x_i^{p-2}, \dots, x_i$. This is because the domain polynomial $\prod_{a \in \mathbb{Z}_p} (x_i - a)$ is in G_1 . Therefore,

$$\begin{aligned} q|_{G_1} &= q'|_{G_1} \cdot x_i^{\alpha_i}|_{G_1} \\ &= q'|_{G_1} \cdot (c_{p-1}x_i^{p-1} + \dots + c_1x_i) \\ &= (q' \cdot c_{p-1}x_i^{p-1})|_{G_1} + \dots + (q' \cdot c_1x_i)|_{G_1} \end{aligned} \quad (17)$$

All of $q' \cdot x_i^j$ in (17) have degree less than q , and hence they have been considered by Algorithm 1 before reaching q . Now, none of $q' \cdot x_i^j$ can be a multiple of the leading monomial of a polynomial in G_2 as otherwise q divides the leading monomial of a polynomial in G_2 . This implies that all $q' \cdot x_i^j$ with $c_j \neq 0$ in (17) are in $B(G_2)$ and we have $q|_{G_1} = \sum_j k_j b_j|_{G_1}$.

We continue by assuming for all $r < i$ we have $\alpha_i \leq p-1$. Note that if $q|_{G_1} = \sum_j k_j b_j|_{G_1}$, then $q|_{G_1} - \sum_j k_j b_j|_{G_1} = 0$ and hence $q - \sum_j k_j b_j \in I(\mathcal{P})$. We proceed by checking, in a systematic way, if there exist coefficients k_j such that $q|_{G_1} - \sum_j k_j b_j|_{G_1} = 0$ holds. We will construct a system of linear equations over \mathbb{R} for coefficients k_j so that this system has a solution if and only if $q|_{G_1} = \sum_j k_j b_j|_{G_1}$.

By Observation 5.25, we have $q|_{G_1} = f_1^{\alpha_1} \dots f_n^{\alpha_n}$ where each f_i is the p -expression associated to x_i . Similarly, for each b_j we have $b_j|_{G_1} = \mathcal{M}_j$ where $\mathcal{M}_j = h_{j1} \cdot h_{j2} \dots h_{jd}$ is a multiplication of at most d (not necessary distinct) p -expressions.

$$f_1^{\alpha_1} \dots f_n^{\alpha_n} = \sum_j k_j b_j|_{G_1} = \sum_j k_j \mathcal{M}_j \quad (18)$$

Recall that degree of q is at most d and, by Lemma 5.28, q can be written as a \mathbb{R} -linear combination of the basis in \mathcal{F}_{n-1} , say \mathcal{L}_q . Similarly, by Lemma 5.28, each product $\mathcal{M}_j = h_{j1} \cdot h_{j2} \dots h_{jd}$ can be written as a \mathbb{R} -linear combination of the basis in \mathcal{F}_{n-1} in polynomial time. Therefore, (18) is equivalent to

$$f_1^{\alpha_1} \dots f_n^{\alpha_n} = \mathcal{L}_q = \sum_j k_j b_j|_{G_1} = \sum_j k_j \mathcal{M}_j = \sum_j k_j \mathcal{L}_j \quad (19)$$

where each \mathcal{L}_j is a \mathbb{R} -linear combination equivalent to \mathcal{M}_j via p -expression basis in \mathcal{F}_n . Rearranging terms in $\sum_j k_j \mathcal{L}_j$ and \mathcal{L}_q yields

$$0 = \sum_j k_j \mathcal{L}_j - \mathcal{L}_q = \sum_j k'_j \mathcal{L}'_j \quad (20)$$

where each k'_j is linear combination of k_j s. Since \mathcal{F}_{n-1} is linearly independent and the left hand side of (20) is a constant we deduce that all k'_j should be zero. Hence, we are left with (possibly more than one) linear equations with respect to k_j over \mathbb{R} . Note that at this point there is not any term with a p -expression. If such a system has a (unique) solution for k_j then we conclude $q|_{G_1} = \sum_j k_j b_j|_{G_1}$, else the equality does not hold.

Note that, by Lemma 5.28, time complexity of finding a \mathbb{R} -linear combination of basis in \mathcal{F}_{n-1} for a multiplication of d p -expressions is $O(np^{O(d)})$. Moreover, we use Lemma 5.28 at most $O(n^{O(d)})$ many times. Hence, the whole process requires $O(n^{O(d)})$ time complexity. \square

Theorem 5.30. *Let Γ be a constraint language where each relation in Γ is expressed as a system of linear equations modulo a prime number p . Then, the $\text{IMP}_d(\Gamma)$ can be solved in polynomial time for fixed d and p .*

6 Omitted proofs from Section 5.4

In this section we give a proof of Theorem 5.26. Also, it turns out in the case $p = 3$ there is another basis of $\mathcal{V}_{n,p}$ that has a particularly clear structure. We give a proof of this result in Section 6.2.

6.1 Linear independence

Recall that we consider the set $\mathcal{V}_{k,p}$ of functions from \mathbb{Z}_p^k to \mathbb{C} as a p^k -dimensional vector space, whose components are values of the function at the corresponding point of \mathbb{Z}_p^k . Let

$$F_k = \left\{ \bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus \beta \mid \alpha_i \in \{0, \dots, p-1\}, i \in [n], \beta \in \{0, \dots, p-2\} \right\}$$

be a collection of linear functions over \mathbb{Z}_p , and let

$$\mathcal{F}_k = F_k \cup \dots \cup F_0 \cup \{1\}.$$

The result we are proving in this section is

Theorem 6.1. *For any k , the collection \mathcal{F}_k of p -expressions is linearly independent as a set of vectors from $\mathcal{V}_{k+1,p}$ and forms a basis of $\mathcal{V}_{k+1,p}$.*

6.1.1 Monomials and projections

As a vector space the set $\mathcal{V}_{k,p}$ has several natural bases. Let $\text{Mon}(k, p)$ denote the set of all monomials $u = x_1^{d_1} \dots x_k^{d_k}$ where $d_i \in \{0, \dots, p-1\}$; and let $\text{cont}(u)$ denote the set $\{i \mid d_i \neq 0\}$. For $\bar{a} \in \mathbb{Z}_p^k$ we denote by $r_{\bar{a}}$ the function given by $r_{\bar{a}}(\bar{a}) = 1$ and $r_{\bar{a}}(\bar{x}) = 0$ when $\bar{x} \neq \bar{a}$. We start with a simple observation.

Lemma 6.2. *The sets $\text{Mon}(k, p)$ and $R(k, p) = \{r_{\bar{a}} \mid \bar{a} \in \mathbb{Z}_p^k\}$ are bases of $\mathcal{V}_{k,p}$.*

As is easily seen, both sets contain p^k elements, $R(k, p)$ obviously spans $\mathcal{V}_{k,p}$. That $\text{Mon}(k, p)$ also spans $\mathcal{V}_{k,p}$ follows from polynomial interpolation properties.

Our goal here is to find yet another basis of $\mathcal{V}_{k,p}$ suitable for our needs, which is the set \mathcal{F}_k defined above.

In this section we view functions from \mathcal{F}_k as elements of $\mathcal{V}_{k+1,p}$. In particular, we will use coordinates of such functions in the bases $\text{Mon}(k+1, p)$ and $R(k+1, p)$. The latter is of course just the collection of values of a function in points from \mathbb{Z}_p^{k+1} , while the former is the polynomial interpolation of a function, which is unique when we restrict ourselves to polynomials of degree at most $p-1$ in each variable.

The number of functions in \mathcal{F}_k is

$$1 + \sum_{\ell=0}^k |F_\ell| = 1 + \sum_{\ell=0}^k p^\ell (p-1) = 1 + (p-1) \sum_{\ell=0}^k p^\ell = 1 + (p-1) \frac{p^{k+1} - 1}{p-1} = p^{k+1}.$$

Thus, we only need to prove that \mathcal{F}_k spans $\mathcal{V}_{k+1,p}$.

We will need a finer partition of sets F_k : for $S \subseteq [k]$ let F_k^S denote the set of functions $\bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus b$ such that $\alpha_i \neq 0$ if and only if $i \in S$.

We prove by induction on $|S|$, $S \subseteq [k+1]$, that any monomial u with $\text{cont}(u) = S$ is in the span of

$$\mathcal{F}_k^S = \bigcup_{\ell \leq k, T \subseteq S \cap [\ell]} F_\ell^T.$$

Since up to renaming the variables \mathcal{F}_k^S can be viewed as $\mathcal{F}_{|S|}$, it suffices to prove the result for $S = [k]$, and our inductive process is actually on k .

For $f \in F_k$ let f' denote the sum of all monomials u of f (with the same coefficients) for which $\text{cont}[u] = [k+1]$. In other words, f' can be viewed as a projection f onto the subspace \mathcal{V}_1 spanned by $\text{Mon}^*(k+1, p) = \{u \in \text{Mon}(k+1, p) \mid \text{cont}(u) = [k+1]\}$ parallel to the subspace \mathcal{V}_2 spanned by $\text{Mon}^+(k+1, p) = \text{Mon}(k+1, p) - \text{Mon}^*(k+1, p)$. Let $F'_k = \{f' \mid f \in F_k\}$. Note that \mathcal{V}_1 is also the subspace of $\mathcal{V}_{n,p}$ spanned by the set $\{r_{\bar{a}} \mid \bar{a} \in (\mathbb{Z}_p^*)^{k+1}\}$, and the dimensionality of \mathcal{V}_1 is $(p-1)^{k+1} = |F_k| = |F'_k|$. Since by the induction hypothesis $f - f'$ is in the span of \mathcal{F}_k , it suffices to prove that vectors in F'_k are linearly independent, and therefore generate \mathcal{V}_1 . This will be proved in the rest of this section.

Proposition 6.3. *The set F'_k is linearly independent.*

We prove Proposition 6.3 by constructing a matrix containing the values of functions from F'_k and find its rank by finding all its eigenvalues. We do it in three steps. First, let F_k^+ denote the superset of F_k that apart from functions from F_k also contains functions of the form $\bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus (p-1)$ for $\alpha_i \in \mathbb{Z}_p^*$, $i \in [k]$. Then, let N_k be the $p(p-1)^k \times p(p-1)^k$ -dimensional matrix whose rows are labeled with $(x_1, \dots, x_{k+1}) \in (\mathbb{Z}_p^*)^k \times \mathbb{Z}_p$ representing values of the arguments of functions from F_k^+ , and the columns are labeled with $f \in F_k^+$. The entry of N_k in row (x_1, \dots, x_{k+1}) and column f is $f(x_1, \dots, x_{k+1})$. In the next section we find the eigenvectors and eigenvalues of N_k . In the second step we use the properties of N_k to study the matrix N_k'' obtained from N_k by replacing every entry of the form $f(x_1, \dots, x_{k+1})$ with the value $f''(x_1, \dots, x_{k+1})$, where f'' is the sum of all the monomials u from f with $x_1, \dots, x_k \in \text{cont}(u)$. We again find the eigenvectors and eigenvalues of N_k'' . Finally, we transform N_k'' to obtain a new matrix N'_k in such a way that the entry of N'_k in the row (x_1, \dots, x_{k+1}) with $x_{k+1} \neq 0$ and column f equals $f'(x_1, \dots, x_{k+1})$. We then finally prove that all the rows of N'_k labeled (x_1, \dots, x_{k+1}) , $x_{k+1} \neq 0$, are linearly independent.

