Satisfiability of Ordering CSPs Above Average Is Fixed-Parameter Tractable

Konstantin Makarychev Yury Makarychev* Yuan Zhou
Microsoft Research TTIC MIT

Abstract

We study the satisfiability of ordering constraint satisfaction problems (CSPs) above average. We show that for every k, the satisfiability above average of ordering CSPs of arity at most k is fixed-parameter tractable. We generalize this result to more general classes of CSPs, including CSPs with predicates defined by linear equations.

---

1 Introduction

In this paper, we study satisfiability of ordering constraint satisfaction problems (CSPs) above the average value. An ordering k-CSP is defined by a set of variables $V = \{x_1, \ldots, x_n\}$ and a set of constraints Π. Each constraint $\pi \in \Pi$ is a disjunction of clauses of the form $x_{\tau_1} < x_{\tau_2} < \cdots < x_{\tau_r}$ for some distinct variables $x_{\tau_1}, \ldots, x_{\tau_r}$ from a k-element subset $V_\pi \subset V$. A linear ordering α of variables $x_1, \ldots, x_n$ satisfies a constraint π if one of the clauses in the disjunction agrees with the linear ordering α. The goal is to find an assignment α that maximizes the number of satisfied constraints.

A classical example of an ordering CSP is the Maximum Acyclic Subgraph problem, in which constraints are of the form "$x_i < x_j$" (the problem has arity 2). Another well-known example is the Betweenness problem, in which constraints are of the form "$(x_i < x_j < x_k)$ or $(x_k < x_j < x_i)$" (the problem has arity 3). Both problems are NP-hard and cannot be solved exactly in polynomial–time unless $P = NP$ [20, 25].

There is a trivial approximation algorithm for ordering CSPs as well as other constraint satisfaction problems: output a random linear ordering of variables $x_1, \ldots, x_n$ (chosen uniformly among all $n!$ linear orderings). Say, if each constraint is just a clause on k variables, this algorithm satisfies each clause with probability $1/k!$ and thus satisfies a $1/k!$ fraction of all constraints in expectation. In 1997, Håstad [19] showed that for some regular (i.e., non-ordering) constraint satisfaction problems the best approximation algorithm is the *random assignment* algorithm. His work raised the following question: for which CSPs are there non-trivial approximation algorithms and for which CSPs is the best approximation algorithm the random assignment? This question has been extensively studied in the literature. Today, there are many known classes of constraint satisfaction problems that do not admit non-trivial approximations assuming the Unique Games or $P \neq NP$ conjectures (see e.g [19, 3, 13, 5]). There are also many constraint satisfaction problems for which we know non-trivial approximation algorithms. Surprisingly, the situation is very different for ordering CSPs: Guruswami, Håstad, Manokaran, Raghavendra, and Charikar [12] showed that *all* ordering k-CSPs do not admit non-trivial approximation assuming the Unique Games Conjecture.

A similar question has been actively studied from the fixed-parameter tractability perspective[1] [1, 7, 8, 17, 21, 22, 26]: Given an instance of a CSP, can we decide whether $OPT \geq AVG + t$ for a fixed parameter t? Here, OPT is the value of the optimal solution for the instance, and AVG is the expected value on a random assignment. In 2011, Alon, Gutin, Kim, Szeider, and Yeo [1] gave the affirmative answer to this question for all (regular) k-CSPs with a constant size alphabet. In [17, 15, 16], Gutin et al. extended this result to 2-arity and 3-arity ordering CSPs. However, the general case of k-arity ordering CSPs has remained open. Below we state the problem formally.

Definition 1.1 (Satisfiability of Ordering CSP Above Average)**.** *Consider an instance $\mathcal{I}$ of arity k and a parameter t. Let $OPT = OPT(\mathcal{I})$ be the number of the constraints satisfied by the optimal solution, and $AVG = AVG(\mathcal{I})$ be the number of constraints satisfied in expectation by a random solution. We need to decide whether $OPT \geq AVG + t$.*

Definition 1.2. *A problem with a parameter t is fixed-parameter tractable if there exists an algorithm for the problem with running time $g(t)poly(n)$, where $g(t)$ is an arbitrary function of t, $poly$ is a fixed polynomial (independent of t), and n is the size of the input.*

In this paper, we prove that satisfiability above average of any ordering CSP of any arity k is fixed-parameter tractable.

Theorem 1.3. *There exists a deterministic algorithm that given an instance $\mathcal{I}$ of an ordering k-CSP on n variables and a parameter t, decides whether $OPT(\mathcal{I}) \geq AVG(\mathcal{I}) + t$ in time $g(t)poly_k(n)$, where g is a function of t, $poly_k$ is a polynomial of n (g and $poly_k$ depend on k). If $OPT(\mathcal{I}) \geq AVG(\mathcal{I}) + t$, then the algorithm also outputs an assignment satisfying at least $AVG(\mathcal{I}) + t$ constraints.*

---

[1]We refer the reader to an excellent survey of results in this area by Gutin and Yeo [18].

Furthermore, we prove that the problem has a *kernel* of size $O_k(t^2)$.

Techniques. Let us examine approaches used previously for ordering CSPs. The algorithms of Gutin et al. [17, 15, 16] work by applying a carefully chosen set of reduction rules to ordering CSPs of arity 2 and 3. These rules heavily depend on the structure of 2 and 3 CSPs. Unfortunately, the structure of ordering CSPs of higher arities is substantially more complex. Here is a quote from [15]: *"it appears technically very difficult to extend results obtained for arities $r = 2$ and 3 to $r > 3$."* In this paper, we do not use such reductions.

The papers [6, 14, 23] use an alternative approach to get an advantage over the random assignment for special families of ordering CSPs. They first reduce the ordering k-CSP to a regular k-CSP with a constant size alphabet, and then work with the obtained regular k-CSP. However, this reduction, generally, does not preserve the value of the CSP. So if for the original ordering CSP instance $\mathcal{I}$ we have $OPT(\mathcal{I}) \geq AVG(\mathcal{I}) + t$, then for the new instance $\mathcal{I}'$ we may have $OPT(\mathcal{I}') \ll AVG(\mathcal{I}) + t$ (we note that $AVG(\mathcal{I}') = AVG(\mathcal{I})$). In this paper, we do not use this reduction either.

Instead, we treat all ordering CSPs as CSPs with the continuous domain: Our goal is to arrange all variables on the interval $[-1, 1]$ so as maximize the number of satisfied constraints. The arrangement of variables uniquely determines their order. Moreover, if we independently assign random values from $[-1, 1]$ to variables x_i, then the induced ordering on x_i's will be uniformly distributed among all $n!$ possible orderings. Thus, our reduction preserves the values of OPT and AVG. However, we can no longer apply Fourier analytic tools used previously in [1, 14, 23]. We cannot use the (standard) Fourier analysis on $[-1, 1]^n$, since we have no control over the Fourier coefficients of the functions we need to analyze. Instead, we work with the Efron—Stein decomposition [10] (see Sections 2 and 3.3). We show that all terms in the Efron—Stein decomposition have a special form. We use this fact to prove that an ordering k-CSP that depends on many variables must have a large variance. Specifically, we show that if a k-CSP instance depends on $C_k t^2$ variables, then the standard deviation of its value from the mean (on a random assignment) is greater than $c_k t$ (for some C_k and $c_k \gg 1$). As is, this claim does not imply that $OPT \geq AVG + t$ since for some assignments the value may be substantially less than $AVG - t$. To finish the proof of the main result, we prove a new hypercontractive inequality, which is an analog of the Bonami Lemma [4]. This inequality is one of the main technical contributions of our paper.

Theorem 1.4 (Bonami Lemma for Efron—Stein Decomposition). *Consider $f \in L_2(\Omega^n, \mu^n)$. Let $f = \sum_S f_S$ be the Efron—Stein decomposition of f. Denote the degree of the decomposition by d. Assume that for every S_1, S_2, S_3, S_4,*

$$\mathbb{E}\left[f_{S_1} f_{S_2} f_{S_3} f_{S_4}\right] \leq C \left(\mathbb{E}\left[f_{S_1}^2\right] \mathbb{E}\left[f_{S_2}^2\right] \mathbb{E}\left[f_{S_3}^2\right] \mathbb{E}\left[f_{S_4}^2\right]\right)^{1/2}. \tag{1}$$

Then

$$\|f\|_4^4 \equiv \mathbb{E}\left[f(X_1, \ldots, X_n)^4\right] \leq 81^d C \|f\|_2^4 \equiv 81^d C \mathbb{E}\left[f(X_1, \ldots, X_n)^2\right]^2. \tag{2}$$

We note that hypercontractive inequalities have been extensively studied under various settings (see e.g., [27, 9, 29, 24]). However, all of them depend on the *mass of the smallest atom in the probability space*. In our case, the smallest atom is polynomially small in n, which is why we cannot apply known hypercontractive inequalities. This is also the reason why we need an extra condition (1) on the function f. Condition (1) is a "local" condition in the sense that all expectations in (1) are over at most $4d$ variables for every $S_1, \ldots, S_4$. Consequently, as we will see below, it is very easy to verify that it holds in many cases (in contrast to (2), which is very difficult to verify directly). Note also that condition (1) is necessary — if it is not satisfied, then the ratio $\|f\|_4 / \|f\|_2$ can be arbitrarily large even for $d = 1$.

Extensions. Once we assume that the domain of every variable is the interval $[-1, 1]$, we might be tempted to write more complex constraints than before such as *"the average of x_1, x_2 and x_3 is at most x_4"*, or

2

"x_1 *lies to the left of the midpoint between* x_2 *and* x_3", or "x_1 *is closer to* x_2 *than to* x_3". Each of these constraints can be written as a system of linear inequalities or a disjunction of clauses, each of which is given by a system of linear inequalities. For instance, "x_1 *lies to the left of the midpoint between* x_2 *and* x_3" can be written as $2x_1 - x_2 - x_3 < 0$. In Appendix E, we extend our results to CSPs in which every constraint is a disjunction of clauses, each of which is a "small" linear program (LP). Namely, each constraint should have arity at most k, only variables that a constraint depends on should appear in the LPs that define it, and all LP coefficients must be integers in the range $\{-b, \ldots, b\}$ (for a fixed b). We call this new class of CSPs (k, b)-LP CSPs.

Definition 1.5. *A (k, b)-LP CSP is defined by a set of variables $V = \{x_1, \ldots, x_n\}$ taking values in the interval $[-1, 1]$ and a set of constraints Π. Each constraint $\pi \in \Pi$ is a disjunction of clauses of the form $Ax < c$, where A is a matrix with integer coefficients in the range $[-b, b]$; c is a vector with integer coefficients in the range $[-b, b]$; the indices of non-zero columns of the matrix A lie in the set V_π of size k (the set V_π is the same for all clauses in π). The goal is to assign distinct real values to variables x_i so to maximize the number of satisfied constraints.*

In fact, we extend our results to a much more general class of valued CSPs – all CSPs whose predicates lie in a lattice of functions with some natural properties (see Appendices D and E for details); but we believe that the subclass of (k, b)-LP CSPs is the most natural example of CSPs in the class. Observe that every ordering k-CSP is a $(k, 1)$-LP CSP since we can write each clause $x_1 < x_2 < \cdots < x_k$ as the system of linear equations $x_i - x_{i+1} < 0$ for $i \in \{1, \ldots, k-1\}$. Similarly, every k-CSP on a finite domain $\{1, \ldots, d\}$ is equivalent to a (k, d)-LP CSP. The reduction works as follows: We break the interval $[-1, 1]$ into d equal subintervals $((2j - d - 2)/d, (2j - d)/d)$ and map every value j to the j-th interval. Then, we replace every condition $x_i = j$ with the equation $x_i \in ((2j - d - 2)/d, (2j - d)/d)$ which can be written as $-dx_i < -(2j - d - 2)$ and $dx_i < (2j - d)$.

Overview. In the next section we give an informal overview of the proof. We formally define the problem and describe the Efron—Stein decomposition in Section 3. Then, in Section 4, we prove several claims about the Efron—Stein decomposition of ordering CSPs. We derive the main results (Theorem 1.3 and Theorem 3.3) in Section 5. Finally, we prove the Bonami Lemma for the Efron—Stein decomposition in Section 6. We generalize our results to all CSPs with a lattice structure in Appendix D and show that (k, b)-LP CSPs (as well as more general "piecewise polynomial" CSPs) have a lattice structure in Appendix E.

