Electronic Colloquium on Computational Complexity, Report No. 7 (2015)

Combinatorial Optimization Algorithms via Polymorphisms

Jonah Brown-Cohen
University of California, Berkeley
California, Berkeley
jonahbc@eecs.berkeley.edu

Prasad Raghavendra *
University of California, Berkeley
California, Berkeley
prasad@eecs.berkeley.edu

January 8, 2015

Abstract

An elegant characterization of the complexity of constraint satisfaction problems has emerged in the form of the the algebraic dichotomy conjecture of [BKJ00]. Roughly speaking, the characterization asserts that a CSP Λ is tractable if and only if there exist certain non-trivial operations known as *polymorphisms* to combine solutions to Λ to create new ones. In an entirely separate line of work, the unique games conjecture yields a characterization of approximability of Max-CSPs. Surprisingly, this characterization for Max-CSPs can also be reformulated in the language of polymorphisms.

In this work, we study whether existence of non-trivial polymorphisms implies tractability beyond the realm of constraint satisfaction problems, namely in the value-oracle model. Specifically, given a function f in the value-oracle model along with an appropriate operation that never increases the value of f, we design algorithms to minimize f. In particular, we design a randomized algorithm to minimize a function $f : [q]^n \to \mathbb{R}$ that admits a fractional polymorphism which is *measure preserving* and has a *transitive symmetry*.

We also reinterpret known results on MaxCSPs and thereby reformulate the unique games conjecture as a characterization of approximability of max-CSPs in terms of their *approximate polymorphisms*.

*Supported by NSF Career Award and Alfred. P. Sloan Fellowship

ISSN 1433-8092

Contents

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. Similar situations arise even when we consider approximations for combinatorial optimization problems. For instance, Min s-t Cut can be solved exactly, while Min 3-way cut is NP-complete and can be approximated to a factor of $\frac{12}{11}$. It is natural to ask why the approximation ratio for Min s-t Cut is 1, while it is $\frac{12}{11}$ for Min 3-way cut.

Over the last decade, unifying theories of tractability have emerged for the class of constraint satisfaction problems (CSP) independently for the exact and the optimization variants. Surprisingly, the two emerging theories for the exact and optimization variants of CSPs appear to coincide! While these candidate theories remain conjectural for now, they successfully explain 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.

Satisfiability. 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 1.1 (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 [Sch78] 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 [FV98] 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 4 [Bul06].

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 [BKJ00] 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 [KS09] 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 for 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 1.2 (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 functions, Horn-SAT has OR functions, and Dual Horn-SAT has the AND functions as polymorphisms. Roughly speaking, Bulatov et al. [BKJ00] conjectured that the existence of non-dictator polymorphisms characterizes CSPs that are tractable. Their work showed that the set of polymorphisms $\mathrm{Poly}(\Lambda)$ of a CSP Λ characterizes the complexity of Λ-Sat. Formally,

Theorem 1.3. *[BKJ00] If two CSPs Λ_1, Λ_2 have the same set of polymorphisms $\mathrm{Poly}(\Lambda_1) = \mathrm{Poly}(\Lambda_2)$, then Λ_1-sat is polynomial-time reducible to Λ_2-sat and vice versa.*

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 [BK12]. A polymorphism $p : [q]^k \to [q]$ is called a *cyclic term* if

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

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

Conjecture 1.4. *([BKJ00, MM08, BK12]) 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 resolved.

Theorem 1.5. *[BKJ00, MM08, BK12] 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 [Bul03], CSPs with no ability to count [BK09] and CSPs with Maltsev operations [BD06]. Recently, Kun and Szegedy reformulated the algebraic dichotomy conjecture using analytic notions similar to influences [KS09].

Polymorphisms in Optimization. Akin to polymorphisms for Λ-Sat, we define the notion of *approximate polymorphisms* for optimization problems. Roughly speaking, an approximate polymorphism is a probability distribution $\mathcal{P}$ over a set of operations of the form $p : [q]^k \to [q]$. In particular, an approximate polymorphism $\mathcal{P}$ can be used to combine k solutions to an optimization problem and produce a probability distribution over new solutions to the same instance. Unlike the case of exact CSPs, here the polymorphism outputs several new solutions. $\mathcal{P}$ is an (c,s)-approximate polymorphism if, given a set of solutions with average value c, the average value of the output solutions is always at least s.

For the sake of exposition, we present an example here. Suppose $f : \{0,1\}^n \to \mathbb{R}^+$ is a *supermodular* function. In this case, f admits a 1-approximate polymorphism described below.

$$\mathcal{P} = \begin{cases} OR(x_1, x_2) & \text{with probability } \frac{1}{2} \\ AND(x_1, x_2) & \text{with probability } \frac{1}{2} \end{cases}$$

Given any two assignments $X^{(1)}$, $X^{(2)}$, the polymorphism $\mathcal{P}$ outputs two solutions $X^{(1)} \vee X^{(2)}$ and $X^{(1)} \wedge X^{(2)}$. The supermodularity of the function f implies

$$\mathrm{val}(X^{(1)}) + \mathrm{val}(X^{(2)}) \leqslant \mathrm{val}(X^{(1)} \vee X^{(2)}) + \mathrm{val}(X^{(1)} \wedge X^{(2)}) \ ,$$

that the average value of solutions output by $\mathcal{P}$ is at least the average value of solutions input to it. The formal definition of an (c, s)-approximate polymorphism is as follows.

Definition 1.6 ((c, s)-approximate polymorphism)**.** *Fix a function $f : [q]^n \to \mathbb{R}^+$. A probability distribution $\mathcal{P}$ over operations $p : [q]^R \to [q]$ is an (c, s)-approximate polymorphism for f if the following holds: for every R assignments $X^{(1)}, X^{(2)}, \ldots, X^{(R)} \in [q]^n$, if $\mathbb{E}_i f(X^{(i)}) \geqslant c$ for all i then,*

$$\mathop{\mathbb{E}}_{p \in \mathcal{P}} [f(p(X^{(1)}, \ldots, X^{(R)}))] \geqslant s$$

Here, $p(X^{(1)}, \ldots, X^{(R)})$ is the assignment obtained by applying the operation p coordinate-wise.

It is often convenient to use a coarser notion of approximation, namely the approximation ratio for all c. In this vein, we define α-approximate polymorphisms.

Definition 1.7 (α-approximate polymorphism)**.** *A probability distribution $\mathcal{P}$ over operations $p : [q]^R \to [q]$ is an α-approximate polymorphism for $f : [q]^n \to \mathbb{R}^+$, if it is an $(c, \alpha \cdot c)$-approximate polymorphism for all $c \geqslant 0$.*

We will refer to a 1-approximate polymorphism as a *fractional polymorphism* along the lines of Cohen et al. [DMP06].

Approximating CSPs. To set the stage for the main result of the paper, we will reformulate prior work on approximation of CSPs in the language of polymorphisms. While the results stated here are not new, their formulation in the language of approximate polymorphisms has not been carried out earlier. This reformulation highlights the the close connection between the algebraic dichotomy conjecture (a theory for exact CSPs) and the unique games conjecture (a theory of approximability of CSPs). We believe this reformulation is also one of the contributions of this work and have devoted Section 7 to give a full account of it. The results we will state now hold for a fairly general class of problems that include both maximization and minimization problems with local bounded payoff functions (referred to as *value CSP* in [CCC+13]). However, for the sake of concreteness, let us restrict our attention to the problem Max-Λ for a CSP Λ.

Problem 1.8 (Max-Λ)**.** *Given an instance $\Im$ of the Λ-CSP, find an assignment that satisfies the maximum number (equivalently fraction) of constraints.*

The formal definition of α-approximate polymorphisms is as follows.

Definition 1.9. *A probability distribution $\mathcal{P}$ over operations $p : [q]^k \to [q]$ is an α-approximate polymorphism for* Max-Λ*, if $\mathcal{P}$ is an $(c, \alpha \cdot c)$-approximate polymorphism for every instance $\Im$ of* Max-Λ.

In the above definition, we treat each instance $\Im$ as a function $\Im : [q]^n \to \mathbb{R}^+$ that we intend to maximize.

As in the case of Λ-SAT, dictator functions are 1-approximate polymorphisms for every MAX-Λ. We will use the analytic notion of *influences* to preclude the dictator functions (see Section 7). Specifically, we define τ-quasirandomness as follows.

Definition 1.10. *An approximate polymorphism $\mathcal{P}$ is (τ, d)-quasirandom if, for every distribution μ on $[q]$,*

$$\mathop{\mathbb{E}}_{p \in \mathcal{P}} \left[\max_i \mathrm{Inf}^{<d}_{i,\mu}(p) \right] \leqslant \tau \quad \forall i \in [k]$$

By analogy to Λ-SAT, it is natural to conjecture that approximate polymorphisms characterize the approximability of every MAX-Λ. Indeed, all known tight approximability results for different Max-CSPs like 3-SAT correspond exactly to the existence of approximate polymorphisms. Moreover, we show that the unique games conjecture [Kho02] is equivalent to asserting that existence of quasirandom approximate polymorphisms characterizes the complexity of approximating every CSP. To state our results formally, let us define some notation. For a CSP Λ define the approximation constant α_Λ as,

$$\alpha_\Lambda \stackrel{\text{def}}{=} \sup \{\alpha \in \mathbb{R} \mid \forall \tau, d > 0 \;\; \exists (\tau, d)\text{-quasirandom, } \alpha\text{-approximate polymorphism for } \Lambda\} \,.$$

Similarly, for each $c > 0$ define the constant $s_\Lambda(c)$ as,

$$s_\Lambda(c) \stackrel{\text{def}}{=} \sup \{s \in \mathbb{R} \mid \forall \tau, d > 0 \;\; \exists (\tau, d)\text{-quasirandom, } (c, s)\text{-approximate polymorphism for } \Lambda\} \,.$$

The connection between unique games conjecture (stated in Section 7) and approximate polymorphisms is summarized below.

Theorem 1.11 (Connection to UGC)**.** *Assuming the unique games conjecture, for every Λ, it is* NP*-hard to approximate* MAX-Λ *better than α_Λ. Moreover, the unique games conjecture is equivalent to the following statement: For every Λ and c, on instances of* MAX-Λ *with value c, it is* NP*-hard to find an assignment with value larger than $s_\Lambda(c)$.*

All unique games based hardness results for MAX-Λ are shown via constructing appropriate dictatorship testing gadgets [KKMO07, Rag08]. The above theorem is a consequence of the connection between approximate polymorphisms and dictatorship testing (see Section 7.4 for details). The connection between approximate polymorphisms and dictatorship testing was observed in the work of Kun and Szegedy [KS09].

Surprisingly, the other direction of the correspondence between approixmate polymorphisms and complexity of CSPs also holds. Formally, we show the following result about a semidefinite programming relaxation for CSPs referred to as the BASIC SDP relaxation (see Section 7.3 for definition).

Theorem 1.12 (Algorithm via Polymorphisms)**.** *For every CSP Λ, the integrality gap of the* BASIC-SDP *relaxation for* MAX-Λ *is at most α_Λ. More precisely, for every instance $\Im$ of* MAX-Λ*, for every c, if the optimum value of the* BASIC SDP *relaxation is at least c, then optimum value for $\Im$ is at least* $\lim_{\varepsilon \to 0} s_\Lambda(c - \varepsilon)$.

In particular, the above theorem implies that the BASIC-SDP relaxation yields an α_Λ (or $s_\Lambda(c)$) approximation to MAX-Λ. This theorem is a consequence of restating the soundness analysis of [Rag08] in the language of approximate polymorphisms. Specifically, one uses the α-approximate polymorphism to construct an α-factor rounding algorithm to the semidefinite program for MAX-Λ (See Section 7.3 for details). In the case of the algebraic dichotomy conjecture, the NP-hardness result is known, while an efficient algorithm for all CSPs via polymorphisms is not known. In contrast, the BASIC SDP relaxation yields an efficient algorithm for all MAX-Λ, but the NP-hardness for max-CSPs is open and equivalent to the unique games conjecture.

The case of MAX-Λ that are solvable exactly (approximation ratio $= 1$) has also received considerable attention in the literature. Generalising the algebraic approach to CSPs, algebraic properties called multimorphisms [CCJK06], fractional polymorphisms [DMP06] and weighted polymorphisms [CCC+13] have been proposed to study the complexity of classes of valued CSPs. The notion of α-approximate polymorphism for $\alpha = 1$ is closely related to these notions (see Section 7.1).

In a recent work, Thapper and Zivny [TZ13] obtained a characterization of MAX-Λ that can be solved exactly. Specifically, they showed that a valued CSP (a generalization of maxCSP) is tractable if and only if it admits a symmetric 1-approximate polymorphism on two inputs. This is further evidence supporting the claim that approximate polymorphisms characterize approximability of max-CSPs.

Beyond CSPs. It is natural to wonder if a similar theory of tractability could be developed for classes of problems beyond constraint satisfaction problems. For instance, it would be interesting to understand the tractability of MINIMUM SPANNING TREE, MINIMUM MATCHING, the intractability of TSP STEINER TREE and the approximability of STEINER TREE, METRIC TSP and NETWORK DESIGN problems.

It appears that tractable problems such as MINIMUM SPANNING TREE and MAXIMUM MATCHING have certain "operations" on the solution space. For instance, the union of two perfect matchings contains alternating cycles that could be used to create new matchings. Similarly, spanning trees form a matroid and therefore given two spanning trees, one could create a sequence of trees by exchanging edges amongst them. Furthermore, some algorithms for these problems crucially exploit these operations. More indirectly, MINIMUM SPANNING TREE and MAXIMUM MATCHING are connected to submodularity via polymatroids. As we saw earlier, submodularity is an example of a 1-approximate polymorphism.

Moreover, here is a heuristic justification for the connection between tractability and existence of operations to combine solutions. Typically, a combinatorial optimization problem like MAXIMUM MATCHING is tractable because of the existence of a linear programming relaxation P and an associated rounding scheme $\mathcal{R}$ (possibly randomized). Given a set of k integral solutions $X^{(1)}, X^{(2)}, \ldots, X^{(k)}$, consider any point in their convex hull, say $y = \frac{1}{k}\sum_{i\in[k]} X^{(i)}$. By convexity, the point y is also a feasible solution to the linear programming relaxation P . Therefore, we could execute the rounding scheme $\mathcal{R}$ on the LP solution y to obtain a distribution over integral solutions. Intuitively, y has less information than $X^{(1)}, \ldots, X^{(k)}$ and therefore the solutions output by the rounding scheme $\mathcal{R}$ should be different from $X^{(1)}, \ldots, X^{(k)}$. This suggests that the rounding scheme $\mathcal{R}$ yields an operation to combine solutions to create new solutions. Recall that in case of CSPs, indeed polymorphisms are used to obtain rounding schemes for the semidefinite programs (see Theorem 1.12). More recently, Barak, Steurer and Kelner [BKS14] show how certain algorithms to *combine* solutions can be used towards rounding sum of squares relaxations.

1.1 Our Results

Algorithmically, one of the fundamental results in combinatorial optimization is polynomial-time minimization of arbitrary submodular functions. Specifically, there exists an efficient algorithm to minimize a submodular function f given access to a value oracle for f [Cun81, Cun83, Cun84, Sch00]. Since submodularity is an example of a fractional polymorphism, it is natural to conjecture that such an algorithm exists whenever f admits a certain fractional polymorphism. Our main result is an efficient algorithm to minimize a function f given access to its value oracle, provided f admits an appropriate fractional polymorphism. Towards stating our result formally, let us define some notation.

Definition 1.13. *A fractional polymorphism is said to be measure-preserving if for each $i \in [q]$ and for every choice of inputs, the fraction of inputs equal to i is always equal to the fraction of outputs equal to i.*

Definition 1.14. *An operation $p : [q]^k \to [q]$ is said to have a transitive symmetry if for all $i, j \in [k]$ there exists a permutation $\sigma_{ij} \in S_k$ of inputs such that $\sigma_{ij}(i) = j$ and $p \circ \sigma = p$. A polymorphism $\mathcal{P}$ is said to be transitive symmetric if every operation $p \in \text{supp}(\mathcal{P})$ is transitive symmetric.*

Notice that cyclic symmetry of the operation is a special case of transitive symmetry. As per our definitions, submodularity is a *measure-preserving* and *transitive symmetric* polymorphism.

Theorem 1.15. *Let $f : [q]^n \to \mathbb{R}^+$ be a function that admits a fractional polymorphism $\mathcal{P}$. If $\mathcal{P}$ is measure preserving and transitive symmetric then there exists an efficient randomized algorithm $\mathcal{A}$ that for every $\varepsilon > 0$, makes $\text{poly}(1/\varepsilon, n)$ queries to the values of f and outputs an $x \in [q]^n$ such that,*

$$f(x) \leqslant \min_{y \in [q]^n} f(y) + \varepsilon \|f\|_\infty$$

Apart from submodularity, here is an example of a *measure preserving* and *transitive symmetric* polymorphism.

$$\mathcal{P} = \begin{cases} \text{Majority}(x_1, x_2, x_3) & \text{with probability } 2/3 \\ \text{Minority}(x_1, x_2, x_3) & \text{with probability } 1/3 \end{cases}$$

Here $\text{Minority}(0, 0, 0) = 0, \text{Minority}(1, 1, 1) = 1$ and $\text{Minority}(x_1, x_2, x_3) = 1 - \text{Majority}(x_1, x_2, x_3)$ on the rest of the inputs. On a larger domain $[q]$, a natural example of a fractional polymorphism would be

$$\mathcal{P} = \begin{cases} \max(x_1, x_2, x_3) & \text{with probability } 1/3 \\ \min(x_1, x_2, x_3) & \text{with probability } 1/3 \\ \text{median}(x_1, x_2, x_3) & \text{with probability } 1/3 \end{cases}$$

More generally, it is very easy to construct examples of fractional polymorphisms that satisfy the hypothesis of Theorem 1.15.

1.2 Related Work

Operations that give rise to tractability in the value oracle model have received considerable attention in the literature. A particularly successful line of work studies generalizations of submodularity over various lattices. In fact, submodular functions given by an oracle can be minimised on distributive lattices [IFF01, Sch00], diamonds [Kui11], and the pentagon [KL08] but the case of general non-distributive lattices remains open.

An alternate generalization of submodularity known as *bisubmodular functions* introduced by Chandrasekaran and Kabadi [CK88] arises naturally both in theoretical and practical contexts (see [FI05, SGB12]). Bisubmodular functions can be minimized in polynomial time given a value oracle over domains of size 3 [FI05, Qi88] but the complexity is open on domains of larger size [HK12].

Fujishige and Murota [FM00] introduced the notion of $L^\sharp$-convex functions – a class of functions that can also be minimized in the oracle model [Mur04]. In recent work, Kolmogorov [Kol10] exhibited efficient algorithms to minimize *strongly tree-submodular* functions on binary trees, which is a common generalization of $L^\sharp$-convex functions and bisubmodular functions.

1.3 Technical Overview

The technical heart of this paper is devoted to understanding the evolution of probability distributions over $[q]^n$ under iterated applications of operations. Fix a probability distribution μ over $[q]^n$. 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. Fix a sequence $\{p_i\}_{i=1}^{\infty}$ of operations with transitive symmetries. Define the dynamical system,

$$\mu_t = p_t(\mu_{t-1})\,,$$

with $\mu_0 = \mu$. We study the properties of the distribution μ_t as $t \to \infty$. Roughly speaking, the key technical insight of this work is that the correlations among the coordinates decay as $t \to \infty$.