6.1.2 Kronecker sum and the eigenvalues of N_k

We first introduce some useful notation.

Let A, B be $q \times r$ and $s \times t$ matrices with entries in \mathbb{Z}_p . The *Kronecker sum* of A and B denoted $A \boxplus B$ is the $qs \times rt$ matrix whose entry in row $is + i'$ and column $jt + j'$ equals $A(i, j) \oplus B(i', j')$. In other words, $A \boxplus B$ is defined the same way as Kronecker product, except using addition modulo p rather than multiplication. We also use $A^{\boxplus k}$ to denote $A \boxplus \dots \boxplus A$.

We consider two matrices, matrix B_p essentially consists of values of unary functions αx on $[p-1]$, except that we rearrange its rows and columns as follows. Let a be a primitive residue modulo p , that is, a generator of \mathbb{Z}_p^* . Then

$$B_p = \begin{pmatrix} 1 & a & a^2 & a^3 & \dots & a^{p-2} \\ a^{p-2} & 1 & a & a^2 & \dots & a^{p-3} \\ a^{p-3} & a^{p-2} & 1 & a & \dots & a^{p-4} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ a & a^2 & a^3 & a^4 & \dots & 1 \end{pmatrix},$$

where a^i denotes exponentiation modulo p . Matrix C_p is again the operation table of addition modulo p with rearranged rows

$$C_p = \begin{pmatrix} 0 & 1 & 2 & \dots & p-1 \\ p-1 & 0 & 1 & \dots & p-2 \\ p-2 & p-1 & 0 & \dots & p-3 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 2 & 3 & \dots & 0 \end{pmatrix}.$$

As is easily seen, the values of functions $\bigoplus_{i=1}^k \alpha_i x_i$, $\alpha_1, \dots, \alpha_k \in \mathbb{Z}_p^*$ on $(\mathbb{Z}_p^*)^k$ can be viewed as $B^{\boxplus k}$; and those of $\bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus b$, $\alpha_1, \dots, \alpha_k \in \mathbb{Z}_p^*$, $b \in \mathbb{Z}_p$, on $x_1, \dots, x_k \in \mathbb{Z}_p^*$, $x_{k+1} \in \mathbb{Z}_p$ can be represented as $N_k = C_p \boxplus B_p^{\boxplus k}$. Next we find the eigenvectors and eigenvalues of N_k .

Recall that a square matrix of the form

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_n \\ a_n & a_1 & a_2 & \dots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \dots & a_{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_2 & a_3 & a_4 & \dots & a_1 \end{pmatrix} \quad (21)$$

is called *circulant*. The eigenvectors and eigenvalues of circulant matrices are well known. We give a brief proof of the following fact for the sake of completeness.

Lemma 6.4. *The eigenvectors of the matrix A in (21) have the form $\vec{v}_\xi = (1, \xi, \xi^2, \dots, \xi^{n-1})$ for n th roots of unity ξ . The eigenvalue corresponding to \vec{v}_ξ is $\mu_\xi = a_1 + a_2\xi + a_3\xi^2 + \dots + a_n\xi^{n-1}$.*

Proof. We compute $A \cdot \vec{v}_\zeta$

$$\begin{aligned}
A \cdot \vec{v}_\zeta &= \begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_n \\ a_n & a_1 & a_2 & \dots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \dots & a_{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_2 & a_3 & a_4 & \dots & a_1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \zeta \\ \zeta^2 \\ \vdots \\ \zeta^{n-1} \end{pmatrix} \\
&= \begin{pmatrix} a_1 + a_2\zeta + a_3\zeta^2 + \dots + a_n\zeta^{n-1} \\ a_n + a_1\zeta + a_2\zeta^2 + \dots + a_{n-1}\zeta^{n-1} \\ a_{n-1} + a_n\zeta + a_1\zeta^2 + \dots + a_{n-2}\zeta^{n-1} \\ \vdots \\ a_2 + a_3\zeta + a_4\zeta^2 + \dots + a_1\zeta^{n-1} \end{pmatrix} \\
&= (a_1 + a_2\zeta + a_3\zeta^2 + \dots + a_n\zeta^{n-1}) \begin{pmatrix} 1 \\ \zeta \\ \zeta^2 \\ \vdots \\ \zeta^{n-1} \end{pmatrix},
\end{aligned}$$

as required. \square

A generalization of circulant matrices is obtained by replacing entries of a circulant matrix by matrices. More precisely, a matrix of the form

$$A = \begin{pmatrix} A_1 & A_2 & A_3 & \dots & A_n \\ A_n & A_1 & A_2 & \dots & A_{n-1} \\ A_{n-1} & A_n & A_1 & \dots & A_{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ A_2 & A_3 & A_4 & \dots & A_1 \end{pmatrix}, \quad (22)$$

where A_1, \dots, A_n are square matrices of the same size is said to be *block-circulant*. Note that N_k can be viewed as a block-circulant matrix:

$$N_k = C_p \boxplus B_p^{\boxplus k} = \begin{pmatrix} B_p^{\boxplus k} & B_p^{\boxplus k} \oplus 1 & \dots & B_p^{\boxplus k} \oplus (p-1) \\ B_p^{\boxplus k} \oplus (p-1) & B_p^{\boxplus k} & \dots & B_p^{\boxplus k} \oplus (p-2) \\ \vdots & \vdots & & \vdots \\ B_p^{\boxplus k} \oplus 2 & B_p^{\boxplus k} \oplus 3 & \dots & B_p^{\boxplus k} \end{pmatrix}.$$

Similarly, $B_p^{\boxplus \ell}$ can also be viewed as a block-circulant matrix:

$$B_p^{\boxplus(\ell-1)} \boxplus B_p = \begin{pmatrix} B_p^{\boxplus(\ell-1)} \oplus 1 & B_p^{\boxplus(\ell-1)} \oplus a & \dots & B_p^{\boxplus(\ell-1)} \oplus a^{p-1} \\ B_p^{\boxplus(\ell-1)} \oplus a^{p-1} & B_p^{\boxplus(\ell-1)} \oplus 1 & \dots & B_p^{\boxplus(\ell-1)} \oplus a^{p-2} \\ \vdots & \vdots & & \vdots \\ B_p^{\boxplus(\ell-1)} \oplus a & B_p^{\boxplus(\ell-1)} \oplus a^2 & \dots & B_p^{\boxplus(\ell-1)} \oplus 1 \end{pmatrix}.$$

In some cases the eigenvectors and eigenvalues of block-circulant matrices can also be found.

Lemma 6.5. Let A be a block-circulant matrix with blocks A_1, \dots, A_n as in (22) such that A_1, \dots, A_n have the same eigenvectors. Then every eigenvector \vec{w} of A has the form $\vec{w}_{v,\xi} = (1, \xi, \dots, \xi^{n-1}) \otimes \vec{v}$, where ξ is an n th root of unity and \vec{v} is an eigenvector of A_1, \dots, A_n . Conversely, for every eigenvector \vec{v} of A_1, \dots, A_n and an n th root of unity ξ , $\vec{w}_{v,\xi}$ is an eigenvector of A . The eigenvalue of A associated with $\vec{w}_{v,\xi}$ is $\mu_{1,\vec{v}} + \mu_{2,\vec{v}}\xi + \mu_{3,\vec{v}}\xi^2 + \dots + \mu_{n,\vec{v}}\xi^{n-1}$, where $\mu_{i,\vec{v}}$ is the eigenvalue of A_i associated with \vec{v} .

Proof. As in the proof of Lemma 6.4 we compute $A \cdot \vec{w}_{v,\xi}$:

$$\begin{aligned} A \cdot \vec{w}_{v,\xi} &= \begin{pmatrix} A_1 & A_2 & A_3 & \dots & A_n \\ A_n & A_1 & A_2 & \dots & A_{n-1} \\ A_{n-1} & A_n & A_1 & \dots & A_{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ A_2 & A_3 & A_4 & \dots & A_1 \end{pmatrix} \cdot \begin{pmatrix} \vec{v} \\ \xi \vec{v} \\ \xi^2 \vec{v} \\ \vdots \\ \xi^{n-1} \vec{v} \end{pmatrix} \\ &= \begin{pmatrix} A_1 \vec{v} + \xi A_2 \vec{v} + \xi^2 A_3 \vec{v} + \dots + \xi^{n-1} A_n \vec{v} \\ A_n \vec{v} + \xi A_1 \vec{v} + \xi^2 A_2 \vec{v} + \dots + \xi^{n-1} A_{n-1} \vec{v} \\ A_{n-1} \vec{v} + \xi A_n \vec{v} + \xi^2 A_1 \vec{v} + \dots + \xi^{n-1} A_{n-2} \vec{v} \\ \vdots \\ A_2 \vec{v} + \xi A_3 \vec{v} + \xi^2 A_4 \vec{v} + \dots + \xi^{n-1} A_1 \vec{v} \end{pmatrix} \\ &= (\mu_{1,\vec{v}} + \xi \mu_{2,\vec{v}} + \xi^2 \mu_{3,\vec{v}} + \dots + \xi^{n-1} \mu_{n,\vec{v}}) \begin{pmatrix} \vec{v} \\ \xi \vec{v} \\ \xi^2 \vec{v} \\ \vdots \\ \xi^{n-1} \vec{v} \end{pmatrix}, \end{aligned}$$

as required. \square

We now apply these techniques to N_k . Let $\vec{v}(\eta) = (1, \eta, \eta^2, \dots, \eta^{p-1})$ for a $(p-1)$ th root of unity, and let \otimes denote Kronecker product.

Lemma 6.6. (1) For any k and any $i, j \in \mathbb{Z}_p$ the matrices $B_p^{\oplus k} \oplus i$ and $B_p^{\oplus k} \oplus j$ have the same eigenvectors, and every eigenvector has the form

$$\vec{v}(\eta_1, \dots, \eta_k) = \vec{v}(\eta_1) \otimes \dots \otimes \vec{v}(\eta_k)$$

for some $(p-1)$ th roots of unity.

