Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 151 (2012)

The Complexity of Somewhat Approximation Resistant Predicates

Subhash Khot* Madhur Tulsiani† Pratik Worah‡

December 5, 2012

Abstract

A boolean predicate $f : \{0,1\}^k \to \{0,1\}$ is said to be *somewhat approximation resistant* if for some constant $\tau > \frac{|f^{-1}(1)|}{2^k}$, given a τ-satisfiable instance of the MAX k-CSP(f) problem, it is NP-hard to find an assignment that *strictly beats* the naive algorithm that outputs a uniformly random assignment. Let $\tau(f)$ denote the supremum over all τ for which this holds. It is known that a predicate is somewhat approximation resistant precisely when its Fourier degree is at least 3. For such predicates, we give a characterization of the *hardness gap* $(\tau(f) - \frac{|f^{-1}(1)|}{2^k})$ up to a factor of $O(k^5)$. We also give a similar characterization of the *integrality gap* for the natural SDP relaxation of MAX k-CSP(f) after $\Omega(n)$ rounds of the Lasserre hierarchy.

1 Introduction

Given a predicate $f : \{0,1\}^k \to \{0,1\}$, an instance of MAX k-CSP(f) problem consists of n boolean variables and m constraints where each constraint is the predicate f applied on some (ordered) subset of k variables and the variables are allowed to appear in negated form. The goal is to find an assignment to the variables that satisfies maximum number of constraints.

Definition 1.1 *Given a predicate $f : \{0,1\}^k \to \{0,1\}$, define the density $\rho(f) := \frac{|f^{-1}(1)|}{2^k}$.*

Definition 1.2 *For a predicate $f : \{0,1\}^k \to \{0,1\}$ and a constant $\tau > \rho(f)$, the predicate is said to be τ-*resistant *if for an arbitrarily small constant $\varepsilon > 0$, it is NP-hard to distinguish instances of MAX k-CSP(f) where a $\tau - \varepsilon$ fraction of constraints can be simultaneously satisfied from those where at most $\rho(f) + \varepsilon$ fraction of the constraints can be simultaneously satisfied.*

A τ-resistant predicate with $\tau = 1$ is more popularly known as *approximation resistant*. There is a substantial body of work on trying to characterize approximation resistant predicates, e.g. [21, 19, 13, 4, 3]. A recent survey by Håstad [22] gives a comprehensive overview of many results in this area. In particular, a celebrated result of his [21] shows that the predicates $x \vee y \vee z$ (i.e. 3SAT) and $x \oplus y \oplus z = 0$ (i.e. 3LIN) are approximation resistant. Such predicates have also been studied in the context of unconditional lower bounds and especially in the context of Linear and Semidefinite hierarchies [20, 12, 24]. In the context of such unconditional lower bounds approximation resistance

*NYU and University of Chicago. khot@cims.nyu.edu Research supported by NSF Expeditions grant CCF-0832795, NSF Waterman Award and BSF grant 2008059.

†Toyota Technological Institute at Chicago. madhurt@ttic.edu.

‡University of Chicago. pworah@cs.uchicago.edu.

ISSN 1433-8092

is synonymous with similar lower bounds in the corresponding proof systems which establishes strong connections with a large body of work in propositional proof complexity [6, 5, 2, 1]. A complete characterization of approximation resistant predicates remains elusive (see [3] however for some progress on this question). In this paper we study a related notion of *somewhat approximation resistance* defined by Håstad [22].

Definition 1.3 *A predicate $f : \{0,1\}^k \to \{0,1\}$ is said to be* somewhat approximation resistant *if there exists some constant $\tau > \rho(f)$ such that the predicate is τ-resistant.*

Definition 1.4 *A predicate $f : \{0,1\}^k \to \{0,1\}$ is said to be* always approximable *if for every constant $\tau > \rho(f)$, there is a constant $\nu > \rho(f)$ and a polynomial time (possibly randomized) algorithm that given an instance of MAX k-CSP(f) where a τ fraction of constraints can be simultaneously satisfied, finds an assignment satisfying ν fraction of the constraints.*

Clearly, the terms *somewhat approximation resistant* and *always approximable* are mutually exclusive (assuming P $\neq$ NP). We recall that any predicate $f : \{0,1\}^k \to \{0,1\}$ has a Fourier representation:

$$f(x) = \sum_{\alpha \in \{0,1\}^k} \hat{f}(\alpha)(-1)^{\alpha \cdot x}.$$

The Fourier degree of f is the maximum Hamming weight $|\alpha|$ such that $\hat{f}(\alpha) \neq 0$. We say that f depends on a variable x_i if that variable *appears* in the above representation (i.e. if there exists $\alpha \in \{0,1\}^k$ such that $\alpha_i = 1, \hat{f}(\alpha) \neq 0$). We note that $\rho(f) = \hat{f}(0) = \sum_\alpha \hat{f}(\alpha)^2$. Håstad [22] shows the following result:[1]

Theorem 1.5 *A predicate f is always approximable if its Fourier degree is at most 2 and somewhat approximation resistant otherwise. Moreover, a function of Fourier degree at most 2 can depend on at most 4 variables.*

In this paper, we focus our attention to the case when f has Fourier degree at least 3 and hence is somewhat approximation resistant.

Definition 1.6 *Let $f : \{0,1\}^k \to \{0,1\}$ be a predicate with Fourier degree at least 3. Define $\tau(f)$ to be the supremum over all τ such that f is τ-resistant.*

The parameter $\tau(f) - \rho(f)$ may be considered as the *hardness gap*. Our goal is to characterize this gap as closely as possible. As we demonstrate, the gap can be as small as $2^{-\Omega(k)}$ for some predicates. Our main result is a characterization of this gap up to a multiplicative factor of $O(k^5)$. Håstad's result [22] gives a lower bound of $\left(\max_{|\alpha| \geq 3} |\hat{f}(\alpha)|\right)$ on the gap $\tau(f) - \rho(f)$.[2] However we show that this bound is too weak for some predicates and a stronger lower bound is $\Omega\left(\frac{1}{k^2} \cdot \sum_{|\alpha| \geq 3} \hat{f}(\alpha)^2\right)$ (clearly there are predicates where the maximum Fourier coefficient at/above level 3 is exponentially far off from the total Fourier mass at/above level 3). However even this bound is not the *correct* one for some predicates and the situation turns out to be a bit subtle. We show that the gap is characterized by two factors:

[1] Håstad provides only a sketch of proofs. The proof details for the first statement in the theorem can be filled in easily. Regarding the second statement, namely that a function of Fourier degree at most 2 can depend on at most 4 variables, the proof follows from [17] (also see Section 3.1 in [18]).

[2] This is the bound that can be inferred from Håstad's proof sketch.

1. Whether f is *close to* or *far from* the class of functions with Fourier degree at most 2. Not surprisingly, this is related to whether the Fourier mass of f at level 3 and above is *low* or *high*.

2. When f is *close to* some function g with Fourier degree at most 2, whether f is *monotonically below* g.

Note that the upper and lower bound on the gap $\tau(f) - \rho(f)$ correspond to an algorithm and a NP-hardness result respectively. We show that our upper and lower bounds also hold in the Lasserre SDP hierarchy in the following sense: The algorithmic upper bound is achieved by a simple SDP relaxation with one round of the natural Lasserre relaxation. On the other hand, for all lower bound results, there is a $\Omega(n)$-level Lasserre integrality gap construction with integrality gap similar to the NP-hardness gap.

1.1 The Main Result

Let $\mathcal{Q}$ denote the set of boolean functions on k variables which are of Fourier degree at most 2. From Theorem 1.5, if $f \in \mathcal{Q}$ then f is always approximable and otherwise it is somewhat approximation resistant. We are interested in the case that $f \notin \mathcal{Q}$. Let $\Delta(f, \mathcal{Q})$ denote the minimum Hamming distance (normalized by a factor 2^k so that it is in the range $[0, 1]$) of f from any function in $\mathcal{Q}$. We now state our main result.

Theorem 1.7 *Let $k \geq 2^{2^{15}}$ and $f : \{0,1\}^k \to \{0,1\}$ be a predicate with Fourier degree at least 3 (and hence $\Delta(f, \mathcal{Q}) > 0$).* [3]

1. *If $\Delta(f, \mathcal{Q}) \geq 1/k^3$, then $\tau(f) \geq \rho(f) + \Omega(1/k^5)$.*
2. *If $\Delta(f, \mathcal{Q}) = \delta \leq 1/k^3$, then let $g \in \mathcal{Q}$ denote the unique function such that $\Delta(f, g) = \delta$.*
   - *(a) If $\exists x \in \{0,1\}^k$ such that $f(x) = 1 \wedge g(x) = 0$ then $\tau(f) \geq \rho(f) + \Omega(1/k)$.*
   - *(b) Otherwise, g is* monotonically above f. *In this case, there is an absolute constant C and a polynomial time algorithm that for any $\varepsilon \geq Ck^3\delta$, given a $\rho(f) + \varepsilon$ satisfiable instance of MAX k-CSP(f), finds an assignment that is $\left(\rho(f) + \Omega(\frac{\varepsilon}{k^2 \log(1/\varepsilon)})\right)$-satisfying. In particular,*
   $$\tau(f) \leq \rho(f) + O(k^3\delta).$$
   Moreover, $\tau(f) \geq \rho(f) + \Omega\left(\frac{\delta}{k^2}\right)$.