2 Proof Overview

Our high-level approach is similar to that developed by Alon et al. [1] and Gutin et al. [17, 16, 15]. As in [17, 16, 15], we design an algorithm that given an instance $\mathcal{I}$ of an ordering CSP does the following:

- It either finds a kernel (another instance of the ordering CSP) $\mathcal{K}$ on $O(t^2)$ variables such that $OPT(\mathcal{I}) = OPT(\mathcal{K})$ and $AVG(\mathcal{I}) = AVG(\mathcal{K})$. Then we can decide whether $OPT \geq AVG + t$ by trying out all possible permutations of variables that $\mathcal{K}$ depends on in time $\exp(O(t^2 \log t))$.

- Or it certifies that $OPT(\mathcal{I}) \geq AVG(\mathcal{I}) + t$.

To this end, we show that either $\mathcal{I}$ depends on at most $O(t^2)$ variables or the variance of $\mathrm{val}_{\mathcal{I}}(\alpha)$ is $\Omega(t^2)$ (where α is chosen uniformly at random). In the former case, the restriction of $\mathcal{I}$ to the variables it depends on is the desired kernel of size $O(t^2)$. In the latter case, $OPT \geq AVG + t$. Though our approach resembles that of [17, 16, 15] at the high level, we employ very different techniques to prove our results.

We extensively use Fourier analysis and, specifically, the Efron—Stein decomposition. Fourier analysis is a very powerful tool, which works especially well with product spaces. The space of feasible solutions of

an ordering CSP is not, however, a product space — it is a discrete space that consists of $n!$ linear orderings of variables $x_1, \ldots, x_n$. To overcome this problem, we define "continuous solutions" for an ordering CSP (see Section 3.2). A solution is an assignment of real values in $[-1, 1]$ to variables $x_1, \ldots, x_n$; that is, it is a point in the product space $[-1, 1]^n$. Each continuous solution defines a combinatorial solution α in a natural way: α orders variables $x_1, \ldots, x_n$ according to the values assigned to them (e.g., if we assign values -0.5, -0.9 and 0.5 to x_1, x_2 and x_3 then $x_2 < x_1 < x_3$ according to α). Thus we get an optimization problem over the product space $[-1, 1]^n$. Denote by $\Phi : [-1, 1]^n \to \mathbb{R}$ its objective function. We consider the Efron—Stein decomposition of Φ: $\Phi = \sum_{S:|S|\leq k} \Phi_S$ (see Section 3.3). Here, informally, Φ_S is the part of Φ that depends on variables x_i with $i \in S$. All functions Φ_S are uncorrelated: $\mathbb{E}\left[\Phi_{S_1}\Phi_{S_2}\right] = 0$ for $S_1 \neq S_2$. We show that each Φ_S is either identically equal to 0 or has variance greater than some positive number, which depends only on k (see Section 4, Lemma 4.4). We now consider two cases.

I. In the first case, there are at most $\Theta_k(t^2)$ terms Φ_S not equal to 0. Using that Φ_S depends only on variables x_i with $i \in \bigcup_{\Phi_S \neq 0} S$ and that there are at most $\Theta_k(t^2)$ sets S such that $\Phi_S \neq 0$, we get that Φ depends on at most $O_k(t^2)$ variables and we are done.

II. In the second case, there are at least $c_k t^2$ terms Φ_S not equal to 0. Since the variance of each term Φ_S is $\Theta_k(1)$ and all terms Φ_S are uncorrelated, the variance of Φ is at least $\Theta_k(t^2)$ (see Theorem 4.5). Therefore, Φ deviates from $AVG = \mathbb{E}\left[\Phi\right]$ by at least $\sqrt{\operatorname{Var}\Phi} = \Theta(t)$. We then show that $\Phi - \mathbb{E}\Phi$ satisfies the conditions of Theorem 1.4 (see Lemma A.1) and the degree of the decomposition is at most k. Thus, the ratio $\|\Phi - \mathbb{E}\Phi\|_4/\|\Phi - \mathbb{E}\Phi\|_2$ is bounded by $O_k(1)$. This implies that $\Pr(\Phi - \mathbb{E}\Phi > t) > 0$. Hence, $OPT \geq AVG + t$.

3 Preliminaries

3.1 Ordering CSP

Consider a set of variables $V = \{x_1, \ldots, x_n\}$. An ordering constraint π on a subset of variables $x_{i_1}, \ldots, x_{i_k}$ is a set of linear orderings of $x_{i_1}, \ldots, x_{i_k}$. A linear ordering α of V satisfies a constraint π on $x_{i_1}, \ldots, x_{i_k}$ if the restriction of α to $x_{i_1}, \ldots, x_{i_k}$ is in π. We say that π depends on variables $x_{i_1}, \ldots, x_{i_k}$.

Definition 3.1. *An instance $\mathcal{I}$ of an ordering constraint satisfaction problem consists of a set of variables $V = \{x_1, \ldots, x_n\}$ and a set of constraints Π; each constraint $\pi \in P$ depends on some subset of variables. A feasible solution to $\mathcal{I}$ is a linear ordering of variables $x_1, \ldots, x_n$. The value $\operatorname{val}(\alpha) = \operatorname{val}_{\mathcal{I}}(\alpha)$ of a solution α is the number of constraints in Π that α satisfies. The goal of the problem is to find a solution of maximum possible value.*

We denote the value of the optimal solution by OPT: $OPT = \max_\alpha \operatorname{val}(\alpha)$. The average value AVG of an instance is the expected value of a solution chosen uniformly at random among $n!$ feasible solutions: $AVG = \mathbb{E}_\alpha[\operatorname{val}(\alpha)]$. We say that $\mathcal{I}$ has arity k if each constraint in $\mathcal{I}$ depends on at most k variables.

Definition 3.2. *In the Satisfiability Above Average Problem, we are given an instance of an ordering CSP of arity k and a parameter t. We need to decide if there is a solution α that satisfies at least $AVG + t$ constraints, or, in other words, if $OPT \geq AVG + t$.*

In this paper, we show that this problem is fixed-parameter tractable. To this end, we design an algorithm that either finds a kernel on $O(t^2)$ variables or certifies that $OPT \geq AVG + t$.

Theorem 3.3. *There is an algorithm that given an instance of an ordering CSP problem of arity k and a parameter t, either finds a kernel on at most $\kappa_k t^2$ variables (where constant κ_k depends only on k) or certifies that $OPT \geq AVG + t$. The algorithm runs in time $O_k(m + n)$ linear in the number of constraints m and variables n (the coefficient in the O-notation depends on k).*

3.2 Ordering CSPs over $[-1, 1]^n$

Consider an instance $\mathcal{I}$ of an ordering CSP on variables $x_1, \ldots, x_n$. Let us say that a continuous feasible solution to $\mathcal{I}$ is an assignment of distinct values $\hat{x}_1, \ldots, \hat{x}_n \in [-1, 1]$ to variables $x_1, \ldots, x_n$. Each continuous solution $\hat{x}_1, \ldots, \hat{x}_n$ defines an ordering α of variables x_i: x_a is less then x_b with respect to α if and only if $\hat{x}_a < \hat{x}_b$. We define the value of a continuous solution $\hat{x}_1, \ldots, \hat{x}_n$ as the value of the corresponding solution (linear ordering) α. We will denote the value of solution $\hat{x}_1, \ldots, \hat{x}_n$ by $\Phi(\hat{x}_1, \ldots, \hat{x}_n)$.

Note that if we sample a continuous solution $\hat{x}_1, \ldots, \hat{x}_n$ uniformly at random, by choosing values $\hat{x}_i$ independently and uniformly from $[-1, 1]$, the corresponding solution α will be uniformly distributed among $n!$ feasible solutions. Therefore,

$$OPT = \max_{\hat{x}_1, \ldots, \hat{x}_n \in [-1, 1]} \Phi(\hat{x}_1, \ldots, \hat{x}_n) \quad \text{and} \quad AVG = \mathbb{E}_{\hat{x}_1, \ldots, \hat{x}_n \in [-1, 1]} \Phi(\hat{x}_1, \ldots, \hat{x}_n).$$

Note that all $\hat{x}_i$ are distinct a.s. and thus a random point in $[-1, 1]^n$ is a feasible continuous solution a.s.

3.3 Efron—Stein Decomposition

The main technical tool in this paper is the Efron—Stein decomposition. We refer the reader to [11, Section 8.3] for a detailed description of the decomposition. Now, we just remind its definition and basic properties.

The Efron—Stein decomposition can be seen as a generalization of the Fourier expansion of Boolean functions on the Hamming cube $\{\pm 1\}^n$. Consider the Fourier expansion of a function $f : \{\pm 1\} \to \mathbb{R}$,

$$f(x_1, \ldots, x_n) = \sum_{S \subset \{1, \ldots, n\}} \hat{f}_S \chi_S(x_1, \ldots, x_n),$$

where $\hat{f}_S$ are Fourier coefficients of f. Informally, the Fourier expansion breaks f into pieces, $\hat{f}_S \chi_S(x_1, \ldots, x_n)$, each of which depends on its own set of variables: The term $\hat{f}_S \chi_S(x_1, \ldots, x_n)$ depends on variables $\{x_i : i \in S\}$ and no other variables.

The Efron—Stein decomposition is an analogue of the Fourier expansion for functions defined on arbitrary product probability spaces. Consider a probability space (Ω, μ) and the product probability space (Ω^n, μ^n). Let $f : \Omega^n \to \mathbb{R}$ be a function (random variable) on Ω^n. Informally, the Efron—Stein decomposition of f is the decomposition of f into the sum of functions f_S, $f = \sum_{S \subset \{1, \ldots, n\}} f_S$, in which $f_S : \Omega^n \to \mathbb{R}$ depends on variables $\{x_i : i \in S\}$.

We formally define the Efron—Stein decomposition as follows. Consider the space $L_2(\Omega^n, \mu^n)$ of functions on Ω^n with bounded second moment. Note that $L_2(\Omega^n, \mu^n) = \bigotimes_{i=1}^n L_2(\Omega, \mu)$. That is, every $f \in L_2(\Omega^n, \mu^n)$ can be represented as $f(x_1, \ldots, x_n) = \sum_j f_1^j(x_1) \cdot f_2^j(x_2) \cdots f_n^j(x_n)$, for some functions $f_i^j \in L_2(\Omega, \mu)$. Let $\Lambda_0 \subset L_2(\Omega, \mu)$ be the one-dimensional space of constant functions on Ω. Let $\Lambda_\perp \subset L_2(\Omega, \mu)$ be the orthogonal complement to Λ_0. That is, $\Lambda_\perp$ is the space of functions $f \in L_2(\Omega, \mu)$ with $\mathbb{E}[f] = 0$. We have, $L_2(\Omega, \mu) = \Lambda_0 \oplus \Lambda_\perp$ and

$$L_2(\Omega^n, \mu^n) = \bigotimes_{i=1}^n L_2(\Omega, \mu) = \bigotimes_{i=1}^n (\Lambda_0 \oplus \Lambda_\perp).$$

Expanding this decomposition, we get a representation of $L_2(\Omega^n, \mu^n)$ as the direct sum of 2^n spaces:

$$L_2(\Omega^n, \mu^n) = \bigoplus_{S \subset \{1, \ldots, n\}} V_S,$$

where V_S is the closed linear span of the set of functions of the form $\prod_{i=1}^n f_i(x_i)$ where $f_i \in \Lambda_\perp$ if $i \in S$, and $f_i \in \Lambda_0$ if $i \notin S$. Since functions in Λ_0 are constants, V_S equals the closed linear span of the set of functions of the form $\prod_{i \in S} f_i(x_i)$ where $f_i \in \Lambda_\perp$.

Consider a function $f \in L_2(\Omega^n, \mu^n)$. Let f_S be the orthogonal projection of f onto V_S. Since the linear spaces V_S are orthogonal, we have $f = \sum_{S \subseteq \{1,\ldots,n\}} f_S$ We call this decomposition the Efron—Stein decomposition of f. We define the degree of f as $\max\{|S| : f_S \neq 0\}$, the size of the largest subset S s.t. f_S is not identically equal to 0 (we let the degree of 0 to be 0).

Let $(X_1, \ldots, X_n)$ be a random element of Ω^n. That is, $X_1, \ldots, X_n$ are n independent random elements of Ω; each of them is distributed according to μ. We write $f = f(X_1, \ldots, X_n)$. We will employ the following properties of the Efron—Stein decomposition (see [11, Section 8.3]).

