Correlation Decay and Tractability of CSPs*

Jonah Brown-Cohen[1] **and Prasad Raghavendra**[2]

1 **University of California, Berkeley, CA, USA**
jonahbc@eecs.berkeley.edu
2 **University of California, Berkeley, CA, USA**
prasad@eecs.berkeley.edu

Abstract

The algebraic dichotomy conjecture of Bulatov, Krokhin and Jeavons yields an elegant characterization of the complexity of constraint satisfaction problems. Roughly speaking, the characterization asserts that a CSP L is tractable if and only if there exist certain non-trivial operations known as polymorphisms to combine solutions to L to create new ones.

In this work, we study the dynamical system associated with repeated applications of a polymorphism to a distribution over assignments. Specifically, we exhibit a correlation decay phenomenon that makes two variables or groups of variables that are not perfectly correlated become independent after repeated applications of a polymorphism.

We show that this correlation decay phenomenon can be utilized in designing algorithms for CSPs by exhibiting two applications:

1. A simple randomized algorithm to solve linear equations over a prime field, whose analysis crucially relies on correlation decay.
2. A sufficient condition for the simple linear programming relaxation for a 2-CSP to be sound (have no integrality gap) on a given instance.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Constraint Satisfaction, Polymorphisms, Linear Equations, Correlation Decay

Digital Object Identifier 10.4230/LIPIcs.ICALP.2016.79

1 Introduction

A vast majority of natural computational problems have been classified to be either polynomial-time solvable or NP-complete. While there is little progress in determining the exact time complexity for fundamental problems like matrix multiplication, it can be argued that a much coarser classification of P vs.~NP-complete has been achieved for a large variety of problems. Notable problems that elude such a classification include factorization or graph isomorphism.

A compelling research direction at this juncture is to understand what causes problems to be easy (in P) or hard (NP-complete). More precisely, for specific classes of problems, does there exist a unifying theory that explains and characterizes why some problems in the class are in P while others are NP-complete? For the sake of concreteness, we will present a few examples.

It is well-known that 2-Sat is polynomial-time solvable, while 3-Sat is NP-complete. However, the traditional proofs of these statements are unrelated to each other and therefore shed little light on what makes 2-Sat easy while 3-Sat NP-complete.

* This work was supported by NSF Career Award & Alfred.P. Sloan Fellowship.

43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016).
Editors: Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi;
Article No. 79; pp. 79:1–79:13

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Over the last decade, a theory of tractability has emerged for the class of constraint satisfaction problems (CSP). While this candidate theory remains conjectural for now, it successfully explains all the existing algorithms and hardness results for CSPs. To set the stage for the results of this paper, we begin with a brief survey of the theory for CSPs.

A constraint satisfaction problem (CSP) Λ is specified by a family of predicates over a finite domain $[q] = \{1, 2, \ldots, q\}$. Every instance of the CSP Λ consists of a set of variables $\mathcal{V}$, along with a set of constraints $\mathcal{C}$ on them. Each constraint in $\mathcal{C}$ consists of a predicate from the family Λ applied to a subset of variables. For a CSP Λ, the associated satisfiability problem Λ-Sat is defined as follows.

▶ **Problem 1** (Λ-Sat)**.** *Given an instance $\Im$ of the CSP Λ, determine whether there is an assignment satisfying all the constraints in $\Im$.*

A classic theorem of Schaefer [11] asserts that among all satisfiability problems over the boolean domain ($\{0, 1\}$), only Linear-Equations-Mod-2, 2-Sat, Horn-Sat, Dual-Horn Sat and certain trivial CSPs are solvable in polynomial time. The rest of the boolean CSPs are NP-complete. The dichotomy conjecture of Feder and Vardi [7] asserts that every Λ-Sat is in P or NP-complete. The conjecture has been shown to hold for CSPs over domains of size up to 3 [4].

In this context, it is natural to question as to what makes certain Λ-Sat problems tractable while the others are NP-complete. Bulatov, Jeavons and Krokhin [6] conjectured a beautiful characterization for tractable satisfiability problems. We will present an informal description of this characterization known as the *algebraic dichotomy conjecture*. We refer the reader to the work of Kun & Szegedy [8] for a more formal description.

To motivate this characterization, let us consider a CSP known as the XOR problem. An instance of the XOR problem consists of a system of linear equations over $\mathbb{Z}_2 = \{0, 1\}$. Fix an instance $\Im$ of XOR over n variables. Given three solutions $X^{(1)}, X^{(2)}, X^{(3)} \in \{0, 1\}^n$ to $\Im$, one can create a new solution $Y \in \{0, 1\}^n$:

$$Y_i = X_i^{(1)} \oplus X_i^{(2)} \oplus X_i^{(3)} \qquad \forall i \in [n]\,.$$

It is easy to check that Y is also a feasible solution to the instance $\Im$. Thus the $XOR : \{0, 1\}^3 \to \{0, 1\}$ yields a way to combine three solutions in to a new solution to the same instance. Note that the function XOR was applied to each bit of the solution individually. An operation of this form that preserves the satisfiability of the CSP is known as a *polymorphism*. Formally, a polymorphism of a CSP Λ-Sat is defined as follows:

▶ **Definition 2** (Polymorphisms)**.** A function $p : [q]^R \to [q]$ is said to be *a polymorphism* for the CSP Λ-Sat, if for every instance $\Im$ of Λ, and R assignments $X^{(1)}, X^{(2)}, \ldots, X^{(R)} \in [q]^n$ that satisfy all constraints in $\Im$, the vector $Y \in [q]^n$ defined below is also a feasible solution.

$$Y_i = p(X_i^{(1)}, X_i^{(2)}, X_i^{(3)}, \ldots, X_i^{(R)}) \qquad \forall i \in [n]\,.$$

Note that the dictator functions $p(x^{(1)}, \ldots, x^{(R)}) = x^{(i)}$ are polymorphisms for every CSP Λ-Sat. These will be referred to as *projections* or trivial polymorphisms. All the tractable cases of boolean CSPs in Schaefer's theorem are characterized by existence of non-trivial polymorphisms. Specifically, 2-SAT has the Majority function, Horn-SAT has the OR function, and Dual Horn-SAT has the AND function as a polymorphism. Roughly speaking, Bulatov et al. [6] conjectured that the existence of non-dictator polymorphisms characterizes CSPs that are tractable. Their work showed that the set of polymorphisms Poly(Λ) of a CSP Λ characterizes the complexity of Λ-Sat. There are many equivalent

ways of formalizing what it means for an operation to be *non-trivial* or *non-dictator*. A particularly simple way to formulate the algebraic dichotomy conjecture arises out of the recent work of Barto and Kozik [2]. A polymorphism $p : [q]^k \to [q]$ is called a *cyclic term* if

$$p(x_1, \ldots, x_k) = p(x_2, \ldots, x_k, x_1) = \ldots = p(x_k, x_1, \ldots, x_{k-1}) \quad \forall x_1, \ldots, x_k \in [q] \,.$$

Note that the above condition strictly precludes the operation p from being a dictator.