For example, let us suppose μ_0 is such that for each $i \in [n]$, the i^{th} coordinate of μ_0 is not perfectly correlated with the first $(i-1)$-coordinates. In this case, we will show that μ_t converges in statistical distance to a product distribution as $t \to \infty$ (see Theorem 3.3). From an algorithmic standpoint, this is very valuable because even if μ_0 has no succinct representation, the limiting distribution $\lim_{t\to\infty} \mu_t$ has a very succinct description. Moreover, since the operations p_i are applied to each coordinate separately, the marginals of the limiting distribution $\lim_{t\to\infty} \mu_t$ are determined entirely by the marginals of the initial distribution μ_0. Thus, since the limiting distribution is a product distribution, it is completely determined by the marginals of the initial distribution μ_0!

Consider an arbitrary probability distribution μ over $[q]^n$. Let $T_{1-\gamma} \circ \mu$ denote the probability distribution over $[q]^n$ obtained by sampling from μ and perturbing each coordinate with a tiny probability γ. For small γ, the statistical distance between μ and $T_{1-\gamma} \circ \mu$ is small, i.e., $\|T_{1-\gamma} \circ \mu - \mu\|_1 \leqslant \gamma n$. However, if we initialize the dynamical system with $\mu_0 = T_{1-\gamma}\mu$ then irrespective of the starting distribution μ, the limiting distribution is always a product distribution (see Corollary 3.4). A brief overview of the correlation decay argument is presented in Section 3.2. 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 6.3). A key bottleneck in the analysis is that the individual marginals change with each iteration thereby changing the fourier spectrum of the operations involved.

Recall that the algebraic dichotomy conjecture for exact CSPs asserts that a CSP Λ admits a cyclic polymorphism if and only if the CSP Λ is in P. It is interesting to note that the correlation decay phenomena applies to cyclic terms. Roughly speaking, the algebraic dichotomy conjecture could be restated in terms of correlation decay in the above described dynamical system. This characterization closely resembles the absorbing subalgebras characterization by Barto and Kozik [BK12] derived using entirely algebraic techniques.

Our approach to prove Theorem 1.15 is as follows. For any function $f : [q]^n \to \mathbb{R}$, one can define a convex extension $\hat{f}$. Let $\blacktriangle_q$ denote the q-dimensional simplex.

Definition 1.16. *(Convex Extension) Given a function $f : [q]^n \to \mathbb{R}$, define its convex extension on $\hat{f} : \blacktriangle_q^n \to \mathbb{R}$ to be*

$$\hat{f}(z) \stackrel{\text{def}}{=} \min_{\substack{p.d.f.\ \mu\ over\ [q]^n \\ \mu_i = z_i}} \mathop{\mathbb{E}}_{x\sim\mu} [f(x)]$$

where the minimization is over all probability distributions μ over $[q]^n$ whose marginals μ_i coincide with z_i.

The convex extension $\hat{f}(z)$ is the minimum expected value of f under all probability distributions over $[q]^n$ whose marginals are equal to z. As the name suggests, $\hat{f}$ is a convex function minimizing which is equivalent to minimizing f. Since $\hat{f}$ is convex, one could appeal to techniques from convex optimization such as the ellipsoid algorithm to minimize $\hat{f}$. However it is in general intractable to even evaluate the value of $\hat{f}$ at a given point in $\blacktriangle_q^n$. In the case of a submodular function f, its convex extension $\hat{f}$ can be evaluated at a given point via a greedy algorithm, and $\hat{f}$ coincides with the function known as the Lovasz-extension of f.

We exhibit a randomized algorithm to evaluate the value of the convex extension $\hat{f}$ when f admits a fractional polymorphism. Given a point $z \in \blacktriangle_q^n$, the randomized algorithm computes the minimum expected value of f over all probability distributions μ whose marginals are equal to z. Let μ be the optimal probability distribution that achieves the minimum. Here μ is an unknown probability distribution which might not even have a compact representation. Consider the probability distribution $T_{1-\gamma} \circ \mu$ obtained by resampling each coordinate independently with probability γ. The probability distribution $T_{1-\gamma} \circ \mu$ is statistically close to μ and therefore has approximately the same expected value of f.

Let μ' be the limiting distribution obtained by iteratively applying the fractional polymorphism $\mathcal{P}$ to the perturbed optimal distribution $T_{1-\gamma} \circ \mu$. Since $\mathcal{P}$ is a fractional polymorphism, the expected value of f does not increase on applying $\mathcal{P}$. Therefore, the limiting probability distribution μ' has an expected value not much larger than the optimal distribution μ. Moreover, since $\mathcal{P}$ is measure preserving, the limiting probability distribution has the same marginals as μ_0! In other words, the limiting distribution μ' has marginals equal to z and achieves almost the same expected value as the unknown optimal distribution μ. By virtue of correlation decay (Corollary 3.4), the limiting distribution μ' admits an efficient sampling algorithm that we use to approximately estimate the value of $\hat{f}(z)$.

2 Background

We first introduce some basic notation. Let $[q]$ denote the alphabet $[q] = \{1, \ldots, q\}$. Furthermore, let $\blacktriangle_q$ denote the standard simplex in $\mathbb{R}^q$, i.e.,

$$\blacktriangle_q = \{x \in \mathbb{R}^q | x_i \geqslant 0 \quad \forall i, \ \sum_i x_i = 1\} .$$

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

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

More generally, an operation can be thought of as a map $p : [q]^k \to \blacktriangle_q$. In particular, we can think of p being given by maps $(p_1 \ldots p_q)$ where $p_i : [q]^k \to \mathbb{R}$ is the indicator

$$p_i(x) = \begin{cases} 1 & : p(x) = i \\ 0 & : p(x) \neq i \end{cases}$$

We next define a method for composing two k-ary operations p_1 and p_2. The idea is to think of each of p_1 and p_2 as nodes with k incoming edges and one outgoing edge. Then we take k copies of p_2 and connect those k outputs to each of the k inputs to p_1. Formally, we define:

Definition 2.1. *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:*

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

Next we state the definition for polymorphisms of an arbitrary cost function $f : [q]^n \to \mathbb{R}$. Intuitively, a polymorphism for f is a probability distribution over operations that, on average, decrease the value of f.

Definition 2.2. *A 1-approximate polymorphism $\mathcal{P}$ for a function $f : [q]^n \to \mathbb{R}$, consists of a probability distribution $\mathcal{P}$ over maps $\mathcal{O} = \{p : [q]^k \to [q]\}$ such that for any set of k assignments $x^{(1)}, \dots, x^{(k)} \in [q]^n$,*

$$\frac{1}{k} \sum_i f(x^{(i)}) \geqslant \mathop{\mathbb{E}}_{p \sim \mathcal{P}} [f(p(x^{(1)}, \dots, x^{(k)}))]$$

Definition 2.3. *For a 1-approximate polymorphism $\mathcal{P}$, $\mathcal{P}^{\otimes r}$ denotes the 1-approximate polymorphism consisting of a distribution over operations defined as, $p_1 \otimes \dots \otimes p_r$ with $p_1, \dots, p_r$ drawn i.i.d from $\mathcal{P}$.*

3 Correlation Decay

In this section we state our main theorem regarding the decay of correlation between random variables under repeated applications of transitive symmetric operations. We begin by defining a quantitative measure of correlation and using it to bound the statistical distance to a product distribution.

3.1 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 choosing real-valued test functions $f, g : \{0,1\} \to \mathbb{R}$ and computing $\mathbb{E}[f(X)g(Y)]$. We would like to define the correlation as the supremum of $\mathbb{E}[f(X)g(Y)]$ over all pairs of appropriately normalized test functions. There are two components to the normalization. First, we require $\mathbb{E}[f(X)] = \mathbb{E}[g(Y)] = 0$ to rule out the possibility of adding a large constant to each test function to increase $\mathbb{E}[f(X)g(Y)]$. Second, we require $\mathbf{Var}[f(X)] = \mathbf{Var}[g(Y)] = 1$ to ensure (by Cauchy-Schwarz) that $\mathbb{E}[f(X)g(Y)] \leqslant 1$.

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.

Definition 3.1. *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*

$$\rho(X,Y) \stackrel{\text{def}}{=} \sup_{\substack{f:[q_1]\to\mathbb{R}, g:[q_2]\to\mathbb{R} \\ \mathbf{Var}[f(X)]=\mathbf{Var}[g(Y)]=1 \\ \mathbb{E}[f(X)]=\mathbb{E}[g(Y)]=0}} \mathbb{E}[f(X)g(Y)]$$

We will interchangeably use the notation $\rho(\mu)$ or $\rho(\Omega_1, \Omega_2)$ to denote the correlation.

Next we show that, as the correlation for a pair of random variables X and Y becomes small, the variables become nearly independent. In particular, we show that $\rho(X,Y)$ can be used to bound the statistical distance of (X,Y) from the product distribution where X and Y are sampled independently.

Lemma 3.2. *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*

$$\|\mu_{XY} - \mu_X \times \mu_Y\|_1 \leqslant \min(q_1, q_2)\rho(X,Y)$$

Proof. Let $\{X_a\}_{a\in[q_1]}$ be indicator variables for the events $X = a$ with $a \in [q_1]$. Similarly, define the indicator variables $\{Y_b\}_{b\in[q_2]}$.

The statistical distance between μ_{XY} and $\mu_X \times \mu_Y$ is given by

$$\begin{aligned}
\|\mu_{XY} - \mu_X \times \mu_Y\|_1 &= \sum_{a\in[q_1],b\in[q_2]} |\,\mathbb{P}[X = a, Y = b] - \mathbb{P}[X = a]\,\mathbb{P}[Y = b]| \\
&= \sum_{a\in[q_1],b\in[q_2]} |\,\mathbb{E}[X_a Y_b] - \mathbb{E}[X_a]\,\mathbb{E}[Y_b]|
\end{aligned}$$

Set $\sigma_{ab} = \operatorname{sign}(\mathbb{E}[X_a Y_b] - \mathbb{E}[X_a]\,\mathbb{E}[Y_b])$ and $Z_a = \sum_{b\in[q_2]} \sigma_{ab} Y_b$.

$$\begin{aligned}
\|\mu_{XY} - \mu_X \times \mu_Y\|_1 &= \sum_{a\in[q_1]} \mathbb{E}[X_a Z_a] - \mathbb{E}[X_a]\,\mathbb{E}[Z_a] \\
&= \sum_{a\in[q_1]} \mathbf{Cov}[X_a, Z_a] \\
&\leqslant \sum_{a\in[q_1]} \sqrt{\mathbf{Var}[X_a]}\sqrt{\mathbf{Var}[Z_a]}\rho(X,Y) \\
&\leqslant \sum_{a\in[q_1]} \sqrt{\mathbf{Var}[X_a]}\rho(X,Y) \qquad (\mathbf{Var}[Z_a] \leqslant 1) \\
&\leqslant q_1 \rho(X,Y)
\end{aligned}$$

The result follows by symmetry.

□

3.2 Correlation Decay

To begin we give an explanation of why one should expect correlation to decay under repeated applications of symmetric operations. Consider the simple example of two boolean random variables X and Y with joint distribution μ. Suppose $X = Y$ with probability $\frac{1}{2} + \gamma$ and that the marginal distributions of both X and Y are uniform on $\{0,1\}$. Let $p : \{0,1\}^k \to \{0,1\}$ be the majority operation on k bits. That is $p(x_1 \dots x_k) = 1$ if and only if the majority of $x_1 \dots x_k$ are one.

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 operations with transitive symmetry. The first is that 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 admits a transitive symmetry to get a bound on the total Fourier mass on degree-one terms. The second obstacle is that, 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 after every application of p.

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. 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. Of course, this particular observation applies only to the OR operation. However, 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 transitive symmetric operations. We now state our main theorem to this effect.

Theorem 3.3. *(Correlation Decay) Let μ be a distribution on $[q]^n$. Let $X_1, \dots, X_n$ be the jointly distributed $[q]$-valued random variables drawn from μ. Let $\rho = \max_i \rho(X_1, \dots, X_{i-1}, X_i) < 1$ and λ be the minimum probability of an atom in the marginal distributions $\{\mu_i\}_{i \in [n]}$. For any $\eta > 0$, the following holds for $r \geqslant \Omega_q\left(\frac{\log \lambda}{\log \rho} \log^2\left(\frac{qn}{\eta}\right)\right)$: If $p_1, \dots, p_r : [q]^k \to [q]$ is a sequence of operations each of which admit a transitive symmetry then,*

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

We defer the proof of the theorem to Section 6. Note that the theorem only applies when $\rho < 1$ i.e. when there are no perfect correlations between the X_i. To ensure that a distribution has no perfect correlations we can introduce a small amount of noise.

Noise. For a probability distribution μ on $[q]^n$ let $T_{1-\gamma} \circ \mu$ denote the probability distribution over $[q]^n$ defined as:

- Sample $X \in [q]^n$ from the distribution μ.

- For each $i \in [n]$, set

$$\tilde{X}_i = \begin{cases} X_i & \text{with probability} 1-\gamma \\ \text{sample from } \mu_i & \text{with probability } \gamma \end{cases}$$

We prove the following corollary that gives correlation decay for distributions with a small amount of added noise.

Corollary 3.4. *Let μ be any probability distribution over $[q]^n$ and let λ denote the minimum probability of an atom in the marginals $\{\mu_i\}_{i\in[n]}$. For any $\gamma, \eta > 0$, given a sequence $p_1, \ldots, p_r : [q]^k \to [q]$ of operations with transitive symmetry with $r > \Omega_q\left(\frac{\log \lambda}{\log(1-\gamma)} \log^2\left(\frac{qn}{\eta}\right)\right)$*

$$\|p_1 \otimes p_2 \otimes \ldots \otimes p_r(T_{1-\gamma} \circ \mu) - p_1 \otimes p_2 \otimes \ldots \otimes p_r(\mu^{\times})\| \leqslant \eta$$

Proof. Let $f : [q] \to \mathbb{R}$ and $g : [q]^{i-1} \to \mathbb{R}$ be functions with $\mathbb{E}[f(\tilde{X}_i)] = \mathbb{E}[g(\tilde{X}_1 \ldots \tilde{X}_{i-1})] = 0$ and $\mathbb{E}[f(\tilde{X}_i)^2] = \mathbb{E}[g(\tilde{X}_1 \ldots \tilde{X}_{i-1})^2] = 1$ such that

$$\mathbb{E}[f(\tilde{X}_i)g(\tilde{X}_1 \ldots \tilde{X}_{i-1})] = \rho(\tilde{X}_i, (\tilde{X}_1 \ldots \tilde{X}_{i-1}))$$

That is, f and g achieve the maximum possible correlation between $\tilde{X}_i$ and $(\tilde{X}_1 \ldots \tilde{X}_{i-1})$. Let Y_i be an independent random sample from the marginal μ_i. Now we expand the above expectation by conditioning on whether or not $\tilde{X}_i$ was obtained by re-sampling from the marginal μ_i.

$$\mathbb{E}[f(\tilde{X}_i)g(\tilde{X}_1 \ldots \tilde{X}_{i-1})] = \mathbb{E}[f(X_i)g(\tilde{X}_1 \ldots \tilde{X}_{i-1})](1-\gamma) + \mathbb{E}[f(Y_i)g(\tilde{X}_1 \ldots \tilde{X}_{i-1})]\gamma$$

By Cauchy-Schwarz inequality, the first term above is bounded by

$$\mathbb{E}[f(X_i)g(\tilde{X}_1 \ldots \tilde{X}_{i-1})](1-\gamma) \leqslant \sqrt{\mathbb{E}[f(X_i)^2]\,\mathbb{E}[g(\tilde{X}_1 \ldots \tilde{X}_{i-1})^2]}(1-\gamma) = 1-\gamma$$

where we have used the fact that $\mathbb{E}[f(X_i)^2] = \mathbb{E}[f(\tilde{X}_i)^2] = 1$. Since Y_i is independent of $\tilde{X}_1 \ldots \tilde{X}_{i-1}$, the second term is

$$\mathbb{E}[f(Y_i)g(\tilde{X}_1 \ldots \tilde{X}_{i-1})]\gamma = \mathbb{E}[f(Y_i)]\,\mathbb{E}[g(\tilde{X}_1 \ldots \tilde{X}_{i-1})]\gamma = 0$$

Thus, we get $\rho(\tilde{X}_i, (\tilde{X}_1 \ldots \tilde{X}_{i-1})) \leqslant 1-\gamma$ for all $i \in [n]$. The result then follows by applying Theorem 3.3 to $T_{1-\gamma} \circ \mu$. $\square$

4 Optimization in the Value Oracle Model

In this section, we will describe an algorithm to minimize a function $f : [q]^n \to \mathbb{R}$ that admits a 1-approximate polymorphism given access to a value oracle for f. We begin by setting up some notation.

Recall that for a finite set A, $\blacktriangle_A$ denotes the set of all probability distributions over A. For notational convenience, we will use the following shorthand for the expectation of the function f over a distribution μ.

Definition 4.1. *(Distributional Extension) Given a function $f : [q]^n \to \mathbb{R}$, define its distributional extension $F : \blacktriangle_{[q]^n} \to \mathbb{R}$ to be*

$$F(\mu) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{x\sim\mu}[f(x)]$$

Given an operation $p : [q]^k \to [q]$ and a probability distribution $\mu \in \blacktriangle_{[q]^n}$, define $p(\mu)$ to be the distribution over $[q]^n$ that is sampled as follows:

- Sample $x_1, x_2, \ldots, x_k \in [q]^n$ independently from the distribution μ.
- Output $p(x_1, x_2, \ldots, x_k)$.

Notice that for each coordinate $i \in [n]$, the i^{th} marginal distribution $(p(\mu))_i$ is given by $p(\mu_i)$. More generally, if a $\mathcal{P}$ denotes a probability distribution over operations then $\mathcal{P}^{\otimes r}(\mu)$ is a distribution over $[q]^n$ that is sampled as follows:

- Sample operations $p_1, \ldots, p_r : [q]^k \to [q]$ independently from the distribution $\mathcal{P}$
- Output a sample from $p_1 \otimes p_2 \otimes \ldots p_r(\mu)$.

Suppose $\mathcal{P}$ is a fractional polymorphism for the function f. By definition of fractional polymorphisms, the average value of f does not increase on applying the operations sampled from $\mathcal{P}$. More precisely, we can prove the following.

