THEORY OF COMPUTING, Volume 18 (3), 2022, pp. 1–29
www.theoryofcomputing.org

SPECIAL ISSUE: APPROX-RANDOM 2016

Near-Optimal NP-Hardness of Approximating MAX k-CSP$_R$

Pasin Manurangsi* Preetum Nakkiran† Luca Trevisan‡

Received April 23, 2017; Revised November 7, 2021; Published February 14, 2022

Abstract. We prove almost optimal hardness for MAX k-CSP$_R$. In MAX k-CSP$_R$, we are given a set of constraints, each of which depends on at most k variables. Each variable can take any value from $1,2,\ldots,R$. The goal is to find an assignment to variables that maximizes the number of satisfied constraints.

We show that, for any $k \geq 2$ and $R \geq 16$, it is NP-hard to approximate MAX k-CSP$_R$ to within factor $k^{O(k)}(\log R)^{k/2}/R^{k-1}$. In the regime where $3 \leq k = o(\log R/\log\log R)$, this ratio improves upon Chan's $O(k/R^{k-2})$ factor NP-hardness of approximation of MAX k-CSP$_R$ (J. ACM 2016). Moreover, when $k = 2$, our result matches the best known hardness result of Khot, Kindler, Mossel and O'Donnell (SIAM J. Comp. 2007). We remark here that NP-hardness of an approximation factor of $2^{O(k)}\log(kR)/R^{k-1}$ is implicit in the (independent) work of Khot and Saket (ICALP 2015), which is better than our ratio for all $k \geq 3$.

In addition to the above hardness result, by extending an algorithm for MAX 2-CSP$_R$ by Kindler, Kolla and Trevisan (SODA 2016), we provide an $\Omega(\log R/R^{k-1})$-approximation

A conference version of this paper appeared in the Proceedings of the 19th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'16) [22]. This material is based upon work supported by the National Science Foundation under Grant No. CCF 1540685.

*Work done while the author was at UC Berkeley.
†Work done while the author was at UC Berkeley.
‡Work done while the author was at UC Berkeley.

ACM Classification: F.2.2

AMS Classification: 68Q17, 68W25

Key words and phrases: constraint satisfaction, hardness of approximation, approximation algorithm

 DOI: 10.4086/toc.2022.v018a003

algorithm for MAX k-CSP$_R$. Thanks to Khot and Saket's result, this algorithm is tight up to a factor of $O(k^2)$ when $k \leq R^{O(1)}$. In comparison, when $3 \leq k$ is a constant, the previously best known algorithm achieves an $O(k/R^{k-1})$-approximation for the problem, which is a factor of $O(k \log R)$ from the inapproximability ratio in contrast to our gap of $O(k^2)$.

1 Introduction

The Maximum Constraint Satisfaction Problem (MAX CSP) is an optimization problem where the inputs are a set of variables, an alphabet, and a set of constraints. Each variable can be assigned any symbol from the alphabet and each constraint depends only on the assignment to a subset of variables. The goal is to find an assignment to the variables that maximizes the number of satisfied constraints.

Many natural optimization problems, such as MAX CUT, MAX k-CUT and MAX k-SAT, can be formulated as MAX CSP. In addition, MAX CSP has been shown to help approximate other seemingly-unrelated problems such as DENSEST k-SUBGRAPH [5]. Due to this, MAX CSP has long been studied by the approximation algorithms community [28, 14, 6, 21, 20, 12]. Furthermore, its relation to PCPs ensures that MAX CSP is also well-studied on the hardness-of-approximation side [26, 10, 27, 16, 1, 13, 11, 4].

The main focus of this paper is on MAX k-CSP$_R$, a family of MAX CSPs where the alphabet is of size R and each constraint depends on only k variables. On the hardness-of-approximation side, most early work focused on Boolean MAX k-CSP. Samorodnitsky and Trevisan first showed that MAX k-CSP$_2$ is NP-hard to approximate to within a factor of $2^{O(\sqrt{k})}/2^k$ [26]. Engebretsen and Holmerin later improved the implicit constant factor in the exponent $O(\sqrt{k})$ but still obtained NP-hardness of a factor of $2^{O(\sqrt{k})}/2^k$ [11]. To break this barrier, Samorodnitsky and Trevisan proved a Unique-Games-hardness (UG-hardness) [15] [1] result; they achieved UG-hardness of an approximation ratio of $O(k/2^k)$, which is tight up to a constant for the Boolean case [27]. Chan later showed that NP-hardness of approximation at this ratio can be achieved, settling the approximability of MAX k-CSP$_2$ [4].

Unlike in the Boolean case, the approximability of MAX k-CSP$_R$ when $R > 2$ is still not resolved. For this case, Engebretsen showed NP-hardness of approximation ratio $R^{O(\sqrt{k})}/R^k$ [10]. UG-hardness of approximation within a factor of $O(kR/R^{k-1})$ was proven by Austrin and Mossel [1] and, independently, by Guruswami and Raghavendra [13]. For the case $k = 2$, results by Khot et al. [16] implicitly demonstrate UG-hardness of approximation within $O(\log R/R)$. This was made explicit in [20]. In the light of the recent resolution of the 2-to-1 Conjecture [8, 17], this inapproximability result is now an NP-hardness result. Moreover, Austrin and Mossel proved UG-hardness of an approximation ratio of $O(k/R^{k-1})$ for infinitely many values of k [1], but only under the condition $k \geq R$. Recently, Chan was able to upgrade these UG-hardness results to NP-hardness [4]. More specifically, Chan showed NP-hardness of approximation within a factor of $O(kR/R^{k-1})$ for every k, R and approximation within a factor of $O(k/R^{k-1})$ for every $k \geq R$. Due to an approximation algorithm with matching approximation ratio by Makarychev and Makarychev [21], Chan's result established tight hardness of approximation for $k \geq R$.

[1] For any $r \in (0,1)$, we say that MAX k-CSP$_R$ is UG-hard to approximate to within an r factor if, for some $0 < \gamma < \zeta < 1$ and $0 < \theta \leq 1$, there is a polynomial-time reduction from the problem of deciding whether a given unique game has value at least ζ or at most γ to the problem of deciding whether a given MAX k-CSP$_R$ instance has value at least θ or at most $\theta \cdot r$. (For the definition of Unique Games, please refer to Section 2.2.) We remark that such a UG-hardness result implies that, if the Unique Games Conjecture holds, then MAX k-CSP$_R$ is NP-hard to approximate to within an r factor.

On the other hand, when $k < R$, Chan's result gives $O(kR/R^{k-1})$ hardness of approximation whereas the best known approximation algorithm achieves only $\Omega(k/R^{k-1})$ approximation ratio [21, 12]. In an attempt to bridge this gap, we prove the following result.

Theorem 1.1 (Main Theorem). *It is NP-hard to approximate* MAX k-CSP$_R$ *to within a* $k^{O(k)}(\log R)^{k/2}/R^{k-1}$ *factor, for any* $k \geq 2$ *and* $R \geq 16$.

Remark 1.2. We remark that the $k^{O(k)}(\log R)^{k/2}/R^{k-1}$ factor in Theorem 1.1 can be greater than one when $R \leq k^c$ for some constant c. In this regime, the hardness result is vacuous.

Range of k,R	NP-Hardness	UG-Hardness	Approximation	References
$k=2$	$O\left(\frac{\log R}{R}\right)$	–	$\Omega\left(\frac{\log R}{R}\right)$	[16, 20]
$3 \leq k < \Theta(\log R)$	$\frac{2^{O(k)}\log(kR)}{R^{k-1}}$	$O\left(\frac{k^2\log(kR)}{R^{k-1}}\right)$	$\Omega\left(\frac{k}{R^{k-1}}\right)$	[19, 21, 12]
$\Theta(\log R) \leq k < R$	$O\left(\frac{k}{R^{k-2}}\right)$	$O\left(\frac{k^2\log(kR)}{R^{k-1}}\right)$	$\Omega\left(\frac{k}{R^{k-1}}\right)$	[4, 19, 21, 12]
$R \leq k$	$O\left(\frac{k}{R^{k-1}}\right)$	–	$\Omega\left(\frac{k}{R^{k-1}}\right)$	[4, 21]
Any $k \geq 2, R \geq 16$	$\frac{k^{O(k)}(\log R)^{k/2}}{R^{k-1}}$	–	$\Omega\left(\frac{\log R}{R^{k-1}}\right)$	this paper

Figure 1: Comparison between our results and previous results. For each range of k,R, we list the previous best hardness-of-approximation factors and the previous best approximation ratio. As mentioned earlier, Khot and Saket's result [19] subsumes our result for every value of k and R whereas our approximation algorithm improves on the previously known algorithm when $3 \leq k = o(\log R)$.

When $k = o(\log R/\log\log R)$, our result improves upon Chan's result in terms of the ratio. As noted in the Abstract, Khot and Saket [19] independently showed UG-hardness of approximation within a factor of[2] $O(k^2\log(kR)/R^{k-1})$ for the problem for any k and R [19]. Furthermore, they can achieve[3] NP-hardness of factor $2^{O(k)}\log(kR)/R^{k-1}$. Both of these ratios are better than the ratio we achieve in Theorem 1.1.

As mentioned earlier, there has also been a long line of work in approximation algorithms for MAX k-CSP$_R$. In the Boolean case, Trevisan first showed a polynomial-time approximation algorithm with approximation ratio $2/2^k$ [28]. Hast later improved the ratio to $\Omega(k/(2^k\log k))$ [14]. Charikar, Makarychev and Makarychev then came up with an $\Omega(k/2^k)$-approximation algorithm [6]. As stated when discussing hardness of approximation of MAX k-CSP$_2$, this approximation ratio is tight up to a constant factor.

For the non-Boolean case, Charikar, Makarychev, and Makarychev's algorithm achieves a ratio of $\Omega(k\log R/R^k)$ for MAX k-CSP$_R$. Makarychev, and Makarychev later improved the approximation

[2] It should be noted that, if the results as stated in [19] are invoked directly, then the hardness of approximation ratio one can get is $O(k^3\log R/R^{k-1})$, which is a factor of $O(k)$ worse than what is stated here. This is because Khot and Saket prove the more general statement that any integrality gap for the standard LP relaxation of MAX k-CSP$_R$ can be translated into a UG-hardness of approximation result at the loss of a factor of $O(k^3\log R)$ in the ratio. However, it is not hard to see that, when one is only interested in hardness of approximation (and not the LP), then a factor of k here can be removed. Please refer to Appendix B for more details regarding this.

[3] In [19], only UG-hardness was stated, but NP-hardness (with a worse ratio) is now possible due to the resolution of the 2-to-1 Conjecture [17]. We briefly discuss this in Section 1.1 and Appendix B.

ratio to $\Omega(k/R^{k-1})$ when $k = \Omega(\log R)$ [21]. Additionally, the algorithm was extended by Goldshlager and Moshkovitz to achieve the same approximation ratio for any k, R [12]. On this front, we show the following result.

Theorem 1.3. *There exists a polynomial-time $\Omega(\log R/R^{k-1})$-approximation algorithm for* MAX k-CSP$_R$.

In comparison to the previously known algorithms, our algorithm gives a better approximation ratio when $3 \le k = o(\log R)$. We remark that our algorithm is just a simple extension of Kindler, Kolla and Trevisan's polynomial-time $\Omega(\log R/R)$-approximation algorithm for Max 2-CSP$_R$ [20].

