Electronic Colloquium on Computational Complexity, Report No. 40 (2012)

Approximation Resistance on Satisfiable Instances for Predicates Strictly Dominating PARITY

Sangxia Huang
KTH — Royal Institute of Technology
sangxia@csc.kth.se

Abstract

In this paper, we study the approximability of Max CSP(P) where P is a Boolean predicate. We prove that assuming Khot's d-to-1 Conjecture, if the set of accepting inputs of P strictly contains all inputs with even (or odd) parity, then it is NP-hard to approximate Max CSP(P) better than the simple random assignment algorithm even on satisfiable instances. This is a generalization of a work by O'Donnell and Wu [15] which proved that it is NP-hard to approximate satisfiable instances of Max CSP(NTW) beyond $\frac{5}{8}+\varepsilon$ for any $\varepsilon > 0$ based on Khot's d-to-1 Conjecture, where NTW is the "Not Two" predicate of size 3.

1 Introduction

A k-bit Constraint Satisfaction Problem (k-CSP) consists of a set of boolean variables, along with boolean constraints each of which involves at most k of these variables. Each boolean constraint is given by some predicate of arity at most k. The Max k-CSP problem is to find a boolean assignment to the variables that maximizes the number of satisfied constraints. A k-CSP is called *satisfiable* if there exists an assignment that satisfies all the constraints simultaneously. We can further restrict the type of predicates in Max k-CSP to some predicate set. Specifically, for a predicate $P : \{0,1\}^k \to \{0,1\}$, the Max CSP($P$) problem is the Max k-CSP problem in which all constraints are of the form $P(l_1, \cdots, l_k)$, where each literal l_i is either a variable or a negated variable. For example, we have the following k-CSP problem on variables $x_1, \cdots, x_n$:

- Max-NTW: constraints are of the form $\mathrm{NTW}(l_i, l_j, l_k)$, where NTW is the 3-ary predicate satisfied if and only if the number of true inputs is not two.
- Max-PARITY: the predicate $\textsc{Parity}(l_{i_1}, \cdots, l_{i_k}) = \neg(\oplus_{j=1}^k l_{i_j})$ that accepts inputs with even parity.

Let $P^{-1}(1)$ be the set of satisfying assignments of P. In this paper, we are interested in predicates P such that $P^{-1}(1)$ properly contains $\textsc{Parity}^{-1}(1)$.

It is known that the Max k-CSP problem is NP-hard for any $k \geq 2$, and as a consequence, a lot of research have been focused on studying approximability of such problems. We say that a (randomized) algorithm has approximation ratio α, if for all instances, the algorithm is guaranteed to find an assignment which (in expectation) satisfies at least $\alpha \cdot Opt$ of the constraints, where Opt is the maximum number of simultaneously satisfied constraints over any assignment.

There is an easy approximation algorithm for Max k-CSP problems: one simply picks a random assignment to the variables. This algorithm has an approximation ratio of $1/2^k$. It

ISSN 1433-8092

was subsequently improved by [4]. As for a specific predicate P of arity k with m satisfying assignments, the above random assignment algorithm achieves a ratio of $m/2^k$.

Given the naiveness of the above algorithm, it may seem that one could do much better in approximating Max k-CSP. Surprisingly, however, it turns out that for certain P, this is the best possible algorithm. In a celebrated result, Håstad [8] showed that for Parity, the Max CSP(P) is NP-hard to approximate within $1/2+\varepsilon$, while the random assignment algorithm achieves $1/2$. Predicates P for which it is hard to approximate the Max CSP(P) problem better than a random assignment are called approximation resistant. A natural and important question is to understand the structure of approximation resistance. For $k = 2$ and $k = 3$, this question is resolved — predicates on 2 variables are never approximation resistant, and a predicate on 3 variables is approximation resistant if and only if it is implied by an XOR of the three variables [8, 23]. For $k = 4$, Hast [7] managed to classify most of the predicates with respect to approximation resistance.

However, for many other classical optimization problems, such as Max-Cut and Max-2SAT, there is still a gap between the best known approximation algorithm and hardness result. To address this, Khot [9] proposed the Unique Games Conjecture (UGC), which states that it is NP-hard to distinguish whether certain Label Cover instance is *almost* satisfiable or far from satisfiable. Assuming the UGC, many optimal approximation lower bound results were proved [12, 10, 11, 14, 1]. Raghavendra proved a powerful result in [16] that assuming the Unique Games Conjecture, if a certain natural SDP relaxation cannot approximate Max CSP(P) to within some factor α, then no polynomial time algorithm can. Also based on the UGC, Austrin and Mossel [2] proved that P is approximation resistant if the set of satsifying assignments $P^{-1}(1)$ contains the support of a pairwise independent distribution.

The case with satisfiable instances seems to be more intriguing as some predicates may behave differently when it comes to satisfiable instances. We call P approximation resistant on satisfiable instances if the best possible algorithm is still the random assignment algorithm even with the promise that it is satisfiable. If predicate P is approximation resistant on satisfiable instances, then clearly it is also approximation resistant. However, the converse may not be true. In fact, it is easy to see that if Max CSP(Parity) is satisfiable, we could find a satisfiable assignment to it, though Parity is approximation resistant on near satisfiable instances. Several other approximation algorithms for satisfiable instances were introduced [22], and in particular, it is known that predicates with fewer than $(k + 1)$ satisfying assignments are never approximation resistant on satisfiable instances. On the other hand, it is proved in [7] that predicates with fewer than $2\lfloor k/2\rfloor + 1$ accepting inputs are not approximation resistant so there is still a small gap between satisfiable and almost-satisfiable instances when k is odd.

Despite the abundance of results on approximation resistance of predicates, the understanding on approximation resistance on satisfiable instances is still quite limited. This situation is not particularly surprising, as there are quite a few differences between satisfiable instances and almost satisfiable instances. One of the differences is that there are approximation resistant predicates which are easy to satisfy on satisfiable instances, for example linear constraint predicates, thus it is possible that the structure of approximation resistance on satisfiable instances is richer than that of approximation resistance. Also, in terms of PCPs, working on satisfiable instances requires perfect completeness in the proofs, making the design and analysis more challenging. Most notably, the Unique Games Conjecture, which has been used to prove a number of optimal results, might not be applicable to satisfiable cases simply because it is non-perfect. To address this, Khot additionally proposed the "d-to-1 Conjectures" [9]. The conjecture states that it is NP-hard to distinguish whether a "d-to-1 Label Cover Instance" is satisfiable or far from satisfiable. The conjectures are parameterized by an integer constant $d \geq 2$. The bigger d is, the less restrictive are d-to-1 Label Cover instances; thus for each d, the d-to-1 Conjecture implies the $(d+1)$-to-1 Conjecture. There are several applications of the d-to-1 Conjectures prior to this work. Dinur, Mossel and Regev [5] proved that the 2-to-1 Conjecture implies hardness of coloring 4-colorable graphs by $O(1)$ colors. O'Donnell and Wu proved a nice result in [15] that Max-NTW is approximation resistant on satisfiable instances assuming the d-to-1 conjecture for some d. Following their approach, Tang [21] studied the approximability of satisfiable Max-$3CSP_q$ where q is a prime greater than 3 and gave a $(1/q + 1/q^2 - 1/q^3)$ lower bound by showing approximation lower bound for a special predicate which could be viewed as a generalization of the NTW in [15]. In [20], Tamaki and Yoshida gave a non-adaptive $(2^k - 1)$-query Long Code test with perfect completeness and soundness $(2q+3)/2^q$ where $q = 2^k - 1$ is the number of queries. Their test was a noised version of Samorodnitsky-Trevisan's Hyper Graph linearity test, and used Invariance-Principle style analysis in the spirit of [15]. However, it is not known if it is possible to combine their tester with d-to-1 Games.

1.1 Our Contribution

In this paper, we prove the following theorem:

Theorem 1.1. *Let $P : \{0,1\}^k \to \{0,1\}$ be a Boolean predicate such that $k \geq 4$ and $P^{-1}(1)$ properly contains* $\text{PARITY}^{-1}(1)$. *Suppose that Khot's d-to-1 Conjecture holds for some finite constant d. Then for any $\varepsilon > 0$, given a satisfiable Max-CSP(P) instance, it is NP-hard to satisfy more than a $|P^{-1}(1)|/2^k + \varepsilon$ fraction of the constraints.*

Based on the classical connection between PCPs and hardness of approximation, the above theorem is equivalent to saying that there is a nonadaptive PCP system for an NP-complete language with perfect completeness and soundness $|P^{-1}(1)|/2^k$, decides to accept or reject based on the predicate P. In the remaining part of this paper, we design such a PCP, and follow the classical approach of showing approximation resistance by arithmetizing the accepting probability and bounding each non-constant terms.

Our result is closely related to [15]. In fact, we can view the NTW predicate as accepting input $(0,0,0)$ in addition to the odd parity predicate. While the analysis of many terms for our predicates are in fact almost identical to [15], there are several difficulties in generalizing their result. First of all, we need to find a suitable distribution. More specifically, having a balanced and "sufficiently" independent testing distribution would ease the analysis for most of the terms. Another difficulty is the analysis of certain terms such as $\mathbf{E}[\prod g(y_i)]$. The similar term in [15] is upperbounded by bounding the expectation under another distribution for which the resulting expectation is no smaller than under the original distribution. To prove that the expectation never decreases when passing to the new distribution, O'Donnell and Wu developed an approach using matrix notation of distributions. This approach is not directly applicable, and we show how to deal with it in our case.

2 Preliminaries

In this section, we recall some basic notions and results. We give definitions for Label Cover problems and state Khot's d-to-1 Conjectures. We also recall some basics in Fourier Analysis of Boolean functions and the Efron-Stein decomposition, a concept closely related to Fourier decompositions. We then define the noise operator and correlation of probability spaces as in

[13], and quote several results on the relation between these concepts. Finally, we describe the overall framework of our PCP system, a generalization of the one in [15].

2.1 Khot's d-to-1 Conjecture

To introduce Khot's d-to-1 Conjecture, we first define the Label-Cover problem.

Definition 2.1. *We define a Label-Cover instance* $\mathcal{L} := (U, V, E, P, R_1, R_2, \Pi)$*. Here* U *and* V *are the two vertex sets of a bipartite graph and* E *is the set of edges between* U *and* V*.* P *is an explicitly given probability distribution on* E*.* R_1 *and* R_2 *are integers with* $1 \le R_1 \le R_2$*.* Π *is a collection of "projections", one for each edge:* $\Pi = \{\pi_e : [R_2] \to [R_1] | e \in E\}$*. A labeling* L *is a mapping* $L : U \to [R_1], V \to [R_2]$*. We say that an edge* $e = (u, v)$ *is "satisfied" by labeling* L *if* $\pi_e(L(v)) = L(u)$*. We define:*

$$Opt(\mathcal{L}) = \max_{L} \Pr_{e=(u,v)\sim P}[\pi_e(L(v)) = L(u)].$$

For Label-Cover problems, we have the following inapproximability theorem of Raz [17].

Theorem 2.2. *For every constant* $\eta > 0$*, there is some constant* $k(\eta) < \infty$ *such that for Label-Cover instance* $\mathcal{L}$ *with* $R_2 \ge k(\eta)$*, it is NP-hard to distinguish the case* $Opt(\mathcal{L}) = 1$ *from the case* $Opt(\mathcal{L}) \ge \eta$*.*

Now we define the d-to-1 Label-Cover.

Definition 2.3. *A projection* $\pi : [R_2] \to [R_1]$ *is said to be "d-to-1" if for each element* $i \in [R_1]$*, we have* $|\pi^{-1}(i)| \le d$*. We say the projection is "exactly d-to-1" if* $R_2 = dR_1$*, and* $|\pi^{-1}(i)| = d$ *for each* i*. The d-to-1 Label-Cover is the Label-Cover in which each projection is d-to-1.*

In Theorem 2.2, one can take the Label-Cover instances to be exactly d-to-1, but d needs to be at least $poly(1/\eta)$. Khot's d-to-1 Conjecture states that one can take d to be a constant independent of η.

Conjecture 2.4. *For each integer* $d \ge 2$*, for every constant* $\eta > 0$*, there is some constant* $k(d, \eta) < \infty$ *such that for d-to-1 Label-Cover instances* $\mathcal{L}$ *with* $R_2 \ge k(d, \eta)$*, it is NP-hard to distinguish the case* $Opt(\mathcal{L}) = 1$ *from the case* $Opt(\mathcal{L}) \le \eta$*.*

2.2 Fourier Analysis and Influences

As is usual in Fourier analysis, from now on we represent "true" by -1 and "false" by 1, instead of 1 and 0. For $S \subseteq [n]$, define $\chi_\sigma : \{-1, 1\}^n \to \mathbb{R}$ as $\chi_S(x_1, \cdots, x_n) = \prod_{i \in S} x_i$. The functions $\{\chi_S\}_{S \subseteq [n]}$ form an orthonormal basis for the $L^2(\{-1, 1\}^n)$, thus every function $f \in L^2(\{-1, 1\}^n)$ can be written as $f(x) = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S(x)$, where $\hat{f} : \mathcal{P}([n]) \to \mathbb{R}$ is defined by $\hat{f}(S) = 2^{-n} \sum_{x \in \{-1,1\}^n} f(x)\chi_S(x)$.

We now define a notion of the influence of a set of coordinates on a function f as in [15].

Definition 2.5. *For a function* $f : \{-1, 1\}^n \to \mathbb{R}$ *and a set of coordinates* $S \subseteq [n]$*, we define the influence of* S *on* f *to be*

$$Inf_S(f) = \sum_{U \supseteq S} \hat{f}(U)^2.$$

Next we recall the Bonami-Beckner operator T_ρ on Boolean functions.

Definition 2.6. *Let $0 \leq \rho \leq 1$. The Bonami-Beckner operator T_ρ is a linear operator mapping functions $g : \{-1,1\}^n \to \mathbb{R}$ into functions $T_\rho g : \{-1,1\}^n \to \mathbb{R}$ as $(T_\rho g)(x) = \mathbf{E}[g(y)]$, where y is formed by setting $y_i = x_i$ with probability ρ and setting y_i to be a uniformly random bit with probability $1-\rho$.*

There is a simple relation between the Fourier decomposition of f and $T_\rho f$.

Claim 2.7.

$$T_\rho f = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \chi_S.$$

The following lemma says that for a noised Boolean function, the sum of influences of constant-size sets is bounded.

Lemma 2.8. *[15] For any function $f : \{-1,1\}^n \to \mathbb{R}$ with $\mathbf{E}[f^2] \leq 1$, and any parameters $0 < \delta \leq 1/2$, $m \in \mathbb{N}$*

$$\sum_{S \subseteq [n], |S| \leq m} \mathit{Inf}_S(T_{1-\delta} f) \leq (m/2\delta)^m.$$

We also need the following Efron-Stein decomposition as is described in [13].

Definition 2.9. *Let (Ω_1, μ_1),...,(Ω_n, μ_n) be discrete probability spaces $(\Omega, \mu) = \prod_{i=1}^n (\Omega_i, \mu_i)$. The Efron-Stein decomposition of $f : \Omega \to \mathbb{R}$ is given by*

$$f(x) = \sum_{S \subseteq [n]} f_S(x_S),$$

where the functions f_S satisfy:

- *f_S depends only on x_S.*
- *For all $S \not\subseteq S'$ and all $x_{S'}$ it holds that $\mathbf{E}[f_S | X_{S'} = x_{S'}] = 0$. That is, as long as we have not conditioned on all the variables that f_S depends on, the conditional expectation is 0.*

It is known that the Efron-Stein decomposition exists and that it is unique. In fact, when $\Omega_i = \{-1,1\}$ for all i, the Efron-Stein decomposition of f is the same as its Fourier decomposition. For general Ω_i's, the Fourier decomposition depends on the choice of Fourier basis, while the Efron-Stein decomposition is invariant under different choices. We use Efron-Stein decomposition in the proof of Lemma B.4.

2.3 Correlation of Probability Spaces

We use an equivalent definition of correlation for probability spaces as defined in [13].

Definition 2.10. *Let $(\Omega \times \Psi, \mu)$ be a correlated probability space, μ is a distribution on the finite product set $\Omega \times \Psi$ and that the marginals of μ on Ω and Ψ have full support. Define the correlation between Ω and Ψ to be*

$$\rho(\Omega, \Psi; \mu) = \max_{\substack{f:\Omega \to \mathbb{R} \\ g:\Psi \to \mathbb{R}}} \{ |\mathbf{E}[fg]| \mid \mathbf{E}[f] = 0, \mathbf{E}[f^2] \leq 1, \mathbf{E}[g] = 0, \mathbf{E}[g^2] \leq 1 \},$$

where the expectation $\mathbf{E}[fg]$ is under μ, and $\mathbf{E}[f]$, $\mathbf{E}[f^2]$, $\mathbf{E}[g]$ and $\mathbf{E}[g^2]$ are under marginals of μ on corresponding variables.

The following lemma is useful for bounding correlation between probability spaces from 1.

Lemma 2.11. *[13] Let $(\Omega \times \Psi, \mu)$ be two correlated spaces such that the probability of the smallest atom in $\Omega \times \Psi$ is at least $\alpha > 0$. Define a bi-partite graph $G = (\Omega, \Psi, E)$ where $(a, b) \in \Omega \times \Psi$ satisfies $(a, b) \in E$ if $\mu(a, b) > 0$. Then if G is connected then $\rho(\Omega, \Psi; \mu) \leq 1 - \alpha^2$.*

Next we recall the definition of the conditional expectation operator.

Definition 2.12. *Let $(\Omega \times \Psi, \mu)$ be two correlated spaces. The conditional expectation operator U associated with (Ω, Ψ) is the operator mapping $f \in L^p(\Psi, \mu)$ to $Uf \in L^p(\Omega, \mu)$ by $(Uf)(x) = \mathbf{E}[f(Y)|X = x]$ for $x \in \Omega$ and $(X, Y) \in \Omega \times \Psi$ is distributed according to μ.*

An important property we need in the analysis is that the Efron-Stein decomposition commutes with the conditional expectation operator.

Proposition 2.13. *[13] Let $(\Omega \times \Psi, \mu) := (\prod \Omega_i \times \prod \Psi_i, \otimes \mu_i)$ be correlated space and let $T := \otimes T_i$ be the conditional expectation operator associated with Ω and Ψ. Suppose $f \in L^2(\Psi)$ has Efron-Stein decomposition $f(x) = \sum_{S \subseteq [n]} f_S(x_S)$. Then the Efron-Stein decomposition of Uf satisfies $(Uf)_S = U(f_S)$ for $S \subseteq [n]$.*

The following result shows that in the above setting, if the correlations between all Ω and Ψ are less than 1, then the L^2 norms of the high-degree terms of Uf are small.

Proposition 2.14. *[13] Assume the setting of Proposition 2.13 and that for all i, we have $\rho(\Omega_i, \Psi_i; \mu_i) \leq \rho_i$. Then for all f, we have $\|U(f_S)\|_2 \leq \left(\prod_{i \in S} \rho_i\right) \|f_S\|_2$.*

2.4 PCP System Framework

Consider a predicate P of arity $(r + 1) \geq 4$. Our PCP system is similar to O'Donnell and Wu's PCP for Max-NTW [15]. For a d-to-1 Label-Cover instance $\mathcal{L} = (U, V, E, \mathcal{P}, R_1, R_2, \Pi)$, a proof consists of a collection of truth tables of Boolean functions, one for each vertex. Let $\mathcal{T}_e$ be the test distribution defined on $\{-1, 1\}^{R_1} \times \prod_{i=1}^{r} \{-1, 1\}^{R_2}$; Our verifier checks the proof as following:

- Pick an edge $e = (u, v)$ from distribution P;
- Generate a tuple $(x, y_1, \cdots, y_r)$ from the test distribution $\mathcal{T}_e$;
- Accept if $P(f_u(x), g_v(y_1), \cdots, g_v(y_r))$.

As usual, we assume the functions f_u and g_v are folded, and therefore we can assume that all the functions h are odd, meaning that $h(-z) = -h(z)$ for all inputs z.

3 The Test Distribution

In this section, we define the test distribution $\mathcal{T}_e$ for predicate P of arity $r + 1$. For a positive integer r, let $[r] := \{1, \cdots, r\}$. For an input α to predicate P, let its coordinates be $\alpha_0, \alpha_1, \cdots, \alpha_r$.

For the picked edge e, we write $d_i = |\pi_e^{-1}(i)|$ for $i \in [R_1]$. The verifier now views f_u as $f_u : \mathcal{X}^1 \times \mathcal{X}^2 \times \cdots \times \mathcal{X}^{R_1} \rightarrow \{-1, 1\}$, where each $\mathcal{X}^i = \{-1, 1\}^{(i)}$. The verifier also views g_v

as a function over an R_1-fold product set $g_v : \mathcal{Y}_j^1 \times \mathcal{Y}_j^2 \times \cdots \times \mathcal{Y}_j^{R_1} \to \{-1, 1\}$, where each $\mathcal{Y}_j^i = \{-1, 1\}^{\pi_e^{-1}(i)}$ for $j \in [r]$. $\mathcal{T}_e$ is a distribution over the following R_1-fold product set

$$\prod_{i=1}^{R_1} \left(\mathcal{X}^i \times \prod_{j=1}^{r} \mathcal{Y}_j^i \right) \simeq \left(\prod_{i=1}^{R_1} \mathcal{X}^i \right) \times \prod_{j=1}^{r} \left(\prod_{i=1}^{R_1} \mathcal{Y}_j^i \right).$$

In fact, for each $i \in [R_1]$, we define a distribution $\mathcal{T}_e^i$ on $\mathcal{X}^i \times \prod_{j=1}^r \mathcal{Y}_j^i$. We let $\mathcal{T}_e$ be the product distribution $\otimes_{i=1}^{R_1} \mathcal{T}_e^i$. The distribution $\mathcal{T}_e^i$ only depends on d_i, and we could simply view it as a distribution on $\{-1, 1\} \times \prod_{i=1}^r \{-1, 1\}^{d_i}$. Further, the distributions for different d_i's are actually constructed in a uniform way, parameterized by d_i. More specifically, we define distributions on $\{-1, 1\} \times \prod_{i=1}^r \{-1, 1\}^D$, which we would write as $\mathcal{X} \times \prod_{i=1}^r \mathcal{Y}_i$ for simplicity. The final distribution is a product of distributions corresponding to each $i \in [R_1]$, substituting D with the respective d_is.

The test distribution is a combination of several simple distributions. The first such distribution is the parity distribution which we denote by $\mathcal{H}$ in its simple form and $\mathcal{H}(D)$ in its *product* form.

Definition 3.1. *Define distribution $\mathcal{H}$ generating $(x, y_1, \cdots, y_r)$ as follows: x, $y_1 \ldots y_{r-1}$ are generated independently and uniformly random, then set y_r to be $x \prod_{i=1}^{r-1} y_i$.*

Define distribution $\mathcal{H}(D)$ on $\mathcal{X} \times \prod_{i=1}^r \mathcal{Y}_i$ as follows: generate x, $y_{i,j}$ for $i \in [D]$, $j \in [r-1]$ independent and uniformly random, then for each $i \in [D]$, $y_{i,r} = x \prod_{j=1}^{r-1} y_{i,j}$.

Next we define the "noise" distributions for $\alpha \in P^{-1}(-1) \setminus \textsc{Parity}^{-1}(-1)$. Note that $\alpha \notin \textsc{Parity}^{-1}(-1)$, thus we have that $\prod_{i=0}^r \alpha_i = -1$, and therefore if we define α' to be the same as α except for the first bit where $\alpha'_0 = -\alpha_0$, then we have $\alpha' \in \textsc{Parity}^{-1}(-1)$. The noise distributions looks very similar to $\mathcal{H}$ and $\mathcal{H}(D)$. In fact, we first generate the parity distribution with some bias on the first bit. The difference now is that in the noise distribution, instead of generating $\alpha' \in \textsc{Parity}^{-1}(-1)$, we always generate α. We also assign different weights so that the marginal of the first bit is uniform. We can view $\mathcal{N}_\alpha$ as a generalization of the distribution $\mathcal{N}$ in [15]. In particular, we distribute probability mass in $\mathcal{N}_\alpha$ over a larger set. One could actually verify that when the size of the predicate is 3, and $\alpha = (1, 1, 1)$, it is exactly the distribution in which we pick $(-1, -1, -1)$ and $(1, 1, 1)$ with probability $\frac{1}{2}$ each.

Definition 3.2. *Define $\mathcal{N}_\alpha$ on $\{-1, 1\}^{r+1}$: generate $y_1, \cdots, y_{r-1}$ independently and uniformly random, and with probability $2^{r-2}/(2^{r-1} - 1)$, set $x = -\alpha_0$, and $x = \alpha_0$ otherwise. Then let $y_r = x \prod_{i=1}^{r-1} y_i$, and flip x if $(x, y_1, \cdots, y_r) = (-\alpha_0, \alpha_1, \cdots, \alpha_r)$.*

Define $\mathcal{N}_{k,\alpha}(D)$ on $\mathcal{X} \times \prod \mathcal{Y}_i$: generate $(x, y_{k,1}, \cdots, y_{k,r}) \sim \mathcal{N}_\alpha$, then generate $(y_{j,1}, \cdots, y_{j,r})$ for all $j \in [D] \setminus \{k\}$ according to parity distribution $\mathcal{H}$ and x. Let $\mathcal{N}_\alpha(D) := (\sum_k \mathcal{N}_{k,\alpha}(D))/D$.

The distribution $\mathcal{N}_\alpha$ has the nice property that the marginal is uniform if we only condition on y_S where $S \subsetneq [r]$.

Lemma 3.3. *Let $S \subsetneq [r]$ be a set of coordinates, and let y_S be an assignment on the coordinates in S. Then $\Pr_{\mathcal{N}_\alpha}[\boldsymbol{y}_S = y_S] = 2^{-|S|}$. Moreover, the marginal distribution over $\mathcal{X}$ is also uniform.*

Proof. If $r \notin S$, then since $\{y_i\}_{i \in S}$ are picked uniformly at random, the statement holds. If $r \in S$, the last step when generating from $\mathcal{N}_\alpha$ only flips x, and this does not change the probability we are considering if we do not condition on x. Before this final flipping, we have $y_r = x \prod_{i=1}^{r-1} y_i$, and we could change the role between y_r and y_j for some $j \notin S$, and conclude that the marginal is uniform.

For the marginal of $\mathcal{X}$, note that in the first step, $x = -\alpha_1$ with probability $2^{r-2}/(2^{r-1} - 1)$, and it would still be the case at the end unless $(x, y_1, \cdots, y_r) = (-\alpha_1, \alpha_2, \cdots, \alpha_{r+1})$ with probability $2^{-(r-1)} \cdot 2^{r-2}/(2^{r-1}-1)$, thus the probability that $x = -\alpha_1$ is $(1-2^{-(r-1)}) \cdot 2^{r-2}/(2^{r-1} - 1) = 1/2$. $\square$

It turns out that all $\mathcal{N}_\alpha$ are suitable for our purpose, therefore we simply pick an arbitrary $\alpha \in P^{-1}(-1) \setminus \textsc{Parity}^{-1}(-1)$ and omit indicating the dependence on α in the rest of this paper. We are now ready to define the test distribution $\mathcal{T}_e$.

Definition 3.4. *For $0 < \gamma < 1$, define distribution $\mathcal{H}_\gamma(D) = (1-\gamma)\mathcal{H}(D) + \gamma\mathcal{N}(D) = (1-\gamma)\mathcal{H}(D) + \gamma(\sum \mathcal{N}_k(D))/D$. For each $i \in [R_1]$, define the $\mathcal{T}_e^i$ to be $\mathcal{H}_\gamma(d_i)$. The test distribution $\mathcal{T}_e$ is the product of these distributions, $\mathcal{T}_e = \otimes_{i=1}^{R_1} \mathcal{T}_e^i$.*

We have the following bound on $\rho(\mathcal{X}, \prod_{i \in S} \mathcal{Y}_i; \mathcal{H}_\gamma(D))$ for $S \subsetneq [r]$. We need to apply Lemma 3.3 and the proof is similar to that of Lemma 5.3 in [15].

Lemma 3.5. *For all $S \subsetneq [r]$, we have $\rho(\mathcal{X}, \prod_{i \in S} \mathcal{Y}_i; \mathcal{H}_\gamma(D)) \leq \gamma$.*

Lemma 3.6. *For any $r_0 \in [r]$, $\alpha \in P^{-1}(-1) \setminus \textsc{Parity}^{-1}(-1)$, $\rho(\mathcal{X} \times \prod_{i \neq r_0} \mathcal{Y}_i, \mathcal{Y}_{r_0}; \mathcal{H}_\gamma(D)) \leq 1 - \beta^2/2$, where $\beta = \gamma \cdot (2^{r-2} - 1)/((2^{r-1} - 1) \cdot 2^{(r-1)D} \cdot D)$ is a lower bound of the least probability of an atom under $\mathcal{H}_\gamma(D)$.*

Proof. For notational simplicity, we only prove this for $r_0 = r$, since all coordinates are entirely symmetric. To apply Lemma 2.11, we only need to show that the distribution is connected. Let $(w_1, \ldots, w_D)$ be an arbitrary right vertex, and $(w'_1, \cdots, w'_D)$ be the right vertex that has $w'_j = w_j$ for $j \neq k$, and $w'_k = \alpha_r$. We claim that for any $k \in [D]$, $(w_1, \cdots, w_D)$ and $(w'_1, \cdots, w'_D)$ are always connected. In fact, if we already have $w_k = \alpha_r$, then we are done. Otherwise, we pick the following left vertex $(x, y_{1,1}, y_{1,2}, \cdots, y_{1,r-1}, y_{2,1}, \cdots, y_{D,r-1})$, such that it is connected to both $(w_1, \cdots, w_D)$ and $(w'_1, \cdots, w'_D)$: $x = \alpha_0$, $(y_{j,1}, y_{j,2}, \cdots, y_{j,r-2}, y_{j,r-1}) = (1, 1, \cdots, 1, \alpha_0 w_j)$ for $j \neq k$, and $(y_{k,1}, y_{k,2}, \cdots, y_{k,r-2}, y_{k,r-1}) = (\alpha_1, \alpha_2, \cdots, \alpha_{r-1})$. It is not hard to see that for any $j \neq k$, $(x, y_{j,1}, \cdots, y_{j,r-1}, w_j) \in \mathcal{H}$, and $(x, y_{k,1}, \cdots, y_{k,r-1}, w_k) \in \mathcal{H}$, $(x, y_{k,1}, \cdots, y_{k,r-1}, w'_k) \in \mathcal{N}_\alpha$. Therefore, it is connected to the right vertex $(w_1, \cdots, w_D)$ due to $\mathcal{H}(D)$, and to $(w'_1, \cdots, w'_D)$ due to $\mathcal{N}_k(D)$.

We conclude from the above claim that all right vertices are connected to the right vertex $(\alpha_r, \cdots, \alpha_r)$. It is also easy to see that all left vertices are at least connected with one right vertex, therefore the bipartite graph is connected. $\square$

Similar to [15], we need to pass to a new distribution $\mathcal{I}_\gamma(D)$ with almost no correlation between $\mathcal{X}$ and $\prod \mathcal{Y}_i$.

Definition 3.7. *Define distribution $\mathcal{I}(D)$ on $\mathcal{X} \times \prod_{i=1}^r \mathcal{Y}_i$ as follows: first draw from $\mathcal{H}(D)$, then uniformly rerandomize the bit $\boldsymbol{x}$. Define $\mathcal{I}_\gamma(D) := (1-\gamma)\mathcal{I}(D) + \gamma\mathcal{N}(D)$.*

We bound $\rho(\mathcal{X}, \prod \mathcal{Y}_i; \mathcal{I}_\gamma(D))$ in the following lemma.

Lemma 3.8. $\rho(\mathcal{X}, \prod_{i=1}^r \mathcal{Y}_i; \mathcal{I}_\gamma(D)) \leq \sqrt{\gamma}$.

The proof of this lemma is almost identical to Lemma 5.4 of [15].

By Proposition 2.13 of [13], bounds in Lemma 3.5, 3.6 and 3.8 holds for products of correlated spaces.

Lemma 3.9.

$$\rho\left(\prod_{j=1}^{R_1}\mathcal{X}^j, \prod_{j=1}^{R_1}\left(\prod_{i\in S}\mathcal{Y}_i^j\right); \bigotimes_{i=1}^{R_1}\mathcal{H}_\gamma(d_i)\right) \le \gamma;$$
$$\rho\left(\prod_{j=1}^{R_1}\left(\mathcal{X}^j \times \prod_{i\neq r_0}\mathcal{Y}_i^j\right), \prod_{j=1}^{R_1}\mathcal{Y}_{r_0}^j; \bigotimes_{j=1}^{R_1}\mathcal{H}_\gamma(d_i)\right) \le 1-\beta^2/2;$$
$$\rho\left(\prod_{j=1}^{R_1}\mathcal{X}^j, \prod_{j=1}^{R_1}\left(\prod_{i=1}^{r}\mathcal{Y}_i^j\right); \bigotimes_{i=1}^{R_1}\mathcal{I}_\gamma(d_i)\right) \le \sqrt{\gamma},$$

where $\beta = \gamma\cdot(2^{r-2}-1)/((2^{r-1}-1)\cdot 2^{(r-1)d}\cdot d)$.

4 Analysis of the Verifier

In this section we give details of the analysis of our verifier.

The completeness analysis is standard.

Theorem 4.1. *The verifier has completeness 1.*

Proof. Let $\mathcal{L}$ be a given d-to-1 Label Cover instance, and $L: U \to [R_1], V \to [R_2]$ a perfect labeling. For each $u \in U$, the prover takes f_u to be the $L(u)$-th dictator function $\chi_{\{L(u)\}}$, and for each $v \in V$, take g_v to be the $L(v)$-th dictator function $\chi_{\{L(v)\}}$. Now for any edge, we have that $\pi_e(L(v)) = L(u)$. It follows from the definition of $\mathcal{T}_e$ that the tuple $(x_{L(u)}, y_{L(v),1}, \cdots, y_{L(v),r})$ generated is in the support of P. Hence the verifier has perfect completeness. □

Next we show soundness.

Theorem 4.2. *For a d-to-1 Label Cover instance $\mathcal{L}$, if the verifier accepts with probability $|P^{-1}(-1)|/2^{r+1}+\varepsilon$, then there is a randomized labeling strategy for $\mathcal{L}$ achieving expected value at least η, a positive constant depending only on d and ε.*

Proof. To analyze the soundness of the verifier, we first arithmetize the probability a given proof passes

$$\begin{aligned}\Pr[\text{verifier accepts}] &= \Pr_{e\sim\mathcal{P}}\Pr_{\mathcal{T}_e}[P(f_u(\boldsymbol{x}), g_v(\boldsymbol{y}_1), \cdots, g_v(\boldsymbol{y}_r))] \\ &= \mathop{\mathbf{E}}_{\substack{e\sim\mathcal{P}\\ \mathcal{T}_e}}\left[\sum_{S\subseteq\{0,\cdots,r\}}\hat{P}(S)\chi_S(f_u(x), g_v(\boldsymbol{y}_1), \cdots, g_v(\boldsymbol{y}_r))\right].\end{aligned}$$

Note that $\hat{P}(\emptyset) = |P^{-1}(1)|/2^{r+1}$. By Lemma 3.3 and the fact that f_u and g_v are odd functions, we conclude that $\mathbf{E}_{e\sim\mathcal{P},(x,\{\boldsymbol{y}_i\})\sim\mathcal{T}_e}[\chi_S(f_u(x), g_v(\boldsymbol{y}_1), \cdots, g_v(\boldsymbol{y}_r))] = 0$ for either $S \subsetneq [r]$ or $S = \{0\}$. Also, by Lemma 3.9, the absolute value of the terms with $S \subsetneq [r]$ and $0 \in S$ are upperbounded by γ. Therefore

$$\begin{aligned}\Pr[\text{verifier accepts}] &\le \frac{|P^{-1}(1)|}{2^{r+1}} + \frac{(2^{r-1}-1)\gamma}{2^r} \\ &+ \mathop{\mathbf{E}}_{\mathcal{T}_e}\left[\hat{P}([r])\prod_{i=1}^{r} g_v(\boldsymbol{y}_i) + \hat{P}([r]\cup\{0\}) f_u(\boldsymbol{x})\prod_{i=1}^{r} g_v(\boldsymbol{y}_i)\right]. \end{aligned} \tag{1}$$

We bound the remaining two terms in the following two theorems:

Theorem 4.3. *For any $e = (u, v)$, and odd function $g_v : \{-1, 1\}^{R_2} \to \{-1, 1\}$, we have $\mathbf{E}_{\mathcal{T}_e}[\prod_{i=1}^{r} g_v(\boldsymbol{y}_i)] \le \gamma$.*

To show the above bound, we apply the matrix approach in [15] to the distribution conditioned on $\boldsymbol{y}_i = y_i$ for $i \in [r-2]$ for any $\{y_i\}_{i\in[r-2]}$. We use the matrix approach to bound the absolute value of the expectation in the conditional distribution. Details are given in Appendix A.

Theorem 4.4. *There exists constants $\delta', \tau > 0$ depending only on d, r and γ, such that the following holds: if for every $i \in [R_1]$ and every odd-cardinality set $S \subseteq \pi_e^{-1}(i)$ we have that $\min\{Inf_i(T_{1-\delta'/2} f_u), Inf_S(T_{1-\delta'/2} g_v)\} \le \tau$, then $|\mathbf{E}_{\mathcal{T}_e}[f_u(\boldsymbol{x}) \prod_{i=1}^{r} g_v(\boldsymbol{y}_i)]| \le (r+3)\sqrt{\gamma}$.*

Proof of Theorem 4.4 is almost the same as [15] and we include it in Appendix B.

Now let us see how Theorem 4.2 follows from Theorem 4.3 and Theorem 4.4. We've proved that under the hypothesis of Theorem 4.4, (1) is upperbounded by

$$\frac{|P^{-1}(1)|}{2^{r+1}} + \frac{(2^{r-1}-1)\gamma}{2^r} + \gamma + (r+3)\sqrt{\gamma} \le \frac{|P^{-1}(1)|}{2^{r+1}} + 2\gamma + (r+3)\sqrt{\gamma}.$$

Equivalently, suppose that the functions f_u and g_v cause the verifier to accept with probability exceeding $|P^{-1}(1)|/2^{r+1} + \varepsilon = |P^{-1}(1)|/2^{r+1} + 2\gamma + 2(r+3)\sqrt{\gamma}$, then $|\mathbf{E}_{\mathcal{T}_e}[f_u(\boldsymbol{x}) \prod_{i=1}^{r} g_v(\boldsymbol{y}_i)]| > 2(r+3)\sqrt{\gamma}$. By an averaging argument, this implies that for at least an $(r+3)\sqrt{\gamma}$ fraction of edges under distribution P over the edges, we have $|\mathbf{E}_{\mathcal{T}_e}[f_u(\boldsymbol{x}) \prod_{i=1}^{r} g_v(\boldsymbol{y}_i)]| > (r+3)\sqrt{\gamma}$. We call such edges "good".

By Theorem 4.4, we know for evey good edge e, there must exist some $i_e \in [R_1]$ and odd cardinality set $S_e \subseteq \pi_e^{-1}(i_e)$ such that $\min\{\mathrm{Inf}_i(T_{1-\delta'/2} f_u), \mathrm{Inf}_S(T_{1-\delta'/2} g_v)\} > \tau$. For each $u \in U$, we define $L_u = \{i \in [R_1] : \mathrm{Inf}_i(T_{1-\delta'/2} f_u) > \tau\}$, for each $v \in V$, we define $L_v = \{j \in [R_2] : j \in S, \mathrm{Inf}_S(T_{1-\delta'/2} g_v) > \tau, |S| \le d, |S| \text{ is odd}\}$. By Lemma 2.8, we know that for g_v, we have $\sum_{|S|\le d} \mathrm{Inf}_S(T_{1-\delta'/2} g_v) \le (d/\delta')^d$. Therefore, at most $(d/\delta')^d/\tau$ sets S can contribute in the definition of L_v, and thus for each $v \in V$, we have $|L_v| \le d \cdot (d/\delta')^d/\tau$. Similarly, we have that for each $u \in U$ $|L_u| \le 1/\delta'\tau$.

Whenever $e = (u, v)$ is a good edge, $i_e \in L_u$ and S_e contributes to L_v. Since S_e is odd, it is nonempty, and thus there exists $j_e \in L_v$ and $i_e \in L_u$ such that $\pi_e(j_e) = i_e$. Therefore, for a good edge, a randomized labeling has at least a $1/(|L_u||L_v|) \ge \tau^2(\delta'/d)^{d+1}$ chance for choosing i_e and j_e and thus satisfying e. Since at least an $(r+3)\sqrt{\gamma}$ fraction (with respect to distribution P) of edges are good, the expected fraction of edge weight satisfied by a randomized labeling exceeds $(r+3)\sqrt{\gamma}\tau^2(\delta'/d)^{d+1}$, a positive constant depending only on d, r and ε, as desired. This completes the proof of Theorem 4.2. □

5 Conclusion

In this work, we generalized O'Donnell and Wu's work [15] on the inapproximability of Max-NTW and showed that for any boolean predicate P of size greater than 3, if its set of accepting input strictly contains all inputs of even (or odd) parity, then assuming the d-to-1 Conjecture for some d, it is NP-hard to approximate Max-CSP(P) better than the random assignment algorithm. While the overall analysis is in a similar flavor to [15], we generalized the design of the test distribution in a new way, and strengthened some part of the analysis so that it worked in general settings.

The question of approximability on satisfiable instances is still widely open. It would be nice to study if there are any similar results as in [2, 16]. An interesting next step would be to try to understand the situation where the predicates imply some general linear constraints, such as the Samorodnitsky-Trevisan predicates in [18, 19].

References

[1] Per Austrin. Balanced Max 2-Sat Might Not be the Hardest. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, STOC '07, pages 189–197, New York, NY, USA, 2007. ACM.

[2] Per Austrin and Elchanan Mossel. Approximation Resistant Predicates from Pairwise Independence. *Computational Complexity*, 18(2):249–271, 2009.

[3] A. Bonami. Ensembles $\Gamma(p)$ dans le dual de D^∞. *Ann. Inst. Fourier*, 18(2):804–915, 1998.

[4] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Approximation Algorithm for the Max k-CSP Problem. *Electronic Colloquium on Computational Complexity (ECCC)*, 13(063), 2006.

[5] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 344–353, New York, NY, USA, 2006. ACM.

[6] Leonard Gross. Logarithmic Sobolev Inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

[7] Gustav Hast. Beating a Random Assignment. In *APPROX-RANDOM*, pages 134–145, 2005.

[8] Johan Håstad. Some Optimal Inapproximability Results. *J. ACM*, 48(4):798–859, 2001.

[9] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, STOC '02, pages 767–775, New York, NY, USA, 2002. ACM.

[10] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal Inapproximability Results for MAX-CUT and Other 2-Variable CSPs? *SIAM J. Comput.*, 37:319–357, April 2007.

[11] Subhash Khot and Ryan O'Donnell. SDP gaps and UGC-hardness for MAXCUTGAIN. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 217–226, Washington, DC, USA, 2006. IEEE Computer Society.

[12] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2-ε. *J. Comput. Syst. Sci.*, 74:335–349, May 2008.

[13] Elchanan Mossel. Gaussian Bounds for Noise Correlation of Functions and Tight Analysis of Long Codes. In *FOCS*, pages 156–165, 2008.

[14] Ryan O'Donnell and Yi Wu. An Optimal SDP Algorithm for Max-Cut, and Equally Optimal Long Code Tests. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 335–344, New York, NY, USA, 2008. ACM.

[15] Ryan O'Donnell and Yi Wu. Conditional hardness for satisfiable 3-CSPs. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 493–502, New York, NY, USA, 2009. ACM.

[16] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *STOC*, pages 245–254, 2008.

[17] Ran Raz. A Parallel Repetition Theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

[18] Alex Samorodnitsky and Luca Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, STOC '00, pages 191–199, New York, NY, USA, 2000. ACM.

[19] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and PCPs. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 11–20, New York, NY, USA, 2006. ACM.

[20] Suguru Tamaki and Yuichi Yoshida. A Query Efficient Non-adaptive Long Code Test with Perfect Completeness. In *APPROX-RANDOM*, pages 738–751, 2010.

[21] Linqing Tang. Conditional Hardness of Approximating Satisfiable Max $3\mathrm{CSP}_q$. In *ISAAC*, pages 923–932, 2009.

[22] Luca Trevisan. Approximating Satisfiable Satisfiability Problems. *Algorithmica*, 28(1):145–172, 2000.

[23] Uri Zwick. Approximation Algorithms for Constraint Satisfaction Problems Involving at Most Three Variables per Constraint. In *SODA*, pages 201–210, 1998.

Appendix

A Analyzing $\mathbf{E}_{\mathcal{T}_e}[\prod_{i=1}^{r} g_v(\boldsymbol{y}_i)]$

In this section, we prove Theorem 4.3. We use the same approach as in [15]. However, the approach in [15] is defined on two variables, while in our case we have r of them. In order to use the matrix notation, we study the conditional distribution where $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{r-2}$ is given. Let $\mathcal{H}(D, \{\boldsymbol{z}_i\}_{i\in[r-2]})$ be the distribution $\mathcal{H}(D)$ conditioned on $\boldsymbol{y}_i = \boldsymbol{z}_i$ for $i = 1, \cdots, r-2$. Similarly, we define $\mathcal{N}(D, \{\boldsymbol{z}_i\}_{i\in[r-2]})$, $\mathcal{N}_k(D, \{\boldsymbol{z}_i\}_{i\in[r-2]})$ and $\mathcal{H}_\gamma(D, \{\boldsymbol{z}_i\}_{i\in[r-2]})$.

Let $M(\mathcal{P})$ be the $2^D \times 2^D$ matrix associated with distribution $\mathcal{P}$ defined on $\{-1,1\}^D \times \{-1,1\}^D$, defined as

$$M(\mathcal{P})_{S,T} = \mathop{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{P}}[\chi_S(\boldsymbol{x})\chi_T(\boldsymbol{y})].$$

For a function $g : \{-1,1\}^D \to \mathbb{R}$, we also identify it with a column vector of length 2^D, with entries indexed by $S \subset [D]$ in the same order as in the matrix notation. The S-th entry of g is $\hat{g}(S)$. Then we have

$$\mathop{\mathbf{E}}_{\mathcal{P}}[g(\boldsymbol{x})g(\boldsymbol{y})] = g^T M(\mathcal{P}) g.$$

Thus to bound $\mathbf{E}_{\mathcal{H}_\gamma(D)}[\prod_{i=1}^{r} g(\boldsymbol{y}_i)]$, all we need is to upperbound the absolute value of

$$g^T \left(\otimes_{i=1}^{R_1} M(\mathcal{H}_\gamma(D, \{\boldsymbol{y}_j\}_{j\in[r-2]}))\right) g.$$

The following proposition is easy to check.

Proposition A.1. *For any* $\{\boldsymbol{y}_i\}_{i\in[r-2]}$*,* $M(\mathcal{H}(D,\{\boldsymbol{y}_i\}_{i\in[r-2]}))_{S,T}$ *is nonzero iff* $S=T$ *and* $|S|$ *and* $|T|$ *are even, in which case it is equal to 1.*

Define the distribution $\mathcal{E}(D)$ on $\mathcal{X}\times\prod_{i=1}^r \mathcal{Y}_i$ which generates pairs $(\boldsymbol{y}_{r-1},\boldsymbol{y}_r)$ by choosing $\boldsymbol{y}_{r-1}$ uniformly at random and setting $\boldsymbol{y}_r=\boldsymbol{y}_{r-1}$, regardless of the values of $\boldsymbol{x}$ and other $\boldsymbol{y}_i$'s. $M(\mathcal{E}(D,\{y_j\}_{j\in[r-2]}))$ is then the identity matrix. Further introduce $\mathcal{E}_\gamma(D)=(1-\gamma)\mathcal{H}(D)+\gamma\mathcal{E}(D)$.

The proof is divided into two steps. First, we show that the absolute value of the expectation under $\mathcal{H}_\gamma(D,\{\boldsymbol{y}_j\}_{j\in[r-2]})$ is upper bounded by the expectation under $\mathcal{E}_\gamma(D,\{\boldsymbol{y}_j\}_{j\in[r-2]})$. Then we derive a bound for the latter. The overall proof is the same as in [15], but we have to make sure that everything still works when conditioning on any value of $\{\boldsymbol{y}_j\}_{j\in[r-2]}$. In fact, conditioning does not change $\mathcal{H}(D)$ at all, while the affect of $\mathcal{N}(D)$ is limited.

To finish the first step, we need the following lemma in matrix algebra.

Lemma A.2 ([15])**.** *Let* A_i *and* B_i *be* $m_i\times m_i$ *matrices,* $i=1..n$*, and suppose that* A_i-B_i *and* A_i+B_i *are positive semidefinite. Then* $\bigotimes_{i=1}^n A_i-\bigotimes_{i=1}^n B_i$ *and* $\bigotimes_{i=1}^n A_i+\bigotimes_{i=1}^n B_i$ *are positive semidefinite.*

Lemma A.3. *For any fixed values of* $\{\boldsymbol{y}_j\}_{j\in[r-2]}$*, the matrices*

$$\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\gamma(d_i,\{\boldsymbol{y}_j\}_{j\in[r-2]}))\pm\bigotimes_{i=1}^{R_1} M(\mathcal{H}_\gamma(d_i,\{\boldsymbol{y}_j\}_{j\in[r-2]}))$$

are positive semidefinitive.

Proof. By Lemma A.2, we only need to show that for each $D\geq 1$, the matrices

$$M(\mathcal{E}_\gamma(D,\{\boldsymbol{y}_i\}_{i\in[r-2]}))\pm M(\mathcal{H}_\gamma(D,\{\boldsymbol{y}_i\}_{i\in[r-2]}))$$

are positive semidefinite. For notational simplicity, we henceforth omit showing the dependence on D and $\{\boldsymbol{y}_i\}_{i\in[r-2]}$.

For the conditional distributions, we still have

$$\mathcal{H}_\gamma(D,\{\boldsymbol{y}_i\}_{i\in[r-2]})=(1-\gamma)\mathcal{H}(D,\{\boldsymbol{y}_i\}_{i\in[r-2]})+\gamma\mathcal{N}(D,\{\boldsymbol{y}_i\}_{i\in[r-2]})$$

and

$$\mathcal{E}_\gamma(D,\{\boldsymbol{y}_i\}_{i\in[r-2]})=(1-\gamma)\mathcal{H}(D,\{\boldsymbol{y}_i\}_{i\in[r-2]})+\gamma\mathcal{E}(D,\{\boldsymbol{y}_i\}_{i\in[r-2]}),$$

therefore $M(\mathcal{E}_\gamma)-M(\mathcal{H}_\gamma)=\gamma(M(\mathcal{E})-M(\mathcal{N}))$. Hence, to show $M(\mathcal{E}_\gamma)-M(\mathcal{H}_\gamma)$ is positive semidefinite, we only need to show it for $M(\mathcal{E})-M(\mathcal{N})$. For any $h:\{-1,1\}^D\to\mathbb{R}$, we have

$$\begin{aligned} h^T M(\mathcal{N})h &= \mathop{\mathbf{E}}_{(\boldsymbol{y}_{r-1},\boldsymbol{y}_r)\sim\mathcal{N}}[h(\boldsymbol{y}_{r-1})h(\boldsymbol{y}_r)] \\ &\leq \sqrt{\mathop{\mathbf{E}}_{(\boldsymbol{y}_{r-1},\boldsymbol{y}_r)\sim\mathcal{N}}[h(\boldsymbol{y}_{r-1})^2]}\sqrt{\mathop{\mathbf{E}}_{(\boldsymbol{y}_{r-1},\boldsymbol{y}_r)\sim\mathcal{N}}[h(\boldsymbol{y}_r)^2]} \end{aligned}$$

by Cauchy-Schwarz. Note that the marginals of $\mathcal{N}$ are uniform by Lemma 3.3 (actually marginals of $(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_{r-1})$ and $(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_{r-2},\boldsymbol{y}_r)$ since we are already conditioning on $\{\boldsymbol{y}_i\}_{i\in[r-2]}$). The marginals are also uniform for $\mathcal{E}$, therefore

$$\begin{aligned} &\sqrt{\mathop{\mathbf{E}}_{(\boldsymbol{y}_{r-1},\boldsymbol{y}_r)\sim\mathcal{N}}[h(\boldsymbol{y}_{r-1})^2]}\sqrt{\mathop{\mathbf{E}}_{(\boldsymbol{y}_{r-1},\boldsymbol{y}_r)\sim\mathcal{N}}[h(\boldsymbol{y}_r)^2]} \\ =\ &\mathbf{E}[h^2]=\mathop{\mathbf{E}}_{(\boldsymbol{y}_{r-1},\boldsymbol{y}_r)\sim\mathcal{E}}[h(\boldsymbol{y}_{r-1})h(\boldsymbol{y}_r)]=h^T M(\mathcal{E})h, \end{aligned}$$

so we've shown that $h^T M(\mathcal{N})h \le h^T M(\mathcal{E})h$ for all h, and hence $M(\mathcal{E}) - M(\mathcal{N})$ is positive semidefinite as needed.

As for the matrix $M(\mathcal{E}_\gamma) + M(\mathcal{H}_\gamma)$, it equals $2(1-\gamma)M(\mathcal{H}) + \gamma(M(\mathcal{E}) + M(\mathcal{N}))$. $M(\mathcal{H})$ is diagonal with only nonnegative numbers on the diagonal, hence it is positive semidefinite.

It remains to show that $M(\mathcal{E}) + M(\mathcal{N})$ also is. The proof is essentially the same: we start with $h^T(-M(\mathcal{N}))h = \mathbf{E}_{(\boldsymbol{y}_{r-1},\boldsymbol{y}_r)\sim\mathcal{N}}[-h(\boldsymbol{y}_{r-1})h(\boldsymbol{y}_r)]$ and the minus sign disappears with the application of Cauchy-Schwarz. □

Lemma A.4.
$$\left| g^T \left(\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\gamma(d_i)) \right) g \right| \le \gamma.$$

Proof. $M(\mathcal{H}(D))$ is a diagonal matrix with (S,S) equal to 0 if $|S|$ is odd and 1 if $|S|$ is even, $M(\mathcal{E}(D))$ is the identity matrix. Thus $M(\mathcal{E}_\gamma(D))$ is a diagonal matrix whose (S,S) entry is equal to 1 if $|S|$ is even and equal to γ if $|S|$ is odd. According to definitions, it follows that

$$\left| g^T \left(\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\gamma(d_i)) \right) g \right| = \sum_{S\subseteq[R_2]} \hat{g}(S)^2 \cdot \gamma^{\#\{i\in[R_1]:|S\cap\pi^{-1}(i)|\text{is odd}\}}. \tag{2}$$

But g is an odd function, and therefore $\hat{g}(S)^2$ is nonzero only if $|S|$ is odd, which implies that $|S\cap\pi^{-1}(i)|$ is odd for at least one i, and hence (2) is upper-bounded by

$$\sum_{S\subseteq[R_2]} \hat{g}(S)^2 \cdot \gamma = \mathbf{E}[g^2] \cdot \gamma = \gamma.$$

□

Theorem A.5. *For any $e = (u,v)$, $g_v : \{-1,1\}^{R_2} \to \{-1,1\}$ is odd implies that for any $\{\boldsymbol{y}_i\}_{i\in[r-2]}$*
$$\left| \mathop{\mathbf{E}}_{\mathcal{T}_e(\{\boldsymbol{y}_i\}_{i\in[r-2]})} [g(\boldsymbol{y}_{r-1})g(\boldsymbol{y}_r)] \right| \le \gamma.$$

Proof. Using the matrix notation, we have

$$\begin{aligned} \mathop{\mathbf{E}}_{\mathcal{T}_e(\{\boldsymbol{y}_i\}_{i\in[r-2]})} [g(\boldsymbol{y}_{r-1})g(\boldsymbol{y}_r)] &= g^T M(\mathcal{T}_e(\{\boldsymbol{y}_i\}_{i\in[r-2]}))g \\ &= g^T \left(\bigotimes_{i=1}^{R_1} M(\mathcal{H}_\gamma(d_i, \{\boldsymbol{y}_i\}_{i\in[r-2]})) \right) g. \end{aligned} \tag{3}$$

We bound (3) by Lemma A.3

$$g^T \left(\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\gamma(d_i)) - \bigotimes_{i=1}^{R_1} M(\mathcal{H}_\gamma(d_i, \{\boldsymbol{y}_i\}_{i\in[r-2]})) \right) g \ge 0,$$
$$g^T \left(\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\gamma(d_i)) + \bigotimes_{i=1}^{R_1} M(\mathcal{H}_\gamma(d_i, \{\boldsymbol{y}_i\}_{i\in[r-2]})) \right) g \ge 0,$$

which implies that

$$-g^T\left(\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\gamma(d_i))\right)g \quad\leq\quad g^T\left(\bigotimes_{i=1}^{R_1} M(\mathcal{H}_\gamma(d_i,\{\boldsymbol{y}_i\}_{i\in[r-2]}))\right)g \leq g^T\left(\bigotimes_{i=1}^{R_1} M(\mathcal{E}_\gamma(d_i))\right)g.$$

By Lemma A.4

$$\left|\underset{\mathcal{T}_e(\{\boldsymbol{y}_i\}_{i\in[r-2]})}{\mathbf{E}}[g(\boldsymbol{y}_{r-1})g(\boldsymbol{y}_r)]\right| \leq \gamma. \tag{4}$$

□

B Analyzing $\mathbf{E}[f_u(\boldsymbol{x})\prod_{i=1}^r g_v(\boldsymbol{y}_i)]$

In this section, we prove Theorem 4.4. The analysis of this term is almost an exact copy of O'Donnell and Wu's approach. We include the full analysis here for reader's convenience.

Let $\mathcal{H}_\gamma := \otimes\mathcal{H}_\gamma(d_i)$, and $\mathcal{I}_\gamma := \otimes\mathcal{I}_\gamma(d_i)$. The overall idea of the proof is to show that

$$\underset{\mathcal{H}_\gamma}{\mathbf{E}}[f_u(\boldsymbol{x})\prod_{i=1}^r g_v(\boldsymbol{y}_i)] \approx \underset{\mathcal{I}_\gamma}{\mathbf{E}}[f_u(\boldsymbol{x})\prod_{i=1}^r g_v(\boldsymbol{y}_i)],$$

and then we bound the right hand side. More precisely, the argument is divided into three steps.

First, we apply the Bonami-Beckner operator to the functions. The following theorem bounds the error introduced in this step.

Theorem B.1. *There are positive constants $\delta \geq \delta' > 0$ depending only on γ and d such that*

$$\left|\underset{\mathcal{H}_\gamma}{\mathbf{E}}[f(\boldsymbol{x})\prod_{i=1}^r g(\boldsymbol{y}_i)] - \underset{\mathcal{H}_\gamma}{\mathbf{E}}[T_{1-\delta'}f(\boldsymbol{x})\prod_{i=1}^r T_{1-\delta}g(\boldsymbol{y}_i)]\right| \leq (r+1)\sqrt{\gamma}.$$

Next, we move from distribution $\mathcal{H}_\gamma$ to $\mathcal{I}_\gamma$.

Theorem B.2. *There exists constants $\tau > 0$ depending only on d, γ and δ', such that the following holds. If for every $i \in [R_1]$ and every odd-cardinality set $S \subseteq \pi_e^{-1}(i)$, we have*

$$\min\{Inf_i(T_{1-\delta'}f_u), Inf_S(T_{1-\delta'}g_v)\} \leq \tau,$$

then

$$\left|\underset{\mathcal{H}_\gamma}{\mathbf{E}}[T_{1-\delta'}f(\boldsymbol{x})\prod_{i=1}^{R_1} T_{1-\delta}g(\boldsymbol{y}_i)] - \underset{\mathcal{I}_\gamma}{\mathbf{E}}[T_{1-\delta'}f(\boldsymbol{x})\prod_{i=1}^{R_1} T_{1-\delta}g(\boldsymbol{y}_i)]\right| \leq \sqrt{\gamma}.$$

And finally, we bound the expectation under $\mathcal{I}_\gamma$.

Theorem B.3.

$$\left|\underset{\mathcal{I}_\gamma}{\mathbf{E}}[T_{1-\delta'}f(\boldsymbol{x})\prod_{i=1}^r T_{1-\delta}g(\boldsymbol{y}_i)]\right| \leq \sqrt{\gamma}.$$

This follows directly from Lemma 3.9.

B.1 Proof of Theorem B.1

We first prove Theorem B.1. The idea is to apply T_ρ to the functions one by one. Intuitively, the Bonami-Beckner operator does not change the low-degree parts of functions too much. For the high-order parts, it follows from Proposition 2.14 that their norms are small and would not affect much.

Lemma B.4. *By taking $\delta > 0$ small enough as a function of δ, d and γ, we ensure that for any $k \in [r]$*

$$\left| \mathop{\mathbf{E}}_{\mathcal{H}_\gamma} [f(\boldsymbol{x}) \prod_{i=1}^{k} g(\boldsymbol{y}_i) \cdot \prod_{i=k+1}^{r} T_{1-\delta} g(\boldsymbol{y}_i)] - \mathop{\mathbf{E}}_{\mathcal{H}_\gamma} [f(\boldsymbol{x}) \prod_{i=1}^{k-1} g(\boldsymbol{y}_i) \cdot \prod_{i=k}^{r} T_{1-\delta} g(\boldsymbol{y}_i)] \right| \leq \sqrt{\gamma}.$$

Proof. Let U be the conditional expectation operator for the correlated probability space

$$((\{-1,1\}^{R_1} \times \prod_{i=1}^{r-1} \{-1,1\}^{R_2}) \times (\{-1,1\}^{R_2}), \mathcal{H}_\gamma),$$

mapping function h on the latter space to the former space by

$$(Uh)(\boldsymbol{x}, \{\boldsymbol{y}_i\}_{i \in [r] \setminus \{k\}}) = \mathop{\mathbf{E}}_{\mathcal{H}_\gamma} [h(\boldsymbol{y}_k) | (\boldsymbol{x}, \{\boldsymbol{y}_i\}_{i \in [r] \setminus \{k\}})].$$

We have

$$\begin{aligned} &\left| \mathop{\mathbf{E}}_{\mathcal{H}_\gamma} [f(\boldsymbol{x}) \prod_{i=1}^{k} g(\boldsymbol{y}_i) \cdot \prod_{i=k+1}^{r} T_{1-\delta} g(\boldsymbol{y}_i)] - \mathop{\mathbf{E}}_{\mathcal{H}_\gamma} [f(\boldsymbol{x}) \prod_{i=1}^{k-1} g(\boldsymbol{y}_i) \cdot \prod_{i=k}^{r} T_{1-\delta} g(\boldsymbol{y}_i)] \right| \\ = &\left| \mathop{\mathbf{E}}_{\mathcal{H}_\gamma} [f(\boldsymbol{x}) \prod_{i=1}^{k-1} g(\boldsymbol{y}_i) \cdot \prod_{i=k+1}^{r} T_{1-\delta} g(\boldsymbol{y}_i) \cdot (id - T_{1-\delta}) g(\boldsymbol{y}_k)] \right| \\ = &\left| \mathop{\mathbf{E}}_{(\boldsymbol{x}, \{\boldsymbol{y}_i\}_{i \in [r] \setminus \{k\}}) \sim \mathcal{H}_\gamma} \Big[f(\boldsymbol{x}) \prod_{i=1}^{k-1} g(\boldsymbol{y}_i) \prod_{i=k+1}^{r} T_{1-\delta} g(\boldsymbol{y}_i) \right. \\ &\left. \qquad \cdot (U(id - T_{1-\delta}) g)(\boldsymbol{x}, \{\boldsymbol{y}_i\}_{i \in [r] \setminus \{k\}}) \Big] \right|. \end{aligned} \tag{5}$$

Consider the function inside the expectation to be a product of two functions on $\mathcal{X} \times \prod_{i=1}^{k-1} \mathcal{Y}_i \times \prod_{i=k+1}^{r} \mathcal{Y}_i$, $F = f \prod_{i=1}^{k-1} g \prod_{i=k+1}^{r} T_{1-\delta} g$ and $G = U(id - T_{1-\delta}) g$. Take the Efron-Stein decomposition of these two functions w.r.t. $\mathcal{H}_\gamma$ on $\mathcal{X} \times \prod_{i=1}^{k-1} \mathcal{Y}_i \times \prod_{i=k+1}^{r} \mathcal{Y}_i$. By orthogonality of the Efron-Stein decomposition and Cauchy-Schwarz,

$$(5) = \left| \sum_{S \subseteq [R_1]} \mathop{\mathbf{E}}_{\substack{(\boldsymbol{x}, \{\boldsymbol{y}_i\}_{i \in [r] \setminus \{k\}}) \\ \sim \mathcal{H}_\gamma}} [F_S(\boldsymbol{x}, \{\boldsymbol{y}_i\}_{i \in [r] \setminus \{k\}}) \cdot G_S(\boldsymbol{x}, \{\boldsymbol{y}_i\}_{i \in [r] \setminus \{k\}})] \right| \tag{6}$$

$$\leq \sqrt{\sum_{S \subseteq [R_1]} \|F_S\|_2^2} \sqrt{\sum_{S \subseteq [R_1]} \|G_S\|_2^2} \leq \sqrt{\sum_{S \subseteq [R_1]} \|G_S\|_2^2}, \tag{7}$$

where the $\|\cdot\|_2$ are with respect to $\mathcal{H}_\gamma$'s marginal on $\mathcal{X} \times \prod_{i \in [r] \setminus k} \mathcal{Y}_i$. The conditional expectation operator U commutes with taking Efron-Stein decomposition, $G_S = UG'_S$, where $G' = (id -$

$T_{1-\delta})g$. Here the Efron-Stein decomposition is with respect to $\mathcal{H}_\gamma$'s marginal distribution on $\mathcal{Y}_k$, namely, the uniform distribution, and it satisfies

$$g_S = \sum_{U\subseteq[R_2]:\pi(U)=S} \hat{g}(U)\chi_U.$$

Applying noise operator also commutes with taking Efron-Stein decomposition, hence $G_S = UG'_S = U(id - T_{1-\delta})g_S$. Substituting this into (7) yields

$$(7) \quad = \quad \sqrt{\sum_{S\subseteq[R_1]} \|U(id-T_{1-\delta})g_S\|_2^2}. \tag{8}$$

Let $\rho_0 = 1-\beta^2/2$ be the bound in Lemma 3.9. Then

$$(8) \quad \leq \quad \sqrt{\sum_{S\subseteq[R_1]} \rho_0^{|S|}\|(id-T_{1-\delta})g_S\|_2^2}, \tag{9}$$

$$\|(id-T_{1-\delta})g_S\|_2^2 \quad = \quad \sum_{U\subseteq[R_2]:\pi(U)=S} \left(1-(1-\delta)^{2|U|}\right)\hat{g}(U)\chi(U) \tag{10}$$

$$\leq \quad \sum_{U\subseteq[R_2]:\pi(U)=S} \left(1-(1-\delta)^{2d|S|}\hat{g}(U)\right)\chi(U) \tag{11}$$

$$= \quad \left(1-(1-\delta)^{2d|S|}\right)\|g_S\|_2^2, \tag{12}$$

therefore

$$(8) \quad \leq \quad \sqrt{\sum_{S\subseteq[R_1]} \rho_0^{|S|}\left(1-(1-\delta)^{2d|S|}\right)\|g_S\|_2^2}. \tag{13}$$

We bound

$$\rho_0^{2|S|}\left(1-(1-\delta)^{2d|S|}\right) \quad \leq \quad \exp\left(-|S|\beta^2\right)\cdot(2d|S|\delta). \tag{14}$$

Choose $\delta > 0$ small enough so that (14) is upper-bounded by γ, and (5) $\leq \sqrt{\gamma}$. □

It remains to prove the following lemma.

Lemma B.5. *By taking $\delta' > 0$ small enough as a function of δ, d, γ and δ, we ensure*

$$\left|\mathop{\mathbf{E}}_{\mathcal{H}_\gamma}[f(\boldsymbol{x})\prod_{i=1}^r T_{1-\delta}g(\boldsymbol{y}_i)] - \mathop{\mathbf{E}}_{\mathcal{H}_\gamma}[T_{1-\delta'}f(\boldsymbol{x})\prod_{i=1}^r T_{1-\delta}g(\boldsymbol{y}_i)]\right| \leq \sqrt{\gamma}.$$

Proof. Define the noised version of $\mathcal{H}_\gamma$, $\mathcal{H}_\gamma^*$ by first generating $(\boldsymbol{x}, \{\boldsymbol{y}_i\}_{i\in[r]}) \sim \mathcal{H}_\gamma$, and then rerandomize each bit in $\boldsymbol{y}_i$ with probability δ. Define $\mathcal{H}_\gamma^*(D)$ on $\{-1,1\}^{R_1} \times \otimes_{i=1}^r\{-1,1\}^{R_2}$ similarly. Then $\mathcal{H}_\gamma^*$ can also be written as a product distribution $\mathcal{H}_\gamma^* = \otimes_{i=1}^{R_1}\mathcal{H}_\gamma^*(d_i)$.

The structure of the proof is similar to Lemma B.4. We need the following correlation lemma whose proof is similar to Lemma 3.9.

Lemma B.6. $\rho(\{-1,1\}, \prod_{i=1}^r\{-1,1\}^D; \mathcal{H}_\gamma^*(D)) \leq 1-\beta^2/2$ *where*

$$\beta = \frac{\gamma\cdot(2^{r-2}-1)\delta^{rD}}{2^{(2r-1)D}D(2^{r-1}-1)}.$$

is a lowerbound of the least probability of an atom in $\mathcal{H}_\gamma^*(D)$.

□

B.2 Proof of Theorem B.2

By definition, we have $\mathcal{H}_\gamma = \otimes\mathcal{H}_\gamma(d_i)$ and $\mathcal{I}_\gamma = \otimes\mathcal{I}_\gamma(d_i)$. The overall plan is to change the distributions in the product one by one from $\mathcal{H}_\gamma(d_i)$ to $\mathcal{I}_\gamma(d_i)$. To this end, we prove the following theorem.

Theorem B.7. *For each $k \in [R_1]$*

$$\left| \mathop{\mathbf{E}}_{\bigotimes_{i=1}^{k-1}\mathcal{I}_\gamma(d_i)\otimes\bigotimes_{i=k}^{R_1}\mathcal{H}_\gamma(d_i)} [T_{1-\delta'}f(\boldsymbol{x})\prod_{i=1}^{r} T_{1-\delta}g(\boldsymbol{y}_i)] - \mathop{\mathbf{E}}_{\bigotimes_{i=1}^{k}\mathcal{I}_\gamma(d_i)\otimes\bigotimes_{i=k+1}^{R_1}\mathcal{H}_\gamma(d_i)} [T_{1-\delta'}f(\boldsymbol{x})\prod_{i=1}^{r} T_{1-\delta}g(\boldsymbol{y}_i)] \right| \le \Delta_k, \tag{15}$$

where

$$\Delta_k := \tau^{\delta'/2(r+1)}\left(2^d\,\mathit{Inf}_k(T_{1-\delta'/2}f) + \sum_{S\subseteq\pi^{-1}(k),|S| \text{ is odd}} \mathit{Inf}_S(T_{1-\delta'/2}g)\right).$$

Proof of Theorem B.2. If we sum over all $k \in [R_1]$, by triangle inequality, we have

$$\begin{aligned}
&\left|\mathop{\mathbf{E}}_{\mathcal{H}_\gamma}[T_{1-\delta'}f(\boldsymbol{x})\prod_{i=1}^{R_1}T_{1-\delta}g(\boldsymbol{y}_i)] - \mathop{\mathbf{E}}_{\mathcal{I}_\gamma}[T_{1-\delta'}f(\boldsymbol{x})\prod_{i=1}^{R_1}T_{1-\delta}g(\boldsymbol{y}_i)]\right| \\
\le\ & \tau^{\delta'/2(r+1)}\left(2^d\sum_{k=1}^{R_1}\mathrm{Inf}_k(T_{1-\delta'/2}f) + \sum_{\substack{S\subseteq\pi^{-1}(k)\\ \text{for some } k\in[R_1]}}\mathrm{Inf}_S(T_{1-\delta'/2}g)\right) \\
\le\ & \tau^{\delta'/2(r+1)}(2^d(1/\delta') + (d/\delta')^d) \\
\le\ & \tau^{\delta'/2(r+1)}2(d/\delta')^d.
\end{aligned}$$

We now choose τ small enough so that the RHS is bounded by $\sqrt{\gamma}$, and this completes the proof. □

Proof of Theorem B.7. We show the theorem for case $k = 1$. We write x' for strings $(x_2, \cdots, x_{R_1})$, y_i' for strings $(y_{i,d_1+1}, \cdots, y_{i,R_2})$. We break up the Fourier expansion of f according to its dependence on x_1:

$$f(x) = F_\emptyset(x') + x_1F_1(x').$$

Similarly, we break up the Fourier expansion of g according to its dependence on the bits $y_1, \cdots, y_{d_1}$:

$$g = \sum_{S\subseteq[d_1]} \chi_S(y_1,\cdots,y_{d_1})G_S(y'),$$

where for any $S \subseteq [d_1]$, we have

$$G_S(y') = \sum_{Q\subseteq[R_2],Q\cap[d_1]=S} \hat{g}(Q)\chi_{Q\setminus S}(y').$$

Since $\hat{g}(Q) = \mathbf{E}_{\boldsymbol{y}}[g(\boldsymbol{y})\chi_Q(\boldsymbol{y})]$, we have

$$G_S(y') = \mathop{\mathbf{E}}_{\boldsymbol{y_1},\boldsymbol{y_2},\cdots,\boldsymbol{y_{d_1}},y'}[g(\boldsymbol{y_1},\boldsymbol{y_2},\cdots,\boldsymbol{y_{d_1}},y')\chi_S(\boldsymbol{y_1},\boldsymbol{y_2},\cdots,\boldsymbol{y_{d_1}})],$$

and therefore G_S is bounded in $[-1,1]$. Similarly, so are $F_\emptyset$ and F_1. We also have the Fourier expansions

$$T_{1-\delta'} f(x) = T_{1-\delta'} F_\emptyset(x') + (1-\delta') x_1 T_{1-\delta'} F_1(x'), \tag{16}$$

$$T_{1-\delta} g(y) = \sum_{S\subseteq[d_1]} (1-\delta)^{|S|} \chi_S(y_1, \cdots, y_{d_1}) T_{1-\delta} G_S(y'). \tag{17}$$

Lemma B.8. *For any functions $F : \mathcal{X} \to \mathbb{R}$, $G_i : \mathcal{Y}_i \to \mathbb{R}$,*

$$\mathop{\mathbf{E}}_{\mathcal{H}_\gamma(D)}[F(\boldsymbol{x})\prod_{i=1}^r G_i(\boldsymbol{y})] - \mathop{\mathbf{E}}_{\mathcal{I}_\gamma(D)}[F(\boldsymbol{x})\prod_{i=1}^r G_i(\boldsymbol{y})] = \sum_{S\subseteq[D], |S| \text{ is odd}} (1-\gamma)\hat{F}(\{1\})\prod_{i=1}^r \hat{G}_i(S). \tag{18}$$

Proof.

$$\text{LHS} = \sum_{\substack{U\subseteq[1]\\ V_i\subseteq[D]}} \hat{F}(U)\prod_{i=1}^r \hat{G}(V_i)\left(\mathop{\mathbf{E}}_{\mathcal{H}_\gamma(D)}[\chi_U(\boldsymbol{x})\prod_{i=1}^r \chi_{V_i}(\boldsymbol{y}_i)] - \mathop{\mathbf{E}}_{\mathcal{I}_\gamma(D)}[\chi_U(\boldsymbol{x})\prod_{i=1}^r \chi_{V_i}(\boldsymbol{y}_i)]\right). \tag{19}$$

Since $\mathcal{H}_\gamma(D) = (1-\gamma)\mathcal{H}(D) + \gamma\mathcal{N}(D)$, $\mathcal{I}_\gamma(D) = (1-\gamma)\mathcal{I}(D) + \gamma\mathcal{N}(D)$, by linearity of expectation, we have

$$\mathop{\mathbf{E}}_{\mathcal{H}_\gamma(D)}[\chi_U(\boldsymbol{x})\prod_{i=1}^r \chi_{V_i}(\boldsymbol{y}_i)] - \mathop{\mathbf{E}}_{\mathcal{I}_\gamma(D)}[\chi_U(\boldsymbol{x})\prod_{i=1}^r \chi_{V_i}(\boldsymbol{y}_i)] \tag{20}$$

$$= (1-\gamma)\left(\mathop{\mathbf{E}}_{\mathcal{H}(D)}[\chi_U(\boldsymbol{x})\prod_{i=1}^r \chi_{V_i}(\boldsymbol{y}_i)] - \mathop{\mathbf{E}}_{\mathcal{I}(D)}[\chi_U(\boldsymbol{x})\prod_{i=1}^r \chi_{V_i}(\boldsymbol{y}_i)]\right). \tag{21}$$

Note that $\mathcal{H}(D)$ and $\mathcal{I}(D)$ have the same marginal distribution on $\prod_{i=1}^r \mathcal{Y}_i$. Therefore, for (21) to be nonzero, U must be nonempty, or $U = \{1\}$. Now we have that

$$\mathop{\mathbf{E}}_{\mathcal{I}(D)}[\chi_U(\boldsymbol{x})\prod_{i=1}^r \chi_{V_i}(\boldsymbol{y}_i)] = \mathop{\mathbf{E}}_{\mathcal{I}(D)}[x]\mathop{\mathbf{E}}_{\mathcal{I}(D)}[\prod_{i=1}^r \chi_{V_i}(\boldsymbol{y}_i)] = 0.$$

It is easy to see that $\mathbf{E}_{\mathcal{H}(D)}[\chi_U(\boldsymbol{x})\prod_{i=1}^r \chi_{V_i}(\boldsymbol{y}_i)]$ is zero unless for all $i \in [r]$, $V_i = V$ for some V. Moreover, $|V|$ must be odd, in which case the expectation is 1. Thus (21) is equal to the RHS. □

We now rewrite the LHS of (15) as

$$\Big|\mathop{\mathbf{E}}_{\mathcal{H}'_\gamma}[\mathop{\mathbf{E}}_{\mathcal{H}_\gamma(d_1)}[T_{1-\delta'} f(\boldsymbol{x})\prod_{i=1}^r T_{1-\delta} g(\boldsymbol{y}_i)] - \mathop{\mathbf{E}}_{\mathcal{I}_\gamma(d_1)}[T_{1-\delta'} f(\boldsymbol{x})\prod_{i=1}^r T_{1-\delta} g(\boldsymbol{y}_i)]]\Big|. \tag{22}$$

By Lemma B.8, the above is equal to

$$|\sum_{S\subseteq[d_1], |S| \text{ is odd}} (1-\gamma)(1-\delta')(1-\delta)^{r|S|}\mathop{\mathbf{E}}_{\mathcal{H}'_\gamma}[T_{1-\delta'} F_1(\boldsymbol{x'})\prod_{i=1}^r T_{1-\delta} G_S(\boldsymbol{y'}_i)]| \tag{23}$$

$$\leq \sum_{S\subseteq[d_1], |S| \text{ is odd}} (1-\delta')(1-\delta)^{r|S|}\mathop{\mathbf{E}}_{\mathcal{H}'_\gamma}[|T_{1-\delta'} F_1(\boldsymbol{x'})\prod_{i=1}^r T_{1-\delta} G_S(\boldsymbol{y'}_i)|] \tag{24}$$

$$\leq \sum_{S\subseteq[d_1], |S| \text{ is odd}} (1-\delta')(1-\delta)^{r|S|}\|T_{1-\delta'} F_1\|_{r+1}\|T_{1-\delta} G_S\|_{r+1}^r, \tag{25}$$

where the last step uses Hölder's Inequality, and the norms $\|\cdot\|_{r+1}$ are with respect to the corresponding marginals of $\mathcal{H}'_\gamma$, which are uniform.

Lemma B.9. *For any function* $f : \{-1,1\}^n \to [-1,1]$ *and* $0 < \eta < 1$,

$$\|T_{1-\eta}f\|_{r+1} \le \|T_{1-\eta/2}f\|_2^{(2+\eta)/(r+1)}.$$

Proof. We first prove that for $f' : \{-1,1\}^n \to [-1,1]$,

$$\|T_{1-\eta'}f'\|_{r+1} \le \|f'\|_2^{(2+2\eta')/(r+1)}.$$

Observe that

$$\begin{aligned} &\|T_{1-\eta'}f'\|_{r+1} = \mathbf{E}[|T_{1-\eta'}f'|^{r+1}]^{1/(r+1)} \\ \le\ & \mathbf{E}[|T_{1-\eta'}f|^{2+2\eta'}]^{1/(r+1)} = \|T_{1-\eta'}f\|_{2+2\eta'}^{(2+2\eta')/(r+1)}. \end{aligned}$$

Since $2+2\eta' \le (1-\eta')^2+1$, by Hypercontractive Inequality of Bonami [3] and Gross [6], the above is upper-bounded by $\|f\|_2^{(2+2\eta')/(r+1)}$. To prove the lemma, observe that

$$\|T_{1-\eta}f\|_{r+1} \le \|T_{1-\eta/2}T_{1-\eta/2}f\|_{r+1} \le \|T_{1-\eta/2}f\|_2^{(2+\eta)/(r+1)}.$$

□

As F_1 and G_S are bounded in $[-1,1]$, we can upper-bound

$$\|T_{1-\delta'}F_1\|_{r+1}\|T_{1-\delta}G_S\|_{r+1}^r \le \|T_{1-\delta'/2}F_1\|_2^{(2+\delta')/(r+1)}\|T_{1-\delta/2}G_S\|_2^{r(2+\delta)/(r+1)}.$$

We express G_S's and F_1's Fourier coefficients with g's and f's original Fourier coefficients

$$\begin{aligned} \|T_{1-\delta/2}G_S\|_2^2 =& \sum_{Q\subseteq[R_2],Q\cap[d_1]=S} (1-\delta/2)^{2|Q|-2|S|}\hat{g}(Q)^2 && (26)\\ \le& \sum_{S\subseteq Q\subseteq[R_2]} (1-\delta/2)^{2|Q|-2|S|}\hat{g}(Q)^2 && (27)\\ \le& (1-\delta/2)^{-2|S|}\cdot \mathrm{Inf}_S(T_{1-\delta'/2}g), && (28)\\ & && (29) \end{aligned}$$

where we used $\delta \ge \delta'$ in the last step.

$$\|T_{1-\delta'/2}F_1\|_2^2 \le (1-\delta'/2)^2 \cdot \mathrm{Inf}_1(T_{1-\delta'/2}f).$$

Plugging these two bounds back, we upper-bound the LHS as following

$$\sum_{S\subseteq[d_1],S \text{ is odd}} \mathrm{Inf}_1(T_{1-\delta'/2}f)^{(2+\delta)/2(r+1)}\cdot \mathrm{Inf}_S(T_{1-\delta'/2}g)^{r(2+\delta)/2(r+1)}, \tag{30}$$

where we also used $\delta \ge \delta'$. By the hypothesis that $\min\{\mathrm{Inf}_1(T_{1-\delta'/2}f), \mathrm{Inf}_S(T_{1-\delta'/2}g)\} \le \tau$, either $\mathrm{Inf}_1(T_{1-\delta'/2}f)^{\delta/2(r+1)} \le \tau^{\delta/2(r+1)}$, or $\mathrm{Inf}_S(T_{1-\delta'/2}g)^{r\delta/2(r+1)} \le \tau^{r\delta/2(r+1)}$ for each S in the sum.

In either case, we can bound the above by

$$\tau^{\delta/2(r+1)} \cdot \sum_{S\subseteq[d_1], S \text{ is odd}} \mathrm{Inf}_1(T_{1-\delta'/2}f)^{1/(r+1)} \cdot \mathrm{Inf}_S(T_{1-\delta'/2}g)^{r/(r+1)} \tag{31}$$

$$\leq \tau^{\delta'/2(r+1)} \cdot \sum_{S\subseteq[d_1], S \text{ is odd}} (\mathrm{Inf}_1(T_{1-\delta'/2}f) + \mathrm{Inf}_S(T_{1-\delta'/2}g)) \tag{32}$$

$$\leq \tau^{\delta'/2(r+1)} \cdot (2^d \mathrm{Inf}_1(T_{1-\delta'/2}f) + \sum_{S\subseteq[d_1], S \text{ is odd}} \mathrm{Inf}_S(T_{1-\delta'/2}g)). \tag{33}$$

□

ECCC ISSN 1433-8092

http://eccc.hpi-web.de