1. $f_S(x_1, \ldots, x_n)$ depends only on variables x_i with $i \in S$.

2. For every two sets S and T, $S \neq T$, we have $\mathbb{E}[f_S f_T] = 0$.

3. Let $S_1, \ldots, S_r$ be subsets of $\{1, \ldots, n\}$. Suppose that there is an index j that belongs to exactly one set S_i. Then $\mathbb{E}[\prod_{i=1}^r f_{S_i}] = 0$.

We will also use the following equivalent and more explicit definition of the Efron—Stein decomposition. For every subset S of indices $\{1, \ldots, n\}$, let

$$f_{\subseteq S} = \mathbb{E}[f(X_1, \ldots, X_n)| \text{ all } X_i \text{ with } i \in S], \tag{3}$$

$$f_S = \sum_{T \subseteq S} (-1)^{|S \setminus T|} f_{\subseteq T}. \tag{4}$$

4 Efron—Stein Decomposition of Ordering CSP Objective

In this section, we study the Efron—Stein decomposition of the function $\Phi(x_1, \ldots, x_n)$. To this end, we represent $\Phi(x_1, \ldots, x_n)$ as a sum of "basic ordering predicates" and then analyze the Efron—Stein of a basic ordering predicate.

4.1 Basic Ordering Predicate

Let $\tau = (\tau_1, \ldots, \tau_r)$ be a tuple of distinct indices from 1 to n. Define the basic ordering predicate ϕ_τ for τ,

$$\phi_\tau(x_1, \ldots, x_n) = \begin{cases} 1, & \text{if } x_{\tau_1} < x_{\tau_2} < \cdots < x_{\tau_r}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that the indicator of each constraint π is a sum of ordering predicates:

$$\sum_{\tau: \text{ ordering } x_{\tau_1} < x_{\tau_2} < \cdots < x_{\tau_r} \text{ is in } \pi} \phi_\tau(x_1, \ldots, x_n),$$

where the sum is over permutations of variables that the constraint π depends on. Since Φ is the sum of indicators of all predicates π in Π, Φ is also a sum of basic ordering predicates ϕ_τ (for some multiset $\mathcal{T}$):

$$\Phi(x_1, \ldots, x_n) = \sum_{\tau \in \mathcal{T}} \phi_\tau(x_1, \ldots, x_n).$$

4.2 Efron—Stein Decomposition of Ordering Predicates

Let $\Omega = [-1, 1]$ and μ be the uniform measure on Ω. We study the Efron—Stein decomposition of a basic ordering predicate ϕ_τ.

Theorem 4.1. *Let τ be a tuple of distinct indices of size $d \leq k$. Denote $g = \phi_\tau$. Consider the Efron—Stein decomposition of g, $g = \sum g_S$, over $[-1, 1]^n$ with uniform measure. There exists a set of polynomials $q_{S,\tau'}$ with integer coefficients of degree at most d such that*

$$g_S(x_1, \ldots, x_n) = \sum_{\tau'} \phi_{\tau'}(x_1, \ldots, x_n) \frac{q_{S,\tau'}(x_1, \ldots, x_n)}{2^d d!},$$

where the summation is over all permutations τ' of S. The polynomial $q_{S,\tau'}$ depends only on variables in $\{x_i : i \in S\}$. It is equal to 0 if S is not a subset of $\{\tau_1, \ldots, \tau_d\}$.

Proof. We may assume without loss of generality that $\tau = \{1, 2, \ldots, d\}$. Since g depends only on variable $x_1, \ldots, x_d$, $g_S \neq 0$ only if $S \subset \{1, \ldots, d\}$. We may therefore assume that $n = d$ for notational convenience.

Denote the elements of S by $s_1 < s_2 < \cdots < s_t$. Define auxiliary variables $X_0 = -1$ and $X_{d+1} = 1$, and let $s_0 = 0$ and $s_{t+1} = d+1$. Let $\mathcal{O}_{ab}$ be the indicator of the event that $X_i < X_j$ for every $a \leq i < j \leq b$. Then $g = \mathcal{O}_{1,d}$. Note that $g = \mathcal{O}_{1,d} = \prod_{i=0}^{t} \mathcal{O}_{s_i, s_{i+1}}$. All events for $\mathcal{O}_{s_i, s_{i+1}}$ (for $i \in \{0, \ldots, t\}$) are independent given variables $X_{s_1}, \ldots, X_{s_t}$. Therefore,

$$g_{\subset S} = \mathbb{E}\left[g \mid X_{s_1}, \ldots, X_{s_t}\right] = \prod_{i=0}^{t} \mathbb{E}\left[\mathcal{O}_{s_i, s_{i+1}} \mid X_{s_1}, \ldots, X_{s_t}\right]. \tag{5}$$

For each i, we have

$$\mathbb{E}\left[\mathcal{O}_{s_i, s_{i+1}} \mid X_{s_1}, \ldots, X_{s_t}\right] = \mathbb{E}\left[\mathcal{O}_{s_i, s_{i+1}} \mid X_{s_i}, X_{s_{i+1}}\right] = \Pr\left(\mathcal{O}_{s_i, s_{i+1}} = 1 \mid X_{s_i}, X_{s_{i+1}}\right).$$

If $X_{s_i} \geq X_{s_{i+1}}$, then $\Pr\left(\mathcal{O}_{s_i, s_{i+1}} = 1 \mid X_{s_i}, X_{s_{i+1}}\right) = 0$. Otherwise,

$$\Pr\left(\mathcal{O}_{s_i, s_{i+1}} = 1 \mid X_{s_i}, X_{s_{i+1}}\right) = \left(\frac{X_{s_{i+1}} - X_{s_i}}{2}\right)^{s_{i+1}-s_i-1} \cdot \frac{1}{(s_{i+1} - s_i - 1)!}.$$

We computed the probability above as follows: Given $X_{s_i} \leq X_{s_{i+1}}$, the probability that $X_j \in [X_{s_i}, X_{s_{i+1}}]$ for all $j \in \{s_i, \ldots, s_{i+1}\}$ equals $\left((X_{s_{i+1}} - X_{s_i})/2\right)^{s_{i+1}-s_i-1}$. Then, given that $X_{s_i} \leq X_{s_{i+1}}$ and $X_j \in [X_{s_i}, X_{s_{i+1}}]$ for all $j \in \{s_i, \ldots, s_{i+1}\}$, the probability that $X_{s_i+1} \leq \cdots \leq X_{s_{i+1}-1}$ equals $1/(s_{i+1} - s_i - 1)!$ as all orderings of $X_{s_i+1}, \ldots, X_{s_{i+1}-1}$ are equally likely. We get

$$\mathbb{E}\left[\mathcal{O}_{s_i, s_{i+1}} \mid X_{s_1}, \ldots, X_{s_t}\right] = I\left\{X_{s_i} < X_{s_{i+1}}\right\} \frac{\left(X_{s_{i+1}} - X_{s_i}\right)^{s_{i+1}-s_i-1}}{2^{s_{i+1}-s_i-1}(s_{i+1} - s_i - 1)!}.$$

Plugging this expression in (5), we obtain the following formula

$$g_{\subset S} = \prod_{i=0}^{t} I\left\{X_{s_i} < X_{s_{i+1}}\right\} \frac{\left(X_{s_{i+1}} - X_{s_i}\right)^{s_{i+1}-s_i-1}}{2^{s_{i+1}-s_i-1}(s_{i+1} - s_i - 1)!}$$

$$= I\left\{X_{s_1} < X_{s_2} < \ldots X_{s_t}\right\} \frac{\prod_{i=0}^{t}\left(X_{s_{i+1}} - X_{s_i}\right)^{s_{i+1}-s_i-1}}{2^{d-|S|} \prod_{i=0}^{t}(s_{i+1} - s_i - 1)!}.$$

Observe that $\prod_{i=0}^{t}(s_{i+1} - s_i - 1)$ divides $(\sum_{i=0}^{t}(s_{i+1} - s_i - 1))! = (d - |S|)!$. Thus

$$\frac{d!}{\prod_{i=0}^{t}(s_{i+1} - s_i - 1)!} 2^{|S|} \prod_{i=0}^{t}\left(X_{s_{i+1}} - X_{s_i}\right)^{s_{i+1}-s_i-1}$$

7

is a polynomial with integer coefficients of degree at most $d - |S|$. Denote this polynomial by p_S. Then,

$$g_{\subset S} = I\{X_{s_1} < X_{s_2} < \ldots X_{s_t}\} \frac{p_S(X_1, \ldots, X_d)}{2^d d!} = \sum_{\tau'} \phi_{\tau'}(X_1, \ldots, X_d) \frac{p_S(X_1, \ldots, X_d)}{2^d d!},$$

where the sum is over all permutations τ' of $\{1, \ldots, d\}$. Using the identity $f_S = \sum_{T \subseteq S} (-1)^{|S \setminus T|} f_{\subseteq T}$, we get a representation of S as

$$f_S = \sum_{\tau'} \phi_{\tau'}(X_1, \ldots, X_d) \frac{q_{S,\tau}(X_1, \ldots, X_d)}{2^d d!},$$

where $q_{S,\tau'}$ are some polynomials with integer coefficients. $\qquad \square$

Since Φ is a sum of some basic ordering predicates (see Section 4.1), we get the following corollary.

Corollary 4.2. *Let $\mathcal{I}$ be an instance of an ordering CSP problem of arity at most k. Let $\Phi(x_1, \ldots, x_n)$ be the value of continuous solution $(x_1, \ldots, x_n)$. Then the Efron—Stein decomposition of Φ has degree at most k. Moreover there exist polynomials $q_{S,\tau}$ with integer coefficients of degree at most k such that*

$$\Phi_S(x_1, \ldots, x_n) = \sum_{\tau \in \mathcal{T}'} \phi_\tau(x_1, \ldots, x_n) \frac{q_{S,\tau}(x_1, \ldots, x_n)}{2^k k!},$$

where the summation is over some set $\mathcal{T}'$ of tuples of indices in S, and $q_{S,\tau}$ depends only on $\{x_i : i \in S\}$.

4.3 Variance of Ordering CSP Objective

In this section, we show that the variance $\mathrm{Var}\,[\Phi] = \Omega(\nu)$ if Φ (non-trivially) depends on at least ν variables.

Claim 4.3. *There exists a sequence of positive numbers α_d such that for every polynomial $f(x_1, \ldots, x_d)$ of degree at most d with integer coefficients we have $\mathbb{E}\left[\phi_{1,\ldots,d} f^2\right] \geq \alpha_d$.*

Proof. Consider the set $\mathcal{Q}$ of polynomials over $x_1, \ldots, x_d$ of degree at most d. Let $\mathcal{Q}_1$ be the set of polynomials in $\mathcal{Q}$, whose largest in absolute value coefficient is equal to 1 or -1.

Denote $V(f) = \mathbb{E}\left[\phi_{1,\ldots,d} f^2\right]$. For every $f \in \mathcal{Q}_1$, we have $V(f) > 0$ since f is not identically equal to 0 on $\{x_1 \leq x_2 \leq \cdots \leq x_d\}$. Note that $\mathcal{Q}_1$ is a compact set and $V(f)$ is a continuous function on it. Therefore, V attains its minimum on Q_1. Let $\alpha_d = \min_{f \in \mathcal{Q}_1} V(f) > 0$.

Now let f be a polynomial with integer coefficients of degree at most d. Denote the absolute value of its largest coefficient (in absolute value) by M. M is a positive integer and thus $M \geq 1$. We have $f/M \in Q_1$ and thus $V(f) = M^2 \cdot V(f/M) \geq M^2 \alpha_d \geq \alpha_d$. $\qquad \square$

Lemma 4.4. *The following claim holds for some positive parameters β_k. Let $\mathcal{I}$ be an instance of arity at most k. Let $\Phi = \sum_S \Phi_S$ be the Efron—Stein decomposition of Φ. Then for every set S either $\Phi_S = 0$ or $\mathbb{E}\left[\Phi_S^2\right] \geq \beta_k$.*

Proof. Let $\beta_d = \alpha_d / (2^k k!)^2 > 0$, where α_d is as in Claim 4.3. Assume that $\Phi_S \neq 0$. By Corollary 4.2,

$$\Phi_S(x_1, \ldots, x_d) = \sum_{\tau \in \mathcal{T}'} \phi_\tau(x_1, \ldots, x_d) \frac{q_{S,\tau}(x_1, \ldots, x_n)}{2^k k!}.$$

Note that all functions $\phi_\tau(x_1, \ldots, x_d) q_{S,\tau}(x_1, \ldots, x_n)/(2^k k!)$ have disjoint support, and, therefore, are pairwise orthogonal. Choose one tuple $\tau \in \mathcal{T}'$ such that $q_{S,\tau} \neq 0$. We have,

$$\mathbb{E}\left[\Phi_S^2\right] \geq \mathbb{E}\left[\phi_\tau q_{S,\tau}^2 / (2^k k!)^2\right] = \mathbb{E}\left[\phi_\tau q_{S,\tau}^2\right] / (2^k k!)^2.$$

By Claim 4.3, $\mathbb{E}\left[\phi_\tau q_{S,\tau}^2\right] \geq \alpha_d$ and hence $\mathbb{E}\left[\Phi_S^2\right] \geq \alpha_d / (2^k k!)^2$. $\qquad \square$

8

We say that Φ depends on the variable x_i if there exist two vectors x and x' that differ only in the i-th coordinate such that $\Phi(x) \neq \Phi(x')$.