(2) For any k and any $j \in \mathbb{Z}_p$ the eigenvalues of the matrix $B_p^{\oplus k} \oplus j$ are

$$\lambda(\eta_1, \dots, \eta_k; j) = \sum_{i_1, \dots, i_k=0}^{p-2} (a^{i_1} \oplus \dots \oplus a^{i_k} \oplus j) \eta_1^{i_1} \dots \eta_k^{i_k}$$

for $(p-1)$ th roots of unity η_1, \dots, η_k .

(3) Let $i_1, \dots, i_s \in [k]$ be such that if $\eta_i \neq 1$ then $i = i_r$, $r \in [s]$, and $s \neq 0$. Then

$$\lambda(\eta_1, \dots, \eta_k; j) = (-1)^{k-s} \lambda(\eta_{i_1}, \dots, \eta_{i_s}; j).$$

Proof. We proceed by induction on k to prove all three claims simultaneously. If $k = 1$ then as $B_p \oplus j$ is a circulant matrix whose first row is $(1 \oplus j, a \oplus j, a^2 \oplus j, \dots, a^{p-2} \oplus j)$, by Lemma 6.4 its eigenvalues have the form

$$\lambda(\eta_1) = (1 \oplus j) + (a \oplus j)\eta_1 + (a^2 \oplus j)\eta_1^2 + \dots + (a^{p-2} \oplus j)\eta_1^{p-2}$$

for a $(p-1)$ th root of unity η_1 , and the corresponding eigenvector is $\vec{v}(\eta_1) = (1, \eta_1, \eta_1^2, \dots, \eta_1^{p-1})$ regardless of j .

Now, suppose that the lemma is true for $k-1$. Also, suppose that every eigenvector of $B_p^{\oplus(k-1)} \oplus j$ has the form $\vec{v}(\eta_2, \dots, \eta_k)$ for $(p-1)$ th roots of unity η_2, \dots, η_k . Then $B_p^{\oplus k} \oplus j$ is a block-circulant matrix with the first row $(B_p^{\oplus(k-1)} \oplus 1 \oplus j, B_p^{\oplus(k-1)} \oplus a \oplus j, B_p^{\oplus(k-1)} \oplus a^2 \oplus j, \dots, B_p^{\oplus(k-1)} \oplus a^{p-1} \oplus j)$. As by the induction hypothesis the blocks in this row have the same eigenvectors, by Lemma 6.5 the eigenvalues of $B_p^{\oplus k} \oplus j$ have the form

$$\mu_{0,\vec{v}} + \mu_{1,\vec{v}}\eta_1 + \mu_{1,\vec{v}}\eta_1^2 + \dots + \mu_{p-2,\vec{v}}\eta_1^{p-2},$$

where \vec{v} is an eigenvector of $B^{\oplus(k-1)}$ and $\mu_{i,\vec{v}}$ is the eigenvalue of $B^{\oplus(k-1)} \oplus a^i \oplus j$ associated with \vec{v} . Thus, plugging in the inductive hypothesis we obtain the result.

To prove item (3) we need to inspect the case when $\eta_k = 1$. In this case

$$\begin{aligned} \lambda(\eta_1, \dots, \eta_k) &= \sum_{i_1, \dots, i_k=0}^{p-2} (a^{i_1} \oplus \dots \oplus a^{i_k} \oplus j) \eta_1^{i_1} \dots \eta_k^{i_k} \\ &= \sum_{i_1, \dots, i_k=0}^{p-2} (a^{i_1} \oplus \dots \oplus a^{i_k} \oplus j) \eta_1^{i_1} \dots \eta_{k-1}^{i_{k-1}} \cdot 1 \\ &= \sum_{i_1, \dots, i_{k-1}=0}^{p-2} \eta_1^{i_1} \dots \eta_{k-1}^{i_{k-1}} \left(\sum_{i_k=0}^{p-2} (a^{i_1} \oplus \dots \oplus a^{i_k} \oplus j) \right) \\ &= \sum_{i_1, \dots, i_{k-1}=0}^{p-2} \eta_1^{i_1} \dots \eta_{k-1}^{i_{k-1}} \left(\frac{p(p-1)}{2} - (a^{i_1} \oplus \dots \oplus a^{i_{k-1}} \oplus j) \right) \\ &= \frac{p(p-1)}{2} \sum_{i_1, \dots, i_{k-1}=0}^{p-2} \eta_1^{i_1} \dots \eta_{k-1}^{i_{k-1}} - \lambda(\eta_1, \dots, \eta_{k-1}) \\ &= -\lambda(\eta_1, \dots, \eta_{k-1}). \end{aligned}$$

The last equality is due to fact that

$$\sum_{i_1, \dots, i_{k-1}=0}^{p-2} \eta_1^{i_1} \dots \eta_{k-1}^{i_{k-1}} = \sum_{i_1=0}^{p-2} \eta_1^{i_1} \dots \sum_{i_{k-1}=0}^{p-2} \eta_{k-1}^{i_{k-1}} = 0.$$

By the induction hypothesis the result follows. \square

Since the matrix N_k can be represented as $C_p \boxplus B_p^{\oplus k}$, its eigenvalues can be found by Lemma 6.5.

Lemma 6.7. *The eigenvalues of N_k can be represented in one of the following forms.*

- for a p th root of unity ξ and $(p-1)$ th roots of unity η_1, \dots, η_k

$$\mu(\eta_1, \dots, \eta_k; \xi) = \sum_{j=0}^{p-1} \xi^j \sum_{i_1, \dots, i_k=0}^{p-2} (a^{i_1} \oplus \dots \oplus a^{i_k} \oplus j) \eta_1^{i_1} \dots \eta_k^{i_k}.$$

- for a p th root of unity ξ and $(p-1)$ th roots of unity η_1, \dots, η_k

$$\mu(\eta_1, \dots, \eta_k; \xi) = P(\xi) \cdot Q(\eta_1, \xi) \cdot \dots \cdot Q(\eta_k, \xi),$$

where $P(\xi) = \frac{p}{\xi-1}$ unless $\xi = 1$, in which case $P(1) = \frac{p(p-1)}{2}$, and

$$Q(\eta, \xi) = \sum_{i=0}^{p-2} \eta^i \xi^{a^i}.$$

Proof. (1) By Lemma 6.5 the eigenvalues of N_k have the form $\lambda_{0, \vec{v}} + \lambda_{1, \vec{v}} \xi + \lambda_{2, \vec{v}} \xi^2 + \dots + \lambda_{p-1, \vec{v}} \xi^{p-1}$, where \vec{v} is an eigenvector of $B_p^{\boxplus k} \oplus i$ for all $i \in \mathbb{Z}_p$, and $\lambda_{i, \vec{v}}$ is the eigenvalue of $B_p^{\boxplus k} \oplus i$ associated with \vec{v} , and ξ is a p th root of unity. By Lemma 6.6 we obtain item (1) of the lemma.

(2) Consider the values $a^{i_1} \oplus \dots \oplus a^{i_k} \oplus j$ in the formula from part (1). For $j = 0, \dots, p-1$ they constitute the set \mathbb{Z}_p regardless of i_1, \dots, i_k , and the sequence, when j grows from 0 to $p-1$, is a sequence of consequent residues modulo p . Therefore

$$\sum_{j=0}^{p-1} (a^{i_1} \oplus \dots \oplus a^{i_k} \oplus j) \xi^j = \xi^{a^{i_1} \oplus \dots \oplus a^{i_k}} \sum_{j=0}^{p-1} j \xi^j,$$

and let $P(\xi) = \sum_{j=0}^{p-1} j \xi^j$. Therefore by part (1)

$$\begin{aligned} \mu(\eta_1, \dots, \eta_k; \xi) &= \sum_{j=0}^{p-1} \xi^j \sum_{i_1, \dots, i_k=0}^{p-2} (a^{i_1} \oplus \dots \oplus a^{i_k} \oplus j) \eta_1^{i_1} \dots \eta_k^{i_k} \\ &= \sum_{i_1, \dots, i_k=0}^{p-2} \xi^{a^{i_1} \oplus \dots \oplus a^{i_k}} P(\xi) \eta_1^{i_1} \dots \eta_k^{i_k} \\ &= P(\xi) \sum_{i_1, \dots, i_k=0}^{p-2} (\eta_1^{i_1} \xi^{a^{i_1}}) \cdot \dots \cdot (\eta_k^{i_k} \xi^{a^{i_k}}) \\ &= P(\xi) \left(\sum_{i_1=0}^{p-2} \eta_1^{i_1} \xi^{a^{i_1}} \right) \cdot \dots \cdot \left(\sum_{i_k=0}^{p-2} \eta_k^{i_k} \xi^{a^{i_k}} \right) \\ &= P(\xi) \cdot Q(\eta_1, \xi) \cdot \dots \cdot Q(\eta_k, \xi). \end{aligned}$$

Finally, we show that $P(\xi)$ has the required form. Since $\sum_{j=0}^{p-1} \xi^j = 0$, we have $P(\xi) = P'(\xi)$,

where $P'(x) = \sum_{j=0}^{p-1} (j+1)x^j$. Then

$$\begin{aligned}
\sum_{j=0}^{p-1} (j+1)x^j &= \frac{d}{dx} \left(\sum_{j=0}^{p-1} x^{j+1} \right) \\
&= \frac{d}{dx} \left(\frac{x^{p+1} - x}{x - 1} \right) \\
&= \frac{((p+1)x^p - 1)(x - 1) - (x^{p+1} - x)}{(x - 1)^2} \\
&= \frac{px^{p+1} - (p+1)x^p + 1}{(x - 1)^2}.
\end{aligned}$$

Since ξ is a p th root of unity, if $\xi \neq 1$ we have

$$P(\xi) = \frac{p\xi - (p+1) + 1}{(\xi - 1)^2} = \frac{p}{\xi - 1}.$$

Finally, $P(1) = \frac{p(p-1)}{2}$, as is easily seen. □

Clearly, the co-rank of N_k equals the multiplicity of the eigenvalue 0. Thus, we need to find the number of combinations of $\xi, \eta_1, \dots, \eta_k$ such that $\mu(\eta_1, \dots, \eta_k; \xi) = 0$. For some of them it is easy.