▶ **Conjecture 3** ([6, 9, 2]). Λ-Sat *is in* P *if* Λ *admits a cyclic term, otherwise* Λ-Sat *is* NP*-complete.*

Surprisingly, one of the implications of the conjecture has already been confirmed.

▶ **Theorem 4** ([6, 9, 2]). Λ-Sat *is* NP*-complete if* Λ *does not admit a cyclic term.*

The algebraic approach to understanding the complexity of CSPs has received much attention, and the algebraic dichotomy conjecture has been verified for several subclasses of CSPs such as conservative CSPs [3], CSPs with no ability to count [1] and CSPs with Maltsev operations [5]. Recently, Kun and Szegedy reformulated the algebraic dichotomy conjecture using analytic notions similar to influences [8].

Despite considerable progress in recent years [2], the algebraic dichotomy conjecture still remains open. Kun & Szegedy suggested the use of analytic techniques towards resolving the conjecture [8], which forms the inspiration for this work. This work demonstrates a phenomenon of correlation decay associated with iterated applications of polymorphisms and then exploits this phenomenon towards designing algorithms for CSPs.

1.1 Correlation Decay

We associate a natural dynamical system with a polymorphism $p : [q]^k \to [q]$ that corresponds to iterated applications of the polymorphism. Towards a formal definition of the dynamical system, let us fix a probability distribution μ over $[q]^n$. It is useful to think of μ as a distribution over assignments to a CSP instance on n variables.

For an operation $p : [q]^k \to [q]$, the distribution $p(\mu)$ over $[q]^n$ is one that is sampled by taking k independent samples from μ and applying the operation p to them. Define the dynamical system $\{\mu_t\}_{t \in \mathbb{N}}$ with $\mu_0 = \mu$,

$$\mu_t \stackrel{\text{def}}{=} p(\mu_{t-1}) \,, \forall t \in \mathbb{N}.$$

Roughly speaking, the key technical insight of this work is that the correlations among the coordinates decay as $t \to \infty$ for a *non-dictator* operation p. For the sake of simplicity, let us restrict our attention to the case of a distribution μ_{XY} on $[q] \times [q]$ (see Section 4 for the general theorem on distributions over $[q]^n$). Let $\mu_{|X}$ and $\mu_{|Y}$ denote the marginals of μ_{XY}. For any distribution Θ, let $\mathrm{supp}(\Theta)$ denote its support. We are ready to state a version of our correlation decay theorem.

▶ **Theorem 5.** *Let* μ_{XY} *be a distribution over* $[q] \times [q]$*. Let* $G_{\mu_{XY}}$ *denote the bipartite graph on vertices* $\mathrm{supp}(\mu_{|X}) \cup \mathrm{supp}(\mu_{|Y})$ *whose edges are given by the support of* μ_{XY}*. For a cyclic term* $p : [q]^k \to [q]$*, consider the dynamical system* $\{\mu_t\}_{t \in \mathbb{N}}$ *defined as* $\mu_0 := \mu_{XY}$*,* $\mu_t := p(\mu_{t-1}) \forall t \in \mathbb{N}$*. If* $G_{\mu_{XY}}$ *is a connected graph then*

$$\lim_{t \to \infty} \|\mu_t - \mu_{t|X} \times \mu_{t|Y}\|_1 = 0 \,.$$

i.e., μ_t *gets closer and closer to a product distribution as* $t \to \infty$*.*

Suppose p is a polymorphism of a CSP Λ and μ_{XY} is a distribution supported over satisfying assignments to an instance of Λ, then for every t, μ_t is also supported over satisfying assignments to Λ. Intuitively, this seems to be at odds with the correlation decay phenomenon: cyclic polymorphisms make variables uncorrelated yet still preserve satisfying assignments.

To resolve this paradox, notice that Theorem 5 requires that the graph $G_{\mu_{XY}}$ be *connected*. Connectivity of $G_{\mu_{XY}}$ corresponds to asserting that there are *no perfect correlations* between X and Y. This lack of perfect correlation can be quantified using the spectrum of the bipartite graph $G_{\mu_{XY}}$. Specifically, one can associate a correlation parameter $\rho(X,Y)$ (see Definition 18) such that $\rho(X,Y) < 1$ if and only if $G_{\mu_{XY}}$ is connected. $\rho(X,Y)$ is closely related to the second-eigenvalue of the adjacency matrix of the bipartite graph $G_{\mu_{XY}}$.

If $G_{\mu_{XY}}$ is disconnected, every connected component of $G_{\mu_{XY}}$ corresponds to a *perfect correlation* between X and Y. If the support of μ consists of all satisfying assignments to a constraint, then the polymorphism p necessarily preserves these *perfect correlations*, i.e., for all $t \in \mathbb{N}$, (X,Y) sampled from μ_t will be such that X and Y belong to the same connected component of $G_{\mu_{XY}}$. Summarizing, our result suggests that a cyclic polymorphism preserves *perfect correlations*, while *imperfect correlations* decay.

Discussion. A brief overview of the correlation decay argument is presented in Section 4. The details of the argument are fairly technical and draw upon various analytic tools such as hypercontractivity, the Berry-Esseen theorem and Fourier analysis (see Section 4). A key bottleneck in the analysis is that the individual marginals change with each iteration thereby changing the fourier spectrum of the operations involved.

Theorem 5 can be thought of as an analytic analogue of a theorem on *absorbing subalgebras* (Theorem 4.11 in [1]), which formed a key ingredient in the breakthrough work of Barto and Kozik [1]. This work of Barto and Kozik showed that a major subclass of CSPs namely *CSPs with no ability to count* can be solved using local consistency. Roughly speaking, *CSPs with no ability to count* are precisely those that don't contain linear equations within them, i.e., these CSPs don't admit gadget reductions from linear equations over a finite field.

Interestingly, we will show that the same correlation decay phenomenon is useful in solving linear equations over prime fields! We will also present an application of correlation decay towards rounding linear programming relaxations for CSPs. We will outline these two applications in the upcoming subsections.