Remark 1.8 *We always have the trivial upper bound $\tau(f) - \rho(f) \leq 1$. Hence in all the cases, $\tau(f) - \rho(f)$ is characterized up to a multiplicative factor of $O(k^5)$ as claimed. In Case (2b), δ could be as small as 2^{-k}, and so $\tau(f) - \rho(f)$ is of the same order. Note that characterizing $\tau(f)$ precisely would in particular completely characterize approximation resistant predicates (with $\tau(f) = 1$) which is open even for the case $k = 4$ [13, 22]. We believe that even characterizing the gap $\tau(f) - \rho(f)$ with a* polylog(k) *factor would require significantly new ideas.*

[3] The current lower bound on k for Theorem 1.7 is large, but it seems to be an artifact of our proof technique and we expect it to hold for smaller values of k. The condition $k \geq 2^{2^{15}}$ arises only in our argument relating $\Delta(f, \mathcal{Q})$ to the Fourier mass of f above level 2 (Section 4).

Remark 1.9 *Whenever Case (1) applies, we have $\rho(f) \geq \frac{1}{k^3}$ (otherwise f would be $\frac{1}{k^3}$-close to the zero-function which is in $\mathcal{Q}$). The functions $g \in \mathcal{Q}$ depend on at most 4 variables and thus $\rho(g) \in \{\frac{\ell}{16} | \ell \in \{0, 1, \ldots, 16\}\}$. Whenever Case (2) applies $\rho(f)$ is $\frac{1}{k^3}$-close to one of these 17 values.*

We can also prove unconditional lower bounds without much extra effort. In this context the notion of NP-hardness is replaced by the notion of the integrality gap which persists even after many levels of Lasserre relaxations.

Definition 1.10 *Given MAX k-CSP(f), we say that f is* τ^*-resistant *for the* Lasserre hierarchy *if for all constant $\varepsilon > 0$, there exists a constant $c = c(\varepsilon) > 0$ and instances with n variables and m constraints, for infinitely many values of n, such that the Lasserre relaxation after $\lfloor cn \rfloor$ rounds has value at least τ^* but the integral optimum is at most $\rho(f) + \varepsilon$.*

Definition 1.11 *Let $f : \{0,1\}^k \to \{0,1\}$ be a predicate with Fourier degree at least 3. Define $\tau^*(f)$ to be the supremum over all τ^* such that f is τ^*-resistant.*

The two notions of τ-resistance (namely Definition 1.11 and 1.6) are very closely related and so we have chosen to use a similar notation for both. Notice that we use a more precise notion of integrality gap which specifies the optimal fractional and integral solution (i.e. the *gap location*) and not just their ratio. Our main result regarding integrality gap mimics the result regarding NP-hardness gap:

Theorem 1.12 *Let $k \geq 2^{2^{15}}$ and $f : \{0,1\}^k \to \{0,1\}$ be a predicate with Fourier degree at least 3 (and hence $\Delta(f, \mathcal{Q}) > 0$).*

1. *If $\Delta(f, \mathcal{Q}) \geq 1/k^3$, then $\tau^*(f) \geq \rho(f) + \Omega(1/k^5)$.*
2. *If $\Delta(f, \mathcal{Q}) = \delta \leq 1/k^3$, then let $g \in \mathcal{Q}$ denote the unique function such that $\Delta(f, g) = \delta$.*
   (a) *If $\exists x \in \{0,1\}^k$ such that $f(x) = 1 \wedge g(x) = 0$ then $\tau^*(f) \geq \rho(f) + \Omega(1/k)$.*
   (b) *Otherwise, g is* monotonically above f. *In this case, SDP rounding of the natural Lasserre relaxation, after just one round, finds an assignment that is $\left(\rho(f) + \Omega(\frac{\varepsilon}{k^2 \log(1/\varepsilon)})\right)$-satisfying if the instance is $\rho(f) + \varepsilon$ satisfiable for any $\varepsilon \geq Ck^3\delta$ and C is an absolute constant. In particular,*
   $$\tau^*(f) \leq \rho(f) + O(k^3\delta).$$
   Moreover, $\tau^*(f) \geq \rho(f) + \Omega\left(\frac{\delta}{k^2}\right)$.

1.2 Overview of the Proof

In this section, we provide a brief sketch of proof of Theorem 1.7, hiding many details however. Our starting point is a recent result of Chan [7] showing that a predicate $L : \{0,1\}^k \mapsto \{0,1\}$ is 1-resistant (i.e. approximation resistant) if $L^{-1}(1)$ is an affine translate of the orthogonal complement of a distance (at least) 3 code. We will call such a predicate a *good* predicate for this section. A useful fact is that sparse good predicates exist (i.e. $|L^{-1}(1)|$ is $O(k^2)$) and are numerous in the sense that an affine translate of the orthogonal complement of a random linear subspace of dimension $k - 2\log_2 k - O(1)$ works with probability 99%.

A predicate $f : \{0,1\}^k \to \{0,1\}$ is said to be τ-correlated with a good predicate L if a uniformly random satisfying assignment for L is also a satisfying assignment for f with probability at least τ (i.e. $|L^{-1}(1) \cap f^{-1}(1)|/|L^{-1}(1)| \geq \tau$). Given a predicate $f : \{0,1\}^k \mapsto \{0,1\}$, we observe that if f is τ-correlated with a good predicate, then f is τ-resistant. The reason is rather straightforward. Chan [7] gives a reduction showing that L is 1-resistant. We take the same reduction but pretend that the predicate used for every constraint is f instead of L and this minor modification suffices to show that f is τ-resistant.[4] For the sake of future reference, we note that any predicate f, not identically zero, always $\Omega\left(\frac{1}{k^2}\right)$-correlates with some good predicate. This is simply because we pick an arbitrary good predicate L with $|L^{-1}(1)| \leq O(k^2)$ and translating it if necessary ensure that $L^{-1}(1) \cap f^{-1}(1) \neq \emptyset$. This gives correlation of at least $1/|L^{-1}(1)|$.

With these observations at hand, our first task is to (approximately) characterize the best possible correlation that a given predicate $f : \{0,1\}^k \mapsto \{0,1\}$ can have with a good predicate. We show that this is related to the Fourier mass of f at level 3 and above, denoted $\gamma_3(f)$, which in turn is related to the distance $\Delta(f, \mathcal{Q})$. In the range of parameters of interest, we show that f is τ-correlated with a good predicate (and hence τ-resistant) with

$$\tau \geq \rho(f) + \Omega\left(\frac{\gamma_3(f)}{k^2}\right),$$

and moreover that

$$\gamma_3(f) = \Theta(\Delta(f, \mathcal{Q})).$$

The first claim uses a (somewhat novel) probabilistic argument showing that a random good predicate works. The second claim uses Fourier analytic techniques from the works of KKL and Friedgut [14, 10][5]. Our lower bound on $\tau(f)$ in Case (1) and Case (2b) of Theorem 1.7 now follow immediately.

We are left with the Case (2a) and the upper bound in Case (2b) of Theorem 1.7. Note that we are in the scenario where there is a function $g \in \mathcal{Q}$ with $\Delta(f, g) = \delta$ for some *tiny* δ.

We illustrate Case (2a) first. For the sake of illustration, assume that $g \equiv 0$, which amounts to saying that $\rho(f) = \delta$. As we noted, f always $\Omega\left(\frac{1}{k^2}\right)$-correlates with a good predicate and hence is $\Omega\left(\frac{1}{k^2}\right)$-resistant. Since $\rho(f) = \delta$ is *tiny*, we have $\tau(f) \geq \rho(f) + \Omega(1/k^2)$ as desired (this is a bit weaker than the bound we actually get/state in Theorem 1.7). The proof for Case (2a) in general is a bit tricky and we refer the reader to Section 5.1. We note that this is the case where the gap $\tau(f) - \rho(f)$ is *large* (i.e. $\Omega(1/k^2)$) even though the Fourier mass at level 3 and above is at most δ which could be as low as 2^{-k}.

Finally, we arrive at the upper bound in Case (2b). Here the algorithm is designed by using an algorithm of Charikar and Wirth [8] as a black-box. Note that we are in the scenario where there is a function $g \in \mathcal{Q}$ with $\Delta(f, g) = \delta$ for a tiny δ and moreover that f *implies* g. Given an instance of $\mathsf{MAX\ k\text{-}CSP}(f)$ that is $(\rho(f) + \varepsilon)$-satisfiable, we begin by pretending that it is an instance of $\mathsf{MAX\ k\text{-}CSP}(g)$ with the predicate f on every constraint replaced by g. Since f implies g and they are close in Hamming distance, the instance remains $(\rho(g) + \varepsilon/2)$-satisfiable as an instance of $\mathsf{MAX\ k\text{-}CSP}(g)$. For the predicate g of Fourier degree at most 2, the algorithm of Charikar and Wirth yields an assignment that is $(\rho(g) + \Omega(\varepsilon/\log(1/\varepsilon)))$-satisfying. This assignment, by itself, might be

[4] One however needs to use some strong uniformity properties of Chan's reduction. Firstly, that in the YES Case, all satisfying assignments to L occur almost equally often. And secondly, that in the NO Case, for any global assignment, the local view at a randomly chosen constraint is almost uniformly random.