Lemma 6.8. *If $\eta_i \neq 1$ for some $i \in [k]$ then $\mu(\eta_1, \dots, \eta_k; 1) = 0$.*

Proof. We use Lemma 6.7. Let $\xi = 1$, and, say, $\eta_1 \neq 1$. Then

$$Q(\eta_1, 1) = \sum_{i=0}^{p-2} \eta_1^i = \frac{\eta_1^{p-1} - 1}{\eta_1 - 1} = 0,$$

as η_1 is a $(p-1)$ th root of unity and $\eta_1 \neq 1$. □

Lemma 6.9. *Let $\xi \neq 1$ be a p th root of unity and η a $(p-1)$ th root of unity. Then $Q(\eta, \xi) \neq 0$.*

Proof. Let χ be a primitive $p(p-1)$ th root of unity. Then η, ξ can be represented as $\eta = \chi^{up}$, $\xi = \chi^{v(p-1)}$ and $Q(\eta, \xi)$ can be rewritten as

$$Q^*(\chi) = \sum_{j=1}^{p-1} \chi^{jup + a^j v(p-1)}.$$

Note that all the arithmetic operations in the exponent including a^j can be treated as regular ones rather than modular, as $\chi^b = \chi^c$ whenever $b \equiv c \pmod{p(p-1)}$. Therefore if there are $\eta, \xi, \xi \neq 1$ such that $Q(\eta, \xi) = 0$, then there exists a primitive $p(p-1)$ th root of unity χ that is also a root of the polynomial

$$Q^*(x) = \sum_{j=1}^{p-1} x^{jup + a^j v(p-1)}.$$

This means that $Q^*(x)$ is divisible by $p(p-1)$ cyclotomic polynomial $C_{p(p-1)}$. The degree of $C_{p(p-1)}$ equals $\varphi(p(p-1))$, where φ is Euler's totient function. In particular, the degree of $C_{p(p-1)}$ is divisible by $p-1$, and so is the degree of Q^* . Since a and $p-1$ are relatively prime with p , it is only possible if u is divisible by $p-1$, that is, $\eta = 1$, in which case, as is easily seen, $Q(1, \xi) = -1$ if $\xi \neq 1$ and $Q(1, 1) = p-1$. \square

The next proposition follows from Lemma 6.9 and the observation that $P(\xi) \neq 0$ whenever ξ is a p th root of unity.

Proposition 6.10. *The rank of N_k is $(p-1)^k + 1$.*

6.1.3 Changed matrices

In this subsection we make the second step in our proof.

Lemma 6.11. *Let $f = \bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus b$ then*

$$f'' = \sum_{S \subseteq [k]} (-1)^{k-|S|} \left(\bigoplus_{i \in S} \alpha_i x_i \oplus x_{k+1} \oplus b \right). \quad (23)$$

Proof. We need to show that $f''(x_1, \dots, x_{k+1}) = 0$ whenever $x_i = 0$ for some $i \in [k]$. In order to do that observe that the terms in (23) can be paired up so that every S containing i is paired with $S - \{i\}$. Then $\bigoplus_{i \in S} \alpha_i x_i \oplus x_{k+1} \oplus b$ and $\bigoplus_{i \in S - \{i\}} \alpha_i x_i \oplus x_{k+1} \oplus b$ appear in (23) with opposite signs, and, as $x_i = 0$ are equal. \square

Let N_k'' denote the matrix constructed the same way as N_k only with f'' , $f \in F^\dagger$, in place of f . More precisely, N_k'' is the $p(p-1)^k \times p(p-1)^k$ -dimensional matrix whose rows are labeled with $(x_1, \dots, x_{k+1}) \in (\mathbb{Z}_p^*)^k \times \mathbb{Z}_p$ representing values of the arguments of functions from F_k^\dagger , and the columns are labeled with $f \in F_k^\dagger$. The entry of N_k'' in row (x_1, \dots, x_{k+1}) and column f is $f''(x_1, \dots, x_{k+1})$.

Using Lemma 6.11 we represent N_k'' as a sum of matrices. Let $f = \bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus b \in F_k^\dagger$ and $S \subseteq [k]$. Then let f_S denote the function $\bigoplus_{i \in S} \alpha_i x_i \oplus x_{k+1} \oplus b$. In other words, by Lemma 6.11

$$f = \sum_{S \subseteq [k]} (-1)^{k-|S|} f_S.$$

By $N_k(S)$, $S \subseteq [k]$, we denote the matrix constructed in a similar way to N_k and N_k'' . Again, its rows are labeled with $(x_1, \dots, x_{k+1}) \in (\mathbb{Z}_p^*)^k \times \mathbb{Z}_p$, and the columns are labeled with $f \in F_k^\dagger$. The entry of $N_k(S)$ in row (x_1, \dots, x_{k+1}) and column f is $f_S(x_1, \dots, x_{k+1})$. It is now easy to see that

$$N_k'' = \sum_{S \subseteq [k]} (-1)^{k-|S|} N_k(S).$$

In order to determine the structure of $N_k(S)$ we need one further observation. Let $\mathbf{0}_\ell$ denote the square ℓ -dimensional matrix whose entries are all 0. Note that for a matrix B and $\mathbf{0}_\ell$

$$B \boxplus \mathbf{0}_\ell = \begin{pmatrix} B & \dots & B \\ \vdots & & \vdots \\ B & \dots & B \end{pmatrix}.$$

Lemma 6.12. Let B an n -dimensional diagonalizable matrix. Then the eigenvectors of $B \boxplus \mathbf{0}_\ell$ are of the form $(\beta_1 \vec{v}, \dots, \beta_\ell \vec{v})$ where v is an eigenvector of B and either $\beta_1 = \dots = \beta_\ell$ or $\beta_1 + \dots + \beta_\ell = 0$. The corresponding eigenvalue in the former case is $\ell\lambda$, where λ is the eigenvalue of B associated with \vec{v} , and 0 in the latter case.

The following lemma establishes the structure of $N_k(S)$, its eigenvalues and eigenvectors.

Lemma 6.13. (a) $N_k(S) = C_p \boxplus D_1 \boxplus \dots \boxplus D_k$, where

$$D_i = \begin{cases} B_p, & \text{if } i \in S, \\ \mathbf{0}_{p-1}, & \text{otherwise.} \end{cases}$$

(b) Every eigenvector of N_k is also an eigenvector of $N_k(S)$.

(c) The eigenvalue $\mu(\eta_1, \dots, \eta_k; S)$ of $N_k(S)$ associated with eigenvector $\vec{v}(\eta_1, \dots, \eta_k; \xi)$ equals

$$\mu(\eta_1, \dots, \eta_k; S) = \begin{cases} 0 & \text{if } \eta_i \neq 1 \text{ for some } i \in [k] - S, \\ \mu(1, \dots, 1; 1), & \text{if } \xi = \eta_1 = \dots = \eta_k = 1, \\ (1-p)^{|[k]-S|} \mu(\eta_1, \dots, \eta_k; \xi), & \text{otherwise.} \end{cases}$$

Proof. We will construct the matrix $N_k(S)$ inductively and prove the three claims of the lemma as we go. Let $N_k(S, \ell)$, $\ell \leq k$, denote the $(p-1)^\ell \times (p-1)^\ell$ -matrix whose rows are labeled with $(x_1, \dots, x_\ell) \in (\mathbb{Z}_p^*)^\ell$, columns are labeled with functions $f = \bigoplus_{i=1}^\ell \alpha_i x_i$. The entry of $N_k(S, \ell)$ in row (x_1, \dots, x_ℓ) and column f is $f_S(x_1, \dots, x_\ell)$. We show that

(a') $N_k(S, \ell) = D_1 \boxplus \dots \boxplus D_\ell$, where the D_i 's are defined as in the lemma.

(b') Every vector of the form $\vec{v}(\eta_1, \dots, \eta_\ell) = \vec{v}(\eta_1) \otimes \dots \otimes \vec{v}(\eta_\ell)$, where η_i is a $(p-1)$ th root of unity is an eigenvector of $N_k(S, \ell) \oplus j$ for $j \in \mathbb{Z}_p$.

(c') The eigenvalue $\mu(\eta_1, \dots, \eta_\ell, S, j)$ of $N_k(S, \ell) \oplus j$ associated with eigenvector $\vec{v}(\eta_1, \dots, \eta_\ell)$ equals

$$\begin{aligned} & \mu(\eta_1, \dots, \eta_\ell, S, j) \\ &= \begin{cases} (p-1)^\ell \cdot j & \text{if } [\ell] \cap S = \emptyset \text{ and } \eta_1 = \dots = \eta_\ell, \\ 0, & \text{if } \eta_i \neq 1 \text{ for some } i \in [\ell] - S, \\ (p-1)^{|\ell|-S|} \lambda(\eta_{i_1}, \dots, \eta_{i_t}; j), & \text{otherwise (see Lemma 6.6),} \end{cases} \\ & \text{where } \{j_1, \dots, j_t\} = [\ell] \cap S. \end{aligned}$$

If $\ell = 1$ then either $N_k(S, 1) = B_p$ if $1 \in S$, or $N_k(S, 1) = \mathbf{0}_{p-1}$ if $1 \notin S$. In the former case we have the result by Lemma 6.6, and in the latter case by Lemma 6.12 every vector of the form $\vec{v}(\eta)$, η is a $(p-1)$ th root of unity, is an eigenvector with eigenvalue $(p-1)j$ if $\eta = 1$ and 0 otherwise.

Suppose the statement is true for some ℓ . If $[\ell+1] \cap S = \emptyset$, the claim is straightforward, as $f_S(x_1, \dots, x_{\ell+1}) = 0$ for all $x_1, \dots, x_{\ell+1} \in \mathbb{Z}_p^*$, and the result follows by Lemma 6.12.

Next, suppose that $[\ell] \cap S \neq \emptyset$, but $\ell+1 \notin S$. In this case the entry of $N_k(S, \ell+1)$ indexed with row $(x_1, \dots, x_{\ell+1})$ and column $f = \bigoplus_{i=1}^{\ell+1} \alpha_i x_i$ is $f_S(x_1, \dots, x_{\ell+1}) = f_S^*(x_1, \dots, x_\ell) = \bigoplus_{i \in S \cap [\ell]} \alpha_i x_i$, where $f^* = \bigoplus_{i=1}^\ell \alpha_i x_i$. This implies that $N_k(S, \ell+1) \oplus j = (N_k(S, \ell) \oplus j) \boxplus \mathbf{0}_{p-1}$. By Lemma 6.12 the eigenvectors of $N_k(S, \ell+1) \oplus j$ are of the two types: $\vec{v}' = (\vec{v}, \dots, \vec{v})$ or

$(\beta_1 \vec{v}, \dots, \beta_{p-1} \vec{v})$ with $\beta_1 + \dots + \beta_{p-1} = 0$, where \vec{v} is an eigenvector of $N_k(S, \ell) \oplus j$. In the first case $\vec{v}' = \vec{v} \otimes \vec{v}(\eta)$ for $\eta = 1$ and by the induction hypothesis has the required form. The corresponding eigenvalue of $N_k(S, \ell + 1) \oplus j$ equals $(p-1)\lambda$, where λ is the eigenvalue of $N_k(S, \ell) \oplus j$ associated with \vec{v} , and so also has the required form. In the latter case $\vec{v} \otimes (1, \eta, \dots, \eta^{p-2})$, $\eta \neq 1$ and \vec{v} is an eigenvector of $N_k(S, \ell) \oplus j$, satisfies the condition $1 + \eta + \dots + \eta^{p-2} = 0$ and has eigenvalue 0. By the induction hypothesis and Lemma 6.6(3) we get the result.