Lemma 4.2. *For every distribution μ on $[q]^n$, and a fractional polymorphism $\mathcal{P}$ of a function $f : [q]^n \to \mathbb{R}$ and $r \in \mathbb{N}$, $F(\mathcal{P}^{\otimes r}(\mu)) \leqslant F(\mu)$ where F is the distributional extension of f*

Proof. We will prove this result by induction on r. First, we prove the result for $r = 1$.

$$\begin{aligned} F(\mathcal{P}(\mu)) &= \mathop{\mathbb{E}}_{x \sim \mathcal{P}(\mu)} [f(x)] \\ &= \mathop{\mathbb{E}}_{p \sim \mathcal{P}} \mathop{\mathbb{E}}_{x_1, \ldots, x_k \sim \mu} [f(p(x_1, \ldots, x_k))] \\ &= \mathop{\mathbb{E}}_{x_1, \ldots, x_k \sim \mu} \left[\mathop{\mathbb{E}}_{p \sim \mathcal{P}} [f(p(x_1, \ldots, x_k))] \right] \\ &\leqslant \mathop{\mathbb{E}}_{x_1, \ldots, x_k \sim \mu} \left[\frac{1}{k} \sum_i f(x_i) \right] = F(\mu) \end{aligned}$$

The last inequality in the above calculation uses the fact that $\mathcal{P}$ is a fractional polymorphism for f. Suppose the assertion is true for r, now we will prove the result for $r + 1$. Observe that the distribution $\mathcal{P}^{\otimes r+1}(\mu)$ can be written as,

$$\mathcal{P}^{\otimes r+1}(\mu) = \mathop{\mathbb{E}}_{p_2, \ldots, p_{r+1} \sim \mathcal{P}} [\mathcal{P}(p_2 \otimes \ldots \otimes p_r(\mu))]$$

where $p_2, \ldots, p_r$ are drawn independently from the distribution $\mathcal{P}$. Hence we can write,

$$\begin{aligned} F(\mathcal{P}^{\otimes r+1}(\mu)) &= \mathop{\mathbb{E}}_{p_2, \ldots, p_{r+1} \sim \mathcal{P}} [F(\mathcal{P}(p_2 \otimes \ldots \otimes p_{r+1}(\mu)))] \\ &\leqslant \mathop{\mathbb{E}}_{p_2, \ldots, p_{r+1} \sim \mathcal{P}} [F(p_2 \otimes \ldots \otimes p_{r+1}(\mu))] \qquad \text{using base case} \\ &= F(\mathcal{P}^{\otimes r}(\mu)) \leqslant F(\mu)\,, \end{aligned}$$

where the last inequality used the induction hypothesis for r. □

Recall that $\mu^\times$ is the product distribution with the same marginals as μ. We will show the following using correlation decay.

Lemma 4.3. *Let $\mathcal{P}$ be a fractional polymorphism with a transitive symmetry for a function $f : [q]^n \to [q]$. Let μ be a probability distribution over $[q]^n$ and let λ denote the minimum probability of an atom in the marginals $\{\mu_i\}_{i \in [n]}$. For $\gamma = \frac{\delta}{2n}$ and $r = \Omega_q\left(\frac{\log \lambda}{\log(1-\gamma)} \log^2\left(\frac{2qn}{\delta}\right)\right)$*

$$F(\mathcal{P}^{\otimes r}(\mu^\times)) \leqslant F(\mu) + \delta \|f\|_\infty$$

Proof. Consider the distribution $T_{1-\gamma} \circ \mu$. By definition of $T_{1-\gamma} \circ \mu$, all the correlations within $T_{1-\gamma} \circ \mu$ are at most $1-\gamma$. Roughly speaking, with repeated applications of the operations from $\mathcal{P}$ all the correlations will vanish. More precisely, by Corollary 3.4, for any sequence of operations $p_1, \ldots, p_r$ with transitive symmetry we have

$$\|p_1 \otimes p_2 \otimes \ldots \otimes p_r(T_{1-\gamma} \circ \mu) - p_1 \otimes p_2 \otimes \ldots \otimes p_r(\mu^{\times})\| \leqslant \frac{\delta}{2} \tag{4.1}$$

Recall that $\mathcal{P}^{\otimes r}(\mu')$ for a distribution μ' is given by, $\mathcal{P}^{\otimes r}(\mu') = \mathbb{E}_{p_1,\ldots,p_r}[p_1 \otimes p_2 \ldots \otimes p_r(\mu')]$. Averaging (4.1) over all choices of $p_1, \ldots p_r$ from $\mathcal{P}$ we get

$$\|\mathcal{P}^{\otimes r}(T_{1-\gamma} \circ \mu) - \mathcal{P}^{\otimes r}(\mu^{\times})\| \leqslant \frac{\delta}{2} \tag{4.2}$$

Now we are ready to finish the proof of the lemma.

$$\begin{aligned} F(\mathcal{P}^{\otimes r}(\mu^{\times})) &\leqslant F(\mathcal{P}^{\otimes r}(T_{1-\gamma} \circ \mu)) + \frac{\delta}{2}\|f\|_\infty \qquad \text{using } ((4.2)) \\ &\leqslant F(T_{1-\gamma} \circ \mu) + \frac{\delta}{2}\|f\|_\infty \qquad \text{using Lemma 4.2} \\ &\leqslant F(\mu) + \delta\|f\|_\infty \qquad \text{using } \|\mu - T_{1-\gamma} \circ \mu\|_1 \leqslant \gamma n \end{aligned}$$

□

Suppose we are looking to minimize the function f or equivalently the distributional extension F. In general, the minima for F could be an arbitrary distribution with no succinct (polynomial-sized) representation. The preceding lemma shows that $\mathcal{P}^{\otimes r}(\mu^{\times})$ has roughly the same value of F. But $\mathcal{P}^{\otimes r}(\mu^{\times})$ not only has a succinct representation, but is efficiently sampleable given the marginals of μ! In the rest of this section, we will use this insight towards minimizing f given a value oracle.

4.1 Convex Extension

Definition 4.4. *(Convex extension) For a function $f : [q]^n \to \mathbb{R}$, the convex extension $\hat{f} : (\blacktriangle_q)^n \to \mathbb{R}$ is defined as,*

$$\hat{f}(w) \stackrel{\text{def}}{=} \min_{\substack{\mu \in \blacktriangle_{[q]^n} \\ \mu_i = w_i \forall i \in [n]}} E_{x \in \mu}[f(x)]\,,$$

where the minimization is over all probability distributions μ over $[q]^n$ whose marginals are given by $w \in (\blacktriangle_q)^n$.

As the name suggests, $\hat{f}$ is a convex function whose minimum is equal to the minimum of f. In general, the convex extension of a function f cannot be computed efficiently since the optimal distribution μ might not even have a succinct representation. In our case, however, we can prove:

Theorem 4.5. *Suppose a function $f : [q]^n \to [q]$ admits a fractional polymorphism $\mathcal{P}$ such that: (1) Each operation $p : [q]^k \to [q]$ in the support of $\mathcal{P}$ has a transitive symmetry, (2) the polymorphism $\mathcal{P}$ is measure preserving. Then there is an algorithm that given $\varepsilon > 0$ and $w \in \blacktriangle_q^n$, runs in time* $\mathrm{poly}(n, \frac{1}{\varepsilon})$ *and computes $\hat{f}(w) \pm \varepsilon\|f\|_\infty$.*

Proof. Given $w \in \blacktriangle_q^n$, we first perturb every coordinate slightly to ensure that the minimum probability in each marginal w_i is bounded away from 0. In particular, we define w' by setting $b = \frac{\varepsilon}{10qn}$ and

$$w'_i(a) = \frac{w_i(a) + b}{1 + qb}$$

for all $a \in [q]$ and $i \in [n]$. Clearly $w' \in \blacktriangle_q^n$ and $w'_i \geqslant \frac{b}{1+qb}$. Now let

$$\mu = \operatorname*{argmin}_{\substack{\mu \in \blacktriangle_{[q]^n} \\ \mu_i = w'_i \forall i \in [n]}} E_{x \in \mu}[f(x)]\,,$$

denote the optimal distribution over $[q]^n$ that achieves the minimum in $\hat{f}(w')$. Fix $\delta = \frac{\varepsilon}{10}$ and note that

$$\frac{\log b}{\log(1-\frac{\delta}{2n})} = O_q\left(n\varepsilon^{-1}\log(n\varepsilon^{-1})\right)$$

Now we claim that for $r = \Omega\left(n\varepsilon^{-1}\log^3(n\varepsilon^{-1})\right)$ we have

$$F(\mathcal{P}^{\otimes r}(\mu^{\times})) - \delta\|f\|_\infty \leqslant \hat{f}(w') = F(\mu) \leqslant F(\mathcal{P}^{\otimes r}(\mu^{\times})\,.$$

The left-hand inequality follows from Lemma 4.3. The right hand side inequality follows because $\mathcal{P}^{\otimes r}(\mu^{\times})$ has its marginals equal to w' since $\mathcal{P}$ is *measure preserving* and μ is the optimal distribution with marginals w'. Moreover, observe that the distribution $\mathcal{P}^{\otimes r}(w')$ can be sampled efficiently as described below.

> Input: $w' \in \blacktriangle_q^n$
> Output: Sample from $\mathcal{P}^{\otimes r}(w')$
>
> - Fix $z_0 = w'$
> - For $i = 1$ to r do,
>   - Sample $p \in \mathcal{P}$ and set $z_i = p(z_{i-1})$.
> - Sample $x \in [q]^n$ by sampling each coordinate independently from z_r.

In order to estimate $\hat{f}(w')$, it is sufficient to sample $\mathcal{P}^{\otimes r}(w)$ independently $O\left(\frac{n}{\varepsilon^2}\right)$ times to compute $F(\mathcal{P}^{\otimes r}(\mu^{\times}))$ to accuracy within $\frac{\varepsilon}{10}\|f\|_\infty$ with high probability. Thus, we can estimate $\hat{f}(w')$ to accuracy $(\delta + \frac{\varepsilon}{10})\,\|f\|_\infty = \frac{\varepsilon}{5}\|f\|_\infty$.

Next let π be the distribution that achieves the optimum for marginals w i.e. $\hat{f}(w) = F(\pi)$. By changing each marginal of π by at most b we obtain a distribution π' with marginals given by w'. By optimality of μ we know

$$\hat{f}(w') = F(\mu) \leqslant F(\pi') = F(\pi) \pm bqn\|f\|_\infty = \hat{f}(w) \pm \frac{\varepsilon}{10}\|f\|_\infty$$

By a symmetric argument we get that $\hat{f}(w) \leqslant \hat{f}(w') \pm \frac{\varepsilon}{10}\|f\|_\infty$. Thus, we conclude that $|\hat{f}(w) - \hat{f}(w')| \leqslant \frac{\varepsilon}{10}\|f\|_\infty$. Therefore we can estimate $\hat{f}(w)$ to accuracy $(\frac{\varepsilon}{10}+\frac{\varepsilon}{5})\|f\|_\infty \leqslant \varepsilon\|f\|_\infty$. In summary, this yields an algorithm running in time $O_q\left(n^2\varepsilon^{-3}\log^3(n\varepsilon^{-1})\right)$ to estimate $\hat{f}(w)$ within an error of $\varepsilon\|f\|_\infty$ with high probability. $\square$

Gradient-Free Minimization. In the previous section, we have demonstrated that the convex extension $\hat{f} : \blacktriangle_q^n \to \mathbb{R}$ can be computed efficiently. In order to complete the proof of our main theorem (Theorem 1.15), we will exhibit an algorithm to minimize the convex function $\hat{f}$. The domain of the convex function $\hat{f}$, namely $\blacktriangle_q^n$ is particularly simple. However, the convex function $\hat{f}$ is not necessarily smooth (differentiable). More importantly, we only have an evaluation oracle for the function $\hat{f}$ with no access to its gradients or subgradients. Gradient-free optimization has been extensively studied (see [NES11, Spa97] and references therein) and there are numerous approaches that have been proposed in literature. At the outset, the idea is to estimate the gradient via one or more function evaluations in the neighborhood. In this work, we will appeal to the randomized algorithms by gaussian perturbations proposed by Nesterov [NES11]. The following result is a restatement of Theorem 5 in the work of Nesterov [NES11].

Theorem 4.6. *Let $\hat{f} : \mathbb{R}^N \to \mathbb{R}$ be a non-smooth convex function given by an evaluation oracle $\mathcal{O}$. Let us suppose f is L-Lipshitz,i.e.,*

$$|\hat{f}(x) - \hat{f}(y)| \leqslant L\|x - y\| \qquad \forall x, y \in \mathbb{R}^N$$

Let $Q \subseteq \mathbb{R}^N$ be a closed convex set given by a projection oracle $\pi_Q : \mathbb{R}^N \to Q$. Let $\Delta(Q) \stackrel{\text{def}}{=} \sup_{x,y \in Q}\|x - y\|$.

There is a randomized algorithm that given $\varepsilon > 0$, with constant probability finds $y \in Q$ such that,

$$|\hat{f}(y) - \min_{x \in Q} \hat{f}(x)| \leqslant \varepsilon \,,$$

using

$$\frac{(N + 4)L^2 \Delta(Q)^2}{\varepsilon^2}$$

calls to the oracle $\mathcal{O}$ and π_Q.

In our setup, the domain $Q = \blacktriangle_q^n \in \mathbb{R}^{nq}$ is a very simple closed convex set with diameter $\Delta(Q) \leqslant \sqrt{2n}$. The convex extension $\hat{f}$ is L-Lipschitz for $L = \|f\|_\infty$. Finally, we turn to the proof of our main theorem, which follows more or less immediately from the results in this section.

Proof of Theorem 1.15. By Theorem 4.5 we can estimate the convex function $\hat{f}(w)$ with sufficient accuracy to apply Theorem 4.6 to approximately minimize $\hat{f}(w)$ in time $\mathrm{poly}(1/\varepsilon, n, \|f\|_\infty)$. Since $\hat{f}(w)$ is the convex extension of f, the output of the algorithm approximately minimizing $\hat{f}(w)$ will also approximately minimize f. □

5 Analytic Background

In this section we introduce the analytic background necessary to prove Theorem 3.3.

5.1 Analysis on Finite Measure Spaces

We begin by introducing the space of functions that will be fundamental to our analysis.

Definition 5.1. *Let μ be a probability distribution on $[q]$. Define $L^2([q], \mu)$ to be the inner product space of all functions $f : [q] \to \mathbb{R}$ with inner product given by*

$$\langle f, g \rangle = \mathop{\mathbb{E}}_{x \sim \mu}[f(x)g(x)] \,.$$

For a function $f \in L^2([q], \mu)$ the L^p-norm of f is defined as

$$\|f\|_p = \mathop{\mathbb{E}}_{x \sim \mu}[f(x)^p]^{\frac{1}{p}} \,.$$

We will also often use the equivalent notation $L^2(\Omega)$ where $\Omega = ([q], \mu)$ is the corresponding probability space. We next introduce a useful basis for representing functions $f : [q]^k \to \mathbb{R}$.

Multilinear Representation. Let μ be a distribution on $[q]$. Pick an orthonormal basis $\phi = \{\phi_0 \equiv 1, \ldots, \phi_{q-1}\}$ for the set of functions from $[q]$ to $\mathbb{R}$ denoted by $L^2([q], \mu)$.

The tensor product of the orthonormal basis ϕ gives an orthonormal basis for functions in $L^2([q]^k, \mu^k)$. Specifically, for a multi-index $\alpha \in \{0, 1, \ldots, q-1\}^k$, let $\phi_\alpha : [q]^k \to \mathbb{R}$ denote the function, $\phi_\alpha(x) = \prod_{i=1}^k \phi_{\alpha_i}(x_i)$. It is easy to see that ϕ_α forms an orthonormal basis for space of functions $L^2([q]^k, \mu^k)$.

We can write any function $p : [q]^k \to \mathbb{R}$ in the ϕ_α basis as

$$p(x) = \sum_{\alpha \in \{0,1,\ldots,q-1\}^k} \hat{p}_\alpha \phi_\alpha(x)$$

where we often refer to the $\hat{p}_\alpha$ as the Fourier coefficients of p. For a multi-index $\alpha \in \{0, 1, \ldots, q-1\}^k$, let $|\alpha| = |\{i|\alpha_i \neq 0\}|$. The degree of a term $\phi_\alpha(x)$ is given by $|\alpha|$.

Let $p : [q]^k \to \blacktriangle_q$ be a k-ary operation represented by q-real valued functions $p = (p_1, \ldots, p_q)$. Associated with every k-ary operation p is a subspace of $L^2(\mu^k)$ given by

$$\mathsf{Span}(p) = \{ \text{ span of } p_1, \ldots, p_q\}$$

Noise Operator. For a function $p : [q]^k \to \mathbb{R}$ and $\rho \in [0, 1]$, define $T_\rho p$ as

$$T_\rho p(x) = \mathop{\mathbb{E}}_{y \sim_\rho x}[p(y)]$$

where $y \sim_\rho x$ denotes the following distribution,

$$y_i = \begin{cases} x_i & \text{with probability } \rho \\ \text{indpendent sample from } \mu & \text{with probability } 1-\rho \end{cases}$$

The multilinear polynomial associated with $T_\rho p$ is given by

$$T_\rho p = \sum_\alpha \hat{p}_\alpha \rho^{|\alpha|} \phi_\alpha(x)$$

5.2 The Conditional Expectation Operator

In this section we introduce the conditional expectation operator associated with a joint probability distribution. The singular values of this operator encode information about correlation, and thus the singular vectors provide a useful basis in which to analyze correlation decay.

Let μ be a joint distribution on $[q_1] \times [q_2]$ with marginals μ_1 and μ_2. Further let $\Omega_1 = ([q_1], \mu_1)$ and $\Omega_2 = ([q_2], \mu_2)$ denote the probability spaces corresponding to μ_1 and μ_2.

Definition 5.2. *The conditional expectation operator $T_\mu : L^2(\Omega_2) \to L^2(\Omega_1)$ is given by*

$$(T_\mu f)(x) = \mathop{\mathbb{E}}_{(X,Y)\sim\mu}[f(Y)|X = x]$$

It follows from the definition that

$$\mathop{\mathbb{E}}_{(X,Y)\sim\mu}[f(X)g(Y)] = \mathop{\mathbb{E}}_{X\sim\mu_1}[f(X)(T_\mu g)(X)]$$

The adjoint operator $T_\mu^* : L^2(\Omega_1) \to L^2(\Omega_2)$ is given by

$$(T_\mu^* g)(y) = \mathop{\mathbb{E}}_{(X,Y)\sim\mu}[g(X)|Y = y]$$

It is possible to choose a singular value decomposition $\{\phi_i, \psi_j\}$ of the operator T_μ such that $\phi_0 = \psi_0 \equiv 1$. Let σ_i denote the singular value corresponding to the pair ϕ_i, ψ_i. That is $T_\mu \psi_i = \sigma_i \phi_i$. It is easy to check the following facts regarding this singular value decomposition.