Theorem 4.5. *Let $\mathcal{I}$ be an instance of arity at most k. Suppose that Φ depends on at least ν variables. Then* $\mathrm{Var}\,[\Phi] \geq \nu \beta_k / k$.

Proof. Consider the Efron—Stein decomposition of Φ. Let $V' = \bigcup_{S:\Phi_S \neq 0} \{x_i : i \in S\}$. Note that Φ depends on all variables in V' and no other variables. Thus, $|V'| \geq \nu$. There are at least ν/k non-empty sets S with $\Phi_S \neq 0$ since each such set S contributes at most k elements to V'. For $S \neq \varnothing$, we have $\mathbb{E}\,[\Phi_S] = 0$ and hence $\mathrm{Var}\,[\Phi_S] = \mathbb{E}\,[\Phi_S^2]$. By Lemma 4.4, $\mathrm{Var}\,[\Phi_S] = \mathbb{E}\,[\Phi_S^2] \geq \beta_k$ if $\Phi_S \neq 0$ and $S \neq \varnothing$. We have,

$$\mathrm{Var}\,[\Phi] = \sum_S \mathrm{Var}\,[\Phi_S] \geq |\{S \neq \varnothing : \Phi_S \neq 0\}|\beta_k \geq (\nu/k)\beta_k. \qquad \square$$

5 Proof of Main Theorems

In this section, we prove Theorems 3.3 and 1.3. We will need the following theorem.

Theorem 5.1 (Corollary 1 from Alon, Gutin, Kim, Szeider, and Yeo [1]). *Let X be a real random variable. Suppose that $\mathbb{E}\,[X] = 0$, $\mathbb{E}\,[X^2] = \sigma^2$, and $\mathbb{E}\,[X^4] < b\sigma^4$ for some $b > 0$. Then $\mathrm{Pr}\,(X \geq \sigma/(2\sqrt{b})) > 0$.*

Proof of Theorem 3.3. Let V' be the set of variables that Φ depends on (see Section 4.3 for definitions). By Theorem 4.5, $\mathrm{Var}\,[\Phi] \geq |V'| \cdot \beta_k / k$ for some absolute constant $\beta_k > 0$. By Lemma A.1, the function $\Phi - \mathbb{E}\Phi$ satisfies condition (1) of the Bonami Lemma for the Efron—Stein Decomposition (Theorem 1.4) with some absolute constant C_k. Hence, $\|\Phi - \mathbb{E}\Phi\|_4^4 \leq 81^k C_k \|\Phi - \mathbb{E}\Phi\|_2^4$. Applying Theorem 5.1 to the random variable $\Phi - \mathbb{E}\Phi$ with $\sigma = (|V'|\beta_k/k)^{1/2}$ and $b = 81^k C_k$, we get that $\mathrm{Pr}\,(\Phi \geq \mathbb{E}\Phi + |V'|^{1/2}/\kappa_k) > 0$, where $\kappa_k^2 = 4kC_k 81^k/\beta_k$. Consequently,

$$OPT = \max_{x \in [-1,1]^n} \Phi(x) \geq \mathbb{E}\Phi + |V'|^{1/2}/\kappa_k = AVG + |V'|^{1/2}/\kappa_k.$$

We are now ready to state the algorithm. The algorithm computes the Efron—Stein decomposition in time $O_k(m+n)$. Then, using the formula $V' = \bigcup_{S:\Phi_S \neq 0} \{x_i : i \in S\}$ (see Theorem 4.5), it finds the set V' also in time $O_k(m+n)$. It considers two cases.

1. If $|V'| \geq \kappa_k t^2$, then the algorithm returns $OPT \geq AVG + t$.

2. Otherwise, if $|V'| < \kappa_k t^2$, the algorithm outputs the restriction of $\mathcal{I}$ to the variables in V'. This is a kernel for $\mathcal{I}$, since Φ depends only on the variables in V'.

$\square$

To prove Theorem 1.3, we need to show how to find an assignment satisfying $AVG + t$ constraints if $|V'| \geq \kappa_k t^2$. This can be easily done using $4k$-rankwise independent permutations. A random permutation $\tilde{\alpha}$ is m-rankwise independent if for every subset $M \subset \{1, \ldots, n\}$ of size m, the order of elements in M induced by $\tilde{\alpha}$ is uniformly distributed (the definition is due to Itoh, Takei, and Tarui [28]). Note that any m-wise independent permutation $\tilde{\alpha}$ is also an m-rankwise independent permutation. Using the result of Alon and Lovett [2], we can obtain a $4k$-wise independent permutation $\tilde{\alpha}$ supported on a set of size $n^{O(k)}$. In Lemma C.1 (in Appendix C), we show that for some permutation α^* in the support of $\tilde{\alpha}$, we have $\mathrm{val}_{\mathcal{I}}(\alpha^*) \geq AVG + t$. Hence, to find an assignment satisfying $AVG + t$ constraints, we need to search for the best permutation in the support of α^*, which can be done in time $n^{O(k)}$.

6   Bonami Lemma

In this section, we prove the Bonami Lemma for the Efron—Stein decomposition (Theorem 1.4) stated in the introduction. Our starting point will be the standard Bonami Lemma for Bernoulli ± 1 random variables. Due to space limitations, we present most of the proofs in Section B.

Lemma 6.1 (see [4, 11]). *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most d. Let $X_1, \ldots, X_n$ be independent unbiased ± 1-Bernoulli variables. Then*

$$\mathbb{E}\left[f(X_1, \ldots, X_n)^4\right] \leq 9^d \mathbb{E}\left[f(X_1, \ldots, X_n)^2\right]^2.$$

We will consider the following probability distribution in this proof. Let Z be a random variable equal to 3 with probability $1/4$ and to -1 with probability $3/4$. Denote by $\mathcal{Z}$ the probability distribution of Z. We first prove a variant of the Bonami Lemma for random variables distributed according to $\mathcal{Z}$.

Lemma 6.2. *Let $f : \{-1,3\}^n \to \mathbb{R}$ be a polynomial of degree at most d. Let $Z_1, \ldots, Z_n$ be independent random variables distributed according to $\mathcal{Z}$. Then*

$$\mathbb{E}\left[f(Z_1, \ldots, Z_n)^4\right] \leq 81^d \mathbb{E}\left[f(Z_1, \ldots, Z_n)^2\right]^2.$$

Now let $f \in L_2(\Omega^n, \mu^n)$ and $f = \sum_S f_S$ be its Efron—Stein decomposition. Define polynomial $M_f \colon \mathbb{R}^n \to \mathbb{R}$ by

$$M_{f,S} = (1/3)^{|S|/2} \mathbb{E}\left[f_S^2\right]^{1/2} \text{ for every } S \subseteq \{1, \ldots, n\},$$

$$M_f(Z_1, \ldots, Z_n) = \sum_{S \subseteq \{1, \ldots, n\}} M_{f,S} \prod_{i \in S} Z_i.$$

We now get bounds for moments of $f(X_1, \ldots, X_n)$ in terms of moments of $M_f(Z_1, \ldots, Z_n)$.

Claim 6.3. *Let Z_i be independent random variables distributed according to $\mathcal{Z}$. We have,*

$$\mathbb{E}\left[M_f(Z_1, \ldots, Z_n)\right] = |\mathbb{E}\left[f(X_1, \ldots, X_n)\right]| \text{ and } \mathbb{E}\left[M_f(Z_1, \ldots, Z_n)^2\right] = \mathbb{E}\left[f(X_1, \ldots, X_n)^2\right].$$

Claim 6.4. *Let f and C be as in the condition of Theorem 1.4. Then $\mathbb{E}\left[f(X_1, \ldots, X_n)^4\right] \leq C\mathbb{E}\left[M_f(Z_1, \ldots, Z_n)^4\right]$.*

Proof. Write,

$$\mathbb{E}\left[f^4\right] = \sum_{S_1, S_2, S_3, S_4} \mathbb{E}\left[f_{S_1} f_{S_2} f_{S_3} f_{S_4}\right], \text{ and}$$

$$\mathbb{E}\left[M_f^4\right] = \sum_{S_1, S_2, S_3, S_4} M_{f,S_1} M_{f,S_2} M_{f,S_3} M_{f,S_4} \mathbb{E}\left[\prod_{i_1 \in S_1} Z_{i_1} \prod_{i_2 \in S_2} Z_{i_2} \prod_{i_3 \in S_3} Z_{i_3} \prod_{i_4 \in S_4} Z_{i_4}\right].$$

To prove the claim, we will show in Appendix B that for every four sets S_1, S_2, S_3, S_4, the following inequality holds, and, therefore, $\mathbb{E}\left[f^4\right] \leq C\mathbb{E}\left[M_f^4\right]$:

$$\mathbb{E}\left[f_{S_1} f_{S_2} f_{S_3} f_{S_4}\right] \leq C M_{f,S_1} M_{f,S_2} M_{f,S_3} M_{f,S_4} \mathbb{E}\left[\prod_{i_1 \in S_1} Z_{i_1} \prod_{i_2 \in S_2} Z_{i_2} \prod_{i_3 \in S_3} Z_{i_3} \prod_{i_4 \in S_4} Z_{i_4}\right]. \tag{6}$$

$\square$

Proof of Theorem 1.4. By Lemma 6.2, we have

$$\|M_f(Z_1, \ldots, Z_n)\|_4^4 \leq 81^d \|M_f(Z_1, \ldots, Z_n)\|_2^4.$$

From Claims 6.3 and 6.4, we get

$$\mathbb{E}\left[f(X_1, \ldots, X_n)^4\right] \leq C\mathbb{E}\left[M_f(Z_1, \ldots, Z_n)^4\right] \leq 81^d C\mathbb{E}\left[M_f(Z_1, \ldots, Z_n)^2\right]^2 = 81^d C\mathbb{E}\left[f(X_1, \ldots, X_n)^2\right]^2.$$

$\square$

Acknowledgement

We thank Matthias Mnich for valuable comments.

References

[1] N. Alon, G. Gutin, E. J. Kim, S. Szeider, and A. Yeo. Solving Max r-SAT Above a Tight Lower Bound. Algorithmica 61 (3), pp. 638–655 (2011).

[2] N. Alon and S. Lovett. Almost k-wise vs. k-wise independent permutations, and uniformity for general group actions. Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques. Springer Berlin Heidelberg, pp. 350–361, 2012.

[3] P. Austrin and E. Mossel. Approximation resistant predicates from pairwise independence. Computational Complexity, 18(2), pp. 249–271, 2009.

[4] A. Bonami. Étude des coefficients Fourier des fonctions de $L^p(G)$. Annales de l' Institut Fourier, 20(2):335–402, 1970.

[5] S. Chan. Approximation resistance from pairwise independent subgroups. STOC 2014, pp. 447–456.

[6] M. Charikar, K. Makarychev, and Y. Makarychev. On the Advantage over Random for Maximum Acyclic Subgraph. FOCS 2007, pp. 625–633.

[7] R. Crowston, M. Fellows, G. Gutin, M. Jones, F. Rosamond, S. Thomassé and A. Yeo. Simultaneously Satisfying Linear Equations Over $\mathbb{F}_2$: MaxLin2 and Max-r-Lin2 Parameterized Above Average. In FSTTCS 2011, LIPICS Vol. 13, 229–240.

[8] R. Crowston, G. Gutin, M. Jones, V. Raman, and S. Saurabh. Parameterized complexity of MaxSat above average. In the Proceedings of LATIN 2012, pp. 184–194.