Finally, let $\ell + 1 \in S$. In this case for any $x_1, \dots, x_{\ell+1} \in \mathbb{Z}_p^*$ and $f = \bigoplus_{i=1}^{\ell+1} \alpha_i x_i$ the entry of $N_k(S, \ell + 1)$ equals

$$f_S(x_1, \dots, x_{\ell+1}) = \bigoplus_{i=1}^{\ell+1} \alpha_i x_i = f_S^*(x_1, \dots, x_\ell) \oplus \alpha_{\ell+1} x_{\ell+1},$$

which implies $N_k(S, \ell + 1) = N_k(S, \ell) \boxplus B_p$ proving (a'). Therefore $N_k(S, \ell + 1) \oplus j$ is a block-circulant matrix and we can apply Lemma 6.5 to show that eigenvectors of $N_k(S, \ell + 1)$ are of the form

$$\vec{v} \otimes (1, \eta, \dots, \eta^{p-2}) = \vec{v} \otimes \vec{v}(\eta),$$

where η is a $(p-1)$ th root of unity and \vec{v} is any eigenvector of $N_k(S, \ell) \oplus j$. The eigenvalue of such a vector can be found using the inductive hypothesis and the last part of the proof of Lemma 6.6 as follows. We have

$$\mu(\eta_1, \dots, \eta_{\ell+1}, S, j) = \mu_{0, \vec{v}} + \mu_{1, \vec{v}} \eta_{\ell+1} + \mu_{1, \vec{v}} \eta_{\ell+1}^2 + \dots + \mu_{p-2, \vec{v}} \eta_{\ell+1}^{p-2},$$

where \vec{v} is an eigenvector of $N_k(S, \ell)$ and $\mu_{i, \vec{v}}$ is the eigenvalue of $N_k(S, \ell) \oplus a^i \oplus j$ associated with \vec{v} . If there is $i \in S \cap [\ell]$ such that $\eta_i \neq 1$, then $\mu(\eta_1, \dots, \eta_{\ell+1}, S, j) = 0$. If $S \cap [\ell] = \emptyset$ and $\eta_1 = \dots = \eta_\ell = 1$ then

$$\mu(\eta_1, \dots, \eta_{\ell+1}, S, j) = \sum_{i=0}^{p-2} (p-1)^\ell (a^i \oplus j) \eta_{\ell+1}^i = (p-1)^\ell \lambda(\eta_{\ell+1}; j).$$

If $S \cap [\ell] = \{i_1, \dots, i_t\} \neq \emptyset$, then by the induction hypothesis

$$\begin{aligned} \mu(\eta_1, \dots, \eta_{\ell+1}, S, j) &= \sum_{i=0}^{p-2} (p-1)^{|\ell|-|S|} \lambda(\eta_{i_1}, \dots, \eta_{i_t}; a^i \oplus j) \eta_{\ell+1}^i \\ &= (p-1)^{|\ell|-|S|} \sum_{i=0}^{p-2} \sum_{j_1, \dots, j_t=0}^{p-2} (a^{j_1} \oplus \dots \oplus a^{j_t} \oplus a^i \oplus j) \eta_{i_1} \dots \eta_{i_t} \eta_{\ell+1}^i \\ &= (p-1)^{|\ell|-|S|} \lambda(\eta_{i_1}, \dots, \eta_{i_t}, \eta_{\ell+1}; j). \end{aligned}$$

We now consider the last step in constructing $N_k(S)$, from $N_k(S, k)$ to $N_k(S)$. As is easily seen, $N_k(S) = C_p \boxplus N_k(S, k)$, implying item (a) of the lemma, and by Lemma 6.5 every vector of the form $(1, \xi, \dots, \xi^{p-1}) \otimes \vec{v}$, where \vec{v} is an eigenvector of $N_k(S, k)$ and ξ is a p th root of unity is an eigenvector of $N_k(S)$. By the induction hypothesis this implies item (b) of the lemma. Finally, again by Lemma 6.5 and the induction hypothesis the eigenvalue associated with the vector $(1, \xi, \dots, \xi^{p-1}) \otimes \vec{v}(\eta_1, \dots, \eta_k)$ equals

$$\begin{aligned} &\mu(\eta_1, \dots, \eta_k; \xi; S) \\ &= \mu(\eta_1, \dots, \eta_k, S, 0) + \mu(\eta_1, \dots, \eta_k, S, 1) \xi + \dots + \mu(\eta_1, \dots, \eta_k, S, p-1) \xi^{p-1}. \end{aligned}$$

If $S = \emptyset$, $\xi \neq 1$, and $\eta_1 = \dots = \eta_k = 1$ then

$$\begin{aligned}\mu(1, \dots, 1; \xi; \emptyset) &= \sum_{j=0}^{p-1} (p-1)^k j \xi^j = (p-1)^k P(\xi) \\ &= (p-1)^k (-1)^k \mu(1, \dots, 1; \xi) \\ &= (1-p)^k \mu(1, \dots, 1; \xi),\end{aligned}$$

as $Q(1, \xi) = -1$, as is easily seen. Also,

$$\mu(1, \dots, 1; 1; \emptyset) = (p-1)^k P(1) = \mu(1, \dots, 1; 1).$$

If $\eta_i \neq 1$ for some $i \in [k] - S$ then $\mu(\eta_1, \dots, \eta_k, S, j) = 0$, and so $\mu(\eta_1, \dots, \eta_k; \xi; S) = 0$. Otherwise if $S = \{i_1, \dots, i_s\}$,

$$\begin{aligned}\mu(\eta_1, \dots, \eta_k; \xi; S) &= (p-1)^{|[k]-S|} \sum_{j=0}^{p-1} \lambda(\eta_{i_1}, \dots, \eta_{i_s}; j) \xi^j \\ &= (p-1)^{|[k]-S|} \mu(\eta_{i_1}, \dots, \eta_{i_s}; \xi)\end{aligned}$$

Finally, by Lemma 6.6 $\mu(\eta_{i_1}, \dots, \eta_{i_s}; \xi) = (-1)^{|[k]-S|} \mu(\eta_1, \dots, \eta_k; \xi)$, if $\xi \neq 1$, $\mu(\eta_{i_1}, \dots, \eta_{i_s}; 1) = 0$ if $\eta_{i_j} \neq 1$ for some j , and $\mu(1, \dots, 1; 1) = (p-1)^{|S|} P(1)$, and the result follows. \square

Now, we are ready to find the eigenvalues of N_k'' .

Lemma 6.14. *Let $\vec{w} = \vec{v}(\eta_1, \dots, \eta_k, \xi)$ and $T \subseteq [k]$ be such that $i \in T$ iff $\eta_i \neq 1$. Then*

$$\mu''(\eta_1, \dots, \eta_k; \xi) = \begin{cases} p^{k-|T|} \mu(\eta_1, \dots, \eta_k; \xi), & \text{if } \xi \neq 1, \\ 0, & \text{if } \xi = 1 \text{ and } \eta_i \neq 1 \text{ for some } i \in [k]. \end{cases}$$

Proof. Assume first that $\xi \neq 1$. By Lemmas 6.11 and 6.13 we have

$$\begin{aligned}\mu''(\eta_1, \dots, \eta_k; \xi) &= \sum_{S \subseteq [k]} (-1)^{|[k]-S|} \mu(\eta_1, \dots, \eta_k; \xi; S) \\ &= \sum_{[k] \supseteq S \supseteq T} (-1)^{|[k]-S|} \mu(\eta_1, \dots, \eta_k; \xi; S) \\ &= \sum_{\ell=0}^{k-|T|} (-1)^\ell \binom{k-|T|}{\ell} (1-p)^\ell \mu(\eta_1, \dots, \eta_k; \xi) \\ &= \mu(\eta_1, \dots, \eta_k; \xi) \sum_{\ell=0}^{k-|T|} \binom{k-|T|}{\ell} (p-1)^\ell \\ &= \mu(\eta_1, \dots, \eta_k; \xi) p^{k-|T|},\end{aligned}$$

as required.

Now let $\xi = 1$. If $T \neq \emptyset$, then $\mu(\eta_1, \dots, \eta_k; \xi; S) = 0$ for any $S \subseteq [k]$. Otherwise, we have

$$\begin{aligned}
\mu'(\eta_1, \dots, \eta_k; \xi) &= \sum_{T \subseteq [k]} (-1)^{k-|T|} \mu(1, \dots, 1; 1; T) \\
&= \sum_{T \subseteq [k]} (-1)^{k-|T|} \mu(1, \dots, 1; 1) \\
&= \mu(1, \dots, 1; 1) \sum_{\ell=0}^{k-1} (-1)^\ell \binom{k}{\ell} \\
&= 0.
\end{aligned}$$

□

Since f' consists of all the monomials u of f with $x_1, \dots, x_{k+1} \in \text{cont}(U)$ and f'' consists of those with $x_1, \dots, x_k \in \text{cont}(u)$, it is easy to see that

$$f'(x_1, \dots, x_k, x_{k+1}) = f''(x_1, \dots, x_k, x_{k+1}) - f''(x_1, \dots, x_k, 0). \quad (24)$$

Let N_k''' denote the matrix obtained from N_k'' by subtracting the row labeled $(x_1, \dots, x_k, 0)$ from every row labeled by $(x_1, \dots, x_k, x_{k+1})$, $x_{k+1} \in \mathbb{Z}_p^*$. By (24) the rows of N_k''' labeled $(x_1, \dots, x_k, x_{k+1})$, $x_{k+1} \in \mathbb{Z}_p^*$ contain the values of $f'(x_1, \dots, x_k, x_{k+1})$. Let N_k' be the submatrix of N_k''' containing only such rows. We need to prove that the columns of N_k' labeled with $f \in F_k$ are linearly independent. We do it by first proving that the rank of N_k' equals $(p-1)^{k+1}$ and then demonstrating that the column labeled $f^{(p-1)} = \bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus (p-1)$ is a linear combination of columns labeled $f^{(b)} = \bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus b$ for $b \in \{0, \dots, p-2\}$.