Fact 5.3. *Let $\{\phi_i, \psi_j\}$ be the above singular value decomposition of T_μ.*

- *The functions ϕ_i and ψ_i form orthonormal bases for $L^2(\Omega_1)$ and $L^2(\Omega_2)$ respectively.*
- $\sigma_0 = 1$ *and* $\sigma_1 = \rho(\mu)$.
- *The expectation over μ of the product of any pair ϕ_i and ψ_i is given by*

$$\mathop{\mathbb{E}}_{(x,y)\sim\mu}[\phi_i(x)\psi_j(y)] = \begin{cases} \sigma_i & : i = j \\ 0 & : i \neq j \end{cases}$$

Thus, for functions $F \in L^2([q_1]^k, \mu_1^k)$ and $G \in L^2([q_2]^k, \mu_2^k)$ we can write their multilinear expansions with respect to tensor powers of the ϕ and ψ bases respectively. In particular, for a multi-index $\alpha \in \{0, \dots, q_1 - 1\}^k$ and $x \in [q_1]^k$ we have the basis functions

$$\phi_\alpha(x) = \prod_i \phi_{\alpha_i}(x_i)$$

with $\psi_\beta(y)$ defined similarly. We then have the following fact

Fact 5.4. *The vectors $\{\phi_\alpha\}_{\alpha\in[q_1]^k}$, $\{\psi_\beta\}_{\beta\in[q_2]^k}$ are a singular value decomposition of $T_\mu^{\otimes k}$ with corresponding singular values $\sigma_\alpha = \prod_i \sigma_{\alpha_i}$. Furthermore $T_{\mu^k} = T_\mu^{\otimes k}$.*

Using the above facts about the singular value basis it follows that

$$\mathop{\mathbb{E}}_{(x,y)\sim\mu^k}[\phi_\alpha(x)\psi_\beta(y)] = \begin{cases} \sigma_\alpha & : \alpha = \beta \\ 0 & : \alpha \neq \beta \end{cases}$$

We can then write the multi-linear expansion of F with respect to ϕ as

$$F(x) = \sum_\alpha \hat{F}_\alpha \phi_\alpha(x)$$

As an immediate consequence of the above discussion we have

Fact 5.5. *Let $F \in L^2([q_1]^k, \mu_1^k)$ and $G \in L^2([q_2]^k, \mu_2^k)$. Then*

$$\mathop{\mathbb{E}}_{(x,y)\sim\mu^k}[F(x)G(y)] = \sum_\alpha \hat{F}_\alpha \hat{G}_\alpha \sigma_\alpha$$

6 Correlation Decay for Symmetric Polymorphisms

In this section we prove Theorem 3.3. The proof has two major components. First, we show that for an operation where the Fourier weight on certain degree-one coefficients is bounded away from one, correlation decreases. Second, we show that for operations admitting a transitive symmetry, this Fourier weight does indeed stay bounded away from one after each application of an operation.

Throughout the section let μ be a joint distribution on $[q] \times [q]^m$ for some m. Let μ_1 and μ_2 be the respective marginal distributions of μ. Further, let $\Omega_1 = ([q], \mu_1)$ and $\Omega_2 = ([q]^m, \mu_2)$ denote the probability spaces corresponding to μ_1 and μ_2. It will be useful to think of q as being a constant and m as possibly being very large.

Further, for an operation $p : [q]^k \to [q]$ we will write $p(\mu)$ for the probability distribution obtained by sampling $(X, Y) \sim \mu^k$ and outputting the pair $(p(X_1, \dots, X_k), p(Y_1, \dots, Y_k)) \in [q] \times [q]^m$. Recall that we have defined $p(Y_1, \dots, Y_k)$ for $Y_i \in [q]^m$ to be the result of applying p on each coordinate of $Y_1, \dots, Y_k$ separately.

6.1 Linearity and Correlation Decay

Given a singular value basis $\{\phi_\alpha\}$ for the space $L^2([q]^k, \mu_1^k)$ we call the set of functions ϕ_α such that $|\alpha| = 1$ (i.e. exactly one coordinate of α is nonzero) the degree-one part of the basis. Note that there are qk such basis functions: q for each of the k coordinates. We will be interested in keeping track of those degree-one basis functions with singular values close to the correlation $\rho(\mu)$.

Definition 6.1. *Let $\{\phi_\alpha\}$ be a singular value basis as above. The linear multi-indices α are given by the set*

$$L_\mu = \{\alpha \mid |\alpha| = 1 \text{ and } \sigma_\alpha \geqslant \rho(\mu)^2\}$$

The linear part of a singular value basis is those ϕ_α with $\alpha \in L_\mu$.

Next we define a quantitative measure of the linearity of a function.

Definition 6.2. *Let $F \in L^2([q]^k, \mu_1^k)$. The linearity of F is defined as*

$$\mathrm{Lin}_\mu(F) = \sum_{\alpha \in L_\mu} \hat{F}_\alpha^2$$

Further, for an operation $p : [q]^k \to [q]$ the linearity of p is given by

$$\mathrm{Lin}_\mu(p) = \sup_{\substack{F \in \mathsf{Span}(p) \\ \|F\|_2 = 1}} \mathrm{Lin}_\mu(F)$$

With these definitions in hand we show that the correlation of μ decays under any operation p that has linearity bounded away from one.

Lemma 6.3. *(Correlation Decay) Let $p : [q]^k \to [q]$ be an operation. Then*

$$\rho(p(\mu)) \leqslant \rho(\mu)\left(1 - \frac{1}{2}(1 - \mathrm{Lin}_\mu(p))(1 - \rho(\mu)^2)\right)$$

Proof. Let $f : [q] \to \mathbb{R}$ and $g : [q]^m \to \mathbb{R}$ be such that $\mathbb{E}_{x \sim \mu_1^k}[f(p(x)] = 0$, $\mathbb{E}_{y \sim \mu_2^k}[g(p(y))] = 0$, $\|f(p)\|_2 = \|g(p)\|_2 = 1$ and

$$\rho(p(\mu)) = \mathop{\mathbb{E}}_{(x,y) \sim \mu^k}[f(p(x))g(p(y))]\,.$$

That is, f and g are the functions achieving the maximum correlation for $p(\mu)$. Now define $F : [q]^k \to \mathbb{R}$ as $F(x) = f(p(x))$. Similarly, define $G : [q]^{mk} \to \mathbb{R}$ as $G(y) = g(p(y))$.

Writing the multilinear expansion of F with respect to the basis ϕ, we have $F = \sum_\alpha \hat{F}_\alpha \phi_\alpha$ where

$$\hat{F}_0 = \mathbb{E}[F] = 0\,, \qquad \sum_\alpha \hat{F}_\alpha^2 = \|F\|_2 = 1\,.$$

Furthermore, since $F \in \mathsf{Span}(p)$ we have

$$\sum_{\alpha \in L_\mu} \hat{F}_\alpha^2 = \mathrm{Lin}_\mu(F) \leqslant \mathrm{Lin}_\mu(p)\,.$$

Similarly, one can write the multilinear expansion of G. Recall that,

$$\mathop{\mathbb{E}}_{(x,y) \sim \mu^k}[f(p(x))g(p(y))] = \mathop{\mathbb{E}}_{(x,y) \sim \mu^k}[F(x)G(y)] = \sum_\alpha \hat{F}_\alpha \hat{G}_\alpha \sigma_\alpha\,.$$

Therefore, we may apply Cauchy-Schwartz inequality to obtain

$$\begin{aligned}
\rho(p(\mu)) &= \sum_\alpha \hat{F}_\alpha \hat{G}_\alpha \sigma_\alpha\,, \\
&\leqslant \left(\sum_\alpha \hat{F}_\alpha^2 \sigma_\alpha^2\right)^{\frac{1}{2}} \left(\sum_\alpha \hat{G}_\alpha^2\right)^{\frac{1}{2}}\,, \\
&= \left(\sum_{\alpha\neq 0} \hat{F}_\alpha^2 \sigma_\alpha^2\right)^{\frac{1}{2}}\,,
\end{aligned}$$

where in the last step we have used both that $\hat{F}_0 = 0$ and $\sum_\alpha \hat{G}_\alpha^2 = 1$. Let $\rho = \rho(\mu)$ and note that for all $\alpha \in L_\mu$ we have $\sigma_\alpha \leqslant \rho$. Further, for non-zero $\alpha \notin L_\mu$ we have $\sigma_\alpha \leqslant \rho^2$. Thus, we can split the above sum to obtain:

$$\begin{aligned}
&\leqslant \left(\rho^2 \sum_{\alpha\in L_\mu} \hat{F}_\alpha^2 + \rho^4 \sum_{\alpha\notin L_\mu} \hat{F}_\alpha^2\right)^{\frac{1}{2}} \\
&\leqslant \rho\left(\mathrm{Lin}_\mu(p) + (1-\mathrm{Lin}_\mu(p))\rho^2\right)^{\frac{1}{2}} \\
&\leqslant \rho\left(1 - (1-\mathrm{Lin}_\mu(p))(1-\rho^2)\right)^{\frac{1}{2}} \\
&\leqslant \rho\left(1 - \frac{1}{2}(1-\mathrm{Lin}_\mu(p))(1-\rho^2)\right)\,.
\end{aligned}$$

□

6.2 Hypercontractivity and the Berry-Esseen Theorem

In light of Lemma 6.3, correlation will always decay under application of an operation p as long as $\mathrm{Lin}_\mu(p)$ is bounded away from one. The main challenge for proving that correlation decays to zero under repeated application of polymorphisms is in controlling the linearity after each application. The intuition for our proof is as follows: If a polymorphism has linearity very close to one, then it is close to a sum of independent random variables, and so the Berry-Esseen Theorem applies to show that $p(\mu_1)$ is nearly Gaussian. However this should be impossible since $p(\mu_1)$ only takes q distinct values. The version of the Berry-Esseen Theorem that we will use can be found as Corollary 59 in Chapter 11 of [O'D14].

Theorem 6.4 (Berry-Esseen [O'D14])**.** *Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be independent random variables with matching first and second moments; i.e. $E[X_i^k] = E[Y_i^k]$ for $i \in [n]$ and $k \in 1, 2$. Let $S_X = \sum_i X_i$ and $S_Y = \sum_i Y_i$. Then for any $\psi : \mathbb{R} \to \mathbb{R}$ which is c-Lipschitz*

$$|\,\mathbb{E}[\psi(S_X)] - \mathbb{E}[\psi(S_Y)]| \leqslant O(c) \cdot \sum_i E[X_i^3] + E[Y_i^3]$$

Note that since the error term in the Berry-Esseen Theorem depends on the L^3- norm of the random variables, we must control the L^3-norms of singular vectors of T_μ. Our main tool to this end will be hypercontractivity of the noise operator T_ρ. We state here a special case of the general hypercontractivity theorem which will suffice for our purposes.

Theorem 6.5 (Hypercontractivity [O'D14])**.** *Let π be a probability distribution on $[q]$ where each outcome has probability at least λ. Let $f \in L^2([q]^k, \pi^k)$ for some $k \in \mathbb{N}$. Then for any $0 \leqslant \rho \leqslant \frac{1}{\sqrt{2}}\lambda^{\frac{1}{6}}$*

$$\|T_\rho f\|_3 \leqslant \|f\|_2$$

where T_ρ is the noise operator.

Our first task will be to relate L^3 norms under the conditional expectation operator T_μ to those under the noise operator T_ρ. However, these operators do not map between the same spaces. To fix this, we instead consider the operator $M_\mu = T_\mu T_\mu^*$ so that $M : L^2(\Omega_1) \to L^2(\Omega_1)$. The following simple lemma gives a hypercontractivity result for M_μ.

Lemma 6.6. *Let $M_\mu = T_\mu T_\mu^*$ and let λ be the minimum probability of an atom in the marginal μ_1 of μ. Let s be a number such that $\rho(\mu)^{2s} \leqslant \frac{1}{\sqrt{2}}\lambda^{\frac{1}{6}}$. For any $k \in \mathbb{N}$, if let $f \in L^2([q]^k, \mu_1^k)$ then*

$$\|M_{\mu^k}^s f\|_3 \leqslant \|f\|_2$$

Proof. Observe first that M_μ is a self-adjoint linear operator and let $\{\sigma_i^2 | i \in [q]\}$ be the eigenvalues and let $\{\phi_i | i \in [q]\}$ denote the eigenfunctions.. Further $M_{\mu^k} = M_\mu^{\otimes k}$ is a self-adjoint operator with eigenvalues $\{\sigma_\alpha^2 | \alpha \in [q]^k\}$ and eigenfunctions $\{\phi_\alpha | \alpha \in [q]^k\}$.

Let $\rho = \rho(\mu)^{2s}$ and define

$$\eta_\alpha = \frac{\sigma_\alpha^{2s}}{\rho^{|\alpha|}}$$

Observe that $\eta_0 = 1$. Next, because $\sigma_j \leqslant \rho(\mu)$ for all $j \neq 0$ and $\sigma_0 = 1$ we have

$$\eta_\alpha = \frac{\prod_i \sigma_{\alpha_i}^{2s}}{\rho^{|\alpha|}} \leqslant \frac{\rho(\mu)^{2s|\alpha|}}{\rho^{|\alpha|}} = 1$$

for all $\alpha \neq 0$. Now define an operator A by setting $A\phi_\alpha = \eta_\alpha \phi_\alpha$ for each $\alpha \in [q]^k$. Since $\{\phi_\alpha\}_{\alpha \in [q]^k}$ form a basis, this uniquely defines A. In particular A is a self-adjoint operator with eigenvalues $\eta_\alpha \leqslant 1$ and corresponding eigenfunctions ϕ_α for $\alpha \in [q]^k$. Note that by construction

$$T_\rho A \phi_\alpha = \sigma_\alpha^{2s} \phi_\alpha = M_{\mu^k}^s \phi_\alpha \,,$$

That is, $T_\rho A$ agrees with $M_{\mu^k}^s$ on the entire ϕ_α basis, and so by linearity $T_\rho A = M_{\mu^k}^s$. Now we compute

$$\|M_{\mu^k}^s f\|_3 = \|T_\rho A f\|_3 \leqslant \|Af\|_2 \,,$$

where the final inequality follows by applying Theorem 6.5 to the function Af. Next since every eigenvalue of A is at most 1 we have

$$\|Af\|_2 \leqslant \|f\|_2 \,.$$

Plugging this into the previous inequality completes the proof. □

We now apply the hypercontractivity theorem in order to control the L^3-norms of singular vectors of $T_{p(\mu)}$ for some operation p. The following lemma gives a trade-off between the L^3-norm of a singular vector of $T_{p(\mu)}$ and the magnitude of the corresponding singular value. Further, the trade-off depends only on properties of the initial distribution μ. This in turn implies that singular vectors with high L^3-norm have low singular values, and thus cannot contribute much to the correlation $\rho(p(\mu))$.

Lemma 6.7. *Let λ be the minimum probability of an atom in the marginal μ_1 of μ. Let s be a number such that $\rho(\mu)^{2s} \leqslant \frac{1}{\sqrt{2}}\lambda^{\frac{1}{6}}$. Let $p : [q]^k \to [q]$ and $\phi_i \in L^2(\Omega_1)$ be a singular vector of $T_{p(\mu)}$ with singular value σ_i for $i \neq 0$. Then*

$$\sigma_i^{2s}\|\phi_i\|_3 \leqslant 1$$

Proof. For any distribution π we will use the notation $M_\pi = T_\pi T_\pi^*$. Since ϕ_i is a singular vector of $T_{p(\mu)}$ with singular value σ_i we have

$$\sigma_i^{2s}\|\phi_i\|_3 = \|M_{p(\mu)}^s \phi_i\|_3 \,.$$

In terms of the dual norm, we can write

$$\begin{aligned}
\|M^s_{p(\mu)}\phi_i\|_3 &= \sup_{\|g\|_{\frac{3}{2}}=1} \mathop{\mathbb{E}}_{z\sim p(\mu_1)}[g(z)M^s_{p(\mu)}\phi_i(z)] \\
&= \sup_{\|g\|_{\frac{3}{2}}=1} \mathop{\mathbb{E}}_{x\sim \mu_1^k}[g(p(x))M^s_{\mu^k}\phi_i(p(x))] \\
&\leqslant \sup_{\|g\|_{\frac{3}{2}}=1} \|g(p)\|_{\frac{3}{2}}\|M^s_{\mu^k}(\phi_i\circ p)\|_3 \\
&= \|M^s_{\mu^k}(\phi_i\circ p)\|_3
\end{aligned}$$

where in the second to last line we have applied Hölders inequality in the space $L^2([q]^k,\mu^k)$. Now note that by Lemma 6.6 ,

$$\|M^s_{\mu^k}(\phi_i\circ p)\|_3 \leqslant \|\phi_i\circ p\|_2 = 1\,.$$

Combining the above inequalities completes the proof. □

Next we will need the following technical lemma regarding discrete distributions taking only few values (e.g. functions $F\in \mathsf{Span}(p)$ for some polymorphism p). Informally, the lemma states that such distributions cannot be close to sums of many i.i.d. random variables.

Lemma 6.8. *Let F be a real-valued random variable taking only q values. Let X_i for $i\in[m]$ be i.i.d. real-valued random variables with $\mathbb{E}[X_i]=0$, $\mathbf{Var}[X_i]\leqslant \frac{1}{m}$ and $\mathbb{E}[|X_i|^3]\leqslant B$. Let $S=\sum_i X_i$. Then for some absolute constant $c>0$,*

$$\mathbb{E}[|F-S|]\geqslant \frac{1}{8q} - cBm\,.$$

Proof. Let $\{a_j\}_{j\in[q]}$ be the set of real values that F takes. Now define the function

$$d(z)=\min_j |z-a_j|$$

which simply measures the distance from z to the nearest a_j. Let $v=\sum_i \mathbf{Var}[X_i]$ and let Y_i be a Gaussian random variable with first two moments matching those of X_i. Note that $\sum_i Y_i$ is a Gaussian with mean zero and variance v. Further observe that d is 1-Lipschitz and so we may apply Theorem 6.4 to obtain

$$\left|\mathop{\mathbb{E}}_{z\sim\mathcal{N}(0,v)}[d(z)] - \mathbb{E}\left[d(S)\right]\right| \leqslant cBm$$

for some absolute constant c.

Next observe that the set of intervals such that $d(z)\leqslant\delta$ has total length $2q\delta$. Thus, the normal distribution $\mathcal{N}(0,v)$ has probability mass at most $2q\delta$ on the region where $d(z)\leqslant\delta$. This implies that

$$\mathop{\mathbb{E}}_{z\sim\mathcal{N}(0,v)}[d(z)]\geqslant \delta(1-2q\delta)$$

Combining this with the previous inequality yields

$$\mathbb{E}[d(S)]\geqslant \delta(1-2q\delta)-cBm$$

Now let $H=F-S$. Since F only takes the values a_j we have

$$\begin{aligned}
\mathbb{E}[d(S)] &= \mathbb{E}[d(F-H)] \\
&= \mathbb{E}[\min_j |F-H-a_j|] \\
&\leqslant \mathbb{E}[|a_j-H-a_j|] = \mathbb{E}[|H|]
\end{aligned}$$

Thus we conclude that

$$\mathbb{E}[|H|] \geqslant \delta(1 - 2q\delta) - cBm$$

Setting $\delta = \frac{1}{4q}$ yields the desired result. $\square$

6.3 Correlation Decay

Throughout the remainder of this section, whenever we refer to a joint probability distribution ν on $[q] \times [q]^m$ we will fix a singular value basis $\{\phi_\alpha, \psi_\alpha\}$ for the operator $T_\nu^{\otimes k}$. Let ν_1 and ν_2 be the respective marginals of ν. Whenever we have a function $F \in L^2([q]^k, \nu_1^k)$ we will write $\hat{F}_\alpha$ to denote the Fourier coefficients of F with respect to the basis ϕ_α.

Now we are ready to show that $\mathrm{Lin}_\mu(p)$ stays bounded away from one for operations p that are transitive symmetric. First, we need the following simple claim which asserts that corresponding degree one Fourier coefficients of p are all equal.

Claim 6.9. *Let $p : [q]^k \to [q]$ be a transitive symmetric operation and let $F \in \mathsf{Span}(p)$. Let ν_1 be any distribution on $[q]$ and fix a Fourier basis ϕ_α for $L^2([q]^k, \nu_1^k)$. Let α, β be multi-indices with $|\alpha| = |\beta| = 1$. Further suppose $\alpha_i = \beta_j$ for the unique pair i and j such that α_i and β_j are non-zero. Then $\hat{F}_\alpha = \hat{F}_\beta$.*

Proof. Let π be a permutation such that $\pi(j) = i$ and $p(x) = p(\pi \circ x)$. Such a permutation always exists since p is transitive symmetric. Now note

$$\hat{F}_\alpha = \mathbb{E}[F(x)\phi_{\alpha_i}(x_i)] = \mathbb{E}[F(\pi \circ x)\phi_{\alpha_i}(x_i)] = \mathbb{E}[F(x)\phi_{\beta_j}(x_j)] = \hat{F}_\beta$$

$\square$

We turn now to the proof of the theorem. The main idea of the proof is to repeatedly apply Lemma 6.3 to reduce correlation while using Lemma 6.7 in conjunction with Lemma 6.8 to control the linearity in each step.