[9] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. Ann. Appl. Probab. 6 (1996), no. 3, 695–750.

[10] B. Efron and C. Stein. The jackknife estimate of variance. Annals of Statistics, 9(3):586–596, 1981.

[11] R. O'Donnell. Analysis of Boolean Functions. Cambridge University Press. 2014. ISBN 9781107038325.

[12] V. Guruswami, J. Håstad, R. Manokaran, P. Raghavendra, and M. Charikar. Beating the random ordering is hard: Every ordering CSP is approximation resistant. SIAM Journal on Computing 40, no. 3 (2011): 878–914.

[13] V. Guruswami and P. Raghavendra. Constraint Satisfaction over a Non-Boolean Domain: Approximation Algorithms and Unique-Games Hardness. APPROX 2008, pp. 77-90.

[14] V. Guruswami and Y. Zhou. Approximating Bounded Occurrence Ordering CSPs. APPROX-RANDOM 2012, pp. 158–169.

[15] G. Gutin, L. van Iersel, M. Mnich, and A. Yeo. All Ternary Permutation Constraint Satisfaction Problems Parameterized Above Average Have Kernels with Quadratic Number of Variables. J. Comput. Syst. Sci. 78 (2012), 151–163.

[16] G. Gutin, E. J. Kim, M. Mnich, and A. Yeo. Betweenness parameterized above tight lower bound. J. Comput. Syst. Sci., 76: 872–878, 2010.

[17] G. Gutin, E.J. Kim, S. Szeider, and A. Yeo. A Probabilistic Approach to Problems Parameterized Above or Below Tight Bounds. J. Comput. Syst. Sci. 77 (2011), 422–429.

[18] G. Gutin and A. Yeo. Constraint satisfaction problems parameterized above or below tight bounds: a survey. In The Multivariate Algorithmic Revolution and Beyond, pp. 257–286, 2012.

[19] J. Håstad. Some optimal inapproximability results. STOC 1997.

[20] R. Karp. Reducibility among combinatorial problems. In Complexity of Computer Computations. New York: Plenum, 1972, pp. 85–103.

[21] E. J. Kim and R. Williams. Improved parameterized algorithms for above average constraint satisfaction. In Parameterized and Exact Computation, pp. 118–131, 2012.

[22] V. Mahajan and V. Raman. Parameterizing above Guaranteed Values: MaxSat and MaxCut. Journal of Algorithms, Vol. 31, Issue 2, May 1999, pp. 335–354.

[23] K. Makarychev. Local Search is Better than Random Assignment for Bounded Occurrence Ordering k-CSPs. STACS 2013, pp. 139–147.

[24] E. Mossel, K. Oleszkiewicz and A. Sen. On Reverse Hypercontractivity Geometric and Functional Analysis, vol 23(3), pp. 1062–1097, 2013.

[25] J. Opatrny. Total ordering problem. SIAM Journal on Computing, 8(1):111–114, Feb. 1979.

[26] I. Razgon and B. O'Sullivan. Almost 2-SAT is fixed-parameter tractable. J. Comput. Syst. Sci. 75(8):435–450, 2009.

[27] M. Talagrand. On Russos approximate 0-1 law , Annals of Probability 22 (1994), 1576–1587.

[28] T. Itoh, Y. Takei, and J. Tarui. On Permutations with Limited Independence. SODA 2000, pp. 137–146.

[29] P. Wolff. Hypercontractivity of simple random variables, Studia Mathematica 180 (2007), pp. 219–326.

A   Bonami Lemma for ordering CSPs

We are going to apply Theorem 1.4 (the Bonami Lemma for the Efron—Stein decomposition) to the function $f = \Phi - \mathbb{E}\left[\Phi\right]$, where Φ is the objective function of the ordering CSP problem. We now show that f satisfies the condition of the theorem with some constant C that depends only on the arity of the CSP.

Lemma A.1. *There exists a sequence of constants C_k such that the following holds. Let $\mathcal{I}$ be an instance of an ordering CSP of arity at most k. Let $f = \Phi - \mathbb{E}\left[\Phi\right]$. Then*

$$\mathbb{E}\left[f_{S_1} f_{S_2} f_{S_3} f_{S_4}\right] \leq C_k \left(\mathbb{E}\left[f_{S_1}^2\right] \mathbb{E}\left[f_{S_2}^2\right] \mathbb{E}\left[f_{S_3}^2\right] \mathbb{E}\left[f_{S_4}^2\right]\right)^{1/2}.$$

Proof. We assume that all sets $S_1, \ldots, S_4$ are non-empty as otherwise both the left and right hand sides of the inequality are equal to 0 (since $f_\varnothing = \mathbb{E}[f] = 0$). Therefore, $f_{S_i} = \Phi_{S_i}$ for $i \in \{1, \ldots, 4\}$.

Note that $|S_1 \cup S_2 \cup S_3 \cup S_4| \leq 4k$. So without loss of generality, we may assume that $S_1, S_2, S_3, S_4 \subset \{1, \ldots, 4k\}$. Let $\mathcal{Q}$ be the set of all functions of $x_1, \ldots, x_{4k}$ of the form

$$\sum \phi_\tau(x_1, \ldots, x_k) q_{S,\tau}(x_1, \ldots, x_k),$$

where $q_{S,\tau}$ are some polynomials of degree at most k (not necessarily with integer coefficients). By Corollary 4.2, $f_{S_1}, f_{S_2}, f_{S_3}, f_{S_4} \in \mathcal{Q}$. Let $\mathcal{Q}_1 = \{h \in \mathcal{Q} : \|h\|_2 = 1\}$. Note that $\mathcal{Q}_1$ is a compact set (since $\mathcal{Q}$ is a finite dimensional space; and $\| \cdot \|_2$ is a norm on it). Therefore, the continuous function $W(g_1, g_2, g_3, g_4) = \mathbb{E}[g_1 g_2 g_3 g_4]$ is bounded when $g_1, g_2, g_3, g_4 \in \mathcal{Q}_1$. Denote its maximum by C_k (note that C_k depends only on k and not on $\mathcal{I}$).

Let $g_i = f_{S_i}/\|f_{S_i}\|_2$. Note that $g_i \in \mathcal{Q}_1$. We have,

$$\mathbb{E}\left[\prod_{i=1}^4 f_{S_i}\right] = \mathbb{E}\left[\prod_{i=1}^4 g_i\right] \cdot \prod_{i=1}^4 \|f_{S_i}\|_2 \leq C_k \cdot \prod_{i=1}^4 \|f_{S_i}\|_2,$$

as required. $\qquad\square$

B   Proofs from Section 6

Proof of Lemma 6.2. Consider $2n$ Bernoulli random variables $Y_1', \ldots, Y_n', Y_1'', \ldots, Y_n''$ (uniformly distributed in $\{-1, 1\}$). Note that random variables $\tilde{Z}_i \equiv Y_i' + Y_i'' + Y_i' Y_i''$ are distributed in the same way as random variables Z_i. Therefore,

$$\mathbb{E}\left[f(Z_1, \ldots, Z_n)^4\right] = \mathbb{E}\left[f(\tilde{Z}_1, \ldots, \tilde{Z}_n)^4\right],$$

$$\mathbb{E}\left[f(Z_1, \ldots, Z_n)^2\right]^2 = \mathbb{E}\left[f(\tilde{Z}_1, \ldots, \tilde{Z}_n)^2\right]^2.$$

Now $f(\tilde{Z}_1, \ldots, \tilde{Z}_n)$ is a polynomial of ± 1 variables $Y_1', \ldots, Y_n', Y_1'', \ldots, Y_n''$ of degree at most $2d$. Applying Lemma 6.1 to $f(\tilde{Z}_1, \ldots, \tilde{Z}_n)$, we get

$$\mathbb{E}\left[f(\tilde{Z}_1, \ldots, \tilde{Z}_n)^4\right] \leq 9^{2d}\mathbb{E}\left[f(\tilde{Z}_1, \ldots, \tilde{Z}_n)^2\right]^2,$$

and, therefore,

$$\mathbb{E}\left[f(Z_1, \ldots, Z_n)^4\right] \leq 81^d \mathbb{E}\left[f(Z_1, \ldots, Z_n)^2\right]^2,$$

as required. $\qquad\square$

Proof of Claim 6.3. Note that $\mathbb{E}[Z_i] = 0$ and thus $\mathbb{E}[M_f(Z_1, \ldots, Z_n)] = M_{f,\varnothing} = |f_\varnothing| = |\mathbb{E}[f]|$.

Also,
$$\mathbb{E}\left[M_f(Z_1, \ldots, Z_n)^2\right] = \sum_{S \subseteq \{1, \ldots, n\}} M_{f,S}^2 \cdot \mathbb{E}[\prod_{i \in S} Z_i^2].$$

Since $\mathbb{E}\left[Z_i^2\right] = 3$, we have $\mathbb{E}\left[\prod_{i \in S} Z_i^2\right] = 3^{|S|}$, and therefore

$$\mathbb{E}\left[M_f(Z_1, \ldots, Z_n)^2\right] = \sum_{S \subseteq \{1, \ldots, n\}} \left((1/3)^{|S|/2}\mathbb{E}\left[f_S^2\right]^{1/2}\right)^2 \cdot 3^{|S|} = \sum_{S \subseteq \{1, \ldots, n\}} \mathbb{E}\left[f_S^2\right] = \mathbb{E}\left[f^2\right].$$

$\qquad\square$

Proof of Inequality (6). We now prove inequality (6).

$$\mathbb{E}\left[f_{S_1} f_{S_2} f_{S_3} f_{S_4}\right] \leq C M_{f,S_1} M_{f,S_2} M_{f,S_3} M_{f,S_4} \mathbb{E}\left[\prod_{i \in S_1} Z_i \prod_{i \in S_2} Z_i \prod_{i \in S_3} Z_i \prod_{i \in S_4} Z_i\right].$$

Note first that if some index j appears in exactly one of the sets S_1, S_2, S_3, and S_4 then the expressions on the left and on the right are equal to 0 (by Property 3 of the Efron—Stein decomposition in Section 3.3), and we are done. So we assume that every index j in $S_1 \cup S_2 \cup S_3 \cup S_4$ appears in at least 2 of the sets S_i. Denote the number of times j appears in sets S_i by $m(j)$.

By the condition of Theorem 1.4 and the definition of coefficients M_{f,S_i},

$$\mathbb{E}\left[f_{S_1} f_{S_2} f_{S_3} f_{S_4}\right] \leq C \left(\mathbb{E}\left[f_{S_1}^2\right] \mathbb{E}\left[f_{S_2}^2\right] \mathbb{E}\left[f_{S_3}^2\right] \mathbb{E}\left[f_{S_4}^2\right]\right)^{1/2} \tag{7}$$

$$= C \cdot 3^{(|S_1|+|S_2|+|S_3|+|S_4|)/2} M_{f,S_1} M_{f,S_2} M_{f,S_3} M_{f,S_4}.$$

On the other hand,

$$\mathbb{E}\left[\prod_{i_1 \in S_1} Z_{i_1} \prod_{i_2 \in S_2} Z_{i_2} \prod_{i_3 \in S_3} Z_{i_3} \prod_{i_4 \in S_4} Z_{i_4}\right] = \mathbb{E}\left[\prod_{r=1}^{4} \prod_{i \in S_r} Z_i\right] = \prod_{i \in \bigcup_r S_r} \mathbb{E}\left[Z_i^{m(i)}\right]$$

$$= \prod_{r=1}^{4} \prod_{i \in S_r} \left(\mathbb{E}\left[Z_i^{m(i)}\right]\right)^{1/m(i)}.$$

We compute $\mathbb{E}\left[Z_i^{m(i)}\right]$ for $m(i) \in \{2, 3, 4\}$. We get $\mathbb{E}Z_i^2 = 3$, $\mathbb{E}Z_i^3 = 6$ and $\mathbb{E}Z_i^4 = 21$. Thus, $(\mathbb{E}Z_i^2)^{1/2} = \sqrt{3}$, $(\mathbb{E}Z_i^3)^{1/3} = \sqrt[3]{6} > \sqrt{3}$ and $(\mathbb{E}Z_i^4)^{1/4} = \sqrt[4]{21} > \sqrt{3}$, and, consequently,

$$\mathbb{E}\left[\prod_{r=1}^{4} \prod_{i \in S_r} Z_i\right] \geq \prod_{r=1}^{4} \prod_{i \in S_r} 3^{1/2} = 3^{(|S_1|+|S_2|+|S_3|+|S_4|)/2}. \tag{8}$$

Since all coefficients $M_{f,S}$ are non-negative, we get from (7) and (8) that inequality (6) holds. $\square$

C Rankwise independent permutations

In this section, we prove the following lemma.