Lemma 6.15. *Let $f = \bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus b$.*

(a)

$$\Sigma' f = \sum_{x_{k+1}=0}^{p-1} f''(x_1, \dots, x_k, x_{k+1}) = 0.$$

(b) *Let $f^{(a)}$ denote the function $f^{(a)} = \bigoplus_{i=1}^k \alpha_i x_i \oplus x_{k+1} \oplus a$ (and so $f = f^{(b)}$). Then*

$$\Sigma f = \sum_{b \in \mathbb{Z}_p} f''^{(b)}(x_1, \dots, x_{k+1}) = 0.$$

Proof. (a) We have

$$\begin{aligned}
\Sigma' f &= \sum_{x_{k+1}=0}^{p-1} f''(x_1, \dots, x_k, x_{k+1}) \\
&= \sum_{x_{k+1}=0}^{p-1} \sum_{S \subseteq [k]} (-1)^{k-|S|} f_S(x_1, \dots, x_{k+1}) \\
&= \sum_{S \subseteq [k]} (-1)^{k-|S|} \sum_{x_{k+1}=0}^{p-1} f_S(x_1, \dots, x_{k+1}).
\end{aligned}$$

Let $S \subseteq [k]$, $x_1, \dots, x_k \in \mathbb{Z}_p^*$, and let us denote $A = f_S(x_1, \dots, x_k, 0) = \bigoplus_{i \in S} \alpha_i x_i \oplus b$. Then

$$\begin{aligned} \sum_{x_{k+1}=0}^{p-1} f_S(x_1, \dots, x_{k+1}) &= \sum_{a=0}^{p-1} (A \oplus a) \\ &= \frac{p(p+1)}{2}. \end{aligned}$$

Now,

$$\Sigma' f = \frac{p(p+1)}{2} \sum_{S \subseteq [k]} (-1)^{k-|S|} = 0.$$

(b) We have

$$\begin{aligned} \Sigma f &= \sum_{b \in \mathbb{Z}_p} f''^{(b)}(x_1, \dots, x_{k+1}) \\ &= \sum_{b \in \mathbb{Z}_p} \sum_{S \subseteq [k]} (-1)^{k-|S|} f_S^{(b)}(x_1, \dots, x_{k+1}) \\ &= \sum_{S \subseteq [k]} (-1)^{k-|S|} \sum_{b \in \mathbb{Z}_p} f_S^{(b)}(x_1, \dots, x_{k+1}). \end{aligned}$$

Let $S \subseteq [k]$, $x_1, \dots, x_{k+1} \in \mathbb{Z}_p^*$, and let us denote $A = \bigoplus_{i \in S} \alpha_i x_i \oplus x_{k+1}$. Then

$$\sum_{b \in \mathbb{Z}_p} f_S^{(b)}(x_1, \dots, x_{k+1}) = \sum_{b \in \mathbb{Z}_p} (A \oplus b) = \sum_{b \in \mathbb{Z}_p} b = \frac{p(p+1)}{2}.$$

Now,

$$\Sigma f = \frac{p(p+1)}{2} \sum_{S \subseteq [k]} (-1)^{k-|S|} = 0.$$

□

We are now in a position to complete the proof of Proposition 6.3. Let $\vec{a}(x_1, \dots, x_k, x_{k+1})$ denote the row of N_k'' labeled with $(x_1, \dots, x_k, x_{k+1})$. Then by Lemma 6.15(a)

$$\vec{a}(x_1, \dots, x_k, 0) = - \sum_{b \in \mathbb{Z}_p^*} \vec{a}(x_1, \dots, x_k, b),$$

and the row of N_k''' labeled with $(x_1, \dots, x_k, x_{k+1})$ is

$$\vec{b}(x_1, \dots, x_k, x_{k+1}) = \vec{a}(x_1, \dots, x_k, x_{k+1}) + \sum_{b \in \mathbb{Z}_p^*} \vec{a}(x_1, \dots, x_k, b).$$

As is easily seen the row $\vec{a}(x_1, \dots, x_k, 0)$ is still a linear combination of $\vec{b}(x_1, \dots, x_k, x_{k+1})$, $x_{k+1} \in \mathbb{Z}_p^*$, implying that the rows of N_k''' labeled $(x_1, \dots, x_{k+1}) \in (\mathbb{Z}_p^*)^{p+1}$ are linearly independent and N_k' has rank $(p-1)^{k+1}$. Finally, by Lemma 6.15(a) the columns of N_k' labeled with $f \in F_k$ are also linear independent.

6.2 Linear equations mod 3

In this section we consider the case where $p = 3$ and provide linearly independent p -expressions that span the space of functions from \mathbb{Z}_3^n to \mathbb{C} . The p -expressions we consider here are different from the ones considered in Theorem 5.26 and we prove they are linearly independent using somewhat a simpler approach.

In this subsection set $p = 3$ and \oplus, \odot denote addition and multiplication modulo 3, respectively. Let x_1, \dots, x_n be variables that take values from the ternary domain $\{0, 1, 2\}$. Here we prove that all the linear expressions of the form

$$(a_1 \odot x_1) \oplus (a_2 \odot x_2) \oplus \dots \oplus (a_n \odot x_n)$$

with $a_i \in \{0, 1, 2\}$ are linearly independent, except the zero expression. For instance, in the case where $n = 1$, the following matrix has rank 2 meaning that x_1 and $2 \odot x_1$ are linearly independent.

$$A = \begin{pmatrix} 0 & x_1 & 2 \odot x_1 \\ 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{pmatrix} \begin{matrix} x_1=0 \\ x_1=1 \\ x_1=2 \end{matrix}$$

Now define sequence of matrices as follows. Set $C_1 = A$ and recursively define C_n to be the following $3^n \times 3^n$ matrix

$$C_n = \begin{pmatrix} C_{n-1} & C_{n-1} & C_{n-1} \\ C_{n-1} & C_{n-1} \oplus \mathbf{1} & C_{n-1} \oplus \mathbf{2} \\ C_{n-1} & C_{n-1} \oplus \mathbf{2} & C_{n-1} \oplus \mathbf{1} \end{pmatrix}$$

Observation 6.16. For any real numbers $a, b \neq 0$ and any integer n we have $\text{rank}(aC_n + \mathbf{b}) = \text{rank}(C_n) + 1$.

Proof. Note that the first row and the first column of C_n contain only zeros. That is

$$C_n = \left(\begin{array}{c|ccc} 0 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array} \right)$$

where B is a $3^n - 1 \times 3^n - 1$ matrix and has the same rank as C_n . Now, $\text{rank}(aC_n + B) = \text{rank}(C_n + \frac{B}{a})$.

$$C_n + \frac{\mathbf{b}}{\mathbf{a}} = \left(\begin{array}{c|ccc} \frac{b}{a} & \frac{b}{a} & \dots & \frac{b}{a} \\ \frac{b}{a} & & & \\ \vdots & & B + \frac{\mathbf{b}}{\mathbf{a}} & \\ \frac{b}{a} & & & \end{array} \right) \rightarrow \left(\begin{array}{c|ccc} \frac{b}{a} & \frac{b}{a} & \dots & \frac{b}{a} \\ 0 & & & \\ \vdots & & B + \mathbf{0} & \\ 0 & & & \end{array} \right) \rightarrow \left(\begin{array}{c|ccc} \frac{b}{a} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array} \right)$$

Hence, $\text{rank}(aC_n + \mathbf{b}) = \text{rank}(C_n + \frac{\mathbf{b}}{\mathbf{a}}) = \text{rank}(C_n) + 1$. □

Lemma 6.17. For any integer n , C_n has rank $3^n - 1$ i.e., all the linear expressions of the form $(a_1 \odot x_1) \oplus (a_2 \odot x_2) \oplus \dots \oplus (a_n \odot x_n)$ with $a_i \in \{0, 1, 2\}$ are linearly independent, except the zero expression.

Proof. The proof is by induction. Clearly, for $n = 1$, the matrix $C_1 = A$ has rank 2. Suppose C_i has rank $3^i - 1$ for all $1 \leq i \leq n$. For a matrix M with 0, 1, 2 entries, let

$$\begin{aligned} p_1(M) &= \frac{3}{2}M \circ M + \frac{5}{2}M + \mathbf{1} \\ p_2(M) &= -\frac{3}{2}M \circ M - \frac{7}{2}M + \mathbf{2} \end{aligned}$$

where \circ denotes the *Hadamard* product or the *element-wise* product of two matrices. Observe that $M \oplus \mathbf{1} = p_1(M)$ and $M \oplus \mathbf{2} = p_2(M)$. Hence, we can write C_{n+1} as follow

$$C_{n+1} = \begin{pmatrix} C_n & C_n & C_n \\ C_n & p_1(C_n) & p_2(C_n) \\ C_n & p_2(C_n) & p_1(C_n) \end{pmatrix}$$

Next, we perform a series of row and column operations to transform C_n into a block-diagonal matrix.

$$\begin{aligned} C_{n+1} &= \begin{pmatrix} C_n & C_n & C_n \\ C_n & p_1(C_n) & p_2(C_n) \\ C_n & p_2(C_n) & p_1(C_n) \end{pmatrix} \rightarrow \begin{pmatrix} C_n & C_n & C_n \\ \mathbf{0} & p_1(C_n) - C_n & p_2(C_n) - C_n \\ \mathbf{0} & p_2(C_n) - C_n & p_1(C_n) - C_n \end{pmatrix} \\ &\rightarrow \begin{pmatrix} C_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & p_1(C_n) - C_n & p_2(C_n) - C_n \\ \mathbf{0} & p_2(C_n) - C_n & p_1(C_n) - C_n \end{pmatrix} \\ &\rightarrow \begin{pmatrix} C_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & p_1(C_n) - C_n & p_2(C_n) - C_n \\ \mathbf{0} & p_1(C_n) + p_2(C_n) - 2C_n & p_1(C_n) + p_2(C_n) - 2C_n \end{pmatrix} \\ &\rightarrow \begin{pmatrix} C_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & p_1(C_n) - C_n & p_2(C_n) - C_n \\ \mathbf{0} & -3C_n + \mathbf{3} & -3C_n + \mathbf{3} \end{pmatrix} \\ &\rightarrow \begin{pmatrix} C_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & p_1(C_n) - C_n & p_1(C_n) + p_2(C_n) - 2C_n \\ \mathbf{0} & -3C_n + \mathbf{3} & -6C_n + \mathbf{6} \end{pmatrix} \\ &\rightarrow \begin{pmatrix} C_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & p_1(C_n) - C_n & -3C_n + \mathbf{3} \\ \mathbf{0} & -3C_n + \mathbf{3} & -6C_n + \mathbf{6} \end{pmatrix} \rightarrow \begin{pmatrix} C_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & p_1(C_n) - C_n - \frac{1}{2}(-3C_n + \mathbf{3}) & -3C_n + \mathbf{3} \\ \mathbf{0} & \mathbf{0} & -6C_n + \mathbf{6} \end{pmatrix} \\ &\rightarrow \begin{pmatrix} C_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & p_1(C_n) - C_n - \frac{1}{2}(-3C_n + \mathbf{3}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -6C_n + \mathbf{6} \end{pmatrix} \\ &\rightarrow \begin{pmatrix} C_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\frac{3}{2}C_n \circ C_n + 3C_n - \frac{1}{2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 6C_n - \mathbf{6} \end{pmatrix} \\ &\rightarrow \begin{pmatrix} C_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -3C_n \circ C_n + 6C_n - \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & C_n - \mathbf{1} \end{pmatrix} \end{aligned}$$