Theorem 6.10. *Let $p_1 \dots p_r$ be transitive symmetric polymorphisms with each $p_i : [q]^k \to [q]$. Let λ be the minimum probability of an atom in the marginal μ_1 of μ. Then for any $\frac{1}{40c} > \varepsilon > 0$ and $r = \Omega_q(\frac{\log \lambda}{\log \rho(\mu)} \log^2(\varepsilon^{-1}))$*

$$\rho(p_1 \otimes p_2 \otimes \cdots \otimes p_r(\mu)) \leqslant \varepsilon$$

Proof. For the analysis we will break $p_1 \dots p_r$ into consecutive segments of length a, where a is a parameter that we will set later. Formally, let $K = k^a$ and let $P_t : [q]^K \to [q]$ be defined as

$$P_t(x) = p_{at+1} \otimes \cdots \otimes p_{at}(x)$$

Let $\mu^{(1)} = \mu$ and $\mu^{(t)} = P_1 \otimes \cdots \otimes P_{t-1}(\mu)$. Observe that each P_t is again a transitive symmetric polymorphism. Next we control $\mathrm{Lin}_{\mu^{(t)}}(P_t)$ only in terms of $\rho(\mu^{(t)})$ and properties of the initial distribution μ.

Let $F \in \mathsf{Span}(P_t)$ so that $\mathbb{E}_{\mu^{(t)}}[F] = 0$ and $\|F\|_2 = 1$. Let $\{\phi_\alpha, \psi_\alpha\}$ be a singular value basis for $T_{\mu^{(t)}}^{\otimes K}$ and let $\hat{F}_\alpha$ be the Fourier coefficients of $F : [q]^K \to \mathbb{R}$ with respect to the ϕ_α basis. Now define the function $l_i : [q] \to \mathbb{R}$ to be

$$l_i(x) = \sum_{\substack{\alpha \in L_{\mu^{(t)}} \\ \alpha_i \neq 0}} \hat{F}_\alpha \phi_{\alpha_i}(x)$$

In words, each l_i is simply the linear part of F corresponding to the ith coordinate. Claim 6.9 implies that for any pair i and j the sums for l_i and l_j are equal term-by-term. Thus, $l_i(x) = l_j(x)$ for all i, j. Now note that

$$F(x) = \sum_i l_i(x_i) + H(x)$$

where $H(x) = \sum_{\alpha \notin L_{\mu^{(t)}}} \hat{F}_\alpha \phi_\alpha(x)$ is the non-linear part of F. Since $\|F\|_2 = 1$ and the l_i are all equal and orthogonal, it follows that $\|l_i\|_2^2 \leqslant \frac{1}{K}$ for all i. This further implies that for any $\alpha \in L_{\mu^{(t)}}$ the coefficient $\hat{F}_\alpha \leqslant \frac{1}{\sqrt{K}}$.

Thus, the $l_i(x_i)$ are i.i.d random variables with mean zero and variance at most $\frac{1}{K}$. Further note that

$$\|l_i\|_3 \leqslant \frac{1}{\sqrt{K}} \sum_{\substack{\alpha \in L_{\mu^{(t)}} \\ \alpha_i \neq 0}} \|\phi_{\alpha_i}\|_3 \,.$$

Let $s = \Omega\left(\frac{\log \lambda}{\log \rho}\right)$. Now we apply Lemma 6.7 with $p = P_1 \otimes \cdots \otimes P_{t-1}$ to obtain

$$\frac{1}{\sqrt{K}} \sum_{\substack{\alpha \in L_{\mu^{(t)}} \\ \alpha_i \neq 0}} \|\phi_{\alpha_i}\|_3 \leqslant \frac{1}{\sqrt{K}} \sum_{\substack{\alpha \in L_{\mu^{(t)}} \\ \alpha_i \neq 0}} \sigma_\alpha^{-2s} \leqslant \frac{q}{\sqrt{K}} \rho(\mu^{(t)})^{-2s} \,,$$

where in the last inequality we have use the fact that $\sigma_\alpha \leqslant \rho(\mu^{(t)})$ for $\alpha \neq 0$, and that the sum has q terms. Therefore, since $F(x)$ is a random variable taking only $[q]$ different values we may apply Lemma 6.8 to obtain

$$\mathbb{E}[|H(x)|] \geqslant \frac{1}{8q} - c\|l_i\|_3^3 K \geqslant \frac{1}{8q} - cq^3 K^{-\frac{1}{2}} \rho(\mu^{(t)})^{-6s} \,.$$

Note that setting $a = 4\log(cq^3\varepsilon^{-6s})$ yields

$$K^{-\frac{1}{2}} = k^{-\frac{1}{2}a} \leqslant (cq^3\varepsilon^{-6s})^{-2} \,.$$

Thus as long as $\rho(\mu^{(t)}) \geqslant \varepsilon$ we have

$$\mathbb{E}[|H(x)|] \geqslant \frac{1}{8q} - (cq^3\varepsilon^{-6s})^{-1} > \frac{1}{10q}$$

where the last inequality follows from $\varepsilon < \frac{1}{40c}$. So, letting $\delta = \mathbb{E}[H(x)^2] \geqslant \mathbb{E}[|H(x)|]^2 = \frac{1}{100q^2}$ we obtain $\mathrm{Lin}_{\mu^{(t)}}(P_t) \leqslant 1 - \delta$.

Now we apply Lemma 6.3 to P_t to obtain

$$\begin{aligned}
\rho(\mu^{(t+1)}) &\leqslant \rho(\mu^{(t)}) \left(1 - \frac{1}{2}(1 - \mathrm{Lin}(\mu^{(t)}))(1 - \rho(\mu^{(t)})^2)\right) \\
&\leqslant \rho(\mu^{(t)}) \left(1 - \frac{\delta}{2}(1 - \rho(\mu^{(t)})^2)\right) \\
&= \left(1 - \frac{\delta}{2}\right) \rho(\mu^{(t)}) + \frac{\delta}{2}\rho(\mu^{(t)})^3
\end{aligned}$$

Solving this recurrence shows that $\rho(\mu^{(t)}) \leqslant \varepsilon$ after $t = O_q(\log(\varepsilon^{-1}))$ steps. Since each step corresponds to a applications of polymorphisms, we get that the total number of applications required is

$$r = \Omega_q(a \log(\varepsilon^{-1})) = \Omega_q(s \log^2(\varepsilon^{-1})).$$

$\square$

We now use the connection given in Lemma 3.2 between correlation and statistical distance to prove Theorem 3.3.

Proof of Theorem 3.3. Let $P = p_1 \otimes p_2 \otimes \cdots \otimes p_r$. The proof that $\|P(\mu) - P(\mu^\times)\|_1 \leqslant \eta$ is by a hybrid argument. We define intermediate distributions $\pi^{(i)}$ by drawing a sample X from μ and then independently re-sampling each of the coordinates $j > i$ from their respective marginals μ_j. Let Y_j be the random variable corresponding to an independent sample from the marginal μ_j. We will use the notation $\pi^{(i)} = (X_1, \ldots, X_i, Y_{i+1}, \ldots, Y_n)$.

Note that since the Y_j are independent of each other and of the X_l we have

$$\|P(\pi^{(i-1)}) - P(\pi^{(i)})\|_1 = \|\Big(P(X_1), \ldots, P(X_{i-1}), P(Y_i)\Big) - \Big(P(X_1), \ldots, P(X_i)\Big)\|_1$$

Let $\varepsilon = \frac{\eta}{qn}$. By Theorem 6.10 we have $\rho\Big(P(X_i), (P(X_1), \ldots, P(X_{i-1}))\Big) \leqslant \varepsilon$. Further, since Y_i is independent of $X_1, \ldots, X_{i-1}$ we may apply Lemma 3.2 to obtain

$$\|\Big(P(X_1), \ldots, P(X_{i-1}), P(Y_i)\Big) - \Big(P(X_1), \ldots, P(X_i)\Big)\|_1 \leqslant q\varepsilon$$

Now since $\pi^{(0)} = \mu^\times$ and $\pi^{(n)} = \mu$ we have by the triangle inequality

$$\|P(\mu) - P(\mu^\times)\|_1 \leqslant \sum_{i=0}^{n-1} \|P(\pi^{(i)}) - P(\pi^{(i+1)})\|_1 \leqslant nq\varepsilon = \eta$$

$\square$

7 Approximate Polymorphisms for CSPs

7.1 Background

Constraint Satisfaction Problems. Fix an alphabet $[q]$. A CSP Λ over $[q]$ is given by a family of payoff functions $\Lambda = \{c : [q]^k \to [-1, 1]\}$. An instance of $\textsc{Max}$-Λ consists of a set of variables $\mathcal{V} = \{X_1, \ldots, X_n\}$ and a set of constraints $\mathcal{C} = \{C_1, \ldots, C_m\}$ where each $C_i(X) = c(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$ for some $c \in \Lambda$. The set $\mathcal{C}$ is equipped with a probability distribution $w : \mathcal{C} \to \mathbb{R}^+$.

The goal is to find an assignment $x \in [q]^{\mathcal{V}}$ that maximizes

$$\mathrm{val}_{\Im}(x) \stackrel{\text{def}}{=} \sum_{C \in \mathcal{C}} w(C) C(x)\,.$$

If the probability distribution w is clear from the context, we will write

$$\mathrm{val}_{\Im}(x) = \mathop{\mathbb{E}}_{C \sim \mathcal{C}} [C(x)]\;.$$

Remark 7.1. *A natural special case of the above definition is the set of unweighted constraint satisfaction problems wherein the weight function $w : \mathcal{C} \to \mathbb{R}_+$ is uniform, i.e.,*

$$\mathrm{val}_{\Im}(x) = \frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} C(x)\,.$$

In terms of approximability, the unweighted case is no easier than the weighted version since for every constant $\eta > 0$, there is a polynomial time reduction from weighted version to its unweighted counterpart with a loss of at most η in the approximation.

Dictatorship tests.

Definition 7.2. *Fix constants* $-1 \leqslant s \leqslant c \leqslant 1$*, an integer* $R \in \mathbb{N}$ *and a family of functions* $\mathcal{F} = \{A : [q]^R \to [q]\}$*. A* (c, s)*-dictatorship test for* MAX-Λ *against the family* $\mathcal{F}$ *consists of an instance* $\Im$ *of* MAX-Λ *with the following properties: The variables of* $\Im$ *are identified with* $[q]^R$ *and therefore an assignment to* $\Im$ *is given by* $p : [q]^R \to [q]$*.*

- *(completeness) For every* $i \in [R]$*, the* i^{th} *dictator assignment* $p_i : [q]^R \to [q]$ *given by* $p_i(x) = x^{(i)}$ *satisfies* $\Im(p_i) \geqslant c$*.*
- *(soundness) For every function* $p \in \mathcal{F}$*, we have* $\Im(p) \leqslant s$

The intimate relation between polymorphisms and dictatorship testing was observed in [KS09], here we spell out the connection to approximate polymorphisms. First, for a function family $\mathcal{F} = \{p : [q]^R \to [q]\}$, we will say an approximate polymorphism $\mathcal{P}$ is *supported on* $\mathcal{F}$ if probability distribution $\mathcal{P}$ is supported only on functions within the family $\mathcal{F}$.

Theorem 7.3. *Given a CSP* Λ*, a natural number* $R \in \mathbb{N}$ *and a finite family of functions* $\mathcal{F} = \{p : [q]^R \to [q]\}$ *closed under permutation of inputs, the following are equivalent:*

- MAX-Λ *does not admit a* (c, s)*-approximate polymorphism supported on* $\mathcal{F}$*.*
- *There exists a* (c, s)*-dictatorship test for* MAX-Λ *against the family* $\mathcal{F}$*.*

Proof. Suppose MAX-Λ does not admit (c, s)-approximate polymorphism supported in $\mathcal{F}$. In other words, for every probability distribution $\mathcal{P}$ over $\mathcal{F}$, there exists an instance $\Im$ and R solutions $X^{(1)}, \ldots, X^{(R)}$ of average value c such that the value of $\mathbb{E}_{p \in \mathcal{P}} \Im(p(X^{(1)}, \ldots, X^{(R)})) \leqslant s$

Notice that the set of probability distributions $\mathcal{P}$ over $\mathcal{F}$ forms a convex set. Furthermore, the set of all instances $\Im$ of MAX-Λ along with R assignments $X^{(1)}, \ldots, X^{(R)}$ with average value at least c, form a convex set. Specifically, given two instances $\Im_1, \Im_2$ and any $\theta \in [0, 1]$, one can construct the instance $\theta\Im_1 + (1 - \theta)\Im_2$ by taking a disjoint union of the two instances weighted appropriately. Since the set of all probability distributions $\mathcal{P}$ over $\mathcal{F}$ is a compact convex set, by min-max theorem there exists a single instance $\Im$ and a set of R solutions for it, that serve as a counterexample against every probability distribution $\mathcal{P}$ over the family $\mathcal{F}$.

We will use the instance $\Im$ and the set of solutions $X^{(1)}, \ldots, X^{(R)}$ to create the following dictatorship test.

- Sample a random constraint C from $\Im$. Let C be a constraint on variables $v_{i_1}, \ldots, v_{i_k}$ in $\Im$.
- Pick a random permutation $\pi : [R] \to [R]$ and set $y_j = (X_{i_j}^{(\pi(1))}, \ldots, X_{i_j}^{(\pi(R))})$ for $j \in [R]$.
- Test the constraint C on $p(y_{i_1}), \ldots, p(y_{i_k})$.

Suppose $p(y) = y^{(i)}$ for some $i \in [R]$. In this case, it is easy to see that over the random choice of constraint C and permutation π, the expected probability of success of the dictatorship test is exactly equal to the average value of the solutions $X^{(1)}, \ldots, X^{(R)}$.

Let $p : [q]^R \to [q]$ denote any function in the family $\mathcal{F}$. If probability of success of p is greater than s then there exist a permutation $\pi : [R] \to [R]$ for which,

$$\Im(p(X^{(\pi(1))}, \ldots, X^{(\pi(R))})) \geqslant s\,.$$

This is a contradiction, since $p \circ \pi$ would also be a function in the family $\mathcal{F}$. Therefore, for every function $p \in \mathcal{F}$, its value on the dictatorship test is at most s. This completes the proof of one direction of the implication.

Now, we will show the other direction, namely that if there exists a (c,s)-dictatorship test then there is no (c,s)-approximate polymorphism. Conversly, suppose there is a (c,s)-dictatorship test against the family $\mathcal{F}$. Consider the instance $\Im$ given by the dictatorship test and the family of R assignments corresponding to the dictator functions, i.e., $p_i(x) = x^{(i)}$. By the completeness of the dictatorship test, for each of the dictator assignments p_i, we will have $\Im(p_i) \geqslant c$. On the other hand, by the soundness of the dictatorship test, for each function p in the family $\mathcal{F}$, $\Im(p) \leqslant s$. This implies that there does not exist a (c,s)-approximate polymorphism supported on the family $\mathcal{F}$.

□

Unique Games Conjecture. For the sake of completeness, we recall the unique games conjecture here.

Definition 7.4. *An instance of Unique Games represented as $\Gamma = (\mathcal{A} \cup \mathcal{B}, E, \Pi, [R])$, consists of a bipartite graph over node sets $\mathcal{A}$,$\mathcal{B}$ with the edges E between them. Also part of the instance is a set of labels $[R] = \{1, \ldots, R\}$, a a set of permutations $\pi_{ab} : [R] \to [R]$ for each edge $e = (a,b) \in E$. An assignment Λ of labels to vertices is said to satisfy an edge $e = (a,b)$, if $\pi_{ab}(\Lambda(a)) = \Lambda(b)$. The objective is to find an assignment Λ that satisfies the maximum number of edges.*

The unique games conjecture of Khot [Kho02] asserts that the unique games CSP is hard to approximate in the following sense.

Conjecture 7.5. *(Unique Games Conjecture [Kho02]) For all constants $\delta > 0$, there exists large enough constant R such that given a bipartite unique games instance $\Gamma = (\mathcal{A} \cup \mathcal{B}, E, \Pi = \{\pi_e : [R] \to [R] \ : \ e \in E\}, [R])$ with number of labels R, it is NP-hard to distinguish between the following two cases:*

- $(1-\delta)$-satisfiable instances: *There exists an assignment Λ of labels that satisfies a $(1-\delta)$-fractional of all edges.*
- Instances that are not δ-satisfiable: *No assignment satisfies more than a δ-fraction of the constraints Π.*

7.2 Quasirandom functions

The function family of interest in this work are those with no influential coordinates. To make the definition precise, we begin by recalling a few analytic notions.

Low Degree Influences. Fix a probability distribution μ over $[q]$. Let $\{\chi_0, \ldots, \chi_{q-1}\}$ be an orthonormal basis for the vector space $L_2([q], \mu)$. Without loss of generality, we can fix a basis such that $\chi_0 = 1$. Given a function $f : [q]^k \to \mathbb{R}$, we can write f as,

$$f = \sum_{\sigma \in \mathbb{N}^k} \hat{f}_\sigma \chi_\sigma ,$$

where $\chi_\sigma(x) \stackrel{\text{def}}{=} \prod_{j=1}^k \chi_{\sigma_j}(x_j)$. Define the degree d influence of the i^{th} coordinate of f under the probability distribution μ as,

$$\mathrm{Inf}_{i,\mu}^{<d}(f) \stackrel{\text{def}}{=} \sum_{\sigma \in \mathbb{N}^k, \sigma_i \neq 0, |\sigma| < d} \hat{f}_\sigma^2 .$$

More generally, for a vector valued function $f : [q]^k \to \mathbb{R}^D$, we will set

$$\mathrm{Inf}_{i,\mu}^{<d}(f) \stackrel{\text{def}}{=} \sum_{j \in [D]} \mathrm{Inf}_{i,\mu}^{<d}(f_j) .$$

A useful property of low-degree influences is that their sum is bounded as expounded in the following lemma.

Lemma 7.6. *For every function $f : [q]^k \to \mathbb{R}$ and every probability distribution μ over $[q]$,*

$$\sum_{i \in [k]} \mathrm{Inf}^{<d}_{i,\mu}(f) \leqslant d \cdot \mathop{\mathbf{Var}}_{\mu}(f)\,,$$

where $\mathbf{Var}_\mu(f)$ denotes the variance of f under the probability distribution μ.

Given an operation $p : [q]^k \to [q]$, we will associate a function $\tilde{p} : [q]^k \to \blacktriangle_q$ given by $\tilde{p}(x) = e_{p(x)}$ where in $e_{p(x)} \in \blacktriangle_q$ is the basis vector along $p(x)^{th}$ coordinate. We will abuse notation and use p to denote both the $[q]$-valued function p and the corresponding real-vector valued function $\tilde{p}$. For example, $\mathrm{Inf}_{i,\mu}(p)$ will denote the influence of the i^{th} coordinate on $\tilde{p}$.