1.2 Solving Linear Systems

The input to the algorithm consists of a linear system $Ax = b$ where $A \in \mathbb{F}_q^{m \times n}$ and $b \in \mathbb{F}_q^n$. Consider the following naive algorithm for solving the linear systems for some $N \in \mathbb{N}$.

1. $\mathcal{S}_0 \leftarrow N$ uniformly random assignments from $\mathbb{F}_q^n$.
2. for $i = 1$ to m do
   - **Selection:** $\mathcal{S}_i \leftarrow \mathcal{S}_{i-1} \cap \{x | \langle A_i, x\rangle = b_i\}$
3. Output an assignment $z \in \mathcal{S}_m$ if $\mathcal{S}_m \neq \emptyset$, else output INFEASIBLE

Clearly, if the algorithm outputs an assignment z then $Az = b$. By definition, the i^{th} *generation* $\mathcal{S}_i$ consists of uniformly random assignments that satisfy the first i equations $\{\langle A_j, x\rangle = b_j | j \leqslant i\}$. Therefore, the $i+1^{st}$ linear function $\langle A_{i+1}, x\rangle$ is either constant over $\mathcal{S}_i$ or takes every value in $\mathbb{F}_q$ with probability roughly $1/q$.

If the linear system $Ax = b$ is satisfiable and is linearly independent, then the expected size of i^{th} generation $\mathcal{S}_i$ is given by $\mathbb{E}\,|\mathcal{S}_i| = \frac{1}{q}\,\mathbb{E}\,|\mathcal{S}_{i-1}|$. Therefore the initial sample size

Input: $A \in \mathbb{F}_q^{m \times n}$, $b \in \mathbb{F}_q^m$

1. $\mathcal{S}_0 \leftarrow N$ uniformly random assignments from $\mathbb{F}_q^n$.
2. for $i = 1$ to m do
   - **Selection:** $\mathcal{T}_i \leftarrow \mathcal{S}_{i-1} \cap \{x | \langle A_i, x \rangle = b_i\}$
   - **Recombination:** Generate $\mathcal{S}_i$ by using the following sampling procedure N times independently.
     - Sample $y^{(1)}, y^{(2)}, y^{(3)}$ i.i.d from $\mathcal{T}_i$ and compute $p(y^{(1)}, y^{(2)}, y^{(3)}) = y^{(1)} - y^{(2)} + y^{(3)}$.
3. Output an assignment $z \in \mathcal{S}_m$ if $\mathcal{S}_m \neq \emptyset$, else output INFEASIBLE

Figure 1 Randomized algorithm for linear equations over $\mathbb{F}_q$.

$|\mathcal{S}_0|$ has to be at least q^m to ensure the correctness of the algorithm, making the runtime exponential.

A natural approach to fix the algorithm is as follows. After each selection step, use the polymorphism associated with linear systems in order to *create* new assignments from the existing sample. For the sake of concreteness, we fix the following polymorphism $p : [q]^3 \to [q]$ for linear systems.

$$p(y^{(1)}, y^{(2)}, y^{(3)}) = y^{(1)} - y^{(2)} + y^{(3)} \, .$$

The details of the algorithm are as shown in Figure 1.

Since p is a polymorphism, the recombination steps don't affect the progress made in the selection steps, i.e., $\mathcal{S}_i$ satisfy the first i equations for each i. While the polymorphism p is useful to maintain the sample size after each selection, the sample size alone is insufficient to ensure the success of the algorithm. We require the sample $\mathcal{S}_i$ to be somewhat similar a uniformly random sample from the set of all solutions to the first i equations.

Here is an alternate take on the issue. A finite sized sample $\mathcal{S}$ of a distribution μ has *spurious correlations* that are absent in μ. For example, in the initial sample $\mathcal{S}_0$, the first two variables $x_1, x_2 \in [q]$ will be close to independent, but there is bound to be some assignment $\alpha, \beta \in [q]^2$ such that $\mathbb{P}_{x \in \mathcal{S}_0}[x_1 = \alpha \wedge x_2 = \beta] \geqslant 1/q^2 + \Omega(1/\sqrt{N})$ due to random deviation. At the i^{th} stage, the sample $\mathcal{S}_i$ has a set of *perfect correlations* induced by the first i equations, but there are additional *spurious correlations* between the variables owing to sample size being bounded.

The magnitudes of spurious correlations in the sample need to be controlled. Otherwise, a spurious correlation could result in $\langle A_i, x \rangle \neq b_i$ for all $x \in \mathcal{S}_{i-1}$ for some i, making $\mathcal{T}_i = \emptyset$ even though the linear system is satisfiable. Each selection step reduces the sample size thereby potentially amplifying the spurious correlations. However, the recombination step exploits the correlation decay phenomena to decrease the spurious correlations. In Section 3, we will show the following.

▶ **Theorem 6.** *For all primes q, the randomized algorithm in Figure 1 with the choice $N = \lceil (150q^4 \ln q) \cdot n \rceil$ satisfies these properties.*

Completeness: *If algorithm returns $z \in \mathbb{F}_q^n$, then z satisfies the linear system $Az = b$.*

Soundness: *If the system $Ax = b$ is feasible, then with probability at least $1 - e^{-n}$ the algorithm will return a solution to the system.*

The algorithm described above is somewhat similar to a deterministic algorithm of Bulatov and Dalmau [5] for CSPs admitting Maltsev polymorphisms in that it maintains a basis for the solution space and updates the basis by including one equation in to the system at each step. We find the randomized algorithm interesting in that it admits a very generic

description that uses little about the structure of the underlying CSP. Moreover, the analysis of the algorithm crucially relies on correlation decay – a phenomenon that seems inherent to all tractable CSPs.

1.3 Rounding Linear Programs via Correlation Decay

Correlation decay can also be used towards rounding linear programming relaxations. For the sake of clarity, let us restrict our attention to 2-CSPs where every constraint has arity two. These results can be generalized to k-CSPs. Further, every CSP can be reduced to a 2-CSP while preserving the existence of cyclic terms.

First, we introduce the BasicLP linear programming relaxation for CSPs of arity two. Let Λ be a CSP of arity two over the alphabet $[q]$, and $\mathcal{I}$ be an instance of Λ. For every variable X in $\mathcal{I}$, the LP associates a probability distribution μ_X over $[q]$. For every constraint $C_i(X,Y)$ in $\mathcal{I}$, the LP associates a probability distribution μ_{XY} over $[q] \times [q]$, supported on satisfying assignments to the constraint C_i. The pairwise distributions μ_{XY} are constrained to be consistent with the marginal distributions μ_X and μ_Y. The BasicLP program for 2-CSPs is described in detail in Section 5.