Lemma C.1. *If $\tilde{\alpha}$ is a random $4k$ rankwise independent permutation and $|V'| \geq \kappa_k t^2$, then for some α^* in the support of $\tilde{\alpha}$, $\mathrm{val}_\mathcal{I}(\alpha^*) \geq AVG + t$.*

Proof. Let α be a permutation uniformly distributed among all $n!$ permutations. The random variables $\mathrm{val}_\mathcal{I}(\alpha)$ and Φ are identically distributed. Hence,

$$\mathrm{Var}[\mathrm{val}_\mathcal{I}(\alpha)] = \mathrm{Var}[\Phi], \quad \text{and} \quad \|\mathrm{val}_\mathcal{I}(\alpha) - AVG\|_4 = \|\Phi - AVG\|_4.$$

Observe, that

$$\mathrm{Var}[\mathrm{val}_\mathcal{I}(\alpha^*)] = \mathrm{Var}[\mathrm{val}_\mathcal{I}(\alpha)], \quad \text{and} \quad \|\mathrm{val}_\mathcal{I}(\alpha) - AVG\|_4 = \|\mathrm{val}_\mathcal{I}(\alpha^*) - AVG\|_4,$$

since for every four predicates $\pi_1, \pi_2, \pi_3, \pi_4 \in \Pi$, we have

$$\mathbb{E}\left[\pi_1(\alpha^*)\pi_2(\alpha^*)\right] = \mathbb{E}\left[\pi_1(\alpha)\pi_2(\alpha)\right];$$
$$\mathbb{E}\left[\pi_1(\alpha^*)\pi_2(\alpha^*)\pi_3(\alpha^*)\pi_4(\alpha^*)\right] = \mathbb{E}\left[\pi_1(\alpha)\pi_2(\alpha)\pi_3(\alpha)\pi_4(\alpha)\right].$$

Hence, as in Theorem 3.3, $\mathrm{Var}[\mathrm{val}_\mathcal{I}(\alpha^*)] \geq |V'|\beta_k/k$ and $\|\mathrm{val}_\mathcal{I}(\alpha^*) - AVG\|_4^4 \leq 81^k C_k \|\mathrm{val}_\mathcal{I}(\alpha^*) - AVG\|_2^4$. Consequently, by Theorem 5.1, $\mathrm{Pr}(\mathrm{val}_\mathcal{I}(\alpha^*) \geq AVG + t) > 0$. This concludes the proof. $\square$

D General Framework

D.1 Filtered A-Lattice of Functions

In this section, we generalize the result of the paper to a more general class of constraint satisfaction problems having a lattice structure. In Appendix E, we show that LP CSPs and valued CSPs with "piece-wise polynomial predicates" (see Appendix E for the defintion) have a lattice structure.

D.2 Discussion

We note that in our proofs we used only few properties of ordering CSPs. Specifically, in Theorem 4.1, we showed that all functions in the Efron—Stein decomposition of the basic ordering predicate are in the set

$$\mathcal{F}_d^{\mathrm{ord}} = \left\{ \sum_\tau \phi_\tau(x_1, \ldots, x_k) \frac{q_{S,\tau}(x_1, \ldots, x_n)}{2^d d!} \right\}.$$

Since $\mathcal{F}_d^{\mathrm{ord}}$ is closed under addition (the sum of any two functions in $\mathcal{F}_d^{\mathrm{ord}}$ is in $\mathcal{F}_d^{\mathrm{ord}}$), we got that all functions in the Efron—Stein decomposition of the ordering CSP objective Φ are also in $\mathcal{F}_d^{\mathrm{ord}}$. Then in the proof of Lemma 4.4, we showed that every non-zero function in $\mathcal{F}_d^{\mathrm{ord}}$ has variance at least β_k (where β_k depends only on k), and this was sufficient to get the result of the paper. To summarize, we only used the following properties of the set of functions $\mathcal{F}_d^{\mathrm{ord}}$:

A. all functions in the Efron—Stein decomposition of each predicate are in $\mathcal{F}_d^{\mathrm{ord}}$,

B. $\mathcal{F}_d^{\mathrm{ord}}$ is closed under addition,

C. every non-zero function in $\mathcal{F}_d^{\mathrm{ord}}$ has variance at least β (for some fixed $\beta > 0$).

D.3 Filtered A-Lattice of Functions

We now formalize properties A, B, and C in the definitions of A-lattice of functions and filtered A-lattice of functions. Recall first the definition of a lattice.

Definition D.1. *Let V be a finite-dimensional space and $\mathcal{L}$ be a subset of V. We say that $\mathcal{L}$ is a lattice in V if for some basis $v_1, \ldots, v_r$ of V, we have $\mathcal{L} = \{\sum_{i=1}^r a_i v_i : a_1, \ldots, a_r \in \mathbb{Z}\}$. We say that $v_1, \ldots, v_r$ is the basis of the lattice $\mathcal{L}$.*

Now we define an A-lattice of functions.

Definition D.2. *Let (Ω, μ) be a probability space. Consider a set $\mathcal{F}$ of bounded (real-valued) functions $f(x_1, \ldots, x_k)$ on Ω; $\mathcal{F} \subset L_\infty(\Omega^k)$. We say that $\mathcal{F}$ is an A-lattice of functions of arity (at most) k on Ω if it satisfies the following properties.*

1. *$\mathcal{F}$ is a lattice in a finite dimensional subspace of $L_\infty(\Omega^k)$.*

2. *If we permute arguments of a function in $\mathcal{F}$, we get a function in $\mathcal{F}$. Specifically, if $f \in \mathcal{F}$ and π is a permutation of $\{1, \ldots, k\}$ then $g(x_1, \ldots, x_k) = f(x_{\pi(1)}, \ldots, x_{\pi(k)}) \in \mathcal{F}$.*

For an A-lattice $\mathcal{F}$, we write that a function $f \in_R \mathcal{F}$ if f is in $\mathcal{F}$ after possibly renaming the arguments of f (in other words, f is in $\mathcal{F}$ as an abstract function from Ω^k to $\mathbb{R}$).[2]

---

[2]For example, let $\mathcal{F}$ be an A-lattice of functions of the form $ax_1 + bx_2$ where $a, b \in \mathbb{Z}$. Then $x_3 + 5x_7 \in_R \mathcal{F}$, since after renaming x_3 to x_1 and x_7 to x_2 we get $x_3 + 5x_2$, which is of the form $ax_1 + bx_2$.

Clearly, every A-lattice $\mathcal{F}$ of functions satisfies property B. Since $\mathcal{F}$ is discrete, it also satisfies property C (we will prove that formally in Claim D.9). We also want to ensure that it satisfies an analog of property A. To this end, we consider the averaging operator A_i, which takes the expectation of a function with respect to variable x_i and require that A_i maps every function in the lattice to a function in the lattice.

Definition D.3. *For $i \in \{1, \ldots, k\}$, let A_i be the averaging operator that maps a function f of arity k to a function $A_i f$ of arity $k - 1$ defined as follows:*

$$A_i f(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) = \int_{\Omega} f(x_1, \ldots, x_k) \, d\mu(x_i)$$
$$= \mathbb{E} f(X_1, \ldots, X_k | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \ldots, X_k = x_k).$$

Definition D.4. *We say that a family of sets $\{\mathcal{F}_\alpha\}_\alpha$ (indexed by integer $\alpha \geq 1$) is a filtered A-lattice of functions of arity (at most) k if it satisfies the following properties.*

1. *$\mathcal{F}_\alpha$ is an A-lattice of functions of arity k on Ω.*

2. *$\{\mathcal{F}_\alpha\}$ is a filtration: $\mathcal{F}_\alpha \subset \mathcal{F}_{\alpha'}$ for $\alpha \leq \alpha'$.*

3. *For every α there exists α', which we denote by $\alpha' = a(\alpha)$, such that the operator A_i maps $\mathcal{F}_\alpha$ to $\mathcal{F}_{\alpha'}$ (for every $i \in \{1, \ldots, k\}$).*

We remark that $\{\mathcal{F}_d^{\mathrm{ord}}\}$ is a filtered A-lattice. We are going to prove that our result for ordering CSPs holds, in fact, for any constraint satisfaction problem with predicates from a filtered A-lattice.

D.4 General A-CSP($\mathcal{F}_\alpha$, α)

Definition D.5. *Consider a probability space (Ω, μ). Let $\mathcal{F}_\alpha$ be a filtered A-lattice of functions and α_0 is an integer. An instance $\mathcal{I}$ of General A-CSP($\mathcal{F}_\alpha$, α_0) consists of a set of variables $x_1, \ldots, x_n$, taking values in Ω, and a set of real-valued constraints of the form $f(x_{i_1}, \ldots, x_{i_k})$ where $f \in_R \mathcal{F}_\alpha$. The objective function $\Phi(x_1, \ldots, x_n)$ is the sum of all the constraints. General A-CSP($\mathcal{F}_\alpha$, α_0) asks to find an assignment to variables $x_1, \ldots, x_n$ that maximizes $\Phi(x_1, \ldots, x_n)$.*

We denote the optimal value of an instance $\mathcal{I}$ by $OPT = \mathrm{ess\,sup}_{x_1, \ldots, x_n \in \Omega} \Phi(x_1, \ldots, x_n)$ and the average by $AVG = \mathbb{E}\left[\Phi(X_1, \ldots, X_n)\right]$, where $X_1, \ldots, X_n$ are independent random elements of Ω distributed according to the probability measure μ.

Remark D.1. *Note that we follow the standard convention that two functions $f, g \in L_\infty(\Omega^k)$ are equal if $f(x_1, \ldots, x_k) \neq g(x_1, \ldots, x_k)$ on a set of measure 0. That is, we identify functions that are equal almost everywhere. Accordingly, we define OPT as the essential supremum of Φ: OPT is equal to the maximum value of M such that*

$$\Pr\left(\Phi \geq M - \varepsilon\right) > 0,$$

for every $\varepsilon > 0$.

We prove a counterpart of Theorem 3.3 for the General A-CSP problem.

Theorem D.6. *There is an algorithm that given an instance of General A-CSP($\mathcal{F}_\alpha$, k, α_0) and a parameter t, either finds a kernel on at most κt^2 variables (where κ depends only on the filtered A-lattice $\{\mathcal{F}_\alpha\}$ and numbers k, α_0) or certifies that $OPT \geq AVG + t$. The algorithm runs in time $O(m + n)$, linear in the number of constraints m and variables n (the coefficient in the O-notation depends the filtered A-lattice $\{\mathcal{F}_\alpha\}$ and numbers k, α_0).*

We assume that computing the sum of two functions in $\mathcal{F}_\alpha$ requires constant time and that computing $A_i f$ requires constant time (the time may depend on α).

We first prove analogues of Theorem 4.1 and Corollary 4.2 for General A-CSP.

Lemma D.7 (cf. Theorem 4.1)**.** *Consider a probability space (Ω, μ) and a filtered A-lattice $\{\mathcal{F}_\alpha\}$. For every α, there exists $\tilde{\alpha}$ so that the following holds. For every $f \in \mathcal{F}_\alpha$ and every $S \subset \{1, \ldots, k\}$, $f_S \in \mathcal{F}_{\tilde{\alpha}}$.*

Proof. Let $\alpha_0 = \alpha$, $\alpha_1 = a(\alpha_0)$ (where a is as in the definition of a filtered A-lattice), $\alpha_2 = a(\alpha_1)$, and so on; $\alpha_i = a(\alpha_{i-1})$. Let $\tilde{\alpha} = \max_{i \in \{0, \ldots, k\}} \alpha_i$. Consider a function $f \in \mathcal{F}_\alpha$ and a set $T \subset \{1, \ldots, k\}$. By (3),

$$f_{\subseteq T} = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid \text{all } X_i \text{ with } i \in T\right].$$

Denote the elements of $\{1, \ldots, k\} \setminus T$ by $i_1 < \cdots < i_t$ (where $t = k - |T|$). Note that

$$f_{\subseteq T} = A_{i_1} A_{i_2} \ldots A_{i_t} f \in \mathcal{F}_{\alpha_t} \subset \mathcal{F}_{\tilde{\alpha}},$$

by the definition of a filtered A-lattice. Now by (4),

$$f_S = \sum_{T \subseteq S} (-1)^{|S \setminus T|} f_{\subseteq T}.$$

Since $\mathcal{F}_{\tilde{\alpha}}$ is a lattice and f_S is a linear combination, with integer coefficients, of functions $f_{\subseteq T}$ (all of which are in $\mathcal{F}_{\tilde{\alpha}}$), f_S is in $\mathcal{F}_{\tilde{\alpha}}$. $\square$

Since the set of functions $\mathcal{F}_{\tilde{\alpha}}$ is closed under addition, we get that for every General A-CSP($\mathcal{F}_\alpha$, α) instance $\mathcal{I}$ with objective function Φ, all functions Φ_S are also in $\mathcal{F}_{\tilde{\alpha}}$.