Definition 7.7. *An approximate polymorphism $\mathcal{P}$ is (τ, d)-quasirandom if for every probability distribution $\mu \in \blacktriangle_q$,*

$$\mathop{\mathbb{E}}_{p \in \mathcal{P}} \left[\max_i \mathrm{Inf}_{i,\mu}(p) \right] \leqslant \tau$$

The following lemma shows that transitive symmetries imply quasirandomness.

Lemma 7.8. *An operation $p : [q]^k \to [q]$ that has a transitive symmetry is $(\frac{qd}{k}, d)$-quasirandom for all $d \in \mathbb{N}$.*

Proof. The lemma is a consequence of two facts, first the sum of degree d influences of a function $p : [q]^k \to [q]$ is always at most qd. Second, if the operation p admits a transitive symmetry, then the degree d influences of each coordinate is the same. $\square$

Noise Operator. For a function $p : [q]^k \to \mathbb{R}$ and $\rho \in [0, 1]$, define $T_\rho p$ as

$$T_\rho p(x) = \mathop{\mathbb{E}}_{y \sim_\rho x}[p(y)]$$

where $y \sim_\rho x$ denotes the following distribution,

$$y_i = \begin{cases} x_i & \text{with probability } \rho \\ \text{independent sample from } \mu & \text{with probability } 1 - \rho\,. \end{cases}$$

The multilinear polynomial associated with $T_\rho p$ is given by

$$T_\rho p = \sum_\sigma \hat{p}_\sigma \rho^{|\sigma|} \chi_\sigma\,.$$

Approximation Thresholds. For each $\tau > 0$ and $d \in \mathbb{N}$, let $\mathcal{F}_{\tau,d}$ denote the family of all (τ, d)-quasirandom functions.

Definition 7.9. *Given a CSP Λ, and a constant $c \in [-1, 1]$ define*

$$s_\Lambda(c) \stackrel{\text{def}}{=} \sup\{s | \forall \tau > 0, d \in \mathbb{N}, \exists \text{ a } (\tau, d)\text{-quasirandom } (c, s)\text{-approximate polymorphism for } \textsc{Max}\text{-}\Lambda\}$$

Moreover, let

$$\alpha_\Lambda \stackrel{\text{def}}{=} \inf_{c \geqslant 0} \frac{s_\Lambda(c)}{c}$$

The following observations are immediate consequences of the definition of s_Λ.

Observation 7.10. *The map $s_\Lambda : [-1, 1] \to [-1, 1]$ is monotonically increasing and $s_\Lambda(c + \varepsilon) \leqslant s_\Lambda(c) + \varepsilon$ for every c, ε such that $c, c + \varepsilon \in (-1, 1)$.*

7.3 Rounding Semidefinite Programs via Polymorphisms

In this section, we will restate the soundness analysis of dictatorship tests in [Rag08] in terms of approximate polymorphisms. Specifically, we will construct a rounding scheme for the BasicSDP relaxation using approximate polymorphisms, and thereby prove Theorem 1.12.

Basic SDP relaxation. Given a Λ-CSP instance $\Im = (\mathcal{V}, \mathcal{C})$, the goal of BasicSDP relaxation is to find a collection of vectors $\{\boldsymbol{b}_{v,a}\}_{v\in\mathcal{V},a\in[q]}$ in a sufficiently high dimensional space and a collection $\{\mu_C\}_{C\in\mathrm{supp}(\mathcal{C})}$ of distributions over local assignments. For each payoff $C \in \mathcal{C}$, the distribution μ_C is a distribution over $[q]^{\mathcal{V}(C)}$ corresponding to assignments for the variables $\mathcal{V}(C)$. We will write $\mathbb{P}_{x\in\mu_C}\{E\}$ to denote the probability of an event E with under the distribution μ_C.

BasicSDP Relaxation

$$\begin{aligned}
\text{maximize} \quad & \mathop{\mathbb{E}}_{C\sim\mathcal{C}} \mathop{\mathbb{E}}_{x\sim\mu_C} C(x) && (\text{ Basic SDP})\\
\text{subject to} \quad & \langle \boldsymbol{b}_{v,i}, \boldsymbol{b}_{v',j}\rangle = \mathop{\mathbb{P}}_{x\sim\mu_C}\Big\{x_v = i, x_{v'} = j\Big\} \\
& \qquad (C \in \mathrm{supp}(\mathcal{C}),\ \ v, v' \in \mathcal{V}(C),\ \ i, j \in [q])\,. \\
& \mu_C \in \blacktriangle([q]^{\mathcal{V}(C)}) \qquad\qquad \forall C \in \mathrm{supp}(\mathcal{C})
\end{aligned}$$

We claim that the above optimization problem can be solved in polynomial time. To show this claim, let us introduce additional real-valued variables $\mu_{C,x}$ for $C \in \mathrm{supp}(\mathcal{C})$ and $x \in [q]^{V(C)}$. We add the constraints $\mu_{C,x} \geqslant 0$ and $\sum_{x\in[q]^{V(C)}} \mu_{C,x} = 1$. We can now make the following substitutions to eliminate the distributions μ_C,

$$\mathop{\mathbb{E}}_{x\sim\mu_C} C(x) = \sum_{x\in[q]^{V(C)}} C(x)\mu_{C,x}\,, \qquad \mathop{\mathbb{P}}_{x\sim\mu_C}\Big\{x_i = a\Big\} = \sum_{\substack{x\in[m]^{V(C)}\\ x_i=a}} \mu_{C,x}\,,$$

$$\mathop{\mathbb{P}}_{x\sim\mu_C}\Big\{x_i = a, x_j = b\Big\} = \sum_{\substack{x\in[m]^{V(C)}\\ x_i=a,x_j=b}} \mu_{C,x}\,.$$

After substituting the distributions μ_C by the scalar variables $\mu_{C,x}$, it is clear that an optimal solution to the relaxation of $\mathcal{C}$ can be computed in time $\mathrm{poly}(m^k, |\,\mathrm{supp}(\mathcal{C})|)$. In the rest of the section, we will show how to obtain an rounding scheme for the above SDP using an approximate polymorphism $\mathcal{P}$.

The overall idea behind the rounding scheme is as follows. Let $(\boldsymbol{V}, \boldsymbol{\mu})$ be a feasible solution to the BasicSDP where $\boldsymbol{V}$ consists of the vectors and $\boldsymbol{\mu}$ consists of the associated local distributions.

Let $\mathcal{P}$ be an (c, s)-approximate polymorphism. If the polymorphism $\mathcal{P}$ is given as input integral assignments whose average value is equal to c, it would output an integral assignment of expected value s. This would be a rounding for BasicSDP relaxation certifying that the on instances with SDP value at least c, the optimum is at least s. However, in general, we do not have access to integral solutions of value c and there might not exist any. The idea is to give as input to $\mathcal{P}$ real-valued assignments such that they have value c, and the polymorphism $\mathcal{P}$ cannot distinguish these real valued assignments from integral assignments. Specifically, these real valued assignments are gaussian random variables obtained by taking random projections of the SDP vectors.

The following is the formal description of the rounding procedure $\mathsf{Round}_{\mathcal{P}}$.

$\mathsf{Round}_{\mathcal{P}}$ Scheme

Input: A Λ-CSP instance $\Im = (\mathcal{V}, \mathcal{C})$ with m variables and an SDP solution $\{\boldsymbol{b}_{v,i}\}, \{\mu_C\}$ with value at least c. A (τ, d)-quasirandom $(c - \eta, s)$-approximate polymorphism $\mathcal{P}$.

Truncation Function Let $f_{\blacktriangle} : \mathbb{R}^q \to \blacktriangle_q$ be a Lipschitz-continous function such that for all $\mathbf{x} \in \blacktriangle_q$, $f_{\blacktriangle}(\mathbf{x}) = \mathbf{x}$.
Rounding Scheme: Fix $\varepsilon = \eta/10k$ where k is arity of Λ.

- Sample R vectors $\zeta^{(1)}, \ldots, \zeta^{(R)}$ with each coordinate being i.i.d normal random variable.
- Sample an operation $p \sim \mathcal{P}$.

For each $v \in \mathcal{V}$ do

- For all $1 \leqslant j \leqslant R$ and $c \in [q]$, compute the projection $g_{v,c}^{(j)}$ of the vector $\boldsymbol{b}_{v,c}$ as follows:
$$g_{v,c}^{(j)} = \langle \mathbf{I}, \boldsymbol{b}_{v,c} \rangle + \left[\langle (\boldsymbol{b}_{v,c} - (\langle \mathbf{I}, \boldsymbol{b}_{v,c} \rangle)\mathbf{I}), \zeta^{(j)} \rangle \right]$$
- Let $\boldsymbol{H}$ denote the multilinear polynomial corresponding to the function $T_{1-\varepsilon}p : \blacktriangle_q^k \to \blacktriangle_q$. Evaluate the function $\boldsymbol{H} = T_{1-\varepsilon}p$ with $g_{v,c}^{(j)}$ as inputs. In other words, compute $\mathbf{z}_v = (z_{v,1}, \ldots, z_{v,q})$ as follows:
$$\mathbf{z}_v = \boldsymbol{H}(\mathbf{g_v})$$
- Round $\mathbf{z}_v$ to $\mathbf{z}_v^* \in \blacktriangle_q$ by using a Lipschitz-continous truncation function $f_{\blacktriangle} : \mathbb{R}^q \to \blacktriangle_q$, i.e., $\mathbf{z}_v^* = f_{\blacktriangle}(\mathbf{z}_v)$.
- Independently for each $v \in \mathcal{V}$, assign a value $j \in [q]$ sampled from the distribution $z_v^* \in \blacktriangle_q$.

Now we will analyze the performance of the above rounding scheme. First, we observe the following fact about approximate polymorphisms.

Lemma 7.11. *Fix an instance $\Im$ of* MAX-Λ *and a distribution Θ over assignments to $\Im$ with $\mathbb{E}_{X\sim\Theta}[\Im(X)] \geqslant c$. Suppose $\mathcal{P}$ is a $(c-\eta, s)$-approximate polymorphism for* MAX-Λ *for some $\eta > 0$, then we have*
$$\mathop{\mathbb{E}}_{p\in\mathcal{P}} \mathop{\mathbb{E}}_{X^{(1)},\ldots,X^{(R)}\sim\Theta} \left[\Im(p(X^{(1)}, \ldots, X^{(R)})) \right] \geqslant s\,.$$
where $X^{(1)}, \ldots, X^{(R)}$ are sampled independently from Θ.

Proof. Let $N \in \mathbb{N}$ be a large positive integer. Let $\Im'$ be an instance consisting of N disjoint copies of $\Im$. Define a distribution Θ' of assignments to $\Im'$ consisting of N-i.i.d samples from Θ. While the distribution Θ only satisfies an average bound on objective value $\mathbb{E}_{X\sim\Theta}[\Im(X)] \geqslant c$, the distribution Θ' will satisfy
$$\mathop{\mathbb{P}}_{Y\sim\Theta'}[\Im'(Y) \geqslant c - \eta] \geqslant 1 - e^{-O(\eta N)}\,.$$
Hence, if we sample $Y^{(1)}, \ldots, Y^{(R)} \sim \Theta'$ then the average value of the assignments is at least $c - \eta$

with probability $1 - e^{-O(\eta N)}$. We are ready to finish the argument as shown below,

$$\begin{aligned}
\mathop{\mathbb{E}}_{p\in\mathcal{P}} \mathop{\mathbb{E}}_{X^{(1)},\ldots,X^{(R)}\sim\Theta} \left[\Im(p(X^{(1)},\ldots,X^{(R)}))\right] &= \mathop{\mathbb{E}}_{p\in\mathcal{P}} \mathop{\mathbb{E}}_{Y^{(1)},\ldots,Y^{(R)}\sim\Theta'} \left[\Im'(p(Y^{(1)},\ldots,Y^{(R)}))\right] \\
&\qquad (\text{ linearity of expectation }) \\
&\geqslant \mathop{\mathbb{E}}_{p\in\mathcal{P}} \mathop{\mathbb{E}}_{Y^{(1)},\ldots,Y^{(R)}\sim\Theta'} \left[\Im'(p(Y^{(1)},\ldots,Y^{(R)})) \;\middle|\; \Im'(Y^{(i)}) \geqslant c-\eta\right] - e^{-O(\eta N)} \\
&\geqslant s - e^{-O(\eta N)}\,.
\end{aligned}$$

Taking limits as $N \to \infty$, the conclusion is immediate. ☐

To analyze the performance of the rounding scheme, we define a set of ensembles of local integral random variables, and global gaussian ensembles as follows.

Definition 7.12. *For every payoff $C \in \mathcal{C}$ of size at most k, the local distribution μ_C is a distribution over $[q]^{\mathcal{V}(C)}$. In other words, the distribution μ_C is a distribution over assignments to the CSP variables in set $\mathcal{V}(C)$. The corresponding local integral ensemble is a set of random variables $\mathcal{L}_C = \{\ell_{v_1},\ldots,\ell_{v_k}\}$ each taking values in $\blacktriangle_q$.*

Definition 7.13. *The global ensemble $\mathcal{G} = \{\boldsymbol{g}_v | v \in \mathcal{V}, j \in [q]\}$ are generated by setting $\boldsymbol{g}_v = \{g_{v,1},\ldots,g_{v,q}\}$ where*

$$g_{v,j} \stackrel{\text{def}}{=} \langle \mathbf{I}, \boldsymbol{b}_{v,j}\rangle + \langle (\boldsymbol{b}_{v,j} - (\langle \mathbf{I}, \boldsymbol{b}_{v,j}\rangle)\mathbf{I}), \zeta\rangle$$

and ζ is a normal Gaussian random vector of appropriate dimension.

It is easy to see that the local and global integral ensembles have matching moments up to degree two.

Observation 7.14. *For any set $C \in \mathcal{C}$, the global ensemble $\mathcal{G}$ matches the following moments of the local integral ensemble $\mathcal{L}_C$*

$$\begin{aligned}
\mathbb{E}[g_{v,j}] = \mathbb{E}[\ell_{v,j}] = \langle \mathbf{I}, \boldsymbol{b}_{v,j}\rangle \qquad\qquad & \mathbb{E}[g_{v,j}^2] = \mathbb{E}[\ell_{v,j}^2] = \langle \mathbf{I}, \boldsymbol{b}_{v,j}\rangle \\
\mathbb{E}[g_{v,j} g_{v,j'}] = \mathbb{E}[\ell_{v,j}\ell_{v,j'}] = 0 \qquad\qquad & \forall j \neq j', v \in \mathcal{V}(C)
\end{aligned}$$

We will appeal to the invariance principle of Mossel et al. [MOO05], Mossel and Issakson [IM09] to argue that a (τ, d)-quasirandom polymorphism cannot distinguish between two distributions that have same first two moments.

Theorem 7.15. *(Invariance Principle [IM09]) Let Ω be a finite probability space with the least non-zero probability of an atom at least $\alpha \leqslant 1/2$. Let $\mathcal{L} = \{\ell_1, \ell_1, \ldots, \ell_m\}$ be an ensemble of random variables over Ω. Let $\mathcal{G} = \{g_1,\ldots,g_m\}$ be an ensemble of Gaussian random variables satisfying the following conditions:*

$$\mathbb{E}[\ell_i] = \mathbb{E}[g_i] \qquad \mathbb{E}[\ell_i^2] = \mathbb{E}[g_i^2] \qquad \mathbb{E}[\ell_i\ell_j] = \mathbb{E}[g_i g_j] \qquad \forall i,j \in [m]$$

Let $K = \log(1/\alpha)$. Let $\mathbf{F} = (F_1,\ldots,F_D)$ denote a vector valued multilinear polynomial and let $H_i = (T_{1-\varepsilon}F_i)$ and $\boldsymbol{H} = (H_1,\ldots,H_D)$. Further let $\mathrm{Inf}_i(\boldsymbol{H}) \leqslant \tau$ and $\mathbf{Var}[H_i] \leqslant 1$ for all i.

If $\Psi : \mathbb{R}^D \to \mathbb{R}$ is a Lipschitz-continous function with Lipschitz constant C_0 (with respect to the L_2 norm). Then,

$$\left|\mathbb{E}\left[\Psi(\boldsymbol{H}(\mathcal{L}^R))\right] - \mathbb{E}\left[\Psi(\boldsymbol{H}(\mathcal{G}^R))\right]\right| \leqslant C_D \cdot C_0 \cdot \tau^{\varepsilon/18K} = o_\tau(1)$$

for some constant C_D depending on D.

The local integral ensemble has an expected payoff equal to c. The local and global ensembles agree on the first two moments. The (c, s)-approximate polymorphism outputs a solution of value at least s when given the local integral ensemble as inputs. Hence, by invariance principle when given these global ensemble of random projections as input to the polymorphism $\mathcal{P}$, it will end up outputting close to integral solutions of value s.