▶ **Definition 7.** An LP relaxation $\mathcal{L}$ for a CSP is *sound* on an instance $\mathcal{I}$ if the feasibility of the LP relaxation $\mathcal{L}$ on $\mathcal{I}$ implies satisfiability of the instance $\mathcal{I}$.

Typically, one shows the soundness of an LP relaxation by a rounding scheme that extracts an assignment to the CSP from the LP solution. We exhibit a sufficient condition for an LP relaxation to be *sound* on an instance $\mathcal{I}$. For a constraint $C_i(X,Y)$, let $G_{\mu_{XY}}$ denote the bipartite graph whose edges are given by the support of the distribution μ_{XY}.

▶ **Theorem 8.** *Let Λ be a 2-CSP that admits a cyclic polymorphism and let $\mathcal{I}$ be an instance Λ. Suppose there exists a solution to the* BasicLP *relaxation for $\mathcal{I}$ such that all the associated graphs $G_{\mu_{XY}}$ are connected, then the* BasicLP *relaxation is* sound *on the instance $\mathcal{I}$, i.e., $\mathcal{I}$ is satisfiable.*

2 Background

We first introduce some basic notation. Let $[q]$ denote the alphabet $[q] = \{1, \ldots, q\}$. For a probability distribution μ on the finite set $[q]$ we will write μ^k to denote the product distribution on $[q]^k$ given by drawing k independent samples from μ.

If μ is a joint probability distribution on $[q]^n$ we will write $\mu_1, \mu_2, \ldots \mu_n$ for the n marginal distributions of μ. Further we will use $\mu^\times$ to denote the product distribution with the same marginals as μ. That is we define $\mu^\times \stackrel{\text{def}}{=} \mu_1 \times \mu_2 \times \cdots \times \mu_n$.

An operation p of arity k is a map $p : [q]^k \to [q]$. For a set of k assignments $x^{(1)}, \ldots, x^{(k)} \in [q]^n$, we will use $p(x^{(1)}, \ldots, x^{(k)}) \in [q]^n$ to be the assignment obtained by applying the operation p on each coordinate of $x^{(1)}, \ldots, x^{(k)}$ separately. More formally, let $x_j^{(i)}$ be the jth coordinate of x_i. We define

$$p(x^{(1)} \ldots x^{(k)}) = \left(p(x_1^{(1)} \ldots x_1^{(k)}), p(x_2^{(1)} \ldots x_2^{(k)}), \ldots, p(x_n^{(1)} \ldots x_n^{(k)})\right).$$

▶ **Definition 9.** For two operations $p_1 : [q]^{k_1} \to [q]$ and $p_2 : [q]^{k_2} \to [q]$, define an operation $p_1 \otimes p_2 : [q]^{k_1 \times k_2} \to [q]$ as follows:

$$p_1 \otimes p_2(\{x_{ij}\}_{i \in [k_1], j \in [k_2]}) = p_1\left(p_2(x_{11}, x_{12}, \ldots, x_{1k_2}), \ldots, p_2(x_{k_1 1}, x_{k_1 2}, \ldots, x_{k_1 k_2})\right)$$

▶ **Lemma 10** (Hoeffding bound)**.** *Suppose $Z_1, \ldots, Z_N$ are complex-valued random variables such that $|Z_i|$ is always bounded by 1. If $Z = \frac{1}{N}\sum_i Z_i$ then, $\mathbb{P}[|Z - \mathbb{E}[Z]| \geqslant \delta] \leqslant 2e^{-\delta^2 N/4}$.*

3 Solving Linear Systems via Correlation Decay

In this section, we will analyze the randomized algorithm for linear equations in Figure 1 and present a proof of Theorem 6. We begin with setting up some notation dealing with affine subspaces of $\mathbb{F}_q^n$.

▶ **Definition 11.** For an affine subspace $V \subseteq \mathbb{F}_q^n$ and $b \in \mathbb{F}_q$, set $V_b^\perp \stackrel{\text{def}}{=} \{w | \forall x \in V, \langle x, w\rangle = b\}$, and let $V^\perp \stackrel{\text{def}}{=} \cup_{b\in\mathbb{F}_q} V_b$.

In the setting of linear equations, correlations can be measured using Fourier analysis. Therefore, we recall the definition of characters over $\mathbb{F}_q^n$.

▶ **Definition 12.** For every $w \in \mathbb{F}_q^n$, the corresponding character χ_w is a function $\chi_w : \mathbb{F}_q^n \to \mathbb{C}$ given by $\chi_w(x) = \omega^{\langle w,x\rangle}$, where ω is a primitive q^{th} root of unity and $\langle w, x\rangle$ denotes the inner product of w and x over $\mathbb{F}_q$.

We will quantify the *spurious correlations* in our sample using the notion of bias as defined below.

▶ **Definition 13** (Bias)**.** For a vector $w \in \mathbb{F}_q^n$ and a multiset $\mathcal{S} \subseteq \mathbb{F}_q^n$, define the bias of w over $\mathcal{S}$ as,

$$\text{bias}_w(\mathcal{S}) = |\mathop{\mathbb{E}}_{x\in\mathcal{S}}[\chi_w(x)]| = \frac{1}{|\mathcal{S}|}|\sum_{x\in\mathcal{S}} \chi_w(x)|\,.$$

An ε-biased sample from an affine subspace V is one in which all the *spurious correlations* are bounded by ε. Formally,

▶ **Definition 14** (ε-biased sample)**.** For $\varepsilon \in [0, 1]$, a multiset of vectors $\mathcal{S} \in \mathbb{F}_q^n$ is a ε-biased sample of an affine subspace $V \subseteq \mathbb{F}_q^n$ if $\mathcal{S} \subseteq V$ and for all $w \notin V^\perp$, $|\text{bias}_w(\mathcal{S})| \leqslant \varepsilon$.