Corollary D.8 (cf. Corollary 4.2)**.** *Consider a probability space (Ω, μ) and a filtered A-lattice $\{\mathcal{F}_\alpha\}$. Let $\mathcal{I}$ be an instance of General A-CSP($\mathcal{F}_\alpha$, α_0) and let Φ be its objective functions. Then for every subset $S \subset \{1, \ldots, n\}$ of size at most k, $\Phi_S \in_R \mathcal{F}_{\tilde{\alpha}}$; for every subset S of size greater than k, $\Phi_S = 0$.*

Furthermore, the Efron—Stein decomposition $\{\Phi_S\}$ of Φ can be computed in time $O(m)$.

D.5 Compactness Properties of Filtered A-Lattices

We now prove counterparts of Claim 4.3 and Lemma A.1 for General A-CSP.

Claim D.9 (cf. Claim 4.3)**.** *Consider a probability space (Ω, μ). Let $\mathcal{F}$ be an A-lattice of functions of arity k on Ω. There exists a positive number β such that for every function $f \in \mathcal{F}$, $\mathbb{E}\left[f^2\right] = \mathbb{E}\left[f(X_1, \ldots, X_k)^2\right] \geq \beta$.*

Proof. Let $f_1, \ldots, f_r$ be the basis of lattice $\mathcal{F}$. Consider the linear span Q of functions $f_1, \ldots, f_r$ (the set of all linear combinations with real coefficients). Vector space Q is finite dimensional. Let $Q_1 = \{\sum a_i f_i : \max_i |a_i| = 1\}$. Note that Q_1 is a compact set. All functions in Q_1 are non-zero (since $f_1, \ldots, f_r$ are linearly independent), and, therefore, $\mathbb{E}\left[g^2\right] > 0$ for every $g \in Q_1$. Since Q_1 is compact, $\min_{g \in Q_1} \mathbb{E}\left[g^2\right] = \inf_{g \in Q_1} \mathbb{E}\left[g^2\right] > 0$. Denote $\beta = \min_{g \in Q_1} \mathbb{E}\left[g^2\right]$.

Now consider a non-zero function $f \in \mathcal{F}$. Write $f = \sum_{i=1}^r a_i f_i$. Let $M = \max_i |a_i|$. Since $f \neq 0$ and all coefficients a_i are integer, $M \geq 1$. Note that $f/M \in Q_1$. We have,

$$\mathbb{E}\left[f^2\right] = M^2 \mathbb{E}\left[(f/M)^2\right] \geq \beta M^2 \geq \beta,$$

as required. $\square$

Lemma D.10 (cf. Lemma A.1)**.** *Consider a probability space (Ω, μ). Let $\mathcal{F}$ be an A-lattice of functions of arity k on Ω. There exists a constant C such that the following holds. Let f_1, f_2, f_3, f_4 be functions of arity at most k; each of them depends on a subset of variables $\{x_1, \ldots, x_n\}$. Assume that $f_1, f_2, f_3, f_5 \in_R \mathcal{F}$ (see Definition D.2). Then*

$$\mathbb{E}\left[f_1 f_2 f_3 f_4\right] \leq C \left(\mathbb{E}\left[f_1^2\right] \mathbb{E}\left[f_2^2\right] \mathbb{E}\left[f_3^2\right] \mathbb{E}\left[f_4^2\right]\right)^{1/2}.$$

Proof. Each function f_i depends on at most k variables among $x_1, \ldots, x_n$. So without loss of generality, we may assume that they depend on a subset of $x_1, \ldots, x_{4k}$. For every subset T of $\{x_1, \ldots, x_{4k}\}$ of size k, let $\mathcal{F}^{(T)}$ be the set of functions f that depend only on variables in T such that $f \in_R \mathcal{F}$. Let $\mathcal{Q}^{(T)}$ be the linear span of $\mathcal{F}^{(T)}$. Since $\mathcal{F}$ is finite dimensional (by the definition of an A-lattice), all vector spaces $\mathcal{Q}^{(T)}$ are also finite dimensional. Finally, let $\mathcal{Q}$ be the set of all functions of $x_1, \ldots, x_{4k}$ of the form $\sum_T f^{(T)}$, where the summation is over all subsets of $x_1, \ldots, x_{4k}$ of size k, and $f^{(T)} \in \mathcal{F}^{(T)}$. Note that $\mathcal{Q}$ is also a finite dimensional vector space of functions.

Now let $\mathcal{Q}_1 = \{h \in \mathcal{Q} : \|h\|_2 = 1\}$. Note that $\mathcal{Q}_1$ is a compact set. Therefore, the continuous function $W(g_1, g_2, g_3, g_4) = \mathbb{E}[g_1 g_2 g_3 g_4]$ is bounded when $g_1, g_2, g_3, g_4 \in \mathcal{Q}_1$. Denote its maximum by C.

Let $g_i = f_i / \|f_i\|_2$. We have,

$$\mathbb{E}\left[\prod_{i=1}^4 f_i\right] = \mathbb{E}\left[\prod_{i=1}^4 g_i\right] \cdot \prod_{i=1}^4 \|f_i\|_2 \le C \cdot \prod_{i=1}^4 \|f_i\|_2,$$

as required. $\qquad\square$

D.6 Variance of A-CSP Objective

We now prove a counterpart of Theorem 4.5 for General A-CSP$(\mathcal{F}_\alpha, k, \alpha)$.

Lemma D.11 (cf. Theorem 4.5). *Consider a probability space (Ω, μ). Let $\{\mathcal{F}_\alpha\}$ be a filtered A-lattice and α_0 be an integer. There exists a number $\beta > 0$, which depends only on $\{\mathcal{F}_\alpha\}$ and α_0 such that the following holds. Let $\mathcal{I}$ be an instance of General A-CSP$(\mathcal{F}_\alpha, \alpha_0)$ and σ be a parameter. Either $\mathcal{I}$ has a kernel on at most $(k/\beta)\sigma^2$ variables or $\mathrm{Var}[\Phi] \ge \sigma^2$. Moreover, there is an algorithm that either finds a kernel on at most $(k/\beta)\sigma^2$ variables or certifies that $\mathrm{Var}[\Phi] \ge \sigma^2$. The algorithm runs in time $O(m+n)$, where n is the number of variables and m is the number of constraints.*

Proof. Let $\tilde{\alpha}$ be as in Lemma D.7. Since $\mathcal{F}_{\tilde{\alpha}}$ is an A-lattice, by Claim D.9, there exists $\beta > 0$ such that $\|f\|_2^2 \ge \beta$ for every non-zero $f \in \mathcal{F}_{\tilde{\alpha}}$. Note that β does not depend on n and t.

Consider the Efron—Stein decomposition of Φ. Let $V' = \bigcup_{S:\Phi_S \ne 0} \{x_i : i \in S\}$. Function Φ depends only on variables in V'. Therefore, the restriction of $\mathcal{I}$ to variables in V' is a kernel for $\mathcal{I}$. Let $\nu = |V'|$. If $\nu < (k/\beta)\sigma^2$, then we are done. So let us assume that $\nu \ge (k/\beta)\sigma^2$. There are at least ν/k non-empty sets S with $\Phi_S \ne 0$ since each such set S contributes at most k variables to V'. Note that $\mathbb{E}[\Phi_S] = 0$ for $S \ne \varnothing$ and hence $\mathrm{Var}[\Phi_S] = \mathbb{E}[\Phi_S^2]$. Since $\Phi_S \in \mathcal{F}_{\tilde{\alpha}}$, $\mathrm{Var}[\Phi_S] = \mathbb{E}[\Phi_S^2] \ge \beta$, if $\Phi_S \ne 0$ and $S \ne \varnothing$. We have,

$$\mathrm{Var}[\Phi] = \sum_S \mathrm{Var}[\Phi_S] \ge |\{S \ne \varnothing : \Phi_S \ne 0\}|\beta \ge (\nu/k)\beta \ge \sigma^2,$$

as required.

Note that we can compute the Efron—Stein decomposition of Φ in time $O(m+n)$ and then find the set V' in time $O(m+n)$. If $|V'| < (k/\beta)\sigma^2$, we output the restriction of $\mathcal{I}$ to V' (which we compute in time $O(m+n)$). Otherwise, we output that $\mathrm{Var}[\Phi] \ge \sigma^2$. $\qquad\square$

D.7 Proof of Theorem D.6

We are ready to prove Theorem D.6.

Proof. Let β be as in Lemma D.11 and C be as in Lemma D.10. Let $\sigma^2 = 4 \cdot 81^k C t^2$. Denote $f = \Phi - \mathbb{E}[\Phi] = \Phi - AVG$. Note that $\mathrm{ess\,sup}_{x \in \Omega^n} f(x_1, \ldots, x_n) = MAX - AVG$.

By Lemma D.11, either $\mathcal{I}$ has a kernel on at most $(k/\beta)\sigma^2 = (4 \cdot 81^k \cdot kC/\beta)t^2$ variables or $\mathrm{Var}[f] \ge \sigma^2$. In the former case, we output the kernel, and we are done. In the latter case, we show that $OPT - AVG \ge t$.

18

Assume that $\mathrm{Var}\,[f] \geq \sigma^2$. By Theorem 1.4 and Lemma D.10, $\|f\|_4^4 \leq 81^k C \|f\|_2^2$. By Theorem 5.1, $\mathrm{Pr}\left(f \geq \sigma/(2 \cdot 9^k \sqrt{C})\right) > 0$, and hence $MAX - AVG \geq \sigma/(2 \cdot 9^k \sqrt{C}) \geq t$.

The algorithm only executes the algorithm from Lemma D.11, so its running time is $O(m + n)$. $\qquad \square$

E   Piecewise Polynomial Predicates

In this section, we present an interesting example of a filtered A-lattice, the set of piecewise polynomial functions. As a corollary, we get that the problem of maximizing the objective over average for a CSP with piecewise polynomial functions is fixed-parameter tractable.

Definition E.1. *Let us say that a subset P of $[-1, 1]^k$ is b-polyhedral if it is defined by a set of linear inequities on $x_1, ..., x_k$, in which all coefficients are bounded by b in absolute value. In other words, P is a b-polyhedral set if for some t there exist a $k \times t$ matrix A and vector c (with t coordinates) such that $P = \{x : Ax < c\}$ (here, the inequality $Ax < c$ is understood coordinate-wise), and every entry of A and coordinate of c is bounded by b in absolute value. We denote the indicator function of a polyhedral set P by I_P.*

Definition E.2. *We denote the set of polynomials $f(x_1, \ldots, x_k)$ with real coefficients of degree at most d by $\mathbb{R}_{\leq d}[x_1, \ldots, x_k]$; we denote the set of polynomials $f(x_1, \ldots, x_k)$ with integer coefficients of degree at most d by $\mathbb{Z}_{\leq d}[x_1, \ldots, x_k]$.*

Definition E.3. *We say that a function $f(x_1, ..., x_k) : [-1, 1]^k \to \mathbb{R}$ is piecewise polynomial on polyhedral sets or (d, b)-PPP if f is the sum of terms of the form $g(x_1, ..., x_k)I_P(x_1, ..., x_k)$, where $g \in \mathbb{Z}_{\leq d}[x_1, \ldots, x_k]$ and P is a b-polyhedral set.*

We note that every (d, b)-PPP function can be written in the following "canonical form". Consider all hyperplanes in $\mathbb{R}^k$ of the form $\langle a, x \rangle = c$, in which $a, c \in \{-b, \ldots, b\}^k$. They partition $[-1, 1]^d$ into polyhedrons. We call these polyhedrons elementary polyhedrons and denote the set of all elementary polyhedrons by $\mathcal{P}_{elem}$. Note that each b-polyhedral set is a union of elementary polyhedrons. Thus we can write every (d, b)-PPP function f as follows:

$$f(x_1, \ldots, x_k) = \sum_{P \in \mathcal{P}_{elem}} I_P(x_1, \ldots, x_k) g_P(x_1, \ldots, x_k), \tag{9}$$

where $g_P \in \mathbb{Z}_{\leq d}[x_1, \ldots, x_k]$.