Theorem 7.16. *For all η and CSP Λ, there exists $\tau, d > 0$ such that the following holds: Suppose $\mathcal{P}$ is a $(c - \eta, s)$-approximate (τ, d)-quasirandom polymorphism for* Max-Λ. *Given an* BasicSDP *solution with objective value at least c, the expected value of the assignment produced by* $\mathsf{Round}_{\mathcal{P}}$ *algorithm is at least $s - \eta$.*

Proof. Let $\mathsf{Round}_{\mathcal{P}}(\boldsymbol{V}, \boldsymbol{\mu})$ denote the expected payoff of the ordering assignment by the rounding scheme $\mathsf{Round}_{\mathcal{P}}$ on the SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ for the Λ-CSP instance $\Im$. Notice that the $\boldsymbol{g}_i$ are nothing but samples of the global ensemble $\mathcal{G}$ associated with $\Im$. For each constraint $C \in \mathcal{C}$, let $\mathcal{G}_C$ denote the subset of random variables from global gaussian ensemble that correspond to variables in $\mathcal{V}(\mathcal{C})$. It will be useful to extend each payoff $C : [q]^k \to [-1, 1]$ to the domain of fractional/distributional assignments $\blacktriangle_q$. Specifically, for $z_1, \ldots, z_k \in \blacktriangle_q$, define $C(z_1, \ldots, z_k) \stackrel{\text{def}}{=} \mathbb{E}_{x_i \sim z_i}[C(x_1, \ldots, x_k)]$ which corresponds to the expected payoff on fixing each input x_i independently from the corresponding distribution $z_i \in \blacktriangle_q$. By definition, the expected payoff is given by

$$\mathsf{Round}_{\mathcal{P}}(\boldsymbol{V}, \boldsymbol{\mu}) = \mathop{\mathbb{E}}_{C \in \mathcal{C}}[C\left(v_1, \ldots, v_k\right)] \tag{7.1}$$

$$= \mathop{\mathbb{E}}_{p \in \mathcal{P}} \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{\mathcal{G}_C^R} \Big[C\Big(f_{\blacktriangle}(\boldsymbol{H}(\mathbf{g}_{v_1}^R)), \ldots, f_{\blacktriangle}(\boldsymbol{H}(\mathbf{g}_{v_k}^R))\Big)\Big] \tag{7.2}$$

Fix a payoff $C \in \mathcal{C}$. Let $\Psi_C : \mathbb{R}^{qk} \to \mathbb{R}$ be a Lipschitz continous function defined as follows:

$$\Psi_C(\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_k) = C\Big(f_{\blacktriangle}(\mathbf{z}_1), \ldots, f_{\blacktriangle}(\mathbf{z}_k)\Big) \qquad \forall \mathbf{z}_1, \ldots \mathbf{z}_k \in \mathbb{R}^q\,.$$

Hence, we can write

$$\mathsf{Round}_{\mathcal{P}}(\boldsymbol{V}, \boldsymbol{\mu}) = \mathop{\mathbb{E}}_{p \in \mathcal{P}} \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{\mathcal{G}_C^R} \Big[\Psi_C\Big(\boldsymbol{H}(\mathbf{g}_{v_1}^R), \ldots, \boldsymbol{H}(\mathbf{g}_{v_k}^R)\Big)\Big]\,. \tag{7.3}$$

For a constraint $\mathcal{C}$, there are k-different marginal distributions in $\mu_{\mathcal{C}}$. Note that since $\mathcal{P}$ is (τ, d)-quasirandom, with probability at least $1 - k\sqrt{\tau}$ over the choice of $p \sim \mathcal{P}$, we will have

$$\max_i \mathrm{Inf}_{i,\mu_j}^{\leqslant d}(p) \leqslant \sqrt{\tau}\,,$$

for each of the marginals μ_j within $\mu_{\mathcal{C}}$. Since the objective value is always between $[-1, 1]$, we get that

$$\mathsf{Round}_{\mathcal{P}}(\boldsymbol{V}, \boldsymbol{\mu}) \geqslant \mathop{\mathbb{E}}_{p \in \mathcal{P}} \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{\mathcal{G}_C^R} \Big[\Psi_C\Big(\boldsymbol{H}(\mathbf{g}_{v_1}^R), \ldots, \boldsymbol{H}(\mathbf{g}_{v_k}^R)\Big) | \max_{i \in [R], j \in [k]} \mathrm{Inf}_{i,\mu_j}^{<d}(p) \leqslant \sqrt{\tau}\Big] - k\sqrt{\tau}$$

Fix $d = (20/\varepsilon)\log(1/\tau)$. For every such p, the polynomial $\boldsymbol{H} = T_{1-\varepsilon}p$, we can conclude that

$$\begin{aligned}
\max_{i \in [R], j \in [k]} \mathrm{Inf}_{i,\mu_j}(\boldsymbol{H}) &\leqslant \max_{i \in [R], j \in [k]} \mathrm{Inf}_{i,\mu_j}^{<d}(T_{1-\varepsilon}p) + (1-\varepsilon)^d \\
&\leqslant \max_{i \in [R], j \in [k]} \mathrm{Inf}_{i,\mu_j}^{<d}(p) + (1-\varepsilon)^d \\
&\leqslant \sqrt{\tau} + (1-\varepsilon)^d \leqslant 2\sqrt{\tau}\,.
\end{aligned}$$

Applying the invariance principle (Theorem 7.15) with the ensembles $\mathcal{L}_C$, $\mathcal{G}_C$, Lipschitz continous functional Ψ and the vector of kq multilinear polynomials given by $(\boldsymbol{H}, \boldsymbol{H}, \ldots, \boldsymbol{H})$ where $\boldsymbol{H} = (H_1, \ldots, H_q)$, we get the required result:

$$\mathsf{Round}_{\mathcal{P}}(\boldsymbol{V}, \boldsymbol{\mu}) \geqslant \mathop{\mathbb{E}}_{p \in \mathcal{P}} \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{\mathcal{L}_C^R} \left[\Psi_C \Big(\boldsymbol{H}(\ell_{v_1}^R), \ldots, \boldsymbol{H}(\ell_{v_k}^R) \Big) \right] - o_\tau(1)$$

where $o_\tau(1)$ is a function that tends to zero as $\tau \to 0$. Since the local ensembles $\mathcal{L}_C$ correspond to the local distributions μ_C over partial assignments $[q]^{\mathcal{V}(C)}$ and $\boldsymbol{H} = T_{1-\varepsilon} p$, we can rewrite the expression as,

$$\mathop{\mathbb{E}}_{p \in \mathcal{P}} \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{\mathcal{L}_C^R} \left[\Psi_C \Big(\boldsymbol{H}(\ell_{v_1}^R), \ldots, \boldsymbol{H}(\ell_{v_k}^R) \Big) \right] = \mathop{\mathbb{E}}_{p \in \mathcal{P}} \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{(X_1, \ldots, X_k) \sim \mu_C^R} \left[\Psi_C \Big(T_{1-\varepsilon} p(X_1), , \ldots, T_{1-\varepsilon} p(X_k) \Big) \right]. \tag{7.4}$$

By definition of Ψ_C, $\Psi_C(\mathbf{z}_1, \ldots, \mathbf{z}_k) = C(\mathbf{z}_1, \ldots, \mathbf{z}_k)$ if $\forall i, \mathbf{z}_i \in \blacktriangle_q$. Summarizing our calculation so far, we have

$$\mathsf{Round}_{\mathcal{P}}(\boldsymbol{V}, \boldsymbol{\mu}) \geqslant \mathop{\mathbb{E}}_{p \in \mathcal{P}} \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{(X_1, \ldots, X_k) \sim \mu_C^R} \left[\Psi_C \Big(T_{1-\varepsilon} p(X_1), , \ldots, T_{1-\varepsilon} p(X_k) \Big) \right] - o_\tau(1)\,.$$

The proof is complete with the following claim which we will prove in the rest of the section.

Claim 7.17.

$$\mathop{\mathbb{E}}_{p \in \mathcal{P}} \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{(X_1, \ldots, X_k) \sim \mu_C^R} \left[\Psi_C \Big(T_{1-\varepsilon} p(X_1), , \ldots, T_{1-\varepsilon} p(X_k) \Big) \right] \geqslant s$$

For each fixed η and $\varepsilon = \eta/k$, the error term $\kappa(\tau, d) \to 0$ as $d \to \infty$ and $\tau \to 0$. Therefore for a sufficiently large d and sufficiently small τ, the error will be smaller than η. □

Proof. (Proof of Claim 7.17) For $X \in [q]^R$, let $Z \sim_{1-\varepsilon} X$ denote the random variable distributed over $[q]^R$ obtained by resampling each coordinate of X from its underlying distribution with probability ε. Notice that $T_{1-\varepsilon} p(X_i) \in \blacktriangle_q$ is a fractional assignment. To sample a value $y \in [q]$ from the $T_{1-\varepsilon} p(X_i)$, we could instead sample $Z_i \sim_{1-\varepsilon} X_i$ and compute $p(Z_i)$. In other words, $y \sim T_{1-\varepsilon} p(X_i)$ is the same as $y = p(Z_i), Z_i \sim_{1-\varepsilon} X_i$. Using the way we defined the payoff function C over fractional assignments, we get

$$\begin{aligned} &\mathop{\mathbb{E}}_{p \in \mathcal{P}} \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{(X_1, \ldots, X_k) \sim \mu_C^R} \left[C \Big(T_{1-\varepsilon} p(X_1), , \ldots, T_{1-\varepsilon} p(X_k) \Big) \right] \\ &\qquad = \mathop{\mathbb{E}}_{p \in \mathcal{P}} \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{(X_1, \ldots, X_k) \sim \mu_C^R} \mathop{\mathbb{E}}_{Z_i \sim_{1-\varepsilon} X_i} \left[C \Big(p(Z_1), \ldots, p(Z_k) \Big) \right] \end{aligned} \tag{7.5}$$

To lower bound the expression, construct an instance $\Im'$ that consists of the constraints in $\Im$ but each over a disjoint set of variables. Consider the distribution Θ for assignments of $\Im'$ wherein the variables corresponding to each constraint $C \in \mathcal{C}$ are sampled independently from μ_C. Let Θ' be the $(1-\varepsilon)$-noisy version of Θ in that it is obtained by sampling from Θ and rerandomizing each coordinate with probability ε.

Notice that

$$\begin{aligned} \mathop{\mathbb{E}}_{Y \sim \Theta'} [\Im'(Y)] &= \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{x_1, \ldots, x_k \sim \mu_C} \mathop{\mathbb{E}}_{z_1, \ldots, z_k \sim_{1-\varepsilon} x_1, \ldots, x_k} [C(x_1, \ldots, x_k)] \\ &\geqslant \mathop{\mathbb{E}}_{C \in \mathcal{C}} \mathop{\mathbb{E}}_{x_1, \ldots, x_k \sim \mu_C} [C(x_1, \ldots, x_k)] - k\varepsilon \\ &\geqslant c - k\varepsilon \end{aligned}$$

where in the last inequality we used the fact that the objective value of the SDP is at least c. Since $\mathcal{P}$ is a $(c-\eta, s)$-approximate polymorphism and $\eta > k\varepsilon$, we can appeal to Lemma 7.11 to conclude that,

$$s \leqslant \mathop{\mathbb{E}}_{p\in\mathcal{P}} \mathop{\mathbb{E}}_{Y^{(1)},\ldots,Y^{(R)}\sim\Theta'} [\Im'(p(Y^{(1)},\ldots,p(Y^{(R)}))]$$

$$\mathop{\mathbb{E}}_{p\in\mathcal{P}} \mathop{\mathbb{E}}_{C\in\mathcal{C}} \mathop{\mathbb{E}}_{(X_1,\ldots,X_k)\sim\mu_C^R} \mathop{\mathbb{E}}_{Z_i\sim_{1-\varepsilon}X_i} \left[C\Big(p(Z_1),\ldots,p(Z_k)\Big)\right] \tag{7.6}$$

The claim is immediate from (7.5) and (7.6). □

Proof. (Proof of Theorem 1.12) By definition of s_Λ, there exists an $(c, s_\Lambda(c))$-approximate polymorphism with degree d influence less than τ for all d, τ. Fix any particular instance $\Im$ of Max-Λ. By applying the rounding scheme $\mathsf{Round}_{\mathcal{P}}$ on a sequence of polymorphisms with influence $\tau \to 0$ and $\eta \to 0$, we get that the SDP integrality gap on $\Im$ is at most $\lim_{\eta\to 0} s_\Lambda(c-\eta)$. □

7.4 Approximate Polymorphisms and Dictatorship Tests

In this section, we will sketch the proof of Theorem 1.11.

Proof. (Proof of Theorem 1.11) For each $\eta, \varepsilon > 0$, there exists τ_0, d_0 such that there are no $(s_\Lambda(c+\varepsilon)+\eta, c+\varepsilon)$-approximate (τ_0, d_0)-quasirandom polymorphisms for Max-Λ. By Theorem 7.18, this implies a unique games hardness to $(c, s_\Lambda(c+\varepsilon)c+\eta)$-approximate Max-$\Lambda$. By making $\varepsilon \to 0$ and $\eta \to 0$, we get a unique games based hardness for $(c, \lim_{\varepsilon\to 0} s_\Lambda(c+\varepsilon))$-approximating Max-$\Lambda$.

By Observation 7.10, we have

$$\lim_{\varepsilon\to 0} s_\Lambda(c+\varepsilon) \leqslant s_\Lambda(c)\,.$$

Hence, the conclusion of Theorem 1.11 follows. □

Theorem 7.18. *Fix constants τ, c and $d \in \mathbb{N}$. Suppose* Max-Λ *does not admit a (c, s)-approximate (τ, d)-quasirandom polymorphism then for all $\eta > 0$, it is unique games hard to $s+\eta$-approximate instances of* Max-Λ *on instances with value at least $c-\eta$.*

Proof. Fix an integer $R \in \mathbb{N}$. For all c, s, the set of (c, s)-approximate polymorphisms $\mathcal{P}$ of arity R form a convex set. Convex combination of two approximate polymorphisms $\mathcal{P}_1$ and $\mathcal{P}_2$ is constructed by taking a convex combination of the underlying distributions over operations. For every (c, s)-approximate polymorphism $\mathcal{P}$, there exists a probability distribution $\mu \in \blacktriangle_q$ such that,

$$\mathop{\mathbb{E}}_{p\in\mathcal{P}} \left[\max_{i\in[R]} \mathrm{Inf}_{i,\mu}^{<d}(p)\right] \geqslant \tau\,.$$

By min-max theorem, there exists a distribution Φ over the simplex $\blacktriangle_q$ such that for every (c, s)-approximate polymorphism $\mathcal{P}$,

$$\mathop{\mathbb{E}}_{p\in\mathcal{P}} \mathop{\mathbb{E}}_{\mu\sim\Phi} \left[\max_{i\in[R]} \mathrm{Inf}_{i,\mu}^{<d}(p)\right] \geqslant \tau \tag{7.7}$$

Define the family of functions $\mathcal{F}_{\tau,d}^{\Phi,R}$ as follows,

$$\mathcal{F}_{\tau,d}^{\Phi,R} \stackrel{\text{def}}{=} \left\{ p : [q]^R \to [q] \;\middle|\; \mathop{\mathbb{E}}_{\mu\sim\Phi} \left[\max_i \mathrm{Inf}_{i,\mu}^{<d}(p)\right] \leqslant \tau \right\}\,.$$

By (7.7), there does not exist (c, s)-approximate polymorphisms supported in $\mathcal{F}_{\tau,d}^{\Phi,R}$. By the connection between approximate polymorphisms and dictatorship tests outlined in Theorem 7.3, this implies that there exists (c, s)-dictatorship tests against the family $\mathcal{F}_{\tau,d}^{\Phi,R}$ for each $R \in \mathbb{N}$. Let $\Im_{\tau,d}^{\Phi,R}$ denote the (c, s)-dictatorship test against the family $\mathcal{F}_{\tau,d}^{\Phi,R}$.

These dictatorship gadgets can be utilized towards carrying out the unique games based hardness reduction. To this end, we will need to show a stronger soundness guarantee for the dictatorship test. Specifically, we will have to show a soundness guarantee against functions that output distributions over $[q]$ (points in $\blacktriangle_q$) instead of elements over $[q]$. The details are described below.

First, we will extend the objective function associated with an instance $\Im$ of MAX-Λ from $[q]$-valued assignments to $\blacktriangle_q$-valued assignments. Given a fractional assignment $z : \mathcal{V} \to \blacktriangle_q$, let $z^{\times}$ denote the product distribution over $[q]^{\mathcal{V}}$ whose marginals given by z. Define $\Im : \blacktriangle_q^{\mathcal{V}} \to \mathbb{R}$ as follows,

$$\Im(z) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{x \sim z^{\times}} [\Im(x)]$$

Notice that the definitions of influences and low-degree influences extend naturally to $\blacktriangle_q$-valued functions. In fact, even for $[q]$-valued functions these notions were defined by expressing them as $\blacktriangle_q$-valued functions. We will show that quasirandom fractional assignments have no larger objective value than quasirandom $[q]$-valued assignments. Specifically, we will show the following.

Lemma 7.19. *For every $\varepsilon > 0$, the following holds for all sufficiently large $R \in \mathbb{N}$. For every function $p : [q]^R \to \blacktriangle_q$ such that $\mathbb{E}_{\mu \sim \Phi}\left[\max_i \mathrm{Inf}_{i,\mu}^{<d}(p)\right] \leqslant \tau/5$,*

$$\Im_{\tau,d}^{\Phi,R}(p) \leqslant s + \varepsilon$$

Proof. The idea is to create a rounded operation $r : [q]^R \to [q]$ by setting for each $x \in [q]^R$,

$$r(x) = \text{ random sample from distribution } p(x)$$

For notational convenience, we will drop the subscripts and superscripts and write $\Im$ for $\Im_{\tau,d}^{\Phi,R}$. It is easy to see that by definition,

$$\mathop{\mathbb{E}}_{r}\left[\Im(r(X^{(1)}, \dots, X^{(R)}))\right] = \Im(p(X^{(1)}, \dots, X^{(R)}))$$

The technical core of the argument shows that if p is *quasirandom* then the rounded function r is *quasirandom* with high probability. This claim is formally stated in Lemma 7.20. By Lemma 7.20, for all sufficiently large $R \in \mathbb{N}$,

$$\mathop{\mathbb{P}}_{r}\left\{\mathop{\mathbb{E}}_{\mu \sim \Phi}[\max_i \mathrm{Inf}_{i,\mu}^{<d}(r)] < \tau\right\} \geqslant 1 - \varepsilon/R\tau$$

Call a rounded function $r : [q]^R \to [q]$ to be *quasirandom*, if $\mathbb{E}_{\mu \sim \Phi}[\max_i \mathrm{Inf}_{i,\mu}^{<d}(r)] < \tau$.

$$\mathop{\mathbb{E}}_{r \sim p}[\Im(r) | r \text{ is quasirandom }] \geqslant \mathop{\mathbb{E}}_{r \sim p}[\Im(r)] - \mathop{\mathbb{P}}_{r}[r \text{ is not quasirandom}] \geqslant \Im(p) - \varepsilon/R\tau$$

For a *quasirandom* function $r : [q]^R \to [q]$, the dictatorship test $\Im$ satisfies $\Im(r) \leqslant s$. This implies that $\Im(p) \leqslant s + \varepsilon$. □

Now, we will outline the details of the unique games hardness reduction. Given a unique games instance $\Gamma = (\mathcal{A}, \mathcal{B}, E, \Pi, [R])$, the reduction produces an instance $\Im_\Gamma$ of MAX-Λ. The variables of $\Im_\Gamma$ are $\mathcal{B} \times [q]^R$. The constraints of $\Im_\Gamma$ can be sampled using the following procedure.

- Sample $a \in \mathcal{A}$ and neighbors $b_1, \dots, b_k \in \mathcal{B}$ independently at random.
- Sample a constraint c from the dictatorship test $\Im_{\tau,d}^{\Phi,R}$. Suppose the constraint is on $x^{(1)}, x^{(2)}, \dots, x^{(k)} \in [q]^R$

- Introduce the constraint C on $\{(b_i, \pi_{ab_i} \circ x^{(i)})\}_{i=1}^k$.

Given an assignment $L : \mathcal{B} \times [q]^R \to [q]$ its value is given by,

$$\Im_\Gamma(L) = \mathop{\mathbb{E}}_{a \in \mathcal{A}} \mathop{\mathbb{E}}_{b_1,\ldots,b_k \in N(a)} \mathop{\mathbb{E}}_{C \in \Im_{\tau,d}^{\Phi,R}} \left[C(L(b_i, \pi_{ab_i} \circ x^{(i)})) \right] \tag{7.8}$$

Completeness. Suppose $\mathcal{L} : \mathcal{A} \cup \mathcal{B} \to [R]$ is a labelling satisfying $(1-\delta)$-fraction of constraints. Consider the labelling $L : \mathcal{B} \times [q]^R \to [q]$ given by $L(b, x) = x_{\mathcal{L}(b)}$. Over the choice of $a, b_1, \ldots, b_k$, with probability at least $(1 - \delta k)$, the labelling $\mathcal{L}$ satisfies each of the k edges $\{(a, b_i) | i \in [k]\}$. In this case, the objective value is at least the completeness c of the dictatorship test $\Im_{\tau,d}^{\Phi,R}$. With the remaining probability, the objective value is at least -1. This implies that the labelling L has an objective value $\geqslant c - 2k\delta$.

Soundness. Suppose $L : \mathcal{B} \times [q]^R \to [q]$ is a labelling with an objective value $s + \eta$. For each $b \in \mathcal{B}$, set $L_b : [q]^R \to [q]$ as $L_b(x) \stackrel{\text{def}}{=} e_{L(b,x)}$ where $e_{L(b,x)} \in \blacktriangle_q$ corresponds to a basis vector along direction $L(b, x)$. For each $a \in \mathcal{A}$, define the fractional assignment $L_a : [q]^R \to \blacktriangle_q$ as $L_a(x) \stackrel{\text{def}}{=} \mathbb{E}_{b \in N(a)}[L_b(\pi_{ab} \circ x)]$.

The objective value in (7.8) can be written as,

$$\Im_\Gamma(L) = \mathop{\mathbb{E}}_{a \in \mathcal{A}} \left[\Im_{\tau,d}^{\Phi,R}(L_a) \right] .$$

Decode an assignment $\ell : \mathcal{A} \cup \mathcal{B} \to [R]$ as follows. Sample a distribution $\mu \sim \Phi$. For each $a \in \mathcal{A}$, define the label set

$$T_a = \{i | \operatorname{Inf}_{i,\mu}^{<d}(L_a) \geqslant \tau/10\}$$

and for each $b \in \mathcal{B}$, define the label set

$$T_b = \{j | \operatorname{Inf}_{j,\mu}^{<d}(L_b) \geqslant \tau/20\} .$$

Since the sum of degree d influences is at most d, we have $|T_a| \leqslant 10d/\tau$ and $|T_b| \leqslant 20d/\tau$ for all $a \in \mathcal{A}$ and $b \in \mathcal{B}$.

For each $a \in \mathcal{A}$, assign $\ell(a)$ to be a random label from T_a if it is non-empty, else assign an arbitrary label from $[R]$. Similarly, for each $b \in \mathcal{B}$, set $\ell(b)$ to be a random label from T_b if it is non-empty, else assign an arbitrary label from $[R]$.

If the objective value of the assignment L is more than $s + \eta$, then for at least $\eta/2$ fraction of $a \in \mathcal{A}$, we have $\Im_{\tau,d}^{\Phi,R}(L_a) > s + \eta/2$. Call such a vertex $a \in \mathcal{A}$ to be *good*. Fix a good vertex $a \in \mathcal{A}$. For every *good* vertex, by the soundness of the dictatorship test we will have,

$$\mathop{\mathbb{E}}_{\mu \sim \Phi}[\max_i \operatorname{Inf}_{i,\mu}^{<d}(L_a)] \geqslant \tau .$$

This implies that over the choice of $\mu \sim \Phi$, with probability at least $\tau/2$, we will have $\max_i \operatorname{Inf}_{i,\mu}^{<d}(L_a) \geqslant \tau/2$ which implies that L_a is non-empty. Suppose $i \in L_a$. Using the fact that $L_a = \mathbb{E}_{b \in N(a)} \pi_{ab} \circ L_b$ and convexity of influences we have,

$$\mathop{\mathbb{E}}_b[\operatorname{Inf}_{\pi_{ab}(i),\mu}^{<d}(L_b)] \geqslant \tau/2 .$$

This implies that for at least $\tau/4$-fraction of neighbors b, we will have $\operatorname{Inf}_{\pi_{ab}(i),\mu}(L_b) \geqslant \tau/4$, i.e., $\pi_{ab}(i) \in T_b$. For every such neighbor b, the labelling ℓ satisfies the edge (a, b) with probability at least $\frac{1}{|T_a|} \cdot \frac{1}{|T_b|} \geqslant \tau^2/200d^2$.

Hence, the fraction of edges satisfied by the labelling ℓ is at least $(\eta/2) \cdot (\tau/2) \cdot (\tau/4) \cdot (\tau^2/200d^2)$ in expectation. By fixing an alphabet size R large enough, the soundness δ of the unique games instance can be made smaller than this fraction.

□

7.5 Quasirandomness under Randomized Rounding

This section is devoted to showing the following lemma which was used in the proof of Theorem 7.18.

Lemma 7.20. *For every τ, d, ε, the following holds for all large enough $R \in \mathbb{N}$: Suppose $p : [q]^R \to \blacktriangle_q$ and a distribution Φ over $\blacktriangle_q$ is such that,*

$$\mathop{\mathbb{E}}_{\mu \sim \Phi}[\max_i \mathrm{Inf}_{i,\mu}^{<d}(p)] < \tau$$

then if $r : [q]^R \to [q]$ is sampled by setting each $r(x) \in [q]$ independently from the distribution $p(x)$,

$$\mathop{\mathbb{P}}_r \left\{ \mathop{\mathbb{E}}_{\mu \sim \Phi}[\max_i \mathrm{Inf}_{i,\mu}^{<d}(r)] < 5\tau \right\} \geqslant 1 - \varepsilon / R\tau$$

To this end, we first show a few simple facts concerning concentration under random sampling.

Random Sampling.

Lemma 7.21. *Given $p : [q]^R \to \blacktriangle_q$ let $r : [q]^R \to [q]$ denote be a function sampled by setting $r(x) \in [q]$ from the distribution $p(x) \in \blacktriangle_q$. Fix a probability distribution $\mu \in \blacktriangle_q$ and a corresponding orthonormal basis $\{\chi_\sigma\}_{\sigma \in [q]^R}$ for $L^2([q]^R, \mu^R)$. For every multi-index $\sigma \in [q]^R$,*

$$\mathop{\mathbb{E}}_R[\hat{r}_\sigma] = \hat{p}_\sigma \,,$$

and

$$\mathop{\mathbb{P}}_r[|\hat{r}_\sigma - \hat{p}_\sigma| > \beta] \leqslant 2\exp\left(-\frac{\beta^2}{\left(\max_{i \in [q]} \mu(i)\right)^R} \right) .$$

Proof. Fix a multi-index $\sigma \in [q]^R$. Let $\chi_\sigma : [q]^R \to \mathbb{R}$ denote the corresponding basis function for $L^2([q]^R, \mu^R)$. Let $\mu^R(x)$ denote the probability of $x \in [q]^R$ under the product distribution μ^R. By definition,

$$\hat{r}_\sigma = \sum_{x \in [q]^n} \mu(x) \chi_\sigma(x) r(x)$$

where $\{r(x)\}_{x \in [q]^R}$ are independent random variables whose expectation is given by $\mathbb{E}\, r(x) = p(x)$. For every x, the random vector $\mu(x)\chi_\sigma(x) r(x)$ has entries bounded in $[0, \mu(x)\chi_\sigma(x)]$. Now we appeal to Hoeffding's inequality stated below.

Lemma 7.22. *(Hoeffding's inequality) Given independent real valued random variables $X_1, \ldots, X_n$ such that $X_i \in [b_i, a_i]$, we have*

$$\mathbb{P}\left[|\sum_i X_i - \mathbb{E}[\sum_i X_i]| \geqslant t \right] \leqslant 2\exp\left(-\frac{t^2}{\sum_i (b_i - a_i)^2} \right)$$

By Hoeffding's inequality, we will have

$$\mathop{\mathbb{P}}_r[|\hat{r}_\sigma - \hat{p}_\sigma| > \beta] \leqslant 2\exp\left(-\frac{\beta^2}{\sum_x \mu(x)^2 \chi_\sigma(x)^2} \right) .$$

The proof is complete once we observe that $\sum_x \mu(x)^2 \chi_\sigma(x)^2 \leqslant \max_x \mu(x) \cdot \sum_x \mu(x) \chi_\sigma^2(x) = \max_x \mu(x) \leqslant (\max_i \mu(i))^R$.

$\square$

Lemma 7.23. *Given a function $r : [q]^R \to \mathbb{R}^d$ and a probability distribution $\mu \in \blacktriangle_q$, for all $i \in [R]$ and $D \in \mathbb{N}$,*

$$\mathrm{Inf}_{i,\mu}^{<D}(r) \leqslant 8 \left(1 - \max_{i \in [q]} \mu(i)\right) \cdot \max_x \|r(x)\|^2$$

Proof. First, note that for each $D \in \mathbb{N}$ we have $\mathrm{Inf}_{i,\mu}^{<D}(r) \leqslant \mathrm{Inf}_{i,\mu}(r)$. Moreover, the influence of the i^{th} coordinate can be written as,

$$\mathrm{Inf}_{i,\mu}(r) = \mathop{\mathbb{E}}_{x_{[n]\setminus i} \sim \mu^{R-1}} \left[\mathop{\mathbf{Var}}_{x_i \sim \mu} r(x) \right] . \tag{7.9}$$

For each fixing of $x_{[n]\setminus i} \in [q]^{R-1}$, we will bound the variance of $r(x)$ over the choice of x_i as follows,

$$\begin{aligned} \mathop{\mathbf{Var}}_{x_i \sim \mu}[r(x)] &= \mathop{\mathbb{E}}_{z,z' \sim \mu} \left[\|r(x_{[n]\setminus i}, z) - r(x_{[n]\setminus i}, z')\|^2 \right] \\ &\leqslant 4 \max_{x \in [q]^R} \|r(x)\|^2 \cdot \mathbb{P}[z \neq z'] \\ &\leqslant 8 \left(\max_{x \in [q]^R} \|r(x)\|^2 \right) \cdot \left(1 - \max_i \mu(i)\right) . \end{aligned}$$

The result follows by using the above bound in (7.9). ☐

Lemma 7.24. *For every τ, d, ε, the following holds for all large enough $R \in \mathbb{N}$: Given $p : [q]^R \to \blacktriangle_q$ let $r : [q]^R \to [q]$ denote a function sampled by setting $r(x) \in [q]$ from the distribution $p(x) \in \blacktriangle_q$. For every distribution $\mu \in \blacktriangle_q$ and $i \in [R]$,*

$$\mathbb{P}[\mathrm{Inf}_{i,\mu}^{<d}(r) > 2\,\mathrm{Inf}_{i,\mu}^{<d}(p) + \tau] < \varepsilon/R^2$$

Proof. By Lemma 7.23, if $\max_{j \in [q]} \mu(j) > 1 - \tau/16$ then we will have $\mathrm{Inf}_{i,\mu}^{<d}(r) < \tau$ for all functions $r : [q]^R \to [q]$. Hence the statement trivially holds if $\max_{j \in [q]} \mu(j) > 1 - \tau/16$.

Suppose $\max_{j \in [q]} \mu(j) \leqslant 1 - \tau/8$. In this case, for every multi-index $\sigma \in [q]^R$, Lemma 7.21 implies that

$$\mathbb{P}[|\hat{r}_\sigma - \hat{p}_\sigma| \geqslant \beta] \leqslant 2 \exp\left(- \frac{\beta^2}{(1 - \tau/8)^R} \right) .$$

By a union bound over all σ with $|\sigma| < d$, we will have

$$\mathbb{P}[|\hat{r}_\sigma - \hat{p}_\sigma| \geqslant \beta \quad \forall |\sigma| < d] \leqslant 2(qR)^d \exp\left(- \frac{\beta^2}{(1 - \tau/8)^R} \right) .$$

Fix $\beta = (qR)^{-d}$. Fix R large enough so that $1/(qR)^d < \tau$ and the above probability bound is smaller than ε/R^2. If $|\hat{r}_\sigma - \hat{p}_\sigma| \leqslant \beta$ for all σ with $|\sigma| < d$, then

$$\mathrm{Inf}_{i,\mu}(r) \leqslant \sum_{\sigma_i \neq 0, |\sigma| < d} (\hat{p}_\sigma + \beta)^2 \leqslant \sum_{\sigma_i \neq 0, |\sigma| < d} 2\hat{p}_\sigma^2 + 2\beta^2 \leqslant 2\,\mathrm{Inf}_{i,\mu}^{<d}(p) + 2(qR)^{-2d}(qR)^d \leqslant 2\,\mathrm{Inf}_{i,\mu}^{<d}(p) + \tau$$

☐

Now we are ready to prove Lemma 7.20.

Proof. (Proof of Lemma 7.20) Fix a distribution $\mu \sim \Phi$. Call a probability distribution $\mu \sim \Phi$ to be *bad* if

$$\max_i \mathrm{Inf}_{i,\mu}^{<d}(r) > 2 \max_i \mathrm{Inf}_{i,\mu}^{<d}(p) + \tau$$

For each distribution $\mu \in \blacktriangle_q$, by Lemma 7.24 and a union bound over $i \in [R]$

$$\mathop{\mathbb{P}}_r[\mu \text{ is bad}] \leqslant \varepsilon/R$$

By Markov's bound,

$$\mathop{\mathbb{P}}_r\left\{\mathop{\mathbb{E}}_{\mu\sim\Phi}\mathbb{I}[\mu \text{ is bad}] > \tau\right\} < \varepsilon/R\tau \tag{7.10}$$

Let us suppose $\mathbb{E}_{\mu\sim\Phi}[\mathbb{I}[\mu \text{ is bad}]] < \tau$. In this case, we conclude that,

$$\mathop{\mathbb{E}}_{\mu\sim\Phi}[\max_i \mathrm{Inf}^{<d}_{i,\mu}(r)] \leqslant \left(2\mathop{\mathbb{E}}_{\mu\sim\Phi}[\max_i \mathrm{Inf}^{<d}_{i,\mu}(p)] + \tau\right) + \tau\cdot 2 \tag{7.11}$$

where we used the fact that for every distribution $\mu \in \blacktriangle_q$ and every function $r : [q]^R \to [q]$ we have $\max_i \mathrm{Inf}^{<d}_{i,\mu}(r) < 2$. The claim follows from (7.10) and (7.11). □

References

[BD06] Andrei A. Bulatov and Víctor Dalmau, *A simple algorithm for mal'tsev constraints*, SIAM J. Comput. **36** (2006), no. 1, 16–27.

[BK09] Libor Barto and Marcin Kozik, *Constraint satisfaction problems of bounded width*, FOCS, 2009, pp. 595–603.

[BK12] ———, *Absorbing subalgebras, cyclic terms, and the constraint satisfaction problem*, Logical Methods in Computer Science **8** (2012), no. 1.

[BKJ00] Andrei A. Bulatov, Andrei A. Krokhin, and Peter Jeavons, *Constraint satisfaction problems and finite algebras*, ICALP, 2000, pp. 272–282.

[BKS14] Boaz Barak, Jonathan A. Kelner, and David Steurer, *Rounding sum-of-squares relaxations*, Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014, 2014, pp. 31–40.

[Bul03] Andrei A. Bulatov, *Tractable conservative constraint satisfaction problems*, LICS, 2003, pp. 321–.

[Bul06] ———, *A dichotomy theorem for constraint satisfaction problems on a 3-element set*, J. ACM **53** (2006), no. 1, 66–120.

[CCC+13] David A. Cohen, Martin C. Cooper, Páidí Creed, Peter G. Jeavons, and Stanislav Zivny, *An algebraic theory of complexity for discrete optimization*, SIAM J. Comput. **42** (2013), no. 5, 1915–1939.

[CCJK06] Cohen, Cooper, Jeavons, and Krokhin, *The complexity of soft constraint satisfaction*, AIJ: Artificial Intelligence **170** (2006).

[CK88] R. Chandrasekaran and Santosh N. Kabadi, *Pseudomatroids*, Discrete Mathematics **71** (1988), no. 3, 205–217.

[Cun81] William H. Cunningham, *On matroid connectivity*, J. Comb. Theory, Ser. B **30** (1981), no. 1, 94–99.

[Cun83] ———, *Decompostiton of submodular functions*, Combinatorica **3** (1983), no. 1, 53–68.

[Cun84] ______, *Testing membership in matroid polyhedra*, J. Comb. Theory, Ser. B **36** (1984), no. 2, 161–188.

[DMP06] Cohen D., Cooper M., and Jeavons P., *An algebraic characterisation of complexity for valued constraints*, Proceedings 12th International Conference on Constraint Programming—CP'06, Lecture Notes in Computer Science, Springer-Verlag, 2006.

[FI05] Satoru Fujishige and Satoru Iwata, *Bisubmodular function minimization*, SIAM J. Discrete Math. **19** (2005), no. 4, 1065–1073.

[FM00] Satoru Fujishige and Kazuo Murota, *Notes on l-/m-convex functions and the separation theorems*, Math. Program. **88** (2000), no. 1, 129–146.

[FV98] 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** (1998), no. 1, 57–104.

[HK12] Anna Huber and Vladimir Kolmogorov, *Towards minimizing k-submodular functions*, Combinatorial Optimization - Second International Symposium, ISCO 2012, Athens, Greece, April 19-21, 2012, Revised Selected Papers, 2012, pp. 451–462.

[IFF01] Satoru Iwata, Lisa Fleischer, and Satoru Fujishige, *A combinatorial strongly polynomial algorithm for minimizing submodular functions*, J. ACM **48** (2001), no. 4, 761–777.

[IM09] Marcus Isaksson and Elchanan Mossel, *Maximally stable gaussian partitions with discrete applications*, August 03 2009.

[Kho02] Subhash Khot, *On the power of unique 2-prover 1-round games*, STOC, 2002, pp. 767–775.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell, *Optimal inapproximability results for max-cut and other 2-variable csps?*, SIAM J. Comput. **37** (2007), no. 1, 319–357.

[KL08] Andrei A. Krokhin and Benoit Larose, *Maximizing supermodular functions on product lattices, with application to maximum constraint satisfaction*, SIAM J. Discrete Math **22** (2008), no. 1, 312–328.

[Kol10] Vladimir Kolmogorov, *Submodularity on a tree: Unifying $L^\natural$-convex and bisubmodular functions*, CoRR **abs/1007.1229** (2010).

[KS09] Gábor Kun and Mario Szegedy, *A new line of attack on the dichotomy conjecture*, STOC, 2009, pp. 725–734.

[Kui11] Fredrik Kuivinen, *On the complexity of submodular function minimisation on diamonds*, Discrete Optimization **8** (2011), no. 3, 459–477.

[MM08] Miklós Maróti and Ralph McKenzie, *Existence theorems for weakly symmetric operations*, Algebra Universalis **59** (2008), 463–489, 10.1007/s00012-008-2122-9.

[MOO05] Elchannan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz, *Noise stability of functions with low influences: Invariance and optimality*, FOCS: IEEE Symposium on Foundations of Computer Science (FOCS), 2005.

[Mos08] 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).

[Mur04] Kazuo Murota, *On steepest descent algorithms for discrete convex functions*, SIAM Journal on Optimization **14** (2004), no. 3, 699–707.

[NES11] Yurii NESTEROV, *Random gradient-free minimization of convex functions*, CORE Discussion Papers 2011001, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), 2011.

[O'D14] Ryan O'Donnell, *Analysis of boolean functions*, Cambridge University Press, 2014.

[Qi88] Liqun Qi, *Directed submodularity, ditroids and directed submodular flows*, Mathematical Programming (1988), no. 1-3, 579–599.

[Rag08] Prasad Raghavendra, *Optimal algorithms and inapproximability results for every csp?*, In STOC '08: Proceedings of the 40th ACM Symposium on Theory of Computing (2008), 245–254.

[Sch78] Thomas J. Schaefer, *The complexity of satisfiability problems*, STOC, 1978, pp. 216–226.

[Sch00] Alexander Schrijver, *A combinatorial algorithm minimizing submodular functions in strongly polynomial time*, J. Comb. Theory, Ser. B **80** (2000), no. 2, 346–355.

[SGB12] Ajit P. Singh, Andrew Guillory, and Jeff Bilmes, *On bisubmodular maximization*, Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2012, La Palma, Canary Islands, April 21-23, 2012, 2012, pp. 1055–1063.

[Spa97] James C. Spall, *A one-measurement form of simultaneous perturbation stochastic approximation*, Automatica **33** (1997), no. 1, 109–112.

[TZ13] Johan Thapper and Stanislav Zivny, *The complexity of finite-valued csps*, Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013, 2013, pp. 695–704.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de