Effect of Selection on Bias

▶ **Lemma 15.** *Let $\mathcal{S}$ be a ε-biased sample from an affine subspace V. For all $w \notin V^\perp$ and $b \in \mathbb{F}_q$, the following holds:*

1. $\mathbb{P}_{z\in\mathcal{S}}[\langle w, z\rangle = b] \in \left[\frac{1}{q} - \varepsilon, \frac{1}{q} + \varepsilon\right]$.
2. *If $\mathcal{T} = \mathcal{S} \cap \{z | \langle w, z\rangle = b\}$ then $\mathcal{T}$ is a $q\varepsilon/(1-q\varepsilon)$-biased sample from the affine subspace $V \cap \{z \in \mathbb{F}_q^n | \langle w, z\rangle = b\}$.*

Proof. Let $\mathbb{I}_{\langle w,z\rangle=b}$ be the indicator of the event that $\langle w, z\rangle = b$. Using the identity $\mathbb{I}_{\langle w,z\rangle=b} = \frac{1}{q}\sum_{\alpha\in\mathbb{F}_q}(\chi_w(z)\omega^{-b})^\alpha$, we can write

$$\mathop{\mathbb{P}}_{z\in\mathcal{S}}[\langle w, z\rangle = b] = \mathop{\mathbb{E}}_{z\in\mathcal{S}}\left[\frac{1}{q}\sum_{\alpha\in\mathbb{F}_q}(\chi_w(z)\omega^{-b})^\alpha\right]\,.$$

Simplifying the above expression using the identity $\chi_w(z)^\alpha = \chi_{\alpha w}(z)$ we get

$$\mathop{\mathbb{P}}_{z\in\mathcal{S}}[\langle w, z\rangle = b] = \frac{1}{q} + \frac{1}{q}\sum_{\alpha\in\mathbb{F}_q/\{0\}} \omega^{-\alpha b} \mathop{\mathbb{E}}_{z\in\mathcal{S}}[\chi_{\alpha w}(z)]\,.$$

Hence,

$$\left|\mathop{\mathbb{P}}_{z\in\mathcal{S}}[\langle w, z\rangle = b] - \frac{1}{q}\right| \leqslant \frac{1}{q}\sum_{\alpha\in\mathbb{F}_q/\{0\}}\left|\mathop{\mathbb{E}}_{z\in S}[\chi_{\alpha w}(z)]\right| \leqslant \frac{1}{q}\sum_{\alpha\in\mathbb{F}_q/\{0\}}|\text{bias}_{\alpha w}(\mathcal{S})| \leqslant \varepsilon$$

Let V' denote the affine subspace $V' = V \cap \{z \in \mathbb{F}_q^n | \langle w, z\rangle = b\}$. Clearly, $\mathcal{T}$ is a subset of V' if $\mathcal{S}$ is a subset of V. By definition, for every $v \in \mathbb{F}_q^n$ the bias over $\mathcal{T}$ is given by

$$\text{bias}_v(\mathcal{T}) = \mathop{\mathbb{E}}_{z\in\mathcal{T}}[\chi_v(z)] = \mathop{\mathbb{E}}_{z\in\mathcal{S}}[\chi_v(z)|\langle w, z\rangle = b] = \frac{1}{\mathbb{P}[\langle w, z\rangle = b]} \mathop{\mathbb{E}}_{z\in\mathcal{S}}\left[\chi_v(z)\mathbb{I}_{\langle w,z\rangle=b}\right] \quad (3.1)$$

Evaluating the expectation inside,

$$\mathop{\mathbb{E}}_{z\in\mathcal{S}}\left[\chi_v(z)\mathbb{I}_{\langle w,z\rangle=b}\right] = \frac{1}{q}\sum_{\alpha\in\mathbb{F}_q}\mathop{\mathbb{E}}_{z\in\mathcal{S}}\left[\chi_v(z)(\chi_w(z)\omega^{-b})^\alpha\right] = \frac{1}{q}\sum_{\alpha\in\mathbb{F}_q}\omega^{-b\alpha}\text{bias}_{v+\alpha w}(\mathcal{S}) \quad (3.2)$$

For a vector $v \notin (V')^\perp$, we claim that $v + \alpha w \notin V^\perp$ for any $\alpha \in \mathbb{F}_q$. Suppose not, then $v + \alpha w \in V^\perp$ which implies that for some $b' \in \mathbb{F}_q$, we have $\langle v + \alpha w, z\rangle = b'$ for all $z \in V$. This implies that for all $z \in V'$, $\langle v, z\rangle = \langle v + \alpha w, z\rangle - \alpha\langle w, z\rangle = b' - \alpha b$ – a constant, a contradiction to the fact that $v \notin (V')^\perp$.

Since $\mathcal{S}$ is an ε-biased sample from V and $v + \alpha w \notin V^\perp$, we have $\text{bias}_{v+\alpha w}(\mathcal{S}) \leqslant \varepsilon$. Using this bound in (3.2) we get $\mathbb{E}_{z\in S}\left[\chi_v(z)\mathbb{I}_{\langle w,z\rangle=b}\right] \leqslant \varepsilon$ .Substituting in (3.1) and using the fact that $\mathbb{P}[\langle w, z\rangle = b] \geqslant 1/q - \varepsilon$, we conclude that $\text{bias}_v(\mathcal{T}) \leqslant \frac{\varepsilon}{(1/q-\varepsilon)}$ for any $v \in V^\perp$. ◀

Recombination Reduces Bias

▶ **Lemma 16.** *For all $i \in [m]$, if the sample $\mathcal{T}_i \in \mathbb{F}_q^n$ is a ε-biased sample of an affine subspace $V \subseteq \mathbb{F}_q^n$, then for all $\delta > 0$, then the sample $\mathcal{S}_i$ generated by recombination is a $\varepsilon^3 + \delta$-biased sample from V with probability at least $1 - 2q^n e^{-\delta^2 N/4}$.*

Proof. The multiset $\mathcal{S}_i$ consists of N-i.i.d samples from a probability distribution. Fix any $w \in V^\perp$. The expected value of the bias w over U is given by,

$$\mathbb{E}\left[\text{bias}_w(U)\right] = \mathbb{E}\left[\frac{1}{N}\sum_{z\in U}\chi_w(z)\right] = \frac{1}{N}\sum_{z\in U}\mathbb{E}\left[\chi_w(z)\right]$$

For every sample $z = p(y^{(1)}, y^{(2)}, y^{(3)})$ in $\mathcal{S}_i$, the expectation of the bias is given by

$$|\mathop{\mathbb{E}}_{y^{(j)}\in\mathcal{T}_i}[\chi_w(y_1 - y_2 + y_3)]| = |\mathop{\mathbb{E}}_{y^{(j)}\in\mathcal{T}_i}[\chi_w(y^{(1)})\overline{\chi_w(y^{(2)})}\chi_w(y^{(3)})]| = |\prod_{j=1}^{3}\mathop{\mathbb{E}}_{y^{(j)}\in\mathcal{T}_i}[\chi_w(y^{(j)}]|$$