Hence, rank of C_{n+1} is $\text{rank}(C_n) + \text{rank}(C_n - \mathbf{1}) + \text{rank}(-3C_n \circ C_n + 6C_n - \mathbf{1})$. Moreover, Observation 6.16 yields

$$\begin{aligned}\text{rank}(C_{n+1}) &= \text{rank}(C_n) + \text{rank}(C_n - \mathbf{1}) + \text{rank}(-3C_n \circ C_n + 6C_n - \mathbf{1}) \\ &= 3^n - 1 + 3^n + \text{rank}(-3C_n \circ C_n + 6C_n - \mathbf{1})\end{aligned}$$

In what follows we prove that $\text{rank}(-3C_n \circ C_n + 6C_n - \mathbf{1}) = 3^n$. Let us define the following two matrices associated to a matrix M with $\{0, 1, 2\}$ entries.

$$M^+ [i, j] = \begin{cases} 0 & \text{if } M[i, j] = 0 \\ 1 & \text{if } M[i, j] = 1 \\ 0 & \text{if } M[i, j] = 2 \end{cases} \quad \text{and} \quad M^{++} [i, j] = \begin{cases} 0 & \text{if } M[i, j] = 0 \\ 0 & \text{if } M[i, j] = 1 \\ 2 & \text{if } M[i, j] = 2 \end{cases}$$

Note that $C_n = C_n^+ + C_n^{++}$. For instance in the case $C_1 = A$ the two matrices A^+ and A^{++} are $A^+ = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $A^{++} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix}$. Here, we simplify the expression $-3C_n \circ C_n + 6C_n - \mathbf{1}$ and write it in terms of C_n^+ and C_n^{++} .

$$\begin{aligned}-3C_n \circ C_n + 6C_n - \mathbf{1} &= -3(C_n^+ + C_n^{++}) \circ (C_n^+ + C_n^{++}) + 6(C_n^+ + C_n^{++}) - \mathbf{1} \\ &= -3(C_n^+ \circ C_n^+ + C_n^{++} \circ C_n^{++}) + 6C_n^+ + 6C_n^{++} - \mathbf{1} \\ &= -3(C_n^+ + 2C_n^{++}) + 6C_n^+ + 6C_n^{++} - \mathbf{1} \\ &= -3C_n^+ - 6C_n^{++} + 6C_n^+ + 6C_n^{++} - \mathbf{1} \\ &= 3C_n^+ - \mathbf{1}\end{aligned}$$

Claim 6.18. *For every positive integer n , the matrix $3C_n^+ - \mathbf{1}$ has full rank. This implies that $-3C_n \circ C_n + 6C_n - \mathbf{1}$ has full rank.*

Proof. For the base case $n = 1$, the matrix $3A^+ - \mathbf{1} = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$ has full rank i.e.,

$\text{rank}(3A^+ - \mathbf{1}) = 3$. For our induction hypothesis suppose the claim is correct for every n . Next we show $3C_{n+1}^+ - \mathbf{1}$ has rank 3^{n+1} .

$$\begin{aligned}
3C_{n+1}^+ &= 3 \begin{pmatrix} C_n^+ & C_n^+ & C_n^+ \\ C_n^+ & (C_n \oplus \mathbf{1})^+ & (C_n \oplus \mathbf{2})^+ \\ C_n^+ & (C_n \oplus \mathbf{2})^+ & (C_n \oplus \mathbf{1})^+ \end{pmatrix} - \mathbf{1} \\
&= \begin{pmatrix} 3C_n^+ - \mathbf{1} & 3C_n^+ - \mathbf{1} & 3C_n^+ - \mathbf{1} \\ 3C_n^+ - \mathbf{1} & 3(C_n \oplus \mathbf{1})^+ - \mathbf{1} & 3(C_n \oplus \mathbf{2})^+ - \mathbf{1} \\ 3C_n^+ - \mathbf{1} & 3(C_n \oplus \mathbf{2})^+ - \mathbf{1} & 3(C_n \oplus \mathbf{1})^+ - \mathbf{1} \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 3C_n^+ - \mathbf{1} & 3C_n^+ - \mathbf{1} & 3C_n^+ - \mathbf{1} \\ \mathbf{0} & 3(C_n \oplus \mathbf{1})^+ - 3C_n^+ & 3(C_n \oplus \mathbf{2})^+ - 3C_n^+ \\ \mathbf{0} & 3(C_n \oplus \mathbf{2})^+ - 3C_n^+ & 3(C_n \oplus \mathbf{1})^+ - 3C_n^+ \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 3C_n^+ - \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 3(C_n \oplus \mathbf{1})^+ - 3C_n^+ & 3(C_n \oplus \mathbf{2})^+ - 3C_n^+ \\ \mathbf{0} & 3(C_n \oplus \mathbf{2})^+ - 3C_n^+ & 3(C_n \oplus \mathbf{1})^+ - 3C_n^+ \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 3C_n^+ - \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (C_n \oplus \mathbf{1})^+ - C_n^+ & (C_n \oplus \mathbf{2})^+ - C_n^+ \\ \mathbf{0} & (C_n \oplus \mathbf{2})^+ - C_n^+ & (C_n \oplus \mathbf{1})^+ - C_n^+ \end{pmatrix}
\end{aligned}$$

For a matrix M with $0, 1, 2$ entries, define $p'_1(M) = \frac{1}{2}M \circ M - \frac{3}{2}M + \mathbf{1}$ and $p'_2(M) = \frac{1}{2}M \circ M - \frac{1}{2}M$ where \circ denotes the Hadamard product or the element-wise product of two matrices. Observe that $(M \oplus \mathbf{1})^+ = p'_1(M)$ and $(M \oplus \mathbf{2})^+ = p'_2(M)$.

$$\begin{aligned}
p'_1(M) &= \frac{1}{2}(M^\dagger + M^{\dagger\dagger}) \circ (M^\dagger + M^{\dagger\dagger}) + \frac{3}{2}(M^\dagger + M^{\dagger\dagger}) + \mathbf{1} = -M^\dagger - \frac{1}{2}M^{\dagger\dagger} + \mathbf{1} \\
p'_2(M) &= \frac{1}{2}(M^\dagger + M^{\dagger\dagger}) \circ (M^\dagger + M^{\dagger\dagger}) - \frac{1}{2}(M^\dagger + M^{\dagger\dagger}) = \frac{1}{2}M^{\dagger\dagger}
\end{aligned}$$

We continue by performing row and column operation to transform $3C_{n+1} - \mathbf{1}$ into a block-

diagonal matrix.

$$\begin{aligned}
& \begin{pmatrix} 3C_n^+ - 1 & 0 & 0 \\ 0 & (C_n \oplus 1)^+ - C_n^+ & (C_n \oplus 2)^+ - C_n^+ \\ 0 & (C_n \oplus 2)^+ - C_n^+ & (C_n \oplus 1)^+ - C_n^+ \end{pmatrix} \\
&= \begin{pmatrix} 3C_n^+ - 1 & 0 & 0 \\ 0 & -2C_n^+ - \frac{1}{2}C_n^{++} + 1 & -C_n^+ + \frac{1}{2}C_n^{++} \\ 0 & -C_n^+ + \frac{1}{2}C_n^{++} & -2C_n^+ - \frac{1}{2}C_n^{++} + 1 \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 3C_n^+ - 1 & 0 & 0 \\ 0 & -2C_n^+ - \frac{1}{2}C_n^{++} + 1 & -C_n^+ + \frac{1}{2}C_n^{++} \\ 0 & -3C_n^+ + 1 & -3C_n^+ + 1 \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 3C_n^+ - 1 & 0 & 0 \\ 0 & -C_n^+ - C_n^{++} + 1 & -C_n^+ + \frac{1}{2}C_n^{++} \\ 0 & 0 & -3C_n^+ + 1 \end{pmatrix} \rightarrow \begin{pmatrix} 3C_n^+ - 1 & 0 & 0 \\ 0 & -C_n^+ - C_n^{++} + 1 & -\frac{3}{2}C_n^+ + \frac{1}{2} \\ 0 & 0 & -3C_n^+ + 1 \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 3C_n^+ - 1 & 0 & 0 \\ 0 & -C_n^+ - C_n^{++} + 1 & 0 \\ 0 & 0 & -3C_n^+ + 1 \end{pmatrix} \rightarrow \begin{pmatrix} 3C_n^+ - 1 & 0 & 0 \\ 0 & C_n^+ + C_n^{++} - 1 & 0 \\ 0 & 0 & 3C_n^+ - 1 \end{pmatrix} \\
&= \begin{pmatrix} 3C_n^+ - 1 & 0 & 0 \\ 0 & C_n - 1 & 0 \\ 0 & 0 & 3C_n^+ - 1 \end{pmatrix}
\end{aligned}$$

By the induction hypothesis, $\text{rank}(3C_n^+ - 1) = 3^n$. Moreover, by the induction hypothesis and Observation 6.16 $\text{rank}(C_n - 1) = 3^n$. As a result, $\text{rank}(3C_{n+1}^+ - 1) = 3^n + 3^n + 3^n = 3^{n+1}$ and $-3C_n \circ C_n + 6C_n - 1$ has full rank. \square

Recall that $\text{rank}(C_{n+1}) = \text{rank}(C_n) + \text{rank}(C_n - 1) + \text{rank}(-3C_n \circ C_n + 6C_n - 1)$. By the induction hypothesis and Observation 6.16, we have $\text{rank}(C_n - 1) = \text{rank}(C_n) + 1 = 3^n$. Moreover, Claim 6.18 yields $\text{rank}(-3C_n \circ C_n + 6C_n - 1) = 3^n$. Hence,

$$\text{rank}(C_{n+1}) = 3^n - 1 + 3^n + 1 + 3^n + 1 = 3^{n+1} - 1.$$

\square

7 Conclusion and future work

The study of CSP-related Ideal Membership Problems is in its infancy, and pretty much all research directions are open. These include expanding the range of tractable IMPs. A number of candidates for such expansions are readily available from the existing results about the CSP. There are however several questions that seem to be more intriguing; they mainly concern with relationship between the IMP, CSP and other problems.