1.1 The role of the 2-to-1 Theorem

In the conference version of this paper [22], our hardness result (Theorem 1.1) was shown conditional on the Unique Games Conjecture (UGC) or the d-to-1 Conjecture [15]. Recent work [9, 8, 2, 17] confirmed the imperfect completeness version of the 2-to-1 conjecture, and this breakthrough upgrades our hardness results to NP-hardness without any additional assumptions. We have incorporated this development into the current version.

The situation is similar for the result of Khot and Saket [19]. In their case, the 2-to-1 Theorem implies NP-hardness of a somewhat worse inapproximability ratio than stated in their paper, but still stronger than our bound. We briefly discuss this in Appendix B.

We emphasize that *Chan's proofs of his NP-hardness results [4] do not depend on the d-to-1 Theorem.*

1.2 Summary of techniques

Our reduction from Unique Games to MAX k-CSP$_R$ follows the reduction of [16] for MAX 2-CSPs. We construct a k-query PCP using a Unique Games "outer verifier," and then design a k-query Long Code test as an "inner verifier." For simplicity, let us think of k as a constant. We essentially construct a k-query *Dictator-vs.-Quasirandom* test for functions $f : [R]^n \to [R]$, with completeness $\Omega(1/(\log R)^{k/2})$ and soundness $O(1/R^{k-1})$. This test satisfies the following completeness and soundness conditions.

- (Completeness) If f is a dictator function, i. e., $f(x) = x_j$ for some coordinate $j \in [n]$, then the test passes with a large probability.

- (Soundness) If f is a balanced function with small low-degree influences, then the test passes with only a small probability.

Our test is structurally similar to the 2-query "noise stability" tests of [16]: first we pick a random $z \in [R]^n$, then we pick k weakly-correlated queries $x^{(1)}, \dots, x^{(k)}$ by choosing each $x^{(i)} \in [R]^n$ as a noisy copy of z, i. e., each coordinate $(x^{(i)})_j$ is chosen as z_j with some probability ρ or uniformly at random otherwise. We accept iff $f(x^{(1)}) = f(x^{(2)}) = \cdots = f(x^{(k)})$. The key technical step is our analysis of the soundness of this test. We need to show that if f is a balanced function with small low-degree influences, then the test passes with probability $O(1/R^{k-1})$. Intuitively, we would like to say that for high enough noise, the values $f(x^{(i)})$ are roughly independent and uniform, so the test passes with probability around $1/R^{k-1}$. To formalize this intuition, we utilize the *Invariance Principle* and *Hypercontractivity*.

More precisely, if we let $f^i(x) : [R]^n \to \{0,1\}$ be the indicator function for $f(x) = i$, then the test accepts iff $f^i(x^{(1)}) = \cdots = f^i(x^{(k)}) = 1$ for some $i \in [R]$. For each $i \in [R]$, this probability can be written as the expectation of the product: $\mathbb{E}[f^i(x^{(1)})f^i(x^{(2)}) \ldots f^i(x^{(k)})]$. Since $x^{(i)}$'s are chosen as noisy copies of z, this expression is related to the k-th norm of a noisy version of f^i. Thus, our problem is reduced to bounding the k-norm of a noisy function $\widetilde{f^i} : [R]^n \to [0,1]$, which has bounded one-norm $\mathbb{E}[\widetilde{f^i}] = 1/R$ since f is balanced. To arrive at this bound, we first apply the Invariance Principle, which essentially states that a low-degree low-influence function on $[R]^n$ behaves on random input similarly to its "Boolean analogue" over domain $\{\pm 1\}^{nR}$. Here "Boolean analogue" refers to the function over $\{\pm 1\}^{nR}$ with matching Fourier coefficients.

Roughly speaking, now that we have transfered to the Boolean domain, we are left to bound the k-norm of a noisy function on $\{\pm 1\}^{nR}$ based on its one-norm. This follows from Hypercontractivity in the Boolean setting, which bounds a higher norm of any noisy function on Boolean domain in terms of a lower norm.

The description above hides several technical complications. For example, when we pass from a function $[R]^n \to [0,1]$ to its "Boolean analogue" $\{\pm 1\}^{nR} \to \mathbb{R}$, the range of the resulting function is no longer bounded to $[0,1]$. This, along with the necessary degree truncation, means we must be careful when bounding norms. Further, Hypercontractivity only allows us to pass from k-norms to $(1+\varepsilon)$-norms for small ε, so we cannot use the known 1-norm directly. For details on how we handle these issues, see Section 3. This allows us to prove soundness of our dictator test without passing through results on Gaussian space, as was done to prove the "Majority is Stablest" conjecture [24] at the core of the [16] 2-query dictator test.

Combining the above test with a standard reduction from Unique Games [16] would immediately give a UG-hardness result. However, we can get NP-hardness because it suffices for us to use Unique Games with completeness 1/2 from [8, 17]. This concludes the overview of our proof of Theorem 1.1.

Lastly, for our approximation algorithm, we simply extend Kindler, Kolla and Trevisan's algorithm by first creating an instance of MAX 2-CSP$_R$ from MAX k-CSP$_R$ by projecting each constraint to all possible subsets of two variables. We then use their algorithm to approximate the instance. Finally, we set our assignment to be the same as that from the KKT algorithm with some probability. Otherwise, we pick the assignment uniformly at random from $[R]$. As we shall show in Section 4, with the right probability, this technique can extend not only the KKT algorithm but any algorithm for MAX k'-CSP$_R$ to an algorithm for MAX k-CSP$_R$ where $k > k'$ with some loss in the advantage over the naive randomized algorithm.

1.3 Organization of the paper

In Section 2, we define notation and list background knowledge that will be used throughout the paper. Next, we prove our hardness of approximation result (Theorem 1.1) in Section 3. In Section 4, we show how to extend Kindler et al.'s algorithm to MAX k-CSP$_R$ and prove Theorem 1.3. Finally, in Section 5, we discuss interesting open questions and directions for future research.

2 Preliminaries

In this section, we introduce notation and list prior results that will be used to prove our results.

Throughout the paper, we use log to denote the logarithm to base 2.

2.1 MAX k-CSP$_R$

We start by giving a formal definition of MAX k-CSP$_R$, which is the main focus of our paper.

Definition 2.1 (MAX k-CSP$_R$)**.** An instance $(\mathcal{X}, \mathcal{C})$ of (weighted) MAX k-CSP$_R$ consists of

- A set $\mathcal{X}$ of variables.
- A set $\mathcal{C} = \{C_1, \dots, C_m\}$ of constraints. Each constraint C_i is a triple (W_i, S_i, P_i) of a positive weight $W_i > 0$ such that $\sum_{i=1}^{m} W_i = 1$, a subset of variables $S_i \subseteq \mathcal{X}$ of size k, and a constraint $P_i : [R]^{S_i} \to \{0,1\}$ that maps each assignment to variables in S_i to $\{0,1\}$. Here we use $[R]^{S_i}$ to denote the set of all functions from S_i to $[R]$, i. e., $[R]^{S_i} = \{\psi : S_i \to [R]\}$.

For each assignment of variables $\varphi : \mathcal{X} \to [R]$, we define its value to be the total weights of the constraints satisfied by this assignment, i. e., $\sum_{i=1}^{m} W_i P_i(\varphi |_{S_i})$. The goal is to find an assignment $\varphi : \mathcal{X} \to [R]$ with maximum value. We sometimes call the optimum the value of $(\mathcal{X}, \mathcal{C})$.

Note that, while some definition of MAX k-CSP$_R$ refers to the unweighted version where $W_1 = \cdots = W_m = 1/m$, Crescenzi, Silvestri and Trevisan showed that the approximability of these two cases are essentially the same [7].[4] Hence, it is enough for us to consider just the weight version.

Throughout the analysis, we assume that $R \geq 16$ and $k \geq 2$. This is without loss of generality, as otherwise our claimed inapproximability (Theorem 1.1) is trivial.

2.2 Unique Games

In this subsection, we give formal definitions for unique games, d-to-1 games and Khot's conjectures about them. First, we give a formal definition of unique games.

Definition 2.2 (Unique Game)**.** A unique game $(V, W, E, n, \{\pi_e\}_{e \in E})$ consists of

- A bipartite graph $G = (V, W, E)$ which is regular[5] on the V side.
- Alphabet size n.
- For each edge $e \in E$, a permutation $\pi_e : [n] \to [n]$.

For an assignment $\varphi : V \cup W \to [n]$, an edge $e = (v, w)$ is satisfied if $\pi_e(\varphi(v)) = \varphi(w)$. The goal is to find an assignment that satisfies as many edges as possible. We define the value of an instance to be the fraction of edges satisfied in the optimum solution.

[4] More specifically, they proved that, if the weighted version is NP-hard to approximate to within ratio r, then the unweighted version is also NP-hard to approximate to within $r - o_n(1)$ where $o_n(1)$ represents a function such that $o_n(1) \to 0$ as $n \to \infty$.

[5] The regularity can be assumed without loss of generality; see, for instance, Lemma 3.4 in [18].

The *Unique Games Conjecture (UGC)*, proposed in Khot's seminal paper [15], states that, for any sufficiently small $\eta, \gamma > 0$, it is NP-hard to distinguish a unique game where at least a $1-\eta$ fraction of constraints can be satisfied from a unique game where at most a γ fraction of constraints can be satisfied. While the UGC itself remains a major open question, recent breakthrough work [9, 8, 2, 17] has shown NP-hardness of Unique Games when the completeness is arbitrarily close to 1/2, as stated formally below.

Theorem 2.3 ([8, 17])**.** *For any $\gamma > 0$ and any $\zeta \in (\gamma, 1/2)$, there exists a constant n such that it is NP-hard to distinguish a unique game with alphabet size n of value at least ζ from one of value at most γ.*

2.3 Fourier expansion

For any function $g : [q]^n \to \mathbb{R}$ over a finite alphabet $[q]$, we define the Fourier expansion of g as follows.

Consider the space of all functions $[q] \to \mathbb{R}$, with the inner-product $\langle u, v\rangle := \mathbb{E}_{x\in[q]}[u(x)v(x)]$, where the expectation is over a uniform $x \in [q]$. Pick an orthonormal basis $l_1, \ldots, l_q$ for this space $l_i : \Sigma \to \mathbb{R}$, such that l_1 is the constant function 1. We can now write g in the tensor-product basis, as

$$g(x_1, x_2, \ldots, x_n) = \sum_{s\in[q]^n} \hat{g}(s) \cdot \prod_{i=1}^{n} l_{s(i)}(x_i).$$

Since we pick $l_1(x) = 1$ for all $x \in [q]$, we also have $\mathbb{E}_{x\in[q]}[l_i(x)] = \langle l_i, l_1\rangle = 0$ for every $i \neq 1$.

Throughout, we use $\hat{g}$ to refer to the Fourier coefficients of a function g.

For functions $g : [q]^n \to \mathbb{R}$, the p-norm is defined as

$$||g||_p = \mathop{\mathbb{E}}_{x\in[q]^n}[|g(x)|^p]^{1/p}.$$

In particular, the squared 2-norm is

$$||g||_2^2 = \mathop{\mathbb{E}}_{x\in[q]^n}[g(x)^2] = \sum_{s\in[q]^n} \hat{g}(s)^2.$$

2.3.1 Noise operators

For $x \in [q]^n$, let $y \overset{\rho}{\leftarrow} x$ denote that y is a ρ-correlated copy of x. That is, each coordinate y_i is independently chosen to be equal to x_i with probability ρ, or chosen uniformly at random otherwise.

Define the noise operator T_ρ acting on any function $g : [q]^n \to \mathbb{R}$ as

$$(T_\rho g)(x) = \mathop{\mathbb{E}}_{y \overset{\rho}{\leftarrow} x}[g(y)].$$

Notice that the noise operator T_ρ acts on the Fourier coefficients on this basis as follows.

$$f(x) = T_\rho g(x) = \sum_{s\in[q]^n} \hat{g}(s) \cdot \rho^{|s|} \cdot \prod_{i=1}^{n} l_{s(i)}(x_i)$$

where $|s|$ is defined as $|\{i \mid s(i) \neq 1\}|$.

2.3.2 Degree truncation

For any function $g : [q]^n \to \mathbb{R}$, let $g^{\leq d}$ denote the $(\leq d)$-degree part of g, i. e.,

$$g^{\leq d}(x) = \sum_{s \in [q]^n, |s| \leq d} \hat{g}(s) \cdot \prod_{i=1}^{n} l_{s(i)}(x_i),$$

and similarly let $g^{>d} : [q]^n \to \mathbb{R}$ denote the $(> d)$-degree part of g, i. e.,

$$g^{>d}(x) = \sum_{s \in [q]^n, |s| > d} \hat{g}(s) \cdot \prod_{i=1}^{n} l_{s(i)}(x_i).$$

Notice that degree-truncation commutes with the noise-operator, so writing $T_\rho g^{\leq d}$ is unambiguous. Also, notice that truncating does not increase 2-norm:

$$||g^{\leq d}||_2^2 = \sum_{s \in [q]^n, |s| \leq d} \hat{g}(s)^2 \leq \sum_{s \in [q]^n} \hat{g}(s)^2 = ||g||_2^2.$$

We frequently use the fact that noisy functions have small high-degree contributions. That is, for any function $g : [q]^n \to [0, 1]$, we have

$$||T_\rho g^{>d}||_2^2 = \sum_{s \in [q]^n, |s| > d} \rho^{2|s|} \hat{g}(s)^2 \leq \rho^{2d} \sum_{s \in [q]^n} \hat{g}(s)^2 = \rho^{2d} ||g||_2^2 \leq \rho^{2d}.$$

2.3.3 Influences

For any function $g : [q]^n \to \mathbb{R}$, the influence of coordinate $i \in [n]$ is defined as

$$\mathrm{Inf}_i[g] = \mathop{\mathbb{E}}_{x \in [q]^n} [\mathrm{Var}_{x_i \in [q]}[g(x) \mid \{x_j\}_{j \neq i}]].$$

This can be expressed in terms of the Fourier coefficients of g as follows:

$$\mathrm{Inf}_i[g] = \sum_{s \in [q]^n :\ s(i) \neq 1} \hat{g}(s)^2.$$

Further, the degree-d influences are defined as

$$\mathrm{Inf}_i^{\leq d}[g] = \mathrm{Inf}_i[g^{\leq d}] = \sum_{\substack{s \in [q]^n : \\ |s| \leq d,\ s(i) \neq 1}} \hat{g}(s)^2.$$

2.3.4 Binary functions

The previous discussion of Fourier analysis can be applied to Boolean functions, by specializing to $q = 2$. However, in this case the Fourier expansion can be written in a more convenient form. Let

$G : \{+1,-1\}^n \to \mathbb{R}$ be a Boolean function over n bits. We can choose orthonormal basis functions $l_1(y) = 1$ and $l_2(y) = y$, so G can be written as

$$G(y) = \sum_{S \subseteq [n]} \hat{G}(S) \prod_{i \in S} y_i$$

for some coefficients $\hat{G}(S)$.

Degree-truncation then results in

$$G^{\leq d}(y) = \sum_{S \subseteq [n] : |S| \leq d} \hat{G}(S) \prod_{i \in S} y_i,$$

and the noise-operator acts as follows:

$$T_\rho G(y) = \sum_{S \subseteq [n]} \hat{G}(S) \rho^{|S|} \prod_{i \in S} y_i.$$

2.3.5 Boolean analogues

To analyze k-CSP$_R$, we will want to translate between functions over $[R]^n$ to functions over $\{\pm 1\}^{nR}$. The following notion of *Boolean analogues* will be useful.

For any function $g : [R]^n \to \mathbb{R}$ with Fourier coefficients $\hat{g}(s)$, define the Boolean analogue of g to be a function $\{g\} : \{\pm 1\}^{n \times R} \to \mathbb{R}$ such that

$$\{g\}(z) = \sum_{s \in [R]^n} \hat{g}(s) \cdot \prod_{i \in [n], s(i) \neq 1} z_{i,s(i)}.$$

Note that

$$||g||_2^2 = \sum_{s \in [R]^n} \hat{g}(s)^2 = ||\{g\}||_2^2,$$

and that

$$\{g^{\leq d}\} = \{g\}^{\leq d}.$$

Moreover, the noise operator acts nicely on $\{g\}$ as follows:

$$T_\rho \{g\} = \{T_\rho g\}.$$

For simplicity, we use T_ρ to refer to both Boolean and non-Boolean noise operators with correlation ρ.

2.4 Invariance Principle and Mollification Lemma

We start with the Invariance Principle in the form of Theorem 3.18 in [24]:

Theorem 2.4 (General Invariance Principle [24])**.** *Let $f : [R]^n \to \mathbb{R}$ be any function such that* $\mathrm{Var}[f] \leq 1$*,* $\mathrm{Inf}_i[f] \leq \delta$ *for all $i \in [n]$, and $\deg(f) \leq d$. Let $F : \{\pm 1\}^{nR} \to \mathbb{R}$ be its Boolean analogue: $F = \{f\}$. Consider any "test-function" $\psi : \mathbb{R} \to \mathbb{R}$ that is $\mathcal{C}^3$, with bounded 3rd-derivative $|\psi'''| \leq C$ everywhere. Then,*

$$\left| \mathop{\mathbb{E}}_{y \in \{\pm 1\}^{nR}} [\psi(F(y))] - \mathop{\mathbb{E}}_{x \in [R]^n} [\psi(f(x))] \right| \leq C 10^d R^{d/2} \sqrt{\delta}.$$

Note that the above version follows directly from Theorem 3.18 and Hypothesis 3 of [24], and the fact that uniform ± 1 bits are $(2,3,1/\sqrt{2})$-hypercontractive as described in [24].

As we shall see later, we will want to apply the Invariance Principle for some functions ψ that are not in $\mathcal{C}^3$. However, these functions will be Lipschitz-continuous with some constant $c \in \mathbb{R}$ (or "c-Lipschitz"), meaning that

$$|\psi(x+\Delta) - \psi(x)| \leq c|\Delta| \quad \text{for all } x, \Delta \in \mathbb{R}.$$

Therefore, similarly to Lemma 3.21 in [24], we can "smooth" it to get a function $\widetilde{\psi}$ that is $\mathcal{C}^3$, and has arbitrarily small pointwise difference to ψ.

Lemma 2.5 (Mollification Lemma [24])**.** *Let $\psi : \mathbb{R} \to \mathbb{R}$ be any c-Lipschitz function. Then for any $\zeta > 0$, there exists a function $\widetilde{\psi} : \mathbb{R} \to \mathbb{R}$ such that*

- $\widetilde{\psi} \in \mathcal{C}^3$,
- $||\widetilde{\psi} - \psi||_\infty \leq \zeta$, *and,*
- $||\widetilde{\psi}'''||_\infty \leq \widetilde{C} c^3/\zeta^2$.

Here $\widetilde{C}$ is a universal constant, not depending on ζ, c.

For completeness, the full proof of the lemma can be found in Section A.1.

Now we state the following version of the Invariance Principle, which will be more convenient to invoke. It can be proved simply by just combining Theorem 2.4 and Lemma 2.5. We give a full proof in Section A.2.

Lemma 2.6 (Invariance Principle)**.** *Let $\psi : \mathbb{R} \to \mathbb{R}$ be one of the following functions:*

1. $\psi_1(t) := |t|$,
2. $\psi_k(t) := \begin{cases} t^k & \text{if } t \in [0,1], \\ 0 & \text{if } t < 0, \\ 1 & \text{if } t \geq 1. \end{cases}$

Let $f : [R]^n \to [0,1]$ be any function with all $\mathrm{Inf}_i^{\leq d}[f] \leq \delta$. Let $F : \{\pm 1\}^{nR} \to \mathbb{R}$ be its Boolean analogue: $F = \{f\}$. Let $f^{\leq d} : [R]^n \to \mathbb{R}$ denote f truncated to degree d, and similarly for $F^{\leq d} : \{\pm 1\}^{nR} \to \mathbb{R}$.

Then, for parameters $d = 10k \log R$ and $\delta = 1/(R^{10+100k\log(R)})$, we have

$$\left| \mathop{\mathbb{E}}_{y \in \{\pm 1\}^{nR}} [\psi(F^{\leq d}(y))] - \mathop{\mathbb{E}}_{x \in [R]^n} [\psi(f^{\leq d}(x))] \right| \leq O(1/R^k).$$

2.5 Hypercontractivity Theorem

Another crucial ingredient in our proof is the Hypercontractivity Theorem (Bonami [3]), which says that, on the $\{\pm 1\}^n$ domain, the operator T_ρ smooths any function so well that the higher norm can be bounded by the lower norm of the original (unsmoothed) function. Here we use the version of the theorem as stated in [25, Chap. 9].

Theorem 2.7 (Hypercontractivity Theorem (Bonami))**.** *Let $1 \le p \le q \le \infty$. For any $\rho \le \sqrt{\frac{p-1}{q-1}}$ and for any function $h : \{\pm 1\}^n \to \mathbb{R}$, the following inequality holds:*

$$||T_\rho h||_q \le ||h||_p.$$

In particular, for choice of parameter $\rho = 1/\sqrt{(k-1)\log R}$, we have

$$||T_{2\rho} h||_k \le ||h||_{1+\varepsilon}. \tag{2.1}$$

where $\varepsilon = 4/\log(R)$.

3 Inapproximability of MAX k-CSP$_R$

In this section, we prove Theorem 1.1. Our proof will be presented in Section 3.3. Before that, we first prove an inequality that is the heart of our soundness analysis in Section 3.2.

3.1 Parameters

We use the following parameters throughout, which we list for convenience here:

- Correlation[6]: $\rho = 1/\sqrt{(k-1)\log R}$
- Degree: $d = 10k\log R$
- Low-degree influences: $\delta = 1/(R^{10+100k\log(R)})$

3.2 Hypercontractivity for noisy low-influence functions

Here we show a version of hypercontractivity for noisy low-influence functions over large domains. Although hypercontractivity does not hold in general for noisy functions over large domains, it turns out to hold in our setting of high-noise and low-influences. The main technical idea is to use the Invariance Principle to reduce functions over larger domains to Boolean functions, then apply Boolean hypercontractivity.

[6]Note that for $k = 2$, this correlation yields a stability of $\approx 1/R$ for the plurality. That is, $\Pr_{z,x,y}[\mathrm{plur}(x_1,\dots,x_n) = \mathrm{plur}(y_1,\dots,y_n)] \approx 1/R$ where each $z_i \in [R]$ is iid uniform, and x_i, y_i are ρ-correlated copies of z_i.

Lemma 3.1 (Main Lemma)**.** *Let $g : [R]^n \to [0,1]$ be any function with $\mathbb{E}_{x\in[R]^n}[g(x)] = 1/R$. Then, for our choice of parameters ρ, d, δ: If $\mathrm{Inf}_i^{\leq d}[g] \leq \delta$ for all i, then*

$$\mathop{\mathbb{E}}_{x\in[R]^n}[(T_\rho g(x))^k] \leq 2^{O(k)}/R^k.$$

Before presenting the full proof, we outline the high-level steps below. Let $f = T_\rho g$, and define Boolean analogues $G = \{g\}$, and $F = \{f\}$. Let $\psi_k : \mathbb{R} \to \mathbb{R}$ be defined as in Lemma 2.6. Then,

$$\begin{aligned}
\mathop{\mathbb{E}}_{x\in[R]^n}[f(x)^k] &\approx \mathbb{E}[\psi_k(f^{\leq d}(x))] && (3.1)\\
&\approx \mathop{\mathbb{E}}_{y\in\{\pm1\}^{nR}}[\psi_k(F^{\leq d}(y))] && \text{(Lemma 2.6: Invariance Principle)} \quad (3.2)\\
&\leq ||F^{\leq d}||_k^k && (\text{Definition of } \psi_k)\\
&= ||T_\rho G^{\leq d}||_k^k && (\text{Definition of } F)\\
&= ||T_{2\rho} T_{1/2} G^{\leq d}||_k^k && \\
&\leq ||T_{1/2} G^{\leq d}||_{1+\varepsilon}^k && (\text{Hypercontractivity, for small } \varepsilon) \quad (3.3)\\
&\approx 2^{O(k)} ||g||_1^k && (\text{Invariance, etc.})\\
&= 2^{O(k)}/R^k. && (\text{Since } \mathbb{E}[|g|] = 1/R)
\end{aligned}$$

Proof of Lemma 3.1. To establish equation (3.1), first notice that

$$\psi_k(f(x)) = \psi_k(f^{\leq d}(x) + f^{>d}(x)) \leq \psi_k(f^{\leq d}(x)) + k|f^{>d}(x)|$$

where the last inequality is because the function ψ_k is k-Lipschitz.

Moreover, since $g(x) \in [0,1]$, we have $f(x) \in [0,1]$, so

$$f(x)^k = \psi_k(f(x)).$$

Thus,

$$\begin{aligned}
\mathbb{E}[f(x)^k] &= \mathbb{E}[\psi_k(f(x))]\\
&\leq \mathbb{E}[\psi_k(f^{\leq d}(x))] + k\,\mathbb{E}[|f^{>d}(x)|]\\
&= \mathbb{E}[\psi_k(f^{\leq d}(x))] + k||f^{>d}||_1\\
&\leq \mathbb{E}[\psi_k(f^{\leq d}(x))] + k||f^{>d}||_2.
\end{aligned}$$

And we can bound the 2-norm of $f^{>d}$, since f is noisy, we have

$$||f^{>d}||_2^2 = ||T_\rho g^{>d}||_2^2 \leq \rho^{2d} \leq O(1/R^{2k}).$$

The last inequality comes from our choice of ρ and d.

So equation (3.1) is established:

$$\mathbb{E}[f(x)^k] \leq \mathbb{E}[\psi_k(f^{\leq d}(x))] + O(k/R^k).$$

Equation (3.2) follows directly from our version of the Invariance Principle (Lemma 2.6), for the function ψ_k:

$$\mathop{\mathbb{E}}_{x\in[R]^n}[\psi_k(f^{\leq d}(x))] \leq \mathop{\mathbb{E}}_{y\in\{\pm1\}^{nR}}[\psi_k(F^{\leq d}(y))] + O(1/R^k).$$

We can now rewrite $\mathbb{E}_{y\in\{\pm1\}^{nR}}[\psi_k(F^{\leq d}(y))]$ as

$$\begin{aligned}\mathop{\mathbb{E}}_{y\in\{\pm1\}^{nR}}[\psi_k(F^{\leq d}(y))] &\leq \mathop{\mathbb{E}}_{y\in\{\pm1\}^{nR}}[|F^{\leq d}(y)|^k] \\ &= ||F^{\leq d}||_k^k \\ &= ||T_\rho G^{\leq d}||_k^k \\ &= ||T_{2\rho} T_{1/2} G^{\leq d}||_k^k.\end{aligned}$$

Now, from the Hypercontractivity Theorem, equation (2.1), we have

$$||T_{2\rho} T_{1/2} G^{\leq d}||_k \leq ||T_{1/2} G^{\leq d}||_{1+\varepsilon}$$

for $\varepsilon = 4/\log R$. This establishes equation (3.3):

$$||T_{2\rho} T_{1/2} G^{\leq d}||_k^k \leq ||T_{1/2} G^{\leq d}||_{1+\varepsilon}^k = \mathbb{E}[|T_{1/2} G^{\leq d}(y)|^{1+\varepsilon}]^{k/(1+\varepsilon)}.$$

To show the remaining steps, we will apply the Invariance Principle once more. Notice that, since $\varepsilon \leq 1$, for all $t \in \mathbb{R} : |t|^{1+\varepsilon} \leq |t| + t^2$. Hence, we can derive the following bound:

$$\begin{aligned}\mathbb{E}[|T_{1/2} G^{\leq d}(y)|^{1+\varepsilon}] &\leq \mathbb{E}[|T_{1/2} G^{\leq d}(y)|] + \mathbb{E}[(T_{1/2} G^{\leq d}(y))^2] \\ \text{(Matching Fourier expansion)} \quad &= \mathbb{E}[|T_{1/2} G^{\leq d}(y)|] + \mathbb{E}[(T_{1/2} g^{\leq d}(x))^2] \\ \text{(Lemma 2.6, Invariance Principle)} \quad &\leq \mathbb{E}[|T_{1/2} g^{\leq d}(x)|] + \mathbb{E}[(T_{1/2} g^{\leq d}(x))^2] + O(1/R^k). \qquad (3.4)\end{aligned}$$

Here we applied our Invariance Principle (Lemma 2.6) for the function ψ_1 as defined in Lemma 2.6. We will bound each of the expectations on the RHS, using the fact that g is balanced, and $T_{1/2} g$ is noisy.

First,

$$\begin{aligned}\mathbb{E}[|T_{1/2} g^{\leq d}(x)|] &= \mathbb{E}[|T_{1/2} g(x) - T_{1/2} g^{>d}(x)|] \\ &\leq \mathbb{E}[|T_{1/2} g(x)|] + \mathbb{E}[|T_{1/2} g^{>d}(x)|] && \text{(Triangle Inequality)} \\ &= ||g||_1 + ||T_{1/2} g^{>d}||_1 \\ &\leq ||g||_1 + ||T_{1/2} g^{>d}||_2 \\ &\leq 1/R + (1/2)^d \\ &= O(1/R). && \text{(By our choice of } d\text{)}\end{aligned}$$

Second,

$$\begin{aligned}\mathbb{E}[(T_{1/2}g^{\le d}(x))^2] &= \sum_{s\in[R]^n,|s|\le d}(1/2)^{2|s|}\hat{g}(s)^2\\ &\le \sum_{s\in[R]^n}(1/2)^{2|s|}\hat{g}(s)^2\\ &= \mathbb{E}[(T_{1/2}g(x))^2]\\ &\le \mathbb{E}[T_{1/2}g(x)] && \text{(Since } g\in[0,1])\\ &= \mathbb{E}[g(x)] = 1/R.\end{aligned}$$

Finally, plugging these bounds into equation (3.4), we find:

$$\begin{aligned}||T_{1/2}G^{\le d}||_{1+\varepsilon}^k &= \mathbb{E}[|T_{1/2}G^{\le d}(y)|^{1+\varepsilon}]^{k/(1+\varepsilon)}\\ &\le (O(1/R))^{k/(1+\varepsilon)}\\ &= 2^{O(k)}/R^{k/(1+\varepsilon)}\\ &\le 2^{O(k)}/R^{k(1-\varepsilon)}\\ &= 2^{O(k)}/R^k. && \text{(Recall } \varepsilon = 4/\log R)\end{aligned}$$

This completes the proof of the main lemma. □

3.3 Reducing Unique Games to MAX k-CSP$_R$

Here we reduce Unique Games to MAX k-CSP$_R$. We will construct a PCP verifier that reads k symbols of the proof (with an alphabet of size R) with the following properties:

- **(Completeness)** If the unique game has value at least ζ, then the verifier accepts an honest proof with probability at least $c = \zeta^k/((\log R)^{k/2}k^{O(k)})$.
- **(Soundness)** If the unique game has value at most $\gamma = 2^{O(k)}\delta^2/(8dR^k)$, then the verifier accepts any (potentially cheating) proof with probability at most $s = 2^{O(k)}/R^{k-1}$.

Since each symbol in the proof can be viewed as a variable and each accepting constraint of the verifier can be viewed as a constraint of MAX k-CSP$_R$, from Theorem 2.3, this PCP implies NP-hardness of approximating MAX k-CSP$_R$ of factor $s/c = k^{O(k)}(\log R)^{k/2}/R^{k-1}$ and, hence, establishes our Theorem 1.1.

3.3.1 The PCP

Given a unique game $(V,W,E,n,\{\pi_e\}_{e\in E})$, the proof is the truth-table of a function $h_w : [R]^n \to [R]$ for each vertex $w \in W$. By folding, we can assume h_w is balanced, i. e., h_w takes on all elements of its range with equal probability: $\Pr_{x\in[R]^n}[h_w(x) = i] = 1/R$ for all $i \in [R]$. [7]

[7]More precisely, if the truth-table provided is of some function $\widetilde{h}_w : [R]^n \to [R]$, we define the "folded" function h_w as $h_w(x_1,x_2,x_3,\dots x_n) := \widetilde{h}_w(0,x_2-x_1,\dots,x_n-x_1)+x_1$, where the $\pm$ is over mod R. Notice that the folded h_w is balanced, and also that folding does not affect dictator functions. Thus we define our PCP in terms of h_w, but simulate queries to h_w using the actual proof $\widetilde{h}_w$.

Notationally, for $x \in [R]^n$, let $(x \circ \pi)$ denote permuting the coordinates of x as: $(x \circ \pi)_i = x_{\pi(i)}$. Also, for an edge $e = (v, w)$, we write $\pi_e = \pi_{v,w}$, and define $\pi_{w,v} = \pi_{v,w}^{-1}$.

The verifier picks a uniformly random vertex $v \in V$, and k independent uniformly random neighbors of v: $w_1, w_2, \ldots, w_k \in W$. Then pick $z \in [R]^n$ uniformly at random, and let $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$ be independent ρ-correlated noisy copies of z (each coordinate x_i chosen as equal to z_i w.p. ρ, or uniformly at random otherwise). The verifier accepts if and only if

$$h_{w_1}(x^{(1)} \circ \pi_{w_1,v}) = h_{w_2}(x^{(2)} \circ \pi_{w_2,v}) = \cdots = h_{w_k}(x^{(k)} \circ \pi_{w_k,v}).$$

To achieve the desired hardness result, we pick $\rho = 1/\sqrt{(k-1)\log R}$.

3.3.2 Completeness analysis

Let the degree of each vertex in V be Δ.

Suppose that the original unique game has an assignment of value at least ζ. Let us call this assignment φ. The honest proof defines h_w at each vertex $w \in W$ as the long code encoding of this assignment, i. e., $h_w(x) = x_{\varphi(w)}$. We can written the verifier acceptance condition as follows:

$$\begin{aligned}\text{The verifier accepts} &\Leftrightarrow h_{w_1}(x^{(1)} \circ \pi_{w_1,v}) = \cdots = h_{w_k}(x^{(k)} \circ \pi_{w_k,v}) \\ &\Leftrightarrow (x^{(1)} \circ \pi_{w_1,v})_{\varphi(w_1)} = \cdots = (x^{(k)} \circ \pi_{w_k,v})_{\varphi(w_k)} \\ &\Leftrightarrow (x^{(1)})_{\pi_{w_1,v}(\varphi(w_1))} = \cdots = (x^{(k)})_{\pi_{w_k,v}(\varphi(w_k))}.\end{aligned}$$

Observe that, if the edges $(v, w_1), \ldots, (v, w_k)$ are satisfied by φ, then $\pi_{w_1,v}(\varphi(w_1)) = \cdots = \pi_{w_k,v}(\varphi(w_k)) = \varphi(v)$. Hence, if the aforementioned edges are satisfied and $x^{(1)}, \ldots, x^{(k)}$ are not perturbed at coordinate $\varphi(v)$, then $(x^{(1)})_{\pi_{w_1,v}(\varphi(w_1))} = \cdots = (x^{(k)})_{\pi_{w_k,v}(\varphi(w_k))}$.

For each $u \in V$, let s_u be the number of satisfied edges touching u. Since $w_1, \ldots, w_k$ are chosen from the neighbors of v independently from each other, the probability that the edges $(v, w_1), (v, w_2), \ldots, (v, w_k)$ are satisfied can be bounded as follows:

$$\begin{aligned}&\Pr_{v,w_1,\ldots,w_k}[(v,w_1),\ldots,(v,w_k) \text{ are satisfied}] \\ &= \sum_{u \in V} \Pr_{w_1,\ldots,w_k}[(v,w_1),\ldots,(v,w_k) \text{ are satisfied} \mid v = u] \Pr[v = u] \\ &= \sum_{u \in V} (s_u/\Delta)^k \Pr[v = u] \\ &= \mathop{\mathbb{E}}_{u \in V}\left[(s_u/\Delta)^k\right] \\ &\geq \mathop{\mathbb{E}}_{u \in V}[s_u/\Delta]^k.\end{aligned}$$

Notice that $\mathbb{E}_{u \in V}[s_u/\Delta]$ is exactly the value of φ, which is at least ζ. As a result,

$$\Pr_{v,w_1,\ldots,w_k}[(v,w_1),\ldots,(v,w_k) \text{ are satisfied}] \geq \zeta^k.$$

Furthermore, it is obvious that the probability that $x_1, \ldots, x_k$ are not perturbed at the coordinate $\varphi(v)$ is ρ^k. As a result, the PCP accepts with probability at least $\zeta^k \rho^k$. When $\rho = 1/\sqrt{(k-1)\log R}$ and ζ is a constant not depending on k and R, the completeness is $1/((\log R)^{k/2} k^{O(k)})$.

3.3.3 Soundness analysis

Suppose that the unique game has value at most $\gamma = 2^{O(k)}\delta^2/(4dR^k)$. We will show that the soundness is $2^{O(k)}/R^{k-1}$.

Suppose for the sake of contradiction that the probability that the verifier accepts $\Pr[accept] > t = 2^{\Omega(k)}/R^{k-1}$ where $\Omega(\cdot)$ hides some large enough constant.

Let $h^i_w(x) : [R]^n \to \{0,1\}$ be the indicator function for $h_w(x) = i$ and let $x \overset{\rho}{\leftarrow} z$ denote that x is a ρ-correlated copy of z. We have

$$\begin{aligned}\Pr[accept] &= \Pr[h_{w_1}(x^{(1)} \circ \pi_{w_1,v}) = \cdots = h_{w_k}(x^{(k)} \circ \pi_{w_k,v})] \\ &= \sum_{i\in[R]} \Pr[i = h_{w_1}(x^{(1)} \circ \pi_{w_1,v}) = \cdots = h_{w_k}(x^{(k)} \circ \pi_{w_k,v})] \\ &= \sum_{i\in[R]} \mathbb{E}[h^i_{w_1}(x^{(1)} \circ \pi_{w_1,v}) \cdots h^i_{w_k}(x^{(k)} \circ \pi_{w_k,v})] \\ &= \sum_{i\in[R]} \mathbb{E}\left[\mathop{\mathbb{E}}_{w_1}[h^i_{w_1}(x^{(1)} \circ \pi_{w_1,v})] \cdots \mathop{\mathbb{E}}_{w_k}[h^i_{w_k}(x^{(k)} \circ \pi_{w_k,v})]\right].\end{aligned}$$

Where the last equality follows since the w_i's are independent, given v.

Now define $g^i_v : [R]^n \to [0,1]$ as

$$g^i_v(x) = \mathop{\mathbb{E}}_{w\sim v}[h^i_w(x \circ \pi_{w,v})]$$

where $w \sim v$ denotes neighbors w of v.

We can rewrite $\Pr[accept]$ as follows:

$$\begin{aligned}\Pr[accept] &= \sum_{i\in[R]} \mathbb{E}[g^i_v(x^{(1)})g^i_v(x^{(2)}) \cdots g^i_v(x^{(k)})] \\ &= \sum_{i\in[R]} \mathbb{E}\left[\mathop{\mathbb{E}}_{x\overset{\rho}{\leftarrow}z}[g^i_v(x)]^k\right] && \text{(Since } x^{(j)}\text{'s are independent given } z\text{)} \\ &= \sum_{i\in[R]} \mathop{\mathbb{E}}_{v,z}[(T_\rho g^i_v(z))^k] \\ &= \mathop{\mathbb{E}}_{v}\left[\sum_{i\in[R]} \mathop{\mathbb{E}}_{z}[(T_\rho g^i_v(z))^k]\right]. && (3.5)\end{aligned}$$

Next, notice that $\sum_{i\in[R]} \mathbb{E}_z[(T_\rho g^i_v(z))^k]$ is simply the probability the verifier accepts given it picks vertex v, and thus this sum is bounded above by 1.

Therefore, since $\Pr[accept] > t$, by equation (3.5), at least $t/2$ fraction of vertices $v \in V$ have

$$\sum_{i\in[R]} \mathop{\mathbb{E}}_{z}[(T_\rho g^i_v(z))^k] \geq t/2.$$

For these "good" vertices, there must exist some $i \in [R]$ for which

$$\mathop{\mathbb{E}}_{z}[(T_\rho g^i_v(z))^k] \geq t/(2R).$$

Then for "good" v and i as above,

$$\mathop{\mathbb{E}}_{z}[(T_\rho g_v^i(z))^k] > 2^{\Omega(k)}/R^k.$$

By Lemma 3.1 (Main Lemma), this means g_v^i has some coordinate j for which

$$\mathrm{Inf}_j^{\leq d}[g_v^i] > \delta$$

for our choice of d, δ as defined in Section 3.1. Pick this j as the label of vertex $v \in V$.

Now to pick the label of a vertex $w \in W$, define the candidate labels as

$$\mathrm{Cand}[w] = \{j \in [n] : \exists\, i \in [R] \text{ s.t. } \mathrm{Inf}_j^{\leq d}[h_w^i] \geq \delta/2\}.$$

Notice that

$$\sum_{j \in [n]} \mathrm{Inf}_j^{\leq d}[h_w^i] = \sum_{s \in [R]^n :\ |s| \leq d} |s| \hat{h}_w^i(s)^2 \leq d \sum_{s : |s| > 0} \hat{h}_w^i(s)^2 = d\ \mathrm{Var}[h_w^i] \leq d.$$

So for each $i \in [R]$, the projection h_w^i can have at most $2d/\delta$ coordinates with influence $\geq \delta/2$. Therefore the number of candidate labels is bounded:

$$|\mathrm{Cand}[w]| \leq 2dR/\delta.$$

Now we argue that picking a random label in Cand$[w]$ for $w \in W$ is in expectation a good decoding. We will show that if we assigned label j to a "good" $v \in V$, then $\pi_{v,w}(j) \in \mathrm{Cand}[w]$ for a constant fraction of neighbors $w \sim v$. Note here that $\pi_{v,w} = \pi_{w,v}^{-1}$.

First, since $g_v^i(x) = \mathbb{E}_{w \sim v}[h_w^i(x \circ \pi_{w,v})]$, the Fourier transform of g_v^i is related to the Fourier transform of the long code labels h_w^i as

$$\hat{g}_v^i(s) = \mathop{\mathbb{E}}_{w \sim v}[\hat{h}_w^i(s \circ \pi_{w,v})].$$

Hence, the influence $\mathrm{Inf}_j^{\leq k}[g_v^i]$ of g_v^i being large implies the expected influence $Inf_{\pi_{v,w}^{-1}(j)}^{\leq d}[h_w^i]$ of its neighbor labels $w \sim v$ is also large as formalized below.

$$
\begin{aligned}
\delta < \mathrm{Inf}_j^{\le k}[g_v^i] &= \sum_{\substack{s\in[R]^n \\ |s|\le k, s_j\neq 1}} \hat{g}_v^i(s)^2 \\
&= \sum \mathop{\mathbb{E}}_{w\sim v}[\hat{h}_w^i(s\circ\pi_{w,v})]^2 \\
&\le \sum \mathop{\mathbb{E}}_{w\sim v}[\hat{h}_w^i(s\circ\pi_{w,v})^2] \\
&= \mathop{\mathbb{E}}_{w\sim v}[\sum_{\substack{s\in[R]^n \\ |s|\le k, s_j\neq 1}} \hat{h}_w^i(s\circ\pi_{w,v})^2] \\
&= \mathop{\mathbb{E}}_{w\sim v}[\sum_{\substack{s\in[R]^n \\ |s|\le k, s_{\pi_{w,v}^{-1}(j)}\neq 1}} \hat{h}_w^i(s)^2] \\
(\text{Since } \pi_{v,w}=\pi_{w,v}^{-1}) &= \mathop{\mathbb{E}}_{w\sim v}[\sum_{\substack{s\in[R]^n \\ |s|\le k, s_{\pi_{v,w}(j)}\neq 1}} \hat{h}_w^i(s)^2] \\
&= \mathop{\mathbb{E}}_{w\sim v}[Inf_{\pi_{v,w}(j)}^{\le d}[h_w^i]]
\end{aligned}
$$

Therefore, at least $\delta/2$ fraction of neighbors $w \sim v$ must have $Inf_{\pi_{v,w}(j)}^{\le d}[h_w^i] \ge \delta/2$, and so $\pi_{v,w}(j) \in \mathrm{Cand}[w]$ for at least $\delta/2$ fraction of neighbors of "good" vertices v.

Finally, recall that at least $(t/2)$ fraction of vertices $v \in V$ are "good". These vertices have at least $(\delta/2)$ fraction of neighbors $w \in W$ with high-influence labels and the matching label $w \in W$ is picked with probability at least $\delta/(2dR)$. Moreover, as stated earlier, we can assume that the graph is regular on V side. Hence, the expected fraction of edges satisfied by this decoding is at least

$$(t/2)(\delta/2)(\delta/2dR) = t\delta^2/(8dR) = 2^{\Omega(k)}\delta^2/(8dR^k) > \gamma,$$

which contradicts our assumption that the unique game has value at most γ. Hence, we can conclude that the soundness is at most $2^{O(k)}/R^{k-1}$ as desired.

4 $\Omega(\log R/R^{k-1})$-approximation algorithm for MAX k-CSP$_R$

Instead of just extending the KKT algorithm to work with MAX k-CSPs, we will show a more general statement that *any* algorithm that approximates MAX CSPs with small arity can be extended to approximate MAX CSPs with larger arities. In particular, we show how to extend any $f(R)/R^{k'}$-approximation algorithm for MAX k'-CSP$_R$ to an $(f(R)/2^{O(\min\{k',k-k'\})})/R^k$-approximation algorithm for MAX k-CSP$_R$ where $k > k'$.

Since the naive algorithm that assigns every variable randomly has an approximation ratio of $1/R^k$, we think of $f(R)$ as the advantage of algorithm A over the randomized algorithm. From this perspective, our extension lemma preserves the advantage up to a factor of $1/2^{O(\min\{k',k-k'\})}$.

The extension lemma and its proof are stated formally below.

Lemma 4.1. *Suppose that there exists a polynomial-time approximation algorithm A for* MAX k'-CSP$_R$ *that outputs an assignment with expected value at least $f(R)/R^{k'}$ times the optimum. For any $k > k'$, we can construct a polynomial-time approximation algorithm B for* MAX k-CSP$_R$ *that outputs an assignment with expected value at least $(f(R)/2^{O(\min\{k',k-k'\})})/R^k$ times the optimum.*

Proof. The main idea of the proof is simple. We turn an instance of MAX k-CSP$_R$ to an instance of MAX k'-CSP$_R$ by constructing $\binom{k}{k'}R^{k-k'}$ new constraints for each original constraint; each new constraint is a projection of the original constraint to a subset of variables of size k'. We then use A to solve the newly constructed instance. Finally, B simply assigns each variable with the assignment from A with a certain probability and assigns it randomly otherwise.

For convenience, let α be $\frac{k-k'}{k}$. We define B on input $(\mathcal{X},\mathcal{C})$ as follows:

1. Create an instance $(\mathcal{X},\mathcal{C}')$ of MAX k'-CSP$_R$ with the same variables and, for each $C = (W,S,P) \in \mathcal{C}$ and for every subset S' of S with $|S'| = k'$ and every $\tau \in [R]^{S-S'}$, create a constraint $C^{S',\tau} = (W',S',P')$ in $\mathcal{C}'$ where $W' = \frac{W}{\binom{k}{k'}R^{k-k'}}$ and $P' : [R]^{S'} \to \{0,1\}$ is defined by
$$P'(\psi) = P(\psi \circ \tau).$$
Here $\psi \circ \tau$ is defined as follows:
$$\psi \circ \tau(x) = \begin{cases} \psi(x) & \text{if } x \in S', \\ \tau(x) & \text{otherwise.} \end{cases}$$
2. Run A on input $(\mathcal{X},\mathcal{C}')$. Denote the output of A by φ_A.
3. For each $x \in \mathcal{X}$, with probability α, pick $\varphi_B(x)$ randomly from $[R]$. Otherwise, let $\varphi_B(x)$ be $\varphi_A(x)$.
4. Output φ_B.

We now show that φ_B has expected value at least $(f(R)/2^{O(\min\{k',k-k'\})})/R^k$ times the optimum.

First, observe that the optimum of $(\mathcal{X},\mathcal{C}')$ is at least $1/R^{k-k'}$ times that of $(\mathcal{X},\mathcal{C})$. To see that this is true, consider any assignment $\varphi : \mathcal{X} \to [R]$ and any constraint $C = (W,S,P)$. Its weighted contribution in $(\mathcal{X},\mathcal{C})$ is $WP(\varphi|_S)$. On the other hand, $\frac{W}{\binom{k}{k'}R^{k-k'}}P(\varphi|_S)$ appears $\binom{k}{k'}$ times in $(\mathcal{X},\mathcal{C}')$, once for each subset $S' \subseteq S$ of size k'. Hence, the value of φ with respect to $(\mathcal{X},\mathcal{C}')$ is at least $1/R^{k-k'}$ times its value with respect to $(\mathcal{X},\mathcal{C})$.

Recall that the algorithm A gives an assignment of expected value at least $f(R)/R^{k'}$ times the optimum of $(\mathcal{X},\mathcal{C}')$. Hence, the expected value of φ_A is at least $f(R)/R^k$ times the optimum of $(\mathcal{X},\mathcal{C})$.

Next, we will compute the expected value of φ_B (with respect to $(\mathcal{X},\mathcal{C})$). We start by computing the expected value of φ_B with respect to a fixed constraint $C = (W,S,P) \in \mathcal{C}$, i. e., $\mathbb{E}_{\varphi_B}[WP(\varphi_B|_S)]$. For each $S' \subseteq S$ of size k', we define $D_{S'}$ as the event where, in Step 3, $\varphi_B(x)$ is assigned to be $\varphi_A(x)$ for all $x \in S'$ and $\varphi_B(x)$ is assigned randomly for all $x \in S - S'$.

Since $D_{S'}$ is disjoint for all $S' \subseteq S$ of size k', we have the following inequality.

$$\begin{aligned}\mathop{\mathbb{E}}_{\varphi_B}[WP(\varphi_B|_S)] &\geq \sum_{\substack{S'\subseteq S\\|S'|=k'}} \Pr[D_{S'}] \mathop{\mathbb{E}}_{\varphi_B}[WP(\varphi_B|_S) \mid D_{S'}] \\ &= \alpha^{k-k'}(1-\alpha)^{k'} \sum_{\substack{S'\subseteq S\\|S'|=k'}} W \mathop{\mathbb{E}}_{\varphi_B}[P(\varphi_B|_S) \mid D_{S'}],\end{aligned}$$

where the equality follows from $\Pr[D_{S'}] = \alpha^{k-k'}(1-\alpha)^{k'}$.

Moreover, since every vertex in $S - S'$ is randomly assigned when $D_{S'}$ occurs, $\mathbb{E}[P(\varphi_B|_S) \mid D_{S'}]$ can be view as the average value of $P((\varphi_A|_{S'}) \circ \tau)$ over all $\tau \in [R]^{S-S'}$. Hence, we can derive the following:

$$\mathop{\mathbb{E}}_{\varphi_B}[P(\varphi_B|_S) \mid D_{S'}] = \frac{1}{R^{k-k'}} \mathop{\mathbb{E}}_{\varphi_A}\left[\sum_{\tau\in[R]^{S-S'}} P((\varphi_A|_{S'}) \circ \tau)\right].$$

As a result, we have

$$\mathop{\mathbb{E}}_{\varphi_B}[WP(\varphi_B|_S)] \geq \frac{\alpha^{k-k'}(1-\alpha)^{k'}}{R^{k-k'}}\left(\mathop{\mathbb{E}}_{\varphi_A}\left[\sum_{\substack{S'\subseteq S\\|S'|=k'}} \sum_{\tau\in[R]^{S-S'}} WP((\varphi_A|_{S'}) \circ \tau)\right]\right). \tag{4.1}$$

By summing equation (4.1) over all constraints $C \in \mathcal{C}$, we arrive at the following inequality.

$$\begin{aligned}&\mathop{\mathbb{E}}_{\varphi_B}\left[\sum_{C=(W,S,P)\in\mathcal{C}} WP(\varphi_B|_S)\right] \\ &\geq \frac{\alpha^{k-k'}(1-\alpha)^{k'}}{R^{k-k'}} \mathop{\mathbb{E}}_{\varphi_A}\left[\sum_{C=(W,S,P)\in\mathcal{C}}\left(\sum_{\substack{S'\subseteq S\\|S'|=k'}} \sum_{\tau\in[R]^{S-S'}} WP((\varphi_A|_{S'}) \circ \tau)\right)\right] \\ &= \binom{k}{k'}\alpha^{k-k'}(1-\alpha)^{k'} \mathop{\mathbb{E}}_{\varphi_A}\left[\sum_{C=(W,S,P)\in\mathcal{C}}\left(\sum_{\substack{S'\subseteq S\\|S'|=k'}} \sum_{\tau\in[R]^{S-S'}} \frac{W}{\binom{k}{k'}R^{k-k'}} P((\varphi_A|_{S'}) \circ \tau)\right)\right] \\ &= \binom{k}{k'}\alpha^{k-k'}(1-\alpha)^{k'} \mathop{\mathbb{E}}_{\varphi_A}\left[\sum_{C'=(W',S',P')\in\mathcal{C}} W'P'(\varphi_A|_{S'})\right]\end{aligned}$$

The first expression is the expected value of φ_B whereas the last is $\binom{k}{k'}\alpha^{k-k'}(1-\alpha)^{k'}$ times the expected value of φ_A. Since the expected value of φ_A is at least $f(R)/R^k$ times the optimum of $(\mathcal{X},\mathcal{C})$, the expected value of φ_B is at least $(\binom{k}{k'}\alpha^{k-k'}(1-\alpha)^{k'})(f(R)/R^k)$ times the optimum of $(\mathcal{X},\mathcal{C})$.

Finally, we substitute $\alpha = \frac{k-k'}{k}$ in to get

$$\binom{k}{k'}\alpha^{k-k'}(1-\alpha)^{k'} = \binom{k}{k'}\left(\frac{k-k'}{k}\right)^{k-k'}\left(\frac{k'}{k}\right)^{k'}.$$

Let $l = \min\{k', k-k'\}$. We then have

$$\begin{aligned}\binom{k}{k'}\left(\frac{k-k'}{k}\right)^{k-k'}\left(\frac{k'}{k}\right)^{k'} &= \binom{k}{l}\left(\frac{k-l}{k}\right)^{k-l}\left(\frac{l}{k}\right)^{l} \\ &\geq \left(\frac{k}{l}\right)^{l}\left(\frac{k-l}{k}\right)^{k-l}\left(\frac{l}{k}\right)^{l} \\ &\geq \left(\frac{k-l}{k}\right)^{k} \\ &= \left((1-l/k)^{0.5k/l}\right)^{2l} \\ &\geq 1/2^{2l},\end{aligned}$$

where the last inequality follows from Bernoulli's inequality and from $l \leq 0.5k$.

Hence, φ_B has expected value at least $(f(R)/2^{O(l)})/R^k$ times the optimum of $(\mathcal{X}, \mathcal{C})$, which completes the proof of this lemma. □

Finally, Theorem 1.3 is an immediate consequence of applying Lemma 4.1 to the algorithm from [20] with $k' = 2$ and $f(R) = \Omega(R \log R)$.

5 Conclusion and open question

In this article we provide a hardness result and an approximation algorithm for MAX k-CSP$_R$. The former is subsumed by independent work of Khot and Saket [19] whereas the latter remains the best known algorithm in the regime $3 \leq k < o(\log R)$. Even with our results, current inapproximability results do not match the best known approximation ratio achievable in polynomial time when $3 \leq k < R$. Hence, it is intriguing to ask what the right ratio that MAX k-CSP$_R$ becomes NP-hard to approximate is. Since Khot and Saket's hardness factor $O(k^2 \log(kR)/R^{k-1})$ [19] does not match Chan's hardness factor $O(k/R^{k-2})$ [4] when $k = R$, it is plausible that there is a k between 3 and $R-1$ such that a drastic change in the hardness factor, and the proof technique, occurs.

A Proofs of preliminary results

For completeness, we prove some of the preliminary results, whose formal proofs were not found in the literature by the authors.

A.1 Mollification Lemma

Below is the proof of the Mollification Lemma. We remark that, while its main idea is explained in [25], the full proof is not shown there. Hence, we provide the proof here for completeness.

Proof. (of Lemma 2.5) Let $p : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^4$ function supported only on $[-1,+1]$, such that $p(y)$ forms a probability distribution. (For example, an appropriately normalized version of $\mathrm{e}^{-1/(1+y^2)}$ for $|y| \leq 1$). Define $p_\lambda(y)$ to be rescaled to have support $[-\lambda,+\lambda]$ for some $\lambda > 0$:

$$p_\lambda(y) := (1/\lambda)p(y/\lambda).$$

Let Y_λ be a random variable with distribution $p_\lambda(y)$, supported on $[-\lambda,+\lambda]$. We will set $\lambda = \zeta/c$.

Now, define

$$\widetilde{\psi} := \mathop{\mathbb{E}}_{Y_\lambda}[\psi(x+Y_\lambda)].$$

This is pointwise close to ψ, since ψ is c-Lipschitz:

$$|\widetilde{\psi}(x) - \psi(x)| = |\mathop{\mathbb{E}}_{Y_\lambda}[\psi(x+Y_\lambda) - \psi(x)]| \leq \mathop{\mathbb{E}}_{Y_\lambda}[|\psi(x+Y_\lambda) - \psi(x)|] \leq \mathop{\mathbb{E}}_{Y_\lambda}[c|Y_\lambda|] \leq c\lambda = \zeta.$$

Further, $\widetilde{\psi}$ is $\mathcal{C}^3$, because $\widetilde{\psi}(x)$ can be written as a convolution:

$$\widetilde{\psi}(x) = (\psi * p_\lambda)(x) \implies \widetilde{\psi}''' = (\psi * p_\lambda)''' = (\psi * p_\lambda''').$$

To see that $\widetilde{\psi}'''$ is bounded, for a fixed $x \in \mathbb{R}$, define the constant $z := \psi(x)$. Then,

$$\begin{aligned}
|\widetilde{\psi}'''(x)| &= |(\psi * p_\lambda''')(x)| \\
&= |(\psi * p_\lambda''' - z' * p_\lambda'')(x)| && (z \text{ is constant, so } z' = 0) \\
&= |(\psi * p_\lambda''' - z * p_\lambda''')(x)| \\
&= |((\psi - z) * p_\lambda''')(x)| \\
&= \left| \int_{-\infty}^{+\infty} p_\lambda'''(y)(\psi(x-y) - z)dy \right| \\
&= \left| \int_{-\infty}^{+\infty} p_\lambda'''(y)(\psi(x-y) - \psi(x))dy \right| \\
&\leq \int_{-\lambda}^{+\lambda} |p_\lambda'''(y)||\psi(x-y) - \psi(x)|dy \\
&\leq ||p_\lambda'''||_\infty \int_{-\lambda}^{+\lambda} |cy|dy && (c\text{-Lipschitz}) \\
&= ||p_\lambda'''||_\infty c\lambda^2.
\end{aligned}$$

Define the universal constant $\widetilde{C} := ||p'''||_\infty$. We have

$$p_\lambda'''(y) = (1/\lambda^4)p'''(y/\lambda) \implies ||p_\lambda'''||_\infty \leq (1/\lambda^4)\widetilde{C}.$$

With our choice of $\lambda = \zeta/c$, this yields $|\widetilde{\psi}'''(x)| \leq \widetilde{C}c^3/\zeta^2$, which completes the proof of Lemma 2.5. □

A.2 Proof of Lemma 2.6

Below we show the proof of Lemma 2.6.

Proof of Lemma 2.6. First, we "mollify" the function ψ to construct a $\mathcal{C}^3$ function $\widetilde{\psi}$, by applying Lemma 2.5 for $\zeta = 1/R^k$. Notice that both choices of ψ are k-Lipschitz. Therefore the Mollification Lemma guarantees that $|\widetilde{\psi}'''(x)| \le \widetilde{C}k^3R^{2k}$ for some universal constant $\widetilde{C}$.

Since $\widetilde{\psi}$ is pointwise close to ψ, with deviation at most $1/R^k$, we have

$$\left| \mathop{\mathbb{E}}_{y\in\{\pm1\}^{nR}}[\psi(F^{\le d}(y))] - \mathop{\mathbb{E}}_{x\in[R]^n}[\psi(f^{\le d}(x))] \right| \le \left| \mathop{\mathbb{E}}_{y\in\{\pm1\}^{nR}}[\widetilde{\psi}(F^{\le d}(y))] - \mathop{\mathbb{E}}_{x\in[R]^n}[\widetilde{\psi}(f^{\le d}(x))] \right| + O(1/R^k).$$

Applying the General Invariance Principle (Theorem 2.4) with the function $\widetilde{\psi}$, we have

$$\left| \mathop{\mathbb{E}}_{y\in\{\pm1\}^{nR}}[\widetilde{\psi}(F^{\le d}(y))] - \mathop{\mathbb{E}}_{x\in[R]^n}[\widetilde{\psi}(f^{\le d}(x))] \right| \le \widetilde{C}k^3R^{2k}10^dR^{d/2}\sqrt{\delta}.$$

By our choice of parameters d, δ, this is $O(1/R^k)$. □

B Khot–Saket dictatorship test

Khot and Saket [19] showed that any integrality gap instance for an LP relaxation of MAX k-CSP$_R$ with completeness c and soundness s can be used to construct a k-query dictatorship test with completeness $\Omega\left(\frac{c}{k^3\log R}\right)$ and soundness $O(s)$. Using a standard reduction (see, e. g., Section 3.3.1), this in turn implies UG-hardness of an approximation ratio of $O(\frac{sk^3\log R}{c})$ for MAX k-CSP$_R$. It is not hard to see that such an integrality gap with $c = 1$ and $s = O(1/R^{k-1})$ for MAX k-CSP$_R$ exists. Plugging this into Khot and Saket's result directly yields UG-hardness of an approximation ratio of $O(k^3\log R/R^{k-1})$ for MAX k-CSP$_R$, as well as a k-query dictatorship test with completeness $\Omega\left(\frac{1}{k^3\log R}\right)$ and soundness $O(1/R^{k-1})$. We remark that their result also immediately implies NP-hardness of an approximation ratio of $O(2^{O(k)}\log R/R^{k-1})$ for MAX k-CSP$_R$ thanks to the new 2-to-1 Theorem; similarly to Section 3.3.2, the extra factor $2^{O(k)}$ comes from the fact that the completeness from Theorem 2.3 is arbitrarily close to 1/2 instead of 1.

Below, we sketch the proof of the aforementioned dictatorship test. In fact, analyzing this test directly (instead of through LP integrality gap as in [19]) improves the completeness slightly, to $\Omega\left(\frac{1}{k^2\log(kR)}\right)$. This also leads to a slightly improved UG-hardness-of-approximation factor $O(k^2\log(kR)/R^{k-1})$.

Notation. For any function $f : [R]^n \to [R]$, and any $i \in [R]$, let $f^i : [R]^n \to \{0,1\}$ denote the indicator function $f^i(x) := 1[f(x) = i]$.

Theorem B.1. *For any $k, R \ge 2$, there exist constants d, δ (depending only on k, R) for a k-query nonadaptive Dictator-vs.-Quasirandom test with the following guarantees:*

- **(Completeness)** *If f is a dictator, then the test passes with probability at least*

$$\rho = \Omega\left(\frac{1}{k^2\log(kR)}\right).$$

- **(Soundness)** *If f has $\mathrm{Inf}_j^{\leq d}[f^i] \leq \delta$ for all coordinates $j \in [n]$ and all projections $i \in [R]$, then the test passes with probability at most*
$$O(1/R^{k-1}).$$

Before sketching the proof of the theorem, we note that it also implies NP-hardness of factor $\frac{2^{O(k)}\log(kR)}{R^{k-1}}$ and UG-hardness of factor $O(k^2\log(kR)/R^{k-1})$ for MAX k-CSP$_R$:

Corollary B.2. *It is NP-hard to approximate* MAX k-CSP$_R$ *to within a $\frac{2^{O(k)}\log(kR)}{R^{k-1}}$ factor, for any $k,R \geq 2$.*

Corollary B.3. *It is UG-hard to approximate* MAX k-CSP$_R$ *to within a $O(k^2\log(kR)/R^{k-1})$ factor, for any $k,R \geq 2$.*

Corollary B.2 and Corollary B.3 follow from applying a standard reduction (similar to Section 3.3.1) from Unique Games using the dictator test from Theorem B.1. Similarly to Section 3.3.2, the extra factor $2^{O(k)}$ in Corollary B.2 comes from the fact that the completeness from Theorem 2.3 is arbitrarily close to 1/2 instead of 1. This results in a completeness of $\rho/2^{O(k)}$ for MAX k-CSP$_R$ after the reduction.

Proof Sketch of Theorem B.1. We may assume that f is balanced ($\mathbb{E}[f^i] = 1/R \forall i$).

The test is defined as follows. We will generate k queries $x^{(1)},\dots x^{(k)} \in [R]^n$. The first coordinates $x_1^{(1)},\dots,x_1^{(k)}$ are picked by:

1. Pick $z \in [R]$.
2. With probability ρ, set all queries to be equal to z at this coordinate:
$$x_1^{(1)} = \cdots = x_1^{(k)} = z.$$
3. Otherwise, set all $x_1^{(1)},\dots,x_1^{(k)} \in [R]$ independently at random.

All coordinates $j \in [n]$ are picked similarly (independently of other coordinates).

The verifier accepts iff
$$f(x^{(1)}) = f(x^{(2)}) = \cdots = f(x^{(k)}).$$

Remark B.4. The difference between this dictatorship test and the test of Theorem B.1 is that here, with probability ρ we set *all* the queries equal to z, instead of setting each query equal to z with probability ρ independently. This means the queries are correlated even given z, so our previous technique of hypercontractivity does not work directly. But it is possible to use a more powerful version of invariance that can deal with this directly.

The completeness analysis is obvious.

Soundness Analysis

$$\begin{aligned}\Pr[\text{accept}] &= \Pr[f(x^{(1)}) = f(x^{(2)}) = \cdots = f(x^{(k)})]\\ &= \sum_{i\in[R]} \Pr[f(x^{(1)}) = f(x^{(2)}) = \cdots = f(x^{(k)}) = i]\\ &= \sum_{i\in[R]} \mathbb{E}[f^i(x^{(1)})f^i(x^{(2)})\dots f^i(x^{(k)})]. \end{aligned} \tag{B.1}$$

We now invoke Proposition 6.4 of [23] to bound this expectation in terms of a related quantity in Gaussian space. In particular, since each coordinate of our queries form a "ρ-correlated space", it follows from Proposition 6.4 that there exists constants [8] δ, d (independent of n) such that if

$$\mathrm{Inf}_j^{\leq d}(f^i) \leq \delta$$

for all coordinates j, then

$$\mathbb{E}\left[\prod_{\ell=1}^{k} f^i(x^{(\ell)})\right] \leq \Gamma_\rho(1/R, \ldots, 1/R) + 1/R^k.$$

Where the function $\Gamma_\rho(1/R)$ is roughly the probability that k ρ-correlated Gaussians all simultaneously satisfy an event that has marginal probability $1/R$ for a single Gaussian (see the formal definition in Definition 1.12 of [23]).

Now we conclude by the quantitative bounds on Γ given by Lemma 2 in [19]. In Lemma 2, set $\varepsilon = 1/k$, and $\mu_i = 1/R$. Then for

$$\rho = \frac{1}{Ck^2 \log(kR)}$$

for some sufficiently large universal constant C, Lemma 2 gives

$$\Gamma_\rho(1/R, \ldots, 1/R) \leq (1+\varepsilon)^{k-1} \prod_{i=1}^{k} \mu_i \leq 4/R^k.$$

Thus, continuing from equation (B.1):

$$\begin{aligned}
\Pr[\text{accept}] &= \sum_{i \in [R]} \mathbb{E}[f^i(x^{(1)}) f^i(x^{(2)}) \ldots f^i(x^{(k)})] \\
&\leq \sum_{i \in [R]} 5/R^k \\
&= O(1/R^{k-1}).
\end{aligned}$$

This completes the soundness analysis. □

Acknowledgments.

We thank Rishi Saket for bringing [19] to our attention. We also thank the anonymous reviewers for providing useful comments and suggestions.

[8] In the notation of Proposition 6.4, we set $\varepsilon = 1/R^k$, and have $\alpha = 1/R^k$, $\delta = \tau$ (given in Prop 6.4), $d = k\log(1/\delta)/\log(R)$.

References

[1] PER AUSTRIN AND ELCHANAN MOSSEL: Approximation resistant predicates from pairwise independence. *Comput. Complexity*, 18(2):249–271, 2009. [doi:10.1007/s00037-009-0272-6] 2

[2] BOAZ BARAK, PRAVESH K. KOTHARI, AND DAVID STEURER: Small-set expansion in shortcode graph and the 2-to-2 conjecture. In *Proc. 10th Innovations in Theoret. Comp. Sci. conf. (ITCS'19)*, pp. 9:1–12. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.ITCS.2019.9] 4, 7

[3] ALINE BONAMI: Étude des coefficients Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier*, 20(2):335–402, 1970. EuDML. 11

[4] SIU ON CHAN: Approximation resistance from pairwise-independent subgroups. *J. ACM*, 63(3):27:1–32, 2016. [doi:10.1145/2873054] 2, 3, 4, 21

[5] MOSES CHARIKAR, MOHAMMADTAGHI HAJIAGHAYI, AND HOWARD KARLOFF: Improved approximation algorithms for Label Cover problems. *Algorithmica*, 61(1):190–206, 2011. [doi:10.1007/s00453-010-9464-3] 2

[6] MOSES CHARIKAR, KONSTANTIN MAKARYCHEV, AND YURY MAKARYCHEV: Near-optimal algorithms for maximum constraint satisfaction problems. *ACM Trans. Algorithms*, 5(3):32:1–14, 2009. [doi:10.1145/1541885.1541893] 2, 3

[7] PIERLUIGI CRESCENZI, RICCARDO SILVESTRI, AND LUCA TREVISAN: On weighted vs unweighted versions of combinatorial optimization problems. *Inform. Comput.*, 167(1):10–26, 2001. Preliminary version in ISTCS'96. [doi:10.1006/inco.2000.3011] 6

[8] IRIT DINUR, SUBHASH KHOT, GUY KINDLER, DOR MINZER, AND MULI SAFRA: Towards a proof of the 2-to-1 Games Conjecture? In *Proc. 50th STOC*, pp. 376–389. ACM Press, 2018. [doi:10.1145/3188745.3188804] 2, 4, 5, 7

[9] IRIT DINUR, SUBHASH KHOT, GUY KINDLER, DOR MINZER, AND MULI SAFRA: On non-optimally expanding sets in Grassmann graphs. *Israel J. Math.*, 243:377–420, 2021. Preliminary version in STOC'18. [doi:10.1007/s11856-021-2164-7] 4, 7

[10] LARS ENGEBRETSEN: The nonapproximability of non-Boolean predicates. *SIAM J. Discr. Math.*, 18(1):114–129, 2004. [doi:10.1137/S0895480100380458] 2

[11] LARS ENGEBRETSEN AND JONAS HOLMERIN: More efficient queries in PCPs for NP and improved approximation hardness of maximum CSP. *Random Struct. Algor.*, 33(4):497–514, 2008. [doi:10.1002/rsa.20226] 2

[12] GIL GOLDSHLAGER AND DANA MOSHKOVITZ: Approximating kCSP for large alphabets. Preprint, 2015. Available at author's website. 2, 3, 4

[13] VENKATESAN GURUSWAMI AND PRASAD RAGHAVENDRA: Constraint satisfaction over a non-Boolean domain: Approximation algorithms and Unique-Games hardness. In *Proc. 11th Internat. Workshop on Approximation Algorithms for Combinat. Opt. Probl. (APPROX'08)*, pp. 77–90. Springer, 2008. [doi:10.1007/978-3-540-85363-3_7] 2

[14] GUSTAV HAST: Approximating Max kCSP – outperforming a random assignment with almost a linear factor. In *Proc. 32nd Internat. Colloq. on Automata, Languages, and Programming (ICALP'05)*, pp. 956–968. Springer, 2005. [doi:10.1007/11523468_77] 2, 3

[15] SUBHASH KHOT: On the power of unique 2-prover 1-round games. In *Proc. 34th STOC*, pp. 767–775. ACM Press, 2002. [doi:10.1145/509907.510017] 2, 4, 7

[16] 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(1):319–357, 2007. [doi:10.1137/S0097539705447372] 2, 3, 4, 5

[17] SUBHASH KHOT, DOR MINZER, AND MULI SAFRA: Pseudorandom sets in Grassmann graph have near-perfect expansion. In *Proc. 59th FOCS*, pp. 592–601. IEEE Comp. Soc., 2018. [doi:10.1109/FOCS.2018.00062] 2, 3, 4, 5, 7

[18] SUBHASH KHOT AND ODED REGEV: Vertex cover might be hard to approximate to within $2-\varepsilon$. *J. Comput. System Sci.*, 74(3):335–349, 2008. [doi:10.1016/j.jcss.2007.06.019] 6

[19] SUBHASH KHOT AND RISHI SAKET: Approximating CSPs using LP relaxation. In *Proc. 42nd Internat. Colloq. on Automata, Languages, and Programming (ICALP'15)*, pp. 822–833. Springer, 2015. [doi:10.1007/978-3-662-47672-7_67] 3, 4, 21, 23, 25

[20] GUY KINDLER, ALEXANDRA KOLLA, AND LUCA TREVISAN: Approximation of non-Boolean 2CSP. In *Proc. 27th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'16)*, pp. 1705–1714. SIAM, 2016. [doi:10.1137/1.9781611974331.ch117] 2, 3, 4, 21

[21] KONSTANTIN MAKARYCHEV AND YURY MAKARYCHEV: Approximation algorithm for non-Boolean Max-k-CSP. *Theory of Computing*, 10(13):341–358, 2014. [doi:10.4086/toc.2014.v010a013] 2, 3, 4

[22] PASIN MANURANGSI, PREETUM NAKKIRAN, AND LUCA TREVISAN: Near-optimal UGC-hardness of approximating Max k-CSP$_R$. In *Proc. 19th Internat. Workshop on Approximation Algorithms for Combinat. Opt. Probl. (APPROX'16)*, pp. 15:1–28. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016. [doi:10.4230/LIPIcs.APPROX-RANDOM.2016.15] 1, 4

[23] ELCHANAN MOSSEL: Gaussian bounds for noise correlation of functions and tight analysis of long codes. In *Proc. 49th FOCS*. IEEE Comp. Soc., 2008. [doi:10.1109/FOCS.2008.44] 25

[24] ELCHANAN MOSSEL, RYAN O'DONNELL, AND KRZYSZTOF OLESZKIEWICZ: Noise stability of functions with low influences: Invariance and optimality. *Ann. Math.*, 171(1):295–341, 2010. [doi:10.4007/annals.2010.171.295] 5, 9, 10

[25] RYAN O'DONNELL: *Analysis of Boolean Functions*. Cambridge Univ. Press, 2014. CUP. 11, 21

[26] ALEX SAMORODNITSKY AND LUCA TREVISAN: A PCP characterization of NP with optimal amortized query complexity. In *Proc. 32nd STOC*, pp. 191–199. ACM Press, 2000. [doi:10.1145/335305.335329] 2

[27] ALEX SAMORODNITSKY AND LUCA TREVISAN: Gowers uniformity, influence of variables, and PCPs. *SIAM J. Comput.*, 39(1):323–360, 2009. Preliminary version in STOC'06. [doi:10.1137/070681612] 2

[28] LUCA TREVISAN: Parallel approximation algorithms by positive linear programming. *Algorithmica*, 21(1):72–88, 1998. [doi:10.1007/PL00009209] 2, 3

AUTHORS

Pasin Manurangsi
Senior Research Scientist
Google Research
Mountain View, CA, USA
pasin@google.com
https://pasin30055.github.io/

Preetum Nakkiran
Postdoctoral researcher
Halıcıoğlu Data Science Institute
University of California, San Diego
La Jolla, CA, USA
preetum@ucsd.edu
https://preetum.nakkiran.org/

Luca Trevisan
Professor of Computer Science
Department of Decision Sciences
Bocconi University
Milan, Italy
l.trevisan@unibocconi.it
https://lucatrevisan.github.io/

ABOUT THE AUTHORS

PASIN MANURANGSI received his Ph. D. from the University of California, Berkeley in 2019; his advisors were Prasad Raghavendra and Luca Trevisan. His thesis was on the intersection between hardness of approximation and fine-grained complexity. Prior to that, he received bachelor's and master's degrees from MIT, where he was advised by Dana Moshkovitz. He is currently a research scientist at Google Research, primarily working on theoretical and practical aspects of algorithmic privacy.

PREETUM NAKKIRAN is a postdoctoral researcher at the University of California, San Diego. He is hosted by Mikhail Belkin, and part of the NSF/Simons Collaboration on the Theoretical Foundations of Deep Learning. He currently works on scientific aspects of machine learning, trying to understand when and why deep learning works. He obtained his Ph. D. in 2021 from Harvard, advised by Madhu Sudan and Boaz Barak. He did his undergraduate work in EECS at the University of California, Berkeley.

LUCA TREVISAN is a professor of computer science at Bocconi University. Luca received his Ph. D. from the Sapienza University of Rome, advised Pierluigi Crescenzi. After graduating, Luca was a postdoc at MIT and at DIMACS, and he was on the faculty of Columbia University, U.C. Berkeley, and Stanford, before returning to Berkeley in 2014 and, at long last, returning to Italy in 2019.

Luca's research is in theoretical computer science, with a focus on computational complexity, on the analysis of algorithms, on the foundations of cryptography, and on topics at the intersection of theoretical computer science and pure mathematics.

Luca received the STOC'97 Danny Lewin (best student paper) award, the 2000 Oberwolfach Prize, a 2000 Sloan Fellowship and an NSF CAREER Award. He was an invited speaker at the 2006 International Congress of Mathematicians. In 2019, he received an ERC Advanced Grant.