Therefore, the expected value of the bias of w over $\mathcal{S}_i$ is, $\mathbb{E}\left[\text{bias}_w(\mathcal{S}_i)\right] \leqslant \text{bias}_w^3(\mathcal{T}_i)$. Since $\mathcal{T}_i$ is a ε-biased sample from V, for each $w \in V^\perp$ $|\text{bias}_w(\mathcal{T}_i)| \leqslant \varepsilon$. Therefore, we get that

$$\begin{aligned}\mathbb{P}\left[|\text{bias}_w(\mathcal{S}_i)| \geqslant \varepsilon^3 + \delta\right] &\leqslant \mathbb{P}\left[|\text{bias}_w(\mathcal{S}_i) - \mathbb{E}[\text{bias}_w(\mathcal{S}_i)]| \geqslant \delta\right] \\ &\leqslant 2e^{-\delta^2 N/4} \qquad \text{(Lemma 10)}\end{aligned}$$

By a union bound over all q^n characters $w \in \mathbb{F}_q^n$, we get the desired result. ◀

Analysis of the Algorithm

Proof of Theorem 6. Fix $\varepsilon = \frac{1}{2q^2}$ and $\delta = \frac{1}{6q}$. Let V_k denote the affine subspace consisting of solutions to the first k equations $\{A_i x = b_i | 1 \leqslant i \leqslant k\}$. We will show the following claim from which Theorem 6 follows immediately.

▶ **Claim 17.** *If $N = \lceil(150q^4 \ln q) \cdot n\rceil$, then with probability at least $1 - e^{-n}$ the following holds: for all $0 \leqslant k \leqslant m$, $\mathcal{S}_k$ is an ε-biased sample from V_k for $\varepsilon = \frac{1}{4q}$.*

The argument is by induction on k. For $k = 0$, the set S_0 consists of N random vectors from $\mathbb{F}_q^n$ and $V_0 = \mathbb{F}_q^n$. By definition, $V_0^\perp = \mathbb{F}_q^n - \{0\}$. For every $w \in \mathbb{F}_q^n - \{0\}$, the bias of w is given by, $\text{bias}_w(\mathcal{S}_0) = \frac{1}{N}\sum_{i=1}^N \chi_w(z_i)$. In particular, it is easy to see that $\mathbb{E}[\text{bias}_w(\mathcal{S}_0)] = 0$. By applying Lemma 10, we get that

$$\mathbb{P}[|\text{bias}_w(\mathcal{S}_0)| \geqslant \varepsilon] \leqslant 2e^{-\varepsilon^2 N/4}\,.$$

By a simple union bound, $\mathcal{S}_0$ is ε-biased sample with probability at least $1 - q^n 2e^{-\varepsilon^2 N/4} = 1 - q^n 2e^{-N/16q^4}$.

Let us suppose $\mathcal{S}_\ell$ is a ε-biased sample from V_ℓ. By Lemma 15, we get that $\mathcal{T}_{\ell+1}$ is a $q\varepsilon/(1-q\varepsilon)$-biased sample from $V_{\ell+1}$. By Lemma 16, with probability at least $1 - 2q^n e^{-\delta^2 N/4}$, the bias of $\mathcal{S}_{\ell+1}$ obtained by recombination is at most,

$$\text{bias}(\mathcal{S}_{\ell+1}) \leqslant \text{bias}(\mathcal{T}_{\ell+1})^3 + \delta \leqslant (q\varepsilon/(1-q\varepsilon))^3 + \delta = \frac{1}{q^3} + \delta < \varepsilon\,.$$

Applying a union bound over all $\ell \in \{0,\ldots,m\}$, with probability at least $1 - 2mq^n e^{-N/144q^2}$, $\mathcal{S}_\ell$ is an ε-biased sample from V_ℓ for all $\ell \in \{1,\ldots,m\}$. Setting $N = \lceil (150q^4 \ln q)\cdot n\rceil$, the claim follows. ◀

4 Correlation Decay

In this section we state our main theorem regarding the decay of correlation between random variables under repeated applications of cyclic operations. Recall that Theorem 5 states the theorem for two variables. Throughout this section we will use this two variable case as a running example. We begin by defining a quantitative measure of correlation and using it to bound the statistical distance to a product distribution.

4.1 Correlation and Statistical Distance

To gain intuition for our measure of correlation consider the example of two boolean random variables X and Y with joint distribution μ. In this case we will measure correlation by taking the supremum over appropriately normalized test functions $f, g : \{0,1\} \to \mathbb{R}$ and computing $\mathbb{E}[f(X)g(Y)]$.

▶ **Definition 18.** Let X, Y be discrete-valued random variables with joint distribution μ. Let $\Omega_1 = ([q_1], \mu_1)$ and $\Omega_2 = ([q_2], \mu_2)$ denote the probability spaces corresponding to X, Y respectively. The correlation $\rho(X, Y)$ is given by

$$\rho(X,Y) \stackrel{\text{def}}{=} \sup_{f,g} \mathbb{E}[f(X)g(Y)]$$

where the supremum runs over all f, g where $\mathbb{E}[f] = \mathbb{E}[g] = 0$ and $\mathbf{Var}[f] = \mathbf{Var}[g] = 1$. We will interchangeably use the notation $\rho(\mu)$ or $\rho(\Omega_1, \Omega_2)$ to denote the correlation.

To see that this notion of correlation makes intuitive sense, suppose X and Y are independent. In this case correlation is zero because $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\,\mathbb{E}[g(Y)] = 0$. Next suppose that $X = Y = 1$ with probability $\frac{1}{2}$ and $X = Y = 0$ with probability $\frac{1}{2}$. In this case we can set $f(1) = g(1) = 1$ and $f(0) = g(0) = -1$ to obtain $\mathbb{E}[f(X)g(Y)] = 1$. This matches up with the intuition that such an X and Y are perfectly correlated. We now give the general definition for our measure of correlation. Next we show that, as the correlation for a pair of random variables X and Y becomes small, the variables become nearly independent.

▶ **Lemma 19.** *Let X, Y be discrete-valued random variables with joint distribution μ_{XY} and respective marginal distributions μ_X and μ_Y. If X takes values in $[q_1]$ and Y takes values in $[q_2]$, then $\|\mu_{XY} - \mu_X \times \mu_Y\|_1 \leqslant \min(q_1, q_2)\rho(X, Y)$*

It turns out that there is also a simple combinatorial condition that is essentially equivalent to a bound on the correlation. First we define a natural bipartite graph associated to a joint distribution.

▶ **Definition 20.** Let X, Y be jointly distributed according to μ as in Definition 18. Define a bipartite graph G_μ on vertex set $([q_1], [q_2])$ by adding an edge (a, b) whenever $\mathbb{P}_\mu[X = a, Y = b] > 0$.