The first one is how the tractability of the IMP can be used in applications such as Nullstellensatz and SOS proofs. The several results we obtain here barely scratch the surface. Establishing connections of this kind seem important, because it would allow for using a much larger toolbox than the usual Gröbner Basis. Note also that constructing an explicit Gröbner Basis beyond

Boolean case is getting very hard very quickly; such techniques may be not very useful in more general cases.

One of the principal techniques in solving the CSP is constraint propagation, that is, the study how local interaction between constraints may tighten them, and even sometimes refute the existence of a solution. In some cases such as IMPs with the dual-discriminator polymorphism, computing the S -polynomial, and therefore constructing a Gröbner Basis is equivalent to establishing so-called arc consistency. This however is not the case in general. On the other hand, constraint propagation is done through some very simple pp-definitions, and so Theorem 1.5(1) and its proof imply that there likely are some parallel constructions with polynomials and ideals.

The main tool for solving restricted degree problems $\text{IMP}_d(\Gamma)$ is constructing a d -truncated Gröbner Basis, in which the degrees of polynomials are also bounded by d . It is interesting what effect restricting the degree of polynomials in a generating set has on the properties of the underlying CSP. More precisely, if $I(\mathcal{P})$ is the ideal corresponding to a CSP instance \mathcal{P} , and $I_d(\mathcal{P})$ is the ideal generated by the reduced Gröbner Basis, then $I_d(\mathcal{P})$ can be translated back, to a less constrained CSP \mathcal{P}' . What is the connection between \mathcal{P} and \mathcal{P}' ? For example, every instance \mathcal{P} of $\text{CSP}(\Gamma)$ where Γ is Boolean and has a semilattice polymorphism, then \mathcal{P} is equivalent to a HORN- or ANTIHORN-SATISFIABILITY. Restricting the degrees of polynomials in $I_d(\mathcal{P})$ is apparently equivalent to removing all clauses of length more than d in \mathcal{P} .

Acknowledgement

The second author would like to thank Reza Dastbastehe for many valuable discussions.

References

- [1] Albert Atserias and Joanna Ochremiak. Proof complexity meets algebra. *ACM Trans. Comput. Log.*, 20(1):1:1–1:46, 2019.
- [2] Libor Barto. Constraint satisfaction problem and universal algebra. *ACM SIGLOG News*, 1(2):14–24, 2014.
- [3] Libor Barto. The constraint satisfaction problem and universal algebra. *Bull. Symb. Log.*, 21(3):319–337, 2015.
- [4] Libor Barto, Andrei A. Krokhin, and Ross Willard. Polymorphisms, and how to use them. In Andrei A. Krokhin and Stanislav Zivný, editors, *The Constraint Satisfaction Problem: Complexity and Approximability*, volume 7 of *Dagstuhl Follow-Ups*, pages 1–44. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.
- [5] Paul Beame, Russell Impagliazzo, Jan Krajíček, Toniann Pitassi, and Pavel Pudlák. Lower bound on hilbert’s nullstellensatz and propositional proofs. In *35th Annual Symposium on Foundations of Computer Science, Santa Fe, New Mexico, USA, 20-22 November 1994*, pages 794–806. IEEE Computer Society, 1994.
- [6] Thomas Becker and Volker Weispfenning. Gröbner bases. In *Gröbner Bases*, pages 187–242. Springer, 1993.

- [7] Arpitha P. Bharathi and Monaldo Mastrolilli. Ideal membership problem and a majority polymorphism over the ternary domain. In Javier Esparza and Daniel Král', editors, *45th International Symposium on Mathematical Foundations of Computer Science, MFCS 2020, August 24–28, 2020, Prague, Czech Republic*, volume 170 of *LIPICs*, pages 13:1–13:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.
- [8] Arpitha P. Bharathi and Monaldo Mastrolilli. Ideal membership problem for boolean minority. *CoRR*, abs/2006.16422, 2020.
- [9] Bruno Buchberger. Bruno buchberger's phd thesis 1965: An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal. *Journal of Symbolic Computation*, 41(3):475 – 511, 2006. Logic, Mathematics and Computer Science: Interactions in honor of Bruno Buchberger (60th birthday).
- [10] Andrei A. Bulatov. A dichotomy theorem for nonuniform csp's. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15–17, 2017*, pages 319–330. IEEE Computer Society, 2017.
- [11] Andrei A. Bulatov and Víctor Dalmau. Towards a dichotomy theorem for the counting constraint satisfaction problem. *Inf. Comput.*, 205(5):651–678, 2007.
- [12] Andrei A. Bulatov, Peter Jeavons, and Andrei A. Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM J. Comput.*, 34(3):720–742, 2005.
- [13] S. Burris and H.P. Sankappanavar. *A course in universal algebra*, volume 78 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1981.
- [14] Samuel R. Buss and Toniann Pitassi. Good degree bounds on nullstellensatz refutations of the induction principle. In Steven Homer and Jin-Yi Cai, editors, *Proceedings of the Eleventh Annual IEEE Conference on Computational Complexity, Philadelphia, Pennsylvania, USA, May 24–27, 1996*, pages 233–242. IEEE Computer Society, 1996.
- [15] Matthew Clegg, Jeff Edmonds, and Russell Impagliazzo. Using the groebner basis algorithm to find proofs of unsatisfiability. In Gary L. Miller, editor, *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22–24, 1996*, pages 174–183. ACM, 1996.
- [16] Martin C. Cooper, David A. Cohen, and Peter Jeavons. Characterising tractable constraints. *Artif. Intell.*, 65(2):347–361, 1994.
- [17] David Cox, John Little, and Donal OShea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer Science & Business Media, 2013.
- [18] Alicia Dickenstein, Noä Fitchas, Marc Giusti, and Carmen Sessa. The membership problem for unmixed polynomial ideals is solvable in single exponential time. *Discret. Appl. Math.*, 33(1-3):73–94, 1991.
- [19] Jean-Charles Faugère, Patrizia M. Gianni, Daniel Lazard, and Teo Mora. Efficient computation of zero-dimensional gröbner bases by change of ordering. *J. Symb. Comput.*, 16(4):329–344, 1993.

- [20] Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through datalog and group theory. *SIAM J. Comput.*, 28(1):57–104, 1998.
- [21] Dima Grigoriev. Tseitin’s tautologies and lower bounds for nullstellensatz proofs. In *39th Annual Symposium on Foundations of Computer Science, FOCS ’98, November 8-11, 1998, Palo Alto, California, USA*, pages 648–652. IEEE Computer Society, 1998.
- [22] Grete Hermann. Die frage der endlich vielen schritte in der theorie der polynomideale. *Mathematische Annalen*, 95(1):736–788, 1926.
- [23] David Hilbert. Ueber die theorie der algebraischen formen. *Mathematische annalen*, 36(4):473–534, 1890.
- [24] David Hilbert. Über die vollen invariantensysteme. *Mathematische annalen*, 42(3):313–373, 1893.
- [25] Peter Jeavons, David A. Cohen, and Marc Gyssens. Closure properties of constraints. *J. ACM*, 44(4):527–548, 1997.
- [26] Christopher Jefferson, Peter Jeavons, Martin J. Green, and M. R. C. van Dongen. Representing and solving finite-domain constraint problems using systems of polynomials. *Annals of Mathematics and Artificial Intelligence*, 67(3):359–382, Mar 2013.
- [27] Phokion G. Kolaitis and Moshe Y. Vardi. Conjunctive-query containment and constraint satisfaction. In Alberto O. Mendelzon and Jan Paredaens, editors, *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 1-3, 1998, Seattle, Washington, USA*, pages 205–213. ACM Press, 1998.
- [28] Monaldo Mastrolilli. The complexity of the ideal membership problem for constrained problems over the boolean domain. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 456–475, 2019.
- [29] Ernst W. Mayr. Membership in plynomial ideals over \mathbb{Q} is exponential space complete. In Burkhard Monien and Robert Cori, editors, *STACS 89, 6th Annual Symposium on Theoretical Aspects of Computer Science, Paderborn, FRG, February 16-18, 1989, Proceedings*, volume 349 of *Lecture Notes in Computer Science*, pages 400–406. Springer, 1989.
- [30] Ernst W Mayr and Albert R Meyer. The complexity of the word problems for commutative semigroups and polynomial ideals. *Advances in mathematics*, 46(3):305–329, 1982.
- [31] Ralph N McKenzie, George F McNulty, and Walter F Taylor. *Algebras, lattices, varieties*, volume 383. American Mathematical Soc., 2018.
- [32] Ryan O’Donnell. SOS is not obviously automatizable, even approximately. In Christos H. Papadimitriou, editor, *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, volume 67 of *LIPICs*, pages 59:1–59:10. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.

- [33] Dona Papert. Congruence relations in semi-lattices. *J. London Math. Soc.*, 39:723–729, 1964.
- [34] George M Phillips. *Interpolation and approximation by polynomials*, volume 14. Springer Science & Business Media, 2003.
- [35] EL Post. The two-valued iterative systems of mathematical logic. number 5 in *annals of math. Studies. Princeton Univ. Press*, 1941.
- [36] Prasad Raghavendra and Benjamin Weitz. On the bit complexity of sum-of-squares proofs. In Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl, editors, *44th International Colloquium on Automata, Languages, and Programming, ICALP 2017, July 10-14, 2017, Warsaw, Poland*, volume 80 of *LIPICs*, pages 80:1–80:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.
- [37] Fred Richman. Constructive aspects of noetherian rings. *Proceedings of the American Mathematical Society*, 44(2):436–441, 1974.
- [38] Thomas J. Schaefer. The complexity of satisfiability problems. In Richard J. Lipton, Walter A. Burkhard, Walter J. Savitch, Emily P. Friedman, and Alfred V. Aho, editors, *Proceedings of the 10th Annual ACM Symposium on Theory of Computing, May 1-3, 1978, San Diego, California, USA*, pages 216–226. ACM, 1978.
- [39] Abraham Seidenberg. Constructions in algebra. *Transactions of the American Mathematical Society*, 197:273–313, 1974.
- [40] Ágnes Szendrei. Clones in universal algebra. *Les presses de L’universite de Montreal*, 1986.
- [41] Marc R.C. van Dongen. *Constraints, Varieties, and Algorithms*. PhD thesis, Department of Computer Science, University College, Cork, Ireland, 2002.
- [42] Dmitriy Zhuk. A proof of CSP dichotomy conjecture. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 331–342. IEEE Computer Society, 2017.