Theorem E.4. *Let $\Omega = [-1, 1]$ and μ be the uniform measure on $[-1, 1]$. Let*

$$\mathcal{F}_\alpha = \{f : (\alpha! f) \text{ is an } (\alpha, \alpha)\text{-PPP function of variables } x_1, \ldots, x_k\}.$$

Then $\mathcal{F}_\alpha$ is a filtered A-lattice of functions.

Proof. First, we prove that each set $\mathcal{F}_\alpha$ is an A-lattice. It follows from (9) that $\mathcal{F}_\alpha$ is a lattice with basis $I_P(x_1, \ldots, x_k) g(x_1, \ldots, x_k)/\alpha!$, where $P \in \mathcal{P}_{elem}$ and $g(x_1, \ldots, x_k)$ is a monomial of degree at most α (i.e., g is of the form $x_1^{r_1} \ldots x_k^{r_k}$). Since every monomial g is bounded on $[-1, 1]^k$, every basis function is bounded, and, therefore, all functions in $\mathcal{F}_\alpha$ are bounded. The definition of $\mathcal{F}_\alpha$ is symmetric with respect to $x_1, \ldots, x_k$, hence if we permute the arguments of any function $f \in \mathcal{F}_\alpha$, we get a function in $\mathcal{F}_\alpha$.

Now we show that $\mathcal{F}_\alpha$ is a filtered A-lattice. The inclusion $\mathcal{F}_\alpha \subset \mathcal{F}_{\alpha'}$ for $\alpha \leq \alpha'$ is immediate. It remains to show that for every α there exists α' such that A_i maps $\mathcal{F}_\alpha$ to $\mathcal{F}_{\alpha'}$. Let $a = \binom{2\alpha}{\alpha}$ and $\alpha' = a^{\alpha+1}$ (we note that, in fact, we can choose a much smaller value of α'; however, we use this value to simplify the exposition). Observe that all integer numbers between 1 and α divide a.

It is sufficient to prove that A_i sends every basis function

$$f(x_1, \ldots, x_k) = I_P(x_1, \ldots, x_k) g(x_1, \ldots, x_k)/\alpha!$$

to $\mathcal{F}_{\alpha'}$. Moreover, since the set of functions $\mathcal{F}_\alpha$ is invariant under permutation of function arguments, we may assume without loss of generality that $i = k$. Denote $g = x_1^{r_1} \ldots x_k^{r_k}$, where $r_1 + \ldots r_k \le \alpha$. Consider the set of linear inequalities L that define polyhedron P. All coefficients in each of the inequalities are bounded by α in absolute value. Let L_0 be those inequalities that do not depend on x_k and L_1 be those that do depend on x_k. We rewrite every inequality in L_1 as follows. Consider an inequality in L_1. Let λ be the coefficient of x_k in it. We multiply the inequality by $a/\lambda \in \mathbb{Z}$, and if $\lambda < 0$, we change the comparison sign in the inequality to the opposite. Finally, we move all terms in the inequality other than ax_k to the right hand side. We get an equivalent inequality of the form either $ax_k > l(1, x_1, \ldots, x_{k-1})$ or $ax_k < u(1, x_1, \ldots, x_{k-1})$, where l and u are linear functions with integer coefficients bounded by αa in absolute value. Denote the inequalities of the form $ax_k > l(1, x_1, \ldots, x_{k-1})$ by

$$ax_k > l_1(1, x_1, \ldots, x_{k-1}), ax_k > l_2(1, x_1, \ldots, x_{k-1}), \ldots, ax_k > l_p(1, x_1, \ldots, x_{k-1}),$$

and the inequalities of the form $ax_k < u(1, x_1, \ldots, x_{k-1})$ by

$$ax_k < u_1(1, x_1, \ldots, x_{k-1}), ax_k < u_2(1, x_1, \ldots, x_{k-1}), \ldots, ax_k < u_q(1, x_1, \ldots, x_{k-1}).$$

Let M_0 be the set of points $x = (x_1, \ldots, x_k)$ such that $l_{j'}(x_1, \ldots, x_k) = l_{j''}(x_1, \ldots, x_k)$ or $u_{j'}(x_1, \ldots, x_k) = u_{j''}(x_1, \ldots, x_k)$ for some $j' \ne j''$. Note that M_0 has measure 0.

Define pq polyhedrons $P_{j_1 j_2}$ in $\mathbb{R}^{k-1}$. For $j_1 \in \{1, \ldots, p\}$ and $j_2 \in \{1, \ldots, q\}$, let $P_{j_1 j_2}$ be the polyhedron defined by the following inequalities:

1. all inequalities in L_0,

2. inequality $l_{j_1}(1, x_1, \ldots, x_{k-1}) < u_{j_2}(1, x_1, \ldots, x_{k-1})$,

3. inequalities $l_j(1, x_1, \ldots, x_{k-1}) < l_{j_1}(1, x_1, \ldots, x_{k-1})$ for every $j \ne j_1$,

4. inequalities $u_j(1, x_1, \ldots, x_{k-1}) > u_{j_2}(1, x_1, \ldots, x_{k-1})$ for every $j \ne j_2$.

Inequalities in items 2–4 are equivalent to the following condition (except for points in M_0):

$$\max_j l_j(1, x_1, \ldots, x_{k-1}) = l_{j_1}(1, x_1, \ldots, x_{k-1}) < u_{j_2}(1, x_1, \ldots, x_{k-1}) = \min_j u_j(1, x_1, \ldots, x_{k-1}). \quad (10)$$

Note that if $(x_1, \ldots, x_k) \in P \setminus M_0$ then $(x_1, \ldots, x_{k-1}) \in P_{j_1 j_2}$ for

$$j_1 = \arg\max_j l_j(1, x_1, \ldots, x_{k-1}) \text{ and } j_2 = \arg\min_j u_j(1, x_1, \ldots, x_{k-1}).$$

Also note that polyhedrons $P_{j_1 j_2}$ are disjoint. Now let

$$h_{j_1 j_2}(x_1, \ldots, x_{k-1}) = \frac{1}{2\alpha! a^{r_k+1}(r_k + 1)} (u_{j_2}(1, x_1, \ldots, x_{k-1})^{r_k+1} - l_{j_1}(1, x_1, \ldots, x_{k-1})^{r_k+1}) x_1^{r_1} \ldots x_{k-1}^{r_{k-1}}.$$

Let $x' = (x_1, \ldots, x_{k-1})$ be a point in $[-1, 1]^{k-1}$. Consider two cases.

Case 1. First, assume that $x' \in P_{j_1 j_2}$ for some j_1 and j_2. Then

$$A_k f(x') = \frac{1}{2\alpha!} \int_{-1}^{1} I_P(x_1, \ldots, x_k) g(x_1, \ldots, x_k) dx_k.$$

The point $(x_1, \ldots, x_k)$ satisfies all inequalities in L_0 since $x' \in P_{j_1 j_2}$. Hence, $I_P(x_1, \ldots, x_k) = 1$ if and only if it satisfies all inequalities in L_1, which are equivalent to

$$\max_j l_j(1, x_1, \ldots, x_{k-1}) \le a x_k \le \min_j u_j(1, x_1, \ldots, x_{k-1}).$$

Combining this with (10), we get that $I_P(x_1, \ldots, x_{k-1}) = 1$ if and only if

$$x_k \in \left[\frac{1}{a} l_{j_1}(1, x_1, \ldots, x_{k-1}), \frac{1}{a} u_{j_2}(1, x_1, \ldots, x_{k-1}) \right].$$

Therefore,

$$
\begin{aligned}
A_k f(x') &= \frac{1}{2\alpha!} \int_{l_{j_1}(1,x_1,\ldots,x_{k-1})/a}^{u_{j_2}(1,x_1,\ldots,x_{k-1})/a} x_1^{r_1} \ldots x_k^{r_k} dx_k \\
&= \frac{1}{2\alpha! a^{r_k+1}(r_k+1)} (u_{j_2}(1, x_1, \ldots, x_{k-1})^{r_k+1} - l_{j_1}(1, x_1, \ldots, x_{k-1})^{r_k+1}) x_1^{r_1} \ldots x_{k-1}^{r_{k-1}} \\
&= h_{j_1 j_2}(x').
\end{aligned}
$$

Case 2. Now assume that $x' \notin P_{j_1 j_2}$ for every j_1 and j_2. Then there is no x_k such that $(x_1, \ldots, x_k) \in P \setminus M_0$. Therefore,

$$A_k f(x') = \frac{1}{2\alpha!} \int_{-1}^{1} I_P(x_1, \ldots, x_k) g(x_1, \ldots, x_k) dx_k = \frac{1}{2\alpha!} \int_{-1}^{1} 0 \cdot g(x_1, \ldots, x_k) dx_k = 0.$$

(The equality holds on a set of full measure; see Remark D.1.)

We conclude that

$$A_k f(x') = \sum_{j_1, j_2} I_{P_{j_1 j_2}} h_{j_1 j_2}(x').$$

All coefficients in the inequalities that define $P_{j_1 j_2}$ are bounded by $2\alpha a$ in absolute value, and $2\alpha! a^{r_k+1}(r_k + 1) h(x') \in \mathbb{Z}_{\le \alpha+1}$. Therefore, $2\alpha! a^{r_k+1}(r_k + 1) A_k f$ is an $(\alpha+1, 2\alpha a)$-PPP function. Thus $A_k f \in \mathcal{F}_{\alpha'}$. $\square$

From Theorems D.6 and E.4, we get the following corollary.

Corollary E.5. *For every k, d and b, there is an algorithm that given an instance of a constraint satisfaction problem on n variables $x_1, \ldots, x_n$ with m real-valued constraints, each of which is a (d, b)-PPP function of arity k, and a parameter t, either finds a kernel on at most κt^2 variables or certifies that $OPT \ge AVG + t$. The algorithm runs in time $O(m + n)$. (The coefficient κ and the coefficient in the O-notation depend only on k, d, and b).*

Proof. Let $\alpha = \max(d, b)$. We apply Theorem D.6 to filtered A-lattice $\mathcal{F}_\alpha$ from Theorem E.4 and get the corollary. $\square$

Since every constraint in a (k, b)-LP CSP problem is a $(0, b)$-PPP function of arity k (see Definition 1.5), we get the following corollary.

Corollary E.6. *For every k and b, there is an algorithm that given an instance of (k, b)-LP CSP either finds a kernel on at most κt^2 variables or certifies that $OPT \geq AVG + t$. The algorithm runs in time $O(m + n)$ (The coefficient κ and the coefficient in the O-notation depend only on k and b).*

Remark E.1. *Note that for an instance of (k, b)-LP CSP, we have*

$$OPT = \max_{x_1,\ldots,x_n \in [-1,1]} \Phi(x_1,\ldots,x_n) = \operatorname*{ess\,sup}_{x_1,\ldots,x_n \in [-1,1]} \Phi(x_1,\ldots,x_n),$$

since all LP constraints are strict. If we were to use non-strict "less-than-or-equal-to" and "greater-than-or-equal-to" LP constraints, we would have to define OPT as $\operatorname{ess\,sup}_{x_1,\ldots,x_n \in [-1,1]} \Phi(x_1,\ldots,x_n)$, and not as $\max_{x_1,\ldots,x_n \in [-1,1]} \Phi(x_1,\ldots,x_n)$, since, in general, $\operatorname{ess\,sup}_{x_1,\ldots,x_n \in [-1,1]} \Phi(x_1,\ldots,x_n)$ might not be equal to $\max_{x_1,\ldots,x_n \in [-1,1]} \Phi(x_1,\ldots,x_n)$. For example, consider an instance of $(2, 1)$-LP CSP with two constraints $x_1 \leq x_2$ and $x_2 \leq x_1$; we have

$$OPT = \operatorname*{ess\,sup}_{x_1,\ldots,x_n \in [-1,1]} \Phi(x_1,\ldots,x_n) = 1,$$

but

$$\max_{x_1,\ldots,x_n \in [-1,1]} \Phi(x_1,\ldots,x_n) = 2.$$

(The maximum is attained on a set of measure 0, where $x_1 = x_2$.)