Now the following lemma from [10] states that $\rho(\mu) < 1$ whenever the graph G_μ is connected.

▶ **Lemma 21** (Lemma 2.9 in [10])**.** *Let μ be a joint distribution where the minimum non-zero probability that μ assigns to any element is α. If G_μ is connected then $\rho(\mu) < 1 - \frac{\alpha^2}{2}$.*

In addition, if G_μ is disconnected, then $\rho(\mu) = 1$. Therefore checking if $\rho(\mu) < 1$ amounts to checking connectivity of G_μ.

4.2 Proof Overview

To begin with, we explain why one should expect correlations to decay under repeated applications of cyclic operations. Consider the simple example of two boolean random variables X and Y with a joint distribution μ. Let the marginal distributions of X and Y be uniform and let us suppose $X = Y$ with probability $\frac{1}{2} + \gamma$ and $X \neq Y$ with the remaining probability. Let $p : \{0,1\}^k \to \{0,1\}$ be the majority operation on k bits.

Next suppose we draw k samples (X_i, Y_i) from μ and evaluate $p(X_1 \dots X_k)$ and $p(Y_1 \dots Y_k)$. Since the marginal distributions of both X and Y are uniform, the same is true for $p(X_1 \dots X_k)$ and $p(Y_1 \dots Y_k)$. However, the probability that $p(X_1 \dots X_k) = p(Y_1 \dots Y_k)$ is strictly less than $\frac{1}{2} + \gamma$. To see why first let $F : \{-1, 1\} \to \{-1, 1\}$ be the majority function where 1 encodes boolean 0 and -1 encodes boolean 1. Note that the probability that $F(X_1 \dots X_k) = F(Y_1 \dots Y_k)$ is given by $\frac{1}{2} + \frac{1}{2}\mathbb{E}[F(X_1 \dots X_k)F(Y_1 \dots Y_k)]$.

Now if we write the Fourier expansion of F the above expectation is

$$\sum_{S,T} \hat{F}_S \hat{F}_T \, \mathbb{E}\left[\prod_{i \in S} X_i \prod_{j \in T} Y_j\right] = \sum_S \hat{F}_S^2 \prod_{i \in S} \mathbb{E}[X_i Y_i] = \sum_S \hat{F}_S^2 (2\gamma)^{|S|}$$

Suppose first that all the non-zero Fourier coefficients $\hat{F}_S$ have $|S| = 1$. In this case the probability that $F(X_1 \dots X_k) = F(Y_1 \dots Y_k)$ stays the same since $\frac{1}{2} + \frac{1}{2}(2\gamma) = \frac{1}{2} + \gamma$. However, in the case of majority, it is well known that $\sum_{|S|=1} \hat{F}_S^2 < 1 - c$ for a constant $c > 0$. Thus, the expectation is in fact given by

$$\mathbb{E}[F(X_1 \dots X_k)F(Y_1 \dots Y_k)] \leqslant (1 - c)(2\gamma) + c(2\gamma)^2 < 2\gamma$$

Thus the probability that $F(X_1 \dots X_k) = F(Y_1 \dots Y_k)$ is strictly less than $\frac{1}{2} + \gamma$. Therefore, if we repeatedly apply the majority operation, we should eventually have that X and Y become very close to independent.

There are two major obstacles to generalizing the above observation to arbitrary cyclic operations. First, for a general operation p, we will not be able to explicitly compute the entire Fourier expansion. Instead, we will have to use the fact that p is cyclic to get a bound

on the total Fourier mass on degree-one terms. Second, unlike in our example, the marginal distributions of X and Y may change after every application of p. This means that the correct Fourier basis to use also changes.

The fact that the marginal distributions change under p causes difficulties even for the simple example of the boolean OR operation on two bits. Consider a highly biased distribution over $0, 1$ given by $X = 1$ with probability ε and $X = 0$ with probability $1 - \varepsilon$. Now consider the function $f(X) = \frac{1}{2}(X_1 + X_2)$. Note that this function agrees with OR except when $X_1 \neq X_2$. Thus, $f(X) = OR(X)$ with probability $1 - 2\varepsilon(1 - \varepsilon) > 1 - 2\varepsilon$. This means that as ε approaches zero, OR approaches a function f with $\sum_{|S|=1} \hat{f}_S^2 = 1$.

Thus, there are distributions for which the correlation decay under the OR operation approaches zero. This means that we cannot hope to prove a universal bound on correlation decay for every marginal distribution, even in this very simple case. The problem for the general case is that as we repeatedly apply some operation p it could be that the marginals converge to some point where p does not result in correlation decay.

It is useful to note that for the OR operation, the probability that $X = 1$ increases under every application. Thus, as long as the initial distribution has a non-negiligible probability that $X = 1$, we will have that correlation does indeed decay in each step. Though this particular observation applies only to the OR operation, our proof in the general case does rely on the fact that, using only properties of the initial distribution of X we can get bounds on correlation decay in every step. In summary, we are able to achieve correlation decay for arbitrary cyclic operations. We now state our main theorem to this effect.

▶ **Theorem 22** (Correlation Decay). *Let μ be a distribution on $[q]^n$. Let $X_1, \ldots, X_n$ be the jointly distributed $[q]$-valued random variables drawn from μ. Further, let $\rho = \max_i \rho((X_1; X_2; \ldots; X_{i-1}), X_i) < 1$ and λ be the minimum probability of an atom in the marginal distributions $\{\mu_i\}_{i\in[n]}$. For any $\eta > 0$ and $r \geqslant \Omega_q\left(\frac{\log \lambda}{\log \rho} \log^2\left(\frac{qn}{\eta}\right)\right)$, if $p_1, \ldots, p_r$ is a sequence of operations each of which are cyclic terms then,*

$$\|p_1 \otimes p_2 \otimes \ldots \otimes p_r(\mu) - p_1 \otimes p_2 \otimes \ldots \otimes p_r(\mu^\times)\|_1 \leqslant \eta\,.$$

We now give a brief outline of the main ideas of the proof. For a cyclic operation p, the degree-one Fourier coefficients with respect to any distribution are all equal. Suppose that for some probability distribution μ, the operation p has nearly all of its Fourier mass on degree one coefficients. Then $p(x)$ is close to a sum of independent random variables. Therefore, a quantitative version of the Central Limit Theorem (in particular a variant of the Berry-Esseen Theorem), implies that $p(x)$ is close to a Gaussian random variable.