[5] This result is similar, though quantitatively incomparable, to a result of Friedgut, Kalai and Naor [11] which relates the Fourier mass above level 1 to the distance from dictator (and constant) functions.

quite bad when viewed as an assignment for MAX k-CSP(f). To correct this, we re-randomize each variable with probability $1 - \frac{1}{2k}$ and show that it now serves as a $(\rho(g) + \Omega(1/k^2 \cdot \varepsilon/\log(1/\varepsilon)))$-satisfying assignment to MAX k-CSP(f) (and $\rho(g) \geq \rho(f)$).

This completes our overview. In the context of Lasserre integrality gaps and Theorem 1.12, our starting point is a result of the second author [23] that is analogous to Chan's NP-hardness result. The *hardness reductions* are now replaced by *integrality gap constructions*, but they are identical in spirit with the same bounds.

It is apparent that most of our techniques are either elementary or borrowed from prior works. We view our main contribution as the appropriate combination of these techniques so that they fit together nicely and yield an (arguably) clean characterization.

2 Preliminaries

2.1 Fourier Analysis on the Boolean Hypercube

Given a function $f : \{0,1\}^k \to \{0,1\}$ we can express it in the Fourier basis, comprising of characters χ_α for $\alpha \in \{0,1\}^k$ and $\chi_\alpha(x) = (-1)^{\alpha \cdot x}$, as follows:

$$f(x) = \sum_{\alpha \in \{0,1\}^k} \hat{f}(\alpha)\chi_\alpha(x).$$

Definition 2.1 ([9]) *Given $f : \{0,1\}^k \to \{0,1\}$, the* influence *of coordinate i, denoted as $\mathrm{Inf}_i(f)$, is defined as:*

$$\mathrm{Inf}_i(f) := \sum_{\alpha : \alpha_i = 1} \hat{f}(\alpha)^2.$$

Definition 2.2 ([9]) *Given $f : \{0,1\}^k \to \mathbb{R}$, The* Noise Operator *$T_\varepsilon$ is defined as follows:*

$$T_\varepsilon(f) := \sum_\alpha \varepsilon^{|\alpha|}\hat{f}(\alpha)\chi(\alpha),$$

where $|\alpha|$ denotes the hamming weight of α.

Lemma 2.3 ([9]) *Given $f : \{0,1\}^k \to \mathbb{R}$, $1 \leq p \leq q$ and $\varepsilon \leq \sqrt{\frac{p-1}{q-1}}$ the Bonami-Beckner inequality states:*

$$||T_\varepsilon f||_q \leq ||f||_p.$$

Definition 2.4 *Given $f : \{0,1\}^k \to \{0,1\}$, we define the* difference *along coordinate i as $f_i(x) := f(x) - f(x + e_i)$, where e_i as usual is the unit vector which is 1 on coordinate i and 0 otherwise.*

Observe that f_i is $\{-1,0,1\}$ valued and that $f_i = 2\sum_{\alpha:\alpha_i=1}\hat{f}(\alpha)\chi_\alpha$. Therefore, for any $p \neq 0$

$$\mathbb{E}\left[|f_i|^p\right] \;=\; \mathbb{E}\left[|f_i|^2\right] \;=\; 4\mathrm{Inf}_i(f).$$

We will let $\gamma_r(f)$ denote the Fourier mass of predicate f at level r or more i.e.

$$\gamma_r(f) \;:=\; \sum_{|\alpha| \geq r} \hat{f}(\alpha)^2\,.$$

2.2 MAX k-CSP(f) and SDP Relaxations

For a predicate $f : \{0,1\}^k \to \{0,1\}$, an instance Φ of MAX k-CSP(f) on variables $x_1, \ldots, x_n$ is given by a set of constraints, each on a k-tuple of variables. We denote the variables in a constraint C by the tuple $\mathbf{x}_C$ and each constraint is of the form $f(\mathbf{x}_C + \mathbf{b}_C)$ where $\mathbf{b}_C \in \{0,1\}^k$ determines which variables are negated in the constraint. The objective is to maximize the fraction of satisfied constraints denoted as $\mathbb{E}_{C\in\Phi}[f(\mathbf{x}_C + \mathbf{b}_C)]$.

The SDP relaxation for MAX k-CSP(f) given by r rounds of the Lasserre hierarchy introduces a vector $\mathbf{V}_{(S,\alpha)}$ for each subset $S \subseteq [n], |S| \leq r$ and each $\alpha \in \{0,1\}^S$. The intended solution is that for an assignment $A : [n] \to \{0,1\}$ to all the variables, we set $\mathbf{V}_{(S,\alpha)} = 1$ if A assigns all the variables in S according to α and 0 otherwise. For any constraint C, we use S_C to denote the *set* of indices for the variables involved in C (note that the we denote the *tuple* of the variables by $\mathbf{x}_C$). For two assignments $\alpha_1 \in \{0,1\}^{S_1}$ and $\alpha_2 \in \{0,1\}^{S_2}$ which agree on $S_1 \cap S_2$, we use $\alpha_1 \circ \alpha_2$ to denote the extension of both assignments to $S_1 \cup S_2$.

Lasserre SDP for MAX k-CSP(f)

$$
\begin{array}{lll}
\text{maximize} & \displaystyle \mathop{\mathbb{E}}_{C\in\Phi} \sum_{\alpha\in\{0,1\}^{S_C}} f(\alpha + \mathbf{b}_C)\cdot \left\|\mathbf{V}_{(S_C,\alpha)}\right\|^2 & \\
\text{subject to} & \left\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \right\rangle = 0 & \forall\ \alpha_1(S_1\cap S_2) \neq \alpha_2(S_1\cap S_2) \\
& \left\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \right\rangle = \left\langle \mathbf{V}_{(S_3,\alpha_3)}, \mathbf{V}_{(S_4,\alpha_4)} \right\rangle & \forall\ S_1\cup S_2 = S_3\cup S_4,\ \alpha_1\circ\alpha_2 = \alpha_3\circ\alpha_4 \\
& \displaystyle\sum_{j\in\{0,1\}} \left\|\mathbf{V}_{(\{i\},j)}\right\|^2 = 1 & \forall i\in[n] \\
& \left\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \right\rangle \geq 0 & \forall S_1,S_2,\alpha_1,\alpha_2 \\
& \left\|\mathbf{V}_{(\emptyset,\emptyset)}\right\| = 1 &
\end{array}
$$

For any set S with $|S| \leq r$, the vectors $\mathbf{V}_{(S,\alpha)}$ induce a probability distribution over $\{0,1\}^S$ such that the assignment $\alpha \in \{0,1\}^S$ appears with probability $\left\|\mathbf{V}_{(S,\alpha)}\right\|^2$. The constraints can be understood by thinking of valid solution as coming from a distribution of assignments for all the variables and of $\left\langle \mathbf{V}_{(S_1,\alpha_1)}, \mathbf{V}_{(S_2,\alpha_2)} \right\rangle$ as the probability of the event that variables in S_1 get value according to α_1 and those in S_2 according to α_2.

2.3 Linear Codes

A linear code over $\mathbb{F}_2^k$ is simply a subspace of $\mathbb{F}_2^k$. We will identify $\mathbb{F}_2^k$ with $\{0,1\}^k$ in the natural fashion and use the two interchangeably.

Definition 2.5 *The* dual space $S^\perp$ *of a linear space* $S \subseteq \mathbb{F}_2^k$ *is defined as:*

$$S^\perp := \left\{x \in \mathbb{F}_2^k : x\cdot y = 0,\ \forall y\in S\right\}.$$

Also, for a linear code S*, we define its* distance *as* $\min\{|x| \ :\ x\in S, x\neq 0\}$.

We will need the existence of low dimensional subspaces S such that $S^\perp$ has distance at least 3.

Claim 2.6 *For all* $t \geq 3$*, there exists a subspace* $S \subseteq \{0,1\}^t$ *such that* $|S| \leq 2t$ *and* $S^\perp$ *is a linear code with distance at least 3.*

Proof: Let r be such that $2^{r-1} - 1 < t \leq 2^r - 1$. Then each element of $[t]$ can be identified with non-empty subset of $[r]$. Let $\mathcal{F} \subseteq 2^{[r]}$ be the family of subsets which corresponds to the elements of $[t]$, such that $\forall \in [r]$, $\{i\} \in \mathcal{F}$. Consider the space S defined by the following set of linearly independent equations over $\mathbb{F}_2$

$$x_T = \sum_{i \in T} x_{\{i\}} \quad \forall T \in \mathcal{F}, |T| > 1\,.$$

Since the number of equations is $t - r$, the size of S is $2^r = 2^{\lceil \log_2(t+1) \rceil} \leq 2t$. Also, since any non-trivial linear combination of the above equations gives an equation with at least 3 variables, $S^\perp$ has distance at least 3. ■

We will also need to count the number of d-dimensional subspaces of $\mathbb{F}_2^k$, which is given by the Gaussian binomial coefficients $\binom{k}{d}_2$ defined as

$$\binom{k}{d}_2 := \frac{\Pi_{i=0}^{d-1}(2^k - 2^i)}{\Pi_{i=0}^{d-1}(2^d - 2^i)}$$

for $0 \leq d \leq k$ and 0 otherwise.

2.4 Boolean Predicates

Definition 2.7 *We say that a predicate* $f : \{0,1\}^k \to \{0,1\}$ τ-correlates *with a predicate* $g : \{0,1\}^k \to \{0,1\}$ *if*

$$\frac{|f^{-1}(1) \cap g^{-1}(1)|}{|g^{-1}(1)|} \geq \tau.$$

Equivalently $\mathbb{E}_{x \in g^{-1}(1)}[f(x)] \geq \tau$.

Definition 2.8 *A* linear predicate $L : \{0,1\}^k \to \{0,1\}$ *corresponds to set of assignments* $L^{-1}(1)$ *which form a affine subspace of* $\mathbb{F}_2^k$. *We call such a predicate* well distributed *if the uniform distribution on* $L^{-1}(1)$ *is a balanced pairwise independent distribution on* $\{0,1\}^k$ *i.e.,* $\forall i \neq j \in [k], b_1, b_2 \in \{0,1\}, \quad \mathbb{P}_{x \in L^{-1}(1)}[x_i = b_1, x_j = b_2] = 1/4$.

The following alternate characterization of well distributed linear predicates is easy to prove.

Claim 2.9 *Let* $L : \{0,1\}^k \to \{0,1\}$ *be a linear predicate such that* $L^{-1}(1) = S + z$ *for a subspace* S *of* $\mathbb{F}_2^k$ *and* $z \in \mathbb{F}_2^k$. *Then* L *is well distributed if and only if* $S^\perp$ *forms a (linear) code of distance at least 3 over* $\{0,1\}^k$.

Proof: Since translation by z does affect the balance and pairwise independence of a distribution, L is well distributed if and only if the uniform distribution on S is balanced and pairwise independent. This is equivalent to the condition that $\mathbb{E}_{x \in S}[\chi_\alpha(x)] = 0$ for all α such that $0 < |\alpha| \leq 2$. Also, $\mathbb{E}_{x \in S}[\chi_\alpha(x)]$ is 1 for $\alpha \in S^\perp$ and 0 otherwise, which implies that the above condition is equivalent to saying $S^\perp$ does not contain any α with $0 < |\alpha| \leq 2$ i.e., it is a code with distance at least 3. ■

3 A Relation to Level 3 Fourier Mass

The following theorem shows that a predicate f with *high* Fourier mass at level 3 and above (i.e. *high* value of $\gamma_3(f)$) has a *high* correlation, say τ, with a well distributed linear predicate. Since a well-distributed linear predicate is 1-resistant, it immediately implies that f is τ-resistant. This argument is used to prove the lower bound on $\tau(f)$ in Case (1) and Case (2b) in Theorem 1.7.

Theorem 3.1 *Let $k \geq 16$ and $f : \{0,1\}^k \to \{0,1\}$ be a predicate. There exists*

$$\tau \geq \sqrt{\rho(f)^2 + \frac{\gamma_3(f)}{100k^2}} \tag{3.1}$$

such that f τ-correlates with some well distributed linear predicate (and hence is τ-resistant as well as τ-resistant for the Lasserre hierarchy).

Proof: Note that $\rho(f) = \hat{f}(0)$ and our statement is equivalent to showing that there exists a subspace S of $\mathbb{F}_2^k$ and $z \in \mathbb{F}_2^k$ such that $S^\perp$ is a distance (at least) 3 code and

$$\mathop{\mathbb{E}}_{x \in S+z}[f(x)] \quad \geq \quad \sqrt{\hat{f}(0)^2 + \frac{\gamma_3(f)}{100k^2}}\,. \tag{3.2}$$

We prove that choosing S, z at random works. Specifically, we show that when S, z are chosen at random appropriately, the square of above inequality holds in expectation. For now fix S, z so that S is a subspace and $S^\perp$ is a distance (at least) 3-code. Let d be the dimension of $S^\perp$. Then, for a basis $\alpha_1, \ldots, \alpha_d$ of $S^\perp$, $S + z$ can be uniquely specified as the set of points satisfying the equations $\alpha_i \cdot x = b_i \;\; \forall i \in [d]$ over $\mathbb{F}_2^k$, where $b_i = \alpha_i \cdot z$. Observe that:

$$\begin{aligned}
\mathop{\mathbb{E}}_{x \in S+z}[f(x)] &= \frac{2^k}{|S|} \cdot \mathop{\mathbb{E}}_{x \in \{0,1\}^k}\left[\mathbb{1}_{\{S+z\}}(x) \cdot f(x)\right] \\
&= 2^d \cdot \mathop{\mathbb{E}}_{x \in \{0,1\}^k}\left[\prod_{i=1}^{d}\left(\frac{1 + (-1)^{\alpha_i \cdot x + b_i}}{2}\right) \cdot f(x)\right] \\
&= \mathop{\mathbb{E}}_{x \in \{0,1\}^k}\left[\prod_{i=1}^{d}(1 + (-1)^{\alpha_i \cdot x + b_i}) \cdot f(x)\right] \\
&= \sum_{\alpha \in S^\perp} (-1)^{\alpha \cdot z} \cdot \hat{f}(\alpha).
\end{aligned}$$

Squaring both sides and taking expectation over a uniformly random choice of z gives (S is still fixed and $S^\perp$ has distance at least 3)

$$\mathop{\mathbb{E}}_{z}\left[\left(\mathop{\mathbb{E}}_{x \in S+z}[f(x)]\right)^2\right] \;=\; \sum_{\alpha \in S^\perp} \hat{f}(\alpha)^2 \;=\; \hat{f}(0)^2 + \sum_{\alpha \in S^\perp, \alpha \neq 0} \hat{f}(\alpha)^2 \;=\; \hat{f}(0)^2 + \sum_{\alpha : |\alpha| \geq 3} \hat{f}(\alpha)^2 \cdot \mathbb{1}_{\{\alpha \in S^\perp\}}\,.$$

Now we consider the expectation over the choice of S which is same as the choice of $S^\perp$. We choose $S^\perp$ to be a random code of dimension $d = \lfloor k - 2\log_2 k - 2 \rfloor$ and distance (at least) 3. Since $\frac{|S^\perp|}{2^k} \approx \frac{1}{4k^2}$ and a random choice of $S^\perp$ behaves as a random subset of $\{0,1\}^k$ with this density, it follows that over the choice of $S^\perp$, every α with $|\alpha| \geq 3$ is in $S^\perp$ with probability $\Omega(1/k^2)$ (i.e. $\mathbb{E}_{S^\perp}\left[\mathbb{1}_{\{\alpha \in S^\perp\}}\right] \geq \Omega(1/k^2)$). This proves the theorem.

Formally, let $\mathcal{C}^d$ denote the set of linear codes in $\mathbb{F}_2^k$ of dimension d and $\mathcal{C}_3^d \subseteq \mathcal{C}^d$ denote the set of codes with distance at least 3. We will assume $d = \lfloor k - 2\log_2 k - 2 \rfloor$ and suppress it for brevity. Choosing $S^\perp$ to be a random code in $\mathcal{C}_3$, for any α with $|\alpha| \geq 3$, we get

$$\mathop{\mathbb{E}}_{S^\perp}\left[\mathbb{1}_{\{\alpha \in S^\perp\}}\right] = \mathop{\mathbb{P}}_{C \in \mathcal{C}_3}[\alpha \in C] = \mathop{\mathbb{P}}_{C \in \mathcal{C}}[\alpha \in C | C \in \mathcal{C}_3] \geq \mathop{\mathbb{P}}_{C \in \mathcal{C}}[\alpha \in C, C \in \mathcal{C}_3].$$

The number of d dimensional codes C is $X = \binom{k}{d}_2$. The number of d dimensional codes C with $\alpha \in C$ is $Y = \binom{k-1}{d-1}_2$, which is obtained by pre-including α and then choosing the basis for C. The number of d dimensional codes C with $\alpha \in C$ and distance at most 2 is at most $Z = k^2 \cdot \binom{k-2}{d-2}_2$, which is obtained by pre-including α and some non-zero vector with Hamming weight at most 2 in choosing a basis of C. Thus the probability above is at least

$$\frac{Y - Z}{X} = \frac{\frac{2^{k-1}-1}{2^{d-1}-1}\binom{k-2}{d-2}_2 - k^2 \cdot \binom{k-2}{d-2}_2}{\frac{(2^k-1)(2^k-2)}{(2^d-1)(2^d-2)}\binom{k-2}{d-2}_2} \geq \frac{1}{2} \cdot \frac{2^{k-d} - k^2}{2^{2(k-d)}} \geq \frac{1}{100k^2},$$

where we used $2\log_2 k + 2 \leq k - d \leq 2\log_2 k + 3$. ■

4 Fourier Spectrum and Closeness to $\mathcal{Q}$

In this section we show that if $\gamma_3(f)$ is sufficiently small then f is close in Hamming distance to a quadratic function $g \in \mathcal{Q}$. In fact the distance $\Delta(f, \mathcal{Q})$ is proportional to $\gamma_3(f)$ whenever $\gamma_3(f) \leq 1/k^3$ (note that we are interested in the case when $\gamma_3(f)$ is polynomially small in $\frac{1}{k}$ which is somewhat atypical situation). This is similar, though incomparable to a result of Friedgut, Kalai and Naor [11] which shows that if $\gamma_2(f)$ is a sufficiently small *constant*, then f is close to a constant function or a dictator (Boolean functions of Fourier degree at most 1). Though our result works for functions of higher Fourier degree, it requires the Fourier mass at higher levels to be polynomially small in $1/k$.

We also prove a version of the main result from this section (Theorem 4.2) in Appendix B, even for the case when the Fourier mass at higher levels (say $\gamma_3(f)$) is a sufficiently small constant. We do get that such a function f must be close to a low-degree function (say $g \in \mathcal{Q}$). However, the distance there is no longer proportional to the Fourier mass as below.

Lemma 4.1 *Let $f : \{0,1\}^k \mapsto \{0,1\}$ be a predicate such that $\gamma_3(f) \leq 1/k^3$ and $k \geq 2^{2^{15}}$. Then*

$$\gamma_3(f) \leq \Delta(f, \mathcal{Q}) \leq C \cdot \gamma_3(f),$$

for an absolute constant C and $C = 128$ works.

We note that the lower bound above holds because a function $g \in \mathcal{Q}$ has no non-zero Fourier coefficient of degree 3 or more and hence

$$\Delta(f, g) = \|f - g\|_2^2 = \sum_\alpha (\hat{f}(\alpha) - \hat{g}(\alpha))^2 \geq \sum_{|\alpha| \geq 3} \hat{f}(\alpha)^2 = \gamma_3(f).$$

The proof of the upper bound above is similar to those in the papers by Kahn, Kalai, Linial [14] and Friedgut [10]. We will work in the general setting when for some integer $r \geq 2$, $\gamma_r(f) \leq 1/k^3$ and prove the following theorem (thus proving the above lemma when $r = 3$).

Theorem 4.2 *Let $r \geq 2$ and $f : \{0,1\}^k \to \{0,1\}$ be a predicate such that $\gamma_r(f) \leq 1/k^3$ and $k \geq 2^{2^{5r}}$. Then there exists $g : \{0,1\}^k \to \{0,1\}$ such that*

- $\Delta(f,g) \leq 2^{r+4}\gamma_r(f)$.
- $deg(g) \leq r-1$.
- *g depends on at most 2^{5r} variables.*

Proof: The proof proceeds in three steps. First we show that under the premise of the theorem, the influences are either *too small* or *too large*.[6] Denoting the set of coordinates with high influence by $I(f)$, we note that $|I(f)|$ is bounded since the total influence is bounded. We next show that most of the Fourier mass of f is contained inside $I(f)$. Truncating from the Fourier representation of f terms not contained in $I(f)$ yields a multi-linear polynomial h. Though h is close to f in ℓ_2-norm, it is non-boolean in general. Finally, we let g to be the indicator function of the event $h \geq \frac{1}{2}$. These steps are more or less standard as we noted.

We set the parameter $\theta = \gamma_r(f)$ for ease of notation. We begin by showing that:

$$\forall i \in [k], \quad \mathrm{Inf}_i(f) \leq 2\theta \quad \text{or} \quad \mathrm{Inf}_i(f) \geq \frac{1}{2^{3r+2}}. \tag{4.1}$$

Let T_ε be the noise operator with $\varepsilon = \frac{1}{\sqrt{2}}$. We recall that the difference function along i^{th} coordinate is $f_i(x) := f(x) - f(x+e_i)$. The Bonami-Beckner inequality implies

$$\left\|T_{\frac{1}{\sqrt{2}}} f_i\right\|_2^2 \leq \|f_i\|_{3/2}^2 = \mathbb{E}\left[|f_i|^{3/2}\right]^{4/3} = 2^{8/3}\ \mathrm{Inf}_i(f)^{4/3}. \tag{4.2}$$

On the other hand (using $\sum_{|\alpha|\geq r} \hat{f}(\alpha)^2 \leq \theta$),

$$\left\|T_{\frac{1}{\sqrt{2}}} f_i\right\|_2^2 = \sum_{\alpha:\alpha_i=1} \frac{4\hat{f}(\alpha)^2}{2^{|\alpha|}} \geq \frac{4}{2^{r-1}} \sum_{\substack{\alpha:\alpha_i=1,\\ |\alpha|\leq r-1}} \hat{f}(\alpha)^2 \geq \frac{4}{2^{r-1}} \cdot (\mathrm{Inf}_i(f) - \theta).$$

Combining the above two inequalities we get,

$$\frac{\mathrm{Inf}_i(f) - \theta}{2^{r-1/3}} \leq \mathrm{Inf}_i(f)^{4/3}.$$

If $\mathrm{Inf}_i(f) \geq 2\theta$, we get $\mathrm{Inf}_i(f)^{4/3} \geq \frac{\mathrm{Inf}_i(f)}{2^{r+2/3}}$ and hence $\mathrm{Inf}_i(f) \geq \frac{1}{2^{3r+2}}$. This proves the claim in Equation (4.1). Let $I(f)$ denote the set of coordinates with high influence, i.e.

$$I(f) := \left\{i \mid \mathrm{Inf}_i(f) \geq \frac{1}{2^{3r+2}}\right\}.$$

Now we observe that the total influence and hence $|I(f)|$ is bounded. Indeed,

$$\sum_{i\in[k]} \mathrm{Inf}_i(f) = \sum_{\alpha} |\alpha| \cdot \hat{f}(\alpha)^2 \leq r \sum_{|\alpha|\leq r-1} \hat{f}(\alpha)^2 + k \sum_{|\alpha|\geq r} \hat{f}(\alpha)^2 \leq r + k\theta \leq r+1,$$

[6] This might be a new and interesting observation. As far as we know, it does not appear explicitly in literature.

and therefore $|I(f)| \leq (r+1) \cdot 2^{3r+2} \leq 2^{5r}$. We next prove that most of the Fourier mass of f is contained inside $I(f)$. Since the Fourier mass at level r or above is already bounded by θ, we only need to consider the mass at least below r and not contained in $I(f)$. Specifically, we prove:

$$\sum_{\substack{\alpha:\alpha\cap\overline{I(f)}\neq\emptyset,\\ |\alpha|\leq r-1}} \hat{f}(\alpha)^2 \leq 2^{r+1}k\theta^{4/3}, \tag{4.3}$$

where we denoted by α also the set of coordinates $\{i|\alpha_i = 1\}$. We use Equation (4.2) again and sum over all $i \in \overline{I(f)}$. Note that

$$\sum_{i\in\overline{I(f)}} \left\|T_{\frac{1}{\sqrt{2}}} f_i\right\|_2^2 = \sum_{\alpha} \frac{1}{2^{|\alpha|}} \cdot 4\hat{f}(\alpha)^2 \cdot |\alpha \cap \overline{I(f)}| \geq \frac{4}{2^{r-1}} \sum_{\substack{\alpha:\alpha\cap\overline{I(f)}\neq\emptyset,\\ |\alpha|\leq r-1}} \hat{f}(\alpha)^2.$$

Since $\mathrm{Inf}_i(f) \leq 2\theta$ for $i \in \overline{I(f)}$ and $|\overline{I(f)}| \leq k$, we immediately obtain

$$\frac{4}{2^{r-1}} \sum_{\substack{\alpha:\alpha\cap\overline{I(f)}\neq\emptyset,\\ |\alpha|\leq r-1}} \hat{f}(\alpha)^2 \leq \sum_{i\in\overline{I(f)}} 2^{8/3}\,\mathrm{Inf}_i(f)^{4/3} \leq k \cdot 2^{8/3} \cdot (2\theta)^{4/3},$$

proving the claim in Equation (4.3). Finally, we let

$$h := \sum_{\alpha\subseteq I(f)} \hat{f}(\alpha)\chi_\alpha$$

and let $g : \{0,1\}^k \to \{0,1\}$ be defined as $g := \mathbb{1}_{\{h\geq 1/2\}}$. Clearly g depends only on the co-ordinates in $I(f)$. We will prove that $\Delta(f,g) \leq 2^{r+4}\theta$ and $deg(g) \leq r-1$. Equation (4.3) and the bound $\sum_{|\alpha|\geq r} \hat{f}(\alpha)^2 \leq \theta$ together imply that

$$||f-h||_2^2 \leq \theta + 2^{r+1}k\theta^{4/3} \leq 2^{r+2}\theta,$$

where we used $\theta \leq 1/k^3$ in the second step. For $x \in \{0,1\}^k$, $f(x) \neq g(x)$ only when $h(x) < 1/2$ and $f(x) = 1$, or when $h(x) \geq 1/2$ and $f(x) = 0$. In both cases $(f(x) - h(x))^2 \geq 1/4$ and so

$$\Delta(f,g) = \mathbb{P}[f \neq g] \leq 4||f-h||_2^2 \leq 2^{r+4}\theta.$$

We finish the proof by showing that $deg(g) \leq r-1$. Suppose on the contrary that there exists a Fourier coefficient in g of degree at least r. Since g depends only on $|I(f)|$ coordinates, this coefficient has magnitude at least $\frac{1}{2^{|I(f)|}}$. Since $2^{r+4}\theta \geq \Delta(f,g) = \|f-g\|_2^2$, the same Fourier coefficient in f has value at least

$$\frac{1}{2^{|I(f)|}} - 2^{(r+4)/2}\sqrt{\theta} \;\geq\; \frac{1}{2^{|I(f)|}} - 2^{(r+4)/2}\frac{1}{k^{3/2}} \geq \frac{1}{k} - \frac{1}{2k} = \frac{1}{2k},$$

contradicting the premise that the Fourier mass of f at or above level r is at most $1/k^3$. Noting that $|I(f)| \leq 2^{5r}$, it is enough to have $k \geq 2^{2^{5r}}$ for our argument to work. ∎

5 Proof of Main Theorem

In this section, we collect the rest of the pieces required in the proof of Theorem 1.7. We first show the hardness of approximating MAX k-CSP(f) when f has good correlation with a well distributed linear predicate (Lemma 5.2). Next, we show that MAX k-CSP(f) is hard to approximate when f is close to a junta g which is *not* monotonically above f (Lemma 5.3). These two statements suffice to prove the required lower bounds on $\tau(f)$ in Theorem 1.7 since we can show that f must have the appropriate correlation with a well distributed linear predicate in cases (1) and (2b), and must be close to a $g \in \mathcal{Q}$ in case (2a). Finally, we give an SDP based approximation algorithm for the case when f is close to a $g \in \mathcal{Q}$ and $g \geq f$.

5.1 Reductions from the hardness of approximating well distributed linear predicates

We now give the reductions from Chan's result [7] on the hardness of approximating well distributed linear predicates. His result shows that a well distributed linear predicate $L : \{0,1\}^k \to \{0,1\}$ is 1-resistant, even on MAX k-CSP(L) instances with certain uniformity properties. These properties concern what the assignments to n variables look when restricted to the k variables in a randomly chosen constraint from the instance. Recall that for an instance Φ of MAX k-CSP(L), a constraint $C \in \Phi$ is of the form $L(x_{i_1} + b_{i_1}, \ldots, x_{i_k} + b_{i_k})$. Let $\mathbf{x}_C$ denote the tuple $(x_{i_1}, \ldots, x_{i_k})$ of the variables in the constraint C and Let $\mathbf{b}_C$ denote the tuple $(b_{i_1}, \ldots, b_{i_k})$. Also, for an assignment $A : [n] \to \{0,1\}$, let $A(\mathbf{x}_C)$ denote $(A(x_{i_1}), \ldots, A(x_{i_k}))$. The following follows easily from the statement of Theorem 5.4 and the proof of Theorem 1.1 in [7].

Theorem 5.1 ([7]) *Let $k \geq 3$ and let $\eta, \varepsilon > 0$ be arbitrarily small constants. $L : \{0,1\}^k \to \{0,1\}$ be a well distributed linear predicate. Then, given an instance Φ of MAX k-CSP(L) on variables $x_1, \ldots, x_n$, it is NP-hard to distinguish between the following two cases:*

Yes: *There exists an assignment $A : [n] \to \{0,1\}$ satisfying $1-\eta$ fraction of the constraints. In fact, for any $z \in L^{-1}(1)$*

$$\frac{1-\eta}{|L^{-1}(1)|} \;\leq\; \mathop{\mathbb{P}}_{C \in \Phi}\left[A(\mathbf{x}_C) + \mathbf{b}_C = z\right] \;\leq\; \frac{1+\eta}{|L^{-1}(1)|}\,.$$

No: *For all assignments $A : [n] \to \{0,1\}$ and all $z \in \{0,1\}^k$, we have*

$$\frac{1-\varepsilon}{2^k} \;\leq\; \mathop{\mathbb{P}}_{C \in \Phi}\left[A(\mathbf{x}_C) + \mathbf{b}_C = z\right] \;\leq\; \frac{1+\varepsilon}{2^k}\,.$$

Thus, the theorem states that in the Yes case, not only are most constraint satisfied, but the tuple $A(\mathbf{x}_C) + \mathbf{b}_C$ looks almost uniformly distributed over $L^{-1}(1)$, over the choice of a random constraint $C \in \Phi$. On the other hand, in the No case, $A(\mathbf{x}_C) + \mathbf{b}_C$ looks almost uniformly distributed over *all of* $\{0,1\}^k$. In particular, this means that the fraction of satisfied constraints is at most $\left|L^{-1}(1)\right|/2^k + \varepsilon$.

Given the above theorem, it is easy to prove that a predicate f which correlates with some well distributed linear predicate must also be hard to approximate.

Lemma 5.2 *Let $k \geq 3$ and let $\eta, \varepsilon > 0$ be arbitrarily small constants. Let $f : \{0,1\}^k \to \{0,1\}$ be a predicate which τ-correlates with some well distributed linear predicate L. Then, given an instance Φ of MAX k-CSP(f) on variables $x_1, \ldots, x_n$, it is NP-hard to distinguish between the following cases:*

Yes: *There exists an assignment* $A : [n] \to \{0,1\}$ *satisfying* $(1-\eta)\cdot\tau$ *fraction of the constraints.*

No: *All assignments* $A : [n] \to \{0,1\}$ *satisfy at most* $(1+\varepsilon)\cdot\rho(f)$ *fraction of the constraints.*

Proof: The proof is by a simple reduction from the hardness of approximating L. Let Φ_L be any instance of MAX k-CSP(L). We can then create an instance of MAX k-CSP(f) such that if one can distinguish between the two cases for Φ, then one can distinguish between the Yes an No cases for Φ_L in Theorem 5.1, which is known to be NP-hard. To construct Φ from Φ_L, we simply replace each constraint $C \in \Phi_L$ of the form $L(\mathbf{x}_C + \mathbf{b}_C)$ by the constraint $f(\mathbf{x}_C + \mathbf{b}_C)$.

We first argue that if we are in the Yes case for Φ_L, then we must be in the Yes case for Φ. Let A be the optimal assignment for Φ_L. We consider the fraction of constraints in Φ satisfied by A.

$$\begin{aligned}
\mathop{\mathbb{E}}_{C\in\Phi}\left[f(A(\mathbf{x}_C)+\mathbf{b}_C)\right] &= \mathop{\mathbb{E}}_{C\in\Phi}\left[\sum_{z\in\{0,1\}^k} f(z)\cdot \mathbb{1}_{\{A(\mathbf{x}_C)+\mathbf{b}_C\}}(z)\right] \\
&\geq \mathop{\mathbb{E}}_{C\in\Phi}\left[\sum_{z\in L^{-1}(1)} f(z)\cdot \mathbb{1}_{\{A(\mathbf{x}_C)+\mathbf{b}_C\}}(z)\right] \\
&= \sum_{z\in L^{-1}(1)} f(z)\cdot \mathop{\mathbb{E}}_{C\in\Phi}\left[\mathbb{1}_{\{A(\mathbf{x}_C)+\mathbf{b}_C\}}(z)\right] \\
&\geq \sum_{z\in L^{-1}(1)} f(z)\cdot \frac{1-\eta}{|L^{-1}(1)|} \\
&\geq \tau\cdot\left|L^{-1}(1)\right|\cdot\frac{1-\eta}{|L^{-1}(1)|} \\
&= \tau\cdot(1-\eta).
\end{aligned}$$

A similar argument gives that when we are in the No case for Φ_L, we must also be in the No case for Φ. For any assignment A, we consider the fraction of satisfied constraints:

$$\begin{aligned}
\mathop{\mathbb{E}}_{C\in\Phi}\left[f(A(\mathbf{x}_C)+\mathbf{b}_C)\right] &= \mathop{\mathbb{E}}_{C\in\Phi}\left[\sum_{z\in\{0,1\}^k} f(z)\cdot \mathbb{1}_{\{A(\mathbf{x}_C)+\mathbf{b}_C\}}(z)\right] \\
&\leq \left(\rho(f)\cdot 2^k\right)\cdot\frac{1+\varepsilon}{2^k} \\
&\leq (1+\varepsilon)\cdot\rho(f)\,.
\end{aligned}$$

Thus, the fraction of satisfied constraints is at most $(1+\varepsilon)\cdot\rho(f)$. ■

Now we consider the case when $f : \{0,1\}^k \to \{0,1\}$ is close to a function g with Fourier degree at most 2, but is not monotonically dominated by it i.e., when $f^{-1}(1)\cap g^{-1}(0) \neq \emptyset$. We show that such an f is $(\rho(f)+\Omega(1/k))$-resistant. In fact we prove the statement below for any function g which is a junta depending only on s variables. This is sufficient because Håstad's result (Theorem 1.5) implies that a Boolean function g of degree 2 must depend on at most 4 of the k variables and hence the required result will follow easily.

Lemma 5.3 *Let k, s be such that $k - s \geq 3$ and let $\varepsilon > 0$ be an arbitrarily small constant. Let $g : \{0,1\}^k \to \{0,1\}$ be an s-junta and let $f : \{0,1\}^k \to \{0,1\}$ be a predicate such that $\Delta(f,g) = \delta \leq 1/(2^{s+2}\cdot k)$ and $f^{-1}(1)\cap g^{-1}(0)\neq\emptyset$. Then, given an instance Φ of MAX k-CSP(f) on variables $x_1,\ldots,x_n$, it is NP-hard to distinguish between the following two cases:*

Yes: *There exists an assignment $A : [n] \to \{0,1\}$ satisfying $\rho(f) + 1/(2^{s+3} \cdot k)$ fraction of the constraints.*

No: *All assignments $A : [n] \to \{0,1\}$ satisfy at most $(1+\varepsilon) \cdot \rho(f)$ fraction of the constraints in Φ.*

Proof: We assume for notational convenience that g depends on the first s of the k variables. For $y \in \{0,1\}^s$, let $f_y : \{0,1\}^{k-s} \to \{0,1\}$ denote the function on $k-s$ variables obtained by setting the first s variables in f according to y. Note that since g depends only on the first s variables, for each y g_y is either identically 1 or 0.

We will need to use well distributed linear predicates with very few accepting assignments. Let S be the subspace on $\{0,1\}^{k-s}$ (of size at most $2(k-s)$) given by Claim 2.6 and let L be the predicate such that $L^{-1}(1) = S$. Since $S^{\perp}$ has distance at least 3, L is well distributed. For each $y \in \{0,1\}^s$ we have

$$\begin{aligned} \mathop{\mathbb{E}}_{z \in \{0,1\}^{k-s}} \mathop{\mathbb{E}}_{z' \in S+z} \left[f_y(z') \right] &= \mathop{\mathbb{E}}_{z \in \{0,1\}^{k-s}} \mathop{\mathbb{E}}_{z' \in S} \left[f_y(z+z') \right] \\ &= \mathop{\mathbb{E}}_{z' \in S} \mathop{\mathbb{E}}_{z \in \{0,1\}^{k-s}} \left[f_y(z+z') \right] \\ &= \rho(f_y) \,. \end{aligned}$$

Thus, for each y, there exists z_y such that $\mathbb{E}_{z' \in S+z_y} \left[f_y(z') \right] \geq \rho(f_y)$. Also, let $x_0 = (y_0, z_0)$ be such that $g(x_0) = 0$ and $f(x_0) = 1$. Then $S + z_0$ is an affine subspace of size at most $2(k-s)$ that contains z_0. Using these, we can now describe the reduction. Since L is a well distributed linear predicate, we know that it is NP-hard to distinguish between the two cases in Theorem 5.1 for a given instance Φ_L of MAX-CSP(L). We describe how to transform an instance of MAX-CSP(L) to an instance of MAX k-CSP(f) such that the two cases of Theorem 5.1 correspond to the two cases in the statement of the lemma.

Let Φ_L be an instance of MAX-CSP(L) as in Theorem 5.1, with parameter η to be chosen later and ε as given. Recall that each constraint $C \in \Phi_L$ is of the form $L(\mathbf{x}_C + \mathbf{b}_C)$. We introduce s fresh variables for each constraint C and denote the s-tuple by $\mathbf{w}_C$. We replace each constraint C by 2^s new constraints. For each $y \in \{0,1\}^s$, we add a constraint C_y defined as

$$C_y \;:=\; f(\mathbf{w}_C + y, \mathbf{x}_C + \mathbf{b}_C + z_y)$$

where z_y is such that $\mathbb{E}_{z' \in S+z_y} \left[f_y(z') \right] \geq \rho(f_y)$ for $y \neq y_0$ and $z_y = z_0$ for $y = y_0$.

We first prove that the Yes case of Theorem 5.1 corresponds to the Yes case of the lemma. Let A be the optimal assignment to the variables in Φ_L. We extend A to an assignment A' for Φ which is the same as A for all the x variables in Φ and assigns all the new variables we introduced ($\mathbf{w}_C$ for each $C \in \Phi_L$) to 0. We show that A' satisfies the required number of constraints.

Claim 5.4 *For $\eta = 1/(2^{s+3} \cdot k)$, the fraction of constraints satisfied by A' is at least $\rho(f) + 1/(2^{s+3} \cdot k)$.*

Proof: The fraction of constraints satisfied is given by

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{C\in\Phi_L}\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[f(A'(\mathbf{w}_C)+y, A'(\mathbf{x}_C)+\mathbf{b}_C+z_y)\right]\\
&=\mathop{\mathbb{E}}_{C\in\Phi_L}\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[f(y, A(\mathbf{x}_C)+\mathbf{b}_C+z_y)\right]\\
&=\mathop{\mathbb{E}}_{C\in\Phi_L}\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[f_y(A(\mathbf{x}_C)+\mathbf{b}_C+z_y)\right]\\
&=\mathop{\mathbb{E}}_{C\in\Phi_L}\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[\sum_{z\in\{0,1\}^{k-s}}\mathbb{1}_{\{A(\mathbf{x}_C)+\mathbf{b}_C+z_y\}}(z)\cdot f_y(z)\right]\\
&\geq\mathop{\mathbb{E}}_{C\in\Phi_L}\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[\sum_{z\in S+z_y}\mathbb{1}_{\{A(\mathbf{x}_C)+\mathbf{b}_C+z_y\}}(z)\cdot f_y(z)\right]\\
&=\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[\sum_{z\in S}\mathop{\mathbb{E}}_{C\in\Phi_L}\left[\mathbb{1}_{\{A(\mathbf{x}_C)+\mathbf{b}_C\}}(z)\right]\cdot f_y(z+z_y)\right]\\
&\geq\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[(1-\eta)\cdot\mathop{\mathbb{E}}_{z\in S}\left[f_y(z+z_y)\right]\right].
\end{aligned}
$$

Note that $\mathbb{E}_{z\in S}\left[f_y(z+z_y)\right]=\mathbb{E}_{z\in S+z_y}\left[f_y(z)\right]$ is at least $\rho(f_y)$ for $y\neq y_0$ and at least $1/|S|\geq 1/2(k-s)\geq 1/2k$ for $y=y_0$. Thus the fraction of satisfied constraints is at least

$$(1-\eta)\cdot\left\{\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[\rho(f_y)\right]-\frac{1}{2^s}\cdot\left(\rho(f_{y_0})-\frac{1}{2k}\right)\right\}$$

where we added and subtracted $(1/2^s)\cdot\rho(f_{y_0})$ to obtain the term $\mathbb{E}_{y\in\{0,1\}^s}\left[\rho(f_y)\right]$ which is equal to $\rho(f)$. Since $g_{y_0}\equiv 0$ and $\Delta(f,g)=\delta$, we must have $\rho(f_{y_0})\leq 2^s\cdot\delta$. Thus, the fraction of satisfied constraints is at least

$$(1-\eta)\cdot\left\{\rho(f)-\delta+\frac{1}{2^{s+1}\cdot k}\right\}$$

By the assumption on δ and η, the above is at least $\rho(f)+1/(2^{s+3}\cdot k)$. ■

Next, we consider an instance Φ_L that corresponds to the No case of Theorem 5.1. We show that any assignment A to the variables of the instance Φ obtained by our reduction satisfies at most $(1+\varepsilon)\cdot\rho(f)$ fraction of the constraints.

Claim 5.5 *Let Φ be as above. Then the fraction of constraints satisfied by any assignment A is at most $(1+\varepsilon)\cdot\rho(f)$.*

Proof: As before, we consider the fraction of satisfied constraints given by

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{C\in\Phi_L}\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[f(A(\mathbf{w}_C)+y, A(\mathbf{x}_C)+\mathbf{b}_C+z_y)\right]\\
&=\mathop{\mathbb{E}}_{C\in\Phi_L}\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[f_{A(\mathbf{w}_C)+y}(A(\mathbf{x}_C)+\mathbf{b}_C+z_y)\right]\\
&=\mathop{\mathbb{E}}_{C\in\Phi_L}\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[\sum_{z\in\{0,1\}^{k-s}}f_{A(\mathbf{w}_C)+y}(z)\cdot\mathbb{1}_{\{A(\mathbf{x}_C)+\mathbf{b}_C+z_y\}}(z)\right]\\
&=\mathop{\mathbb{E}}_{C\in\Phi_L}\left[\sum_{z\in\{0,1\}^{k-s}}\mathop{\mathbb{E}}_{y\in\{0,1\}^s}\left[f_y(z)\right]\cdot\mathbb{1}_{\{A(\mathbf{x}_C)+\mathbf{b}_C+z_y\}}(z)\right],
\end{aligned}
$$

where we used the fact that $\mathbb{E}_{y\in\{0,1\}^s}\left[f_{A(\mathbf{w}_C)+y}(z)\right] = \mathbb{E}_{y\in\{0,1\}^s}\left[f_y(z)\right]$. Since we are in the No case of Theorem 5.1, we have that $\mathbb{E}_{C\in\Phi_L}\left[\mathbb{1}_{\{A(\mathbf{x}_C)+\mathbf{b}_C+z_y\}}(z)\right]$ is at most $(1+\varepsilon)/2^{k-s}$. Hence, the above expression is at most

$$(1+\varepsilon)\cdot \mathop{\mathbb{E}}_{y\in\{0,1\}^s}\ \mathop{\mathbb{E}}_{z\in\{0,1\}^{k-s}}\left[f_y(z)\right] = (1+\varepsilon)\cdot\rho(f)\,.$$

This gives that the fraction of satisfied constraints is at most $(1+\varepsilon)\cdot\rho(f)$ ■

Using the above two claims, we prove the lemma by choosing $\eta = 1/(2^{s+3}\cdot k)$. ■

5.2 Proofs of Lower Bounds on $\tau(f)$

Using the previous section, we can now prove the lower bounds on $\tau(f)$ in Theorem 1.7.

- Case 1: $\Delta(f,\mathcal{Q}) \geq 1/k^3$ implies by Lemma 4.1 that $\gamma_3 \geq 1/(12k^3)$. Theorem 3.1 then gives that f is τ-correlated with some well-distributed predicate for

  $$\tau \geq \sqrt{\rho(f)^2 + \Omega\left(\frac{1}{k^5}\right)} \geq \rho(f) + \Omega(1/k^5)\,.$$

  Lemma 5.2 then implies that f must be τ-resistant and hence $\tau(f) \geq \rho(f) + \Omega(1/k^5)$.

- Case 2b: In this case, Lemma 4.1 again gives $\gamma_3 \geq \Omega(\delta)$ and the Theorem 3.1 again gives that f must τ-correlate with a well distributed linear predicate for

  $$\tau \geq \sqrt{\rho(f)^2 + \Omega\left(\frac{\delta}{k^2}\right)} \geq \rho(f) + \Omega(\delta/k^2)$$

  As before, an application of Lemma 5.2 completes the proof.

- Case 2a: In this case $\Delta(f,g) = \delta \leq 1/k^3$ for some $g \in \mathcal{Q}$, which must be a 4-junta by Theorem 1.5. Then, if $f^{-1}(1)\cap g^{-1}(0) \neq \emptyset$, Lemma 5.3 gives that $\tau(f) \geq \rho(f) + \Omega(1/k)$.

The remaining part i.e., the upper bound in case (2b) of Theorem 1.7 will follow from the algorithm in Section 5.4.

5.3 Integrality Gaps in the Lasserre Hierarchy

We briefly sketch below the proofs of lower bounds on $\tau^*(f)$ i.e. Theorem 1.12. To obtain the required lower bounds, we need to prove integrality gap results (for $\Omega(n)$ rounds of the Lasserre hierarchy) analogous to the NP-hardness results in Section 5.1. We start with the following analogue of Theorem 5.1.

Theorem 5.6 ([23]) *Let $k \geq 3$ and $\varepsilon > 0$. Let $L : \{0,1\}^k \to \{0,1\}$ be a well distributed linear predicate. Then there exists a constant $c = c(\varepsilon)$ such that for every large enough n, there is an instance Φ of MAX k-CSP(L) on n variables, with the following properties:*

1. *The value of the SDP relaxation obtained obtained by $\lfloor cn \rfloor$ rounds of the Lasserre hierarchy is equal to 1. In fact,* for each *$C \in \Phi$ and $\alpha \in \{0,1\}^{S_C}$ such that $L(\alpha + \mathbf{b}_C) = 1$, we have that $\left\|\mathbf{V}_{(S_C,\alpha)}\right\|^2 = 1/\left|L^{-1}(1)\right|$.*

2. *For all assignments $A : [n] \to \{0,1\}$ and all $z \in \{0,1\}^k$, we have*

$$\frac{1-\varepsilon}{2^k} \;\leq\; \mathop{\mathbb{P}}_{C \in \Phi}\left[A(\mathbf{x}_C) + \mathbf{b}_C = z\right] \;\leq\; \frac{1+\varepsilon}{2^k}\,.$$

The second property, which is identical to the No case in Theorem 5.1, is not explicitly stated in [23] but it is easy to prove. The version of Theorem 5.6 in [23] is for random instances of MAX k-CSP(L) where each tuples $\mathbf{x}_C$ and string $\mathbf{b}_C$ is chosen uniformly at random, and the second property is easy to verify for such instances.

To prove an analogue of Lemma 5.2 in the Lasserre hierarchy, we need to construct an instance Φ_f of MAX k-CSP(f), given that f τ-correlates with L, for which the values of the SDP relaxation is at least τ and property 2 (Theorem 5.6) holds for the instance. We start with an instance φ of MAX k-CSP(L) as given by Theorem 5.6 and replace each constraint $L(\mathbf{x}_C + \mathbf{b}_C)$ in Φ by the constraint $f(\mathbf{x}_C + \mathbf{b}_C)$ to obtain Φ_f (as in the proof of Lemma 5.2). Moreover, we use the *same* vectors which form an optimal SDP solution for Φ to give a solution for Φ_f. The value of this SDP solution for Φ_f is equal to

$$\mathop{\mathbb{E}}_{C \in \Phi_f}\left[\sum_{\alpha \in \{0,1\}^{S_C}} f(\alpha + \mathbf{b}_C) \cdot \left\|\mathbf{V}_{(S_C,\alpha)}\right\|^2\right] = \mathop{\mathbb{E}}_{C \in \Phi_f}\left[\sum_{\alpha \in L^{-1}(1)+\mathbf{b}_C} f(\alpha + \mathbf{b}_C) \cdot (1/|L^{-1}(1)|)\right] = \tau\,.$$

Therefore we get the following statement analogous to Lemma 5.2.

Lemma 5.7 *Let $k \geq 3$ and let $\varepsilon > 0$ be an arbitrarily small constant. Let $f : \{0,1\}^k \to \{0,1\}$ be a predicate which τ-correlates with some well distributed linear predicate L. Then there exists a constant $c = c(\varepsilon)$, such that for every large enough n there is an instance Φ of MAX k-CSP(f) on n variables satisfying the following:*

- *The value of the relaxation obtained by $\lfloor cn \rfloor$ rounds of the Lasserre hierarchy is at least τ.*
- *All assignments $A : [n] \to \{0,1\}$ satisfy at most $(1+\varepsilon) \cdot \rho(f)$ fraction of the constraints.*

Theorem 3.1 and Lemma 5.7 together imply the lower bounds in case (1) and case (2b) of Theorem 1.12.

Similarly, to prove an analogue of Lemma 5.3, we start with an instance Φ of MAX k-CSP(L) for L as chosen in Lemma 5.3. We use the same reduction and replace each constraint $C \in \Phi$ by 2^s constraints of the form $C_y := f(\mathbf{w}_C + y, \mathbf{x}_C + \mathbf{b}_C + z_y)$ for a fresh set of variables $\mathbf{w}_C$ for each C, for $y \in \{0,1\}^s$ and z_y as chosen in Lemma 5.3. We now need to define the SDP vectors for any set S of variables and $\alpha \in \{0,1\}^S$. Let $S = S_1 \cup S_2$ where S_1 is the set of original variables from Φ and S_2 is the set of variables of the form $\mathbf{w}_c$ which we added during the reduction. Correspondingly, let $\alpha = \alpha_1 \circ \alpha_2$. Since the intended solution is to assign all the $\mathbf{w}_C$ variables to 0, we take $\mathbf{V}_{(S,\alpha)} = \mathbf{V}_{(S_1,\alpha_1)}$ if $\alpha_2 = 0^{|S_2|}$ and $\mathbf{V}_{(S,\alpha)} = 0$ otherwise. It is easy to verify that all SDP constraints are satisfied and that the SDP value is at least $\rho(f) + \Omega(1/k)$, thereby proving the lower bound in case (2a) of Theorem 1.12.

5.4 An SDP Rounding Algorithm

We now provide an SDP rounding algorithm based on the algorithm by Charikar and Wirth [8] to prove the upper bound in case (2b) of Theorem 1.7. For f and δ as in case (2b), and Φ

which is a $(\rho(f)+\varepsilon)$-satisfiable instance of $\mathsf{MAX\ k\text{-}CSP}(f)$, the algorithm below yields a non-trivial approximation when $\varepsilon = \Omega(k^3 \cdot \delta)$. This gives $\tau(f) \leq \rho(f) + O(k^3 \cdot \delta)$.

Håstad [22] observed that algorithm of Charikar and Wirth [8], which rounds an SDP relaxation for maximizing a *homogeneous* quadratic objective function, can in fact be used for approximating $\mathsf{MAX\ k\text{-}CSP}(g)$ for any $g : \{0,1\}^k \to \{0,1\}$ which has Fourier degree at most 2 (by rounding the standard SDP relaxation). This observation gives the following lemma for which we provide a proof in the appendix.

Lemma 5.8 *Let $g \in \mathcal{Q}$. Then there exists a randomized polynomial time algorithm for rounding the standard SDP relaxation of $\mathsf{MAX\ k\text{-}CSP}(g)$, which given an instance Φ with SDP value $\rho(g)+\varepsilon$, outputs an assignment A satisfying at least $\rho(g) + \frac{c\cdot\varepsilon}{\log(1/\varepsilon)}$ fraction of the constraints in expectation. Here c is an absolute constant.*

We now proceed to the main theorem for this section. The proof will essentially replace an instance Φ of $\mathsf{MAX\ k\text{-}CSP}(f)$ by an appropriate instance Φ_g of $\mathsf{MAX\ k\text{-}CSP}(g)$ and use the algorithm in Lemma 5.8 to find an assignment A_g for Φ_g. Our assignment for Φ will be obtained from A_g by a simple transformation which trades-off the approximation factor to avoid the bad situation where A_g ends up falsifying many constraints in Φ while still satisfying many constraints in Φ_g. We show the following:

Theorem 5.9 *Let $f : \{0,1\}^k \to \{0,1\}$ be a predicate such that there exists another predicate $g \in \mathcal{Q}$ satisfying $g \geq f$ and $\Delta(f,g) = \delta \leq 1/k^3$. Then there exists a randomized polynomial time algorithm, which given an instance of $\mathsf{MAX\ k\text{-}CSP}(f)$ in which $\rho(f)+\varepsilon$ fraction of constraints can be satisfied for $\varepsilon = \Omega(k^3 \cdot \delta)$, finds an assignment such that $\mathbb{E}_A\left[\mathsf{val}_\Phi(A)\right] \geq \rho(f) + \frac{c\cdot\varepsilon}{8k^2\log(\frac{1}{\varepsilon})}$.*