Next, since $p(x)$ is an operation on $[q]$ it only takes q different values. This should then give us a contradiction: a random variable taking only q different values cannot be close to a continuous random variable like a Gaussian. Unfortunately there is a problem with this argument. The error term in the Berry-Esseen theorem depends on the L^3-norm of the independent random variables. Thus, we must control the L^3-norms of the Fourier basis for p under the distribution μ in order for the previous argument to work.

Now the problem is that, even in the case of the OR operation, the L^3-norms of vectors in the Fourier basis can become arbitrarily large as μ changes under repeated applications of the operation. So, we are forced to prove that the elements of the Fourier basis that have high L^3-norm somehow have very small contribution to the correlation. The main idea here is that the correlation of a joint distribution μ is determined by the singular values of a certain linear operator T_μ known as the conditional expectation operator.

We establish a trade-off between the L^3-norm of the singular vectors of T_μ and the correlation contributed by their corresponding singular values. In particular we show that,

for any singular vector v of T_μ with large L^3-norm, the corresponding singular value must be small. This in turn implies that we need only look at the elements of the Fourier basis with small L^3-norm, as all the other elements do not contribute to the correlation of μ.

Our proof relating L^3-norms to singular values relies heavily on the fact that the operator T_μ is *hypercontractive*. Briefly, hypercontractivity is a property that allows us to bound $\|T_\mu f\|_3 \leqslant \|f\|_2$ under certain conditions on μ. If f is a singular vector of T_μ with singular value σ and unit L^2-norm, we then have $\sigma\|f\|_3 \leqslant 1$. This is precisely the sort of trade-off between the L^3-norm of f and the corresponding singular value that we use in our proof. We defer the details of the proof of the theorem to the full version.

5 Soundness of a LP relaxation

In this section, we use correlation decay to give a sufficient condition for when linear programming can be used to solve a CSP with a cyclic polymorphism. For clarity we state and prove everything in this section for CSPs where every constraint has arity two. First we introduce the basic LP relaxation for CSPs of arity two.

Let Λ be a CSP of arity two over the alphabet $[q]$, and $\mathcal{I}$ be an instance of Λ. For every variable X in $\mathcal{I}$ and element $a \in [q]$ we introduce an LP variable $\mu_X(a)$, which can be thought of as the probability that X is assigned a. For every constraint $C_i(X,Y)$ in $\mathcal{I}$ and every pair of elements $a, b \in [q]$ we introduce an LP variable $\mu_{XY}(a,b)$, which can be thought of as the probability that the pair of variables (X,Y) are assigned the values (a,b). The basic LP relaxation for instance $\mathcal{I}$ is then given by the following LP feasibility problem.

BasicLP Relaxation

$\sum_{a\in[q]} \mu_X(a) = 1$	$\forall X$ (μ_X is a probability distribution)
$\sum_{a,b\in[q]} \mu_{XY}(a,b) = 1$	$\forall X,Y$ (μ_{XY} is a probability distribution)
$\sum_{b\in[q]} \mu_{XY}(a,b) = \mu_X(a)$	$\forall b, C_i(X,Y)$ (local consistency for X)
$\sum_{a\in[q]} \mu_{XY}(a,b) = \mu_Y(b)$	$\forall b, C_i(X,Y)$ (local consistency for Y)
$\mu_{XY}(a,b) = 0$	$\forall C_i(X,Y), a, b$ s.t. $C_i(a,b) = 0$ (μ_{XY} satisfies $C_i(X,Y)$)

Proof of Theorem 8. Let p be a cyclic polymorphism of Λ. For each constraint $C_i(X,Y)$, since $G_{\mu_{XY}}$ is connected, Theorem 5 implies that as $k \to \infty$,

$$\|p^{\otimes k}(\mu_{XY}) - p^{\otimes k}(\mu_X) \times p^{\otimes k}(\mu_Y)\|_1 \to 0$$

Now independently sample the value of every variable V from the distribution $p^{\otimes k}(\mu_V)$. The joint distribution of values for every pair (X,Y) is precisely the product distribution $p^{\otimes k}(\mu_X) \times p^{\otimes k}(\mu_Y)$. Thus, for every constraint $C_i(X,Y)$, the distribution of the values for (X,Y) can be made arbitrarily close to the distribution $p^{\otimes k}(\mu_{XY})$ by taking k large enough. Since p is a polymorphism of Λ and μ_{XY} is a distribution on satisfying assignments to $C_i(X,Y)$, we have that $p^{\otimes k}(\mu_{XY})$ is a distribution on satisfying assignments.

Therefore, for large enough k, there will be a non-zero probability that every constraint $C_i(X,Y)$ is satisfied. In particular, this implies that the instance $\mathcal{I}$ is satisfiable. ◀

References

1. Libor Barto and Marcin Kozik. Constraint satisfaction problems of bounded width. In *FOCS*, pages 595–603, 2009. doi:10.1109/FOCS.2009.32.
2. Libor Barto and Marcin Kozik. Absorbing subalgebras, cyclic terms, and the constraint satisfaction problem. *Logical Methods in Computer Science*, 8(1), 2012. doi:10.2168/LMCS-8(1:7)2012.
3. Andrei A. Bulatov. Tractable conservative constraint satisfaction problems. In *LICS*, pages 321–330, 2003. doi:10.1109/LICS.2003.1210072.
4. Andrei A. Bulatov. A dichotomy theorem for constraint satisfaction problems on a 3-element set. *J. ACM*, 53(1):66–120, 2006. doi:10.1145/1120582.1120584.
5. Andrei A. Bulatov and Víctor Dalmau. A simple algorithm for mal'tsev constraints. *SIAM J. Comput.*, 36(1):16–27, 2006. doi:10.1137/050628957.
6. Andrei A. Bulatov, Andrei A. Krokhin, and Peter Jeavons. Constraint satisfaction problems and finite algebras. In *ICALP*, pages 272–282, 2000. doi:10.1007/3-540-45022-X_24.
7. Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic snp and constraint satisfaction: A study through datalog and group theory. *SIAM J. Comput.*, 28(1):57–104, 1998. doi:10.1137/S0097539794266766.
8. Gábor Kun and Mario Szegedy. A new line of attack on the dichotomy conjecture. In *STOC*, pages 725–734, 2009. doi:10.1145/1536414.1536512.
9. Miklós Maróti and Ralph McKenzie. Existence theorems for weakly symmetric operations. *Algebra Universalis*, 59:463–489, 2008. doi:10.1007/s00012-008-2122-9.
10. E. Mossel. Gaussian bounds for noise correlation of functions. *In FOCS'08: Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, 2008.
11. Thomas J. Schaefer. The complexity of satisfiability problems. In *STOC*, pages 216–226, 1978. doi:10.1145/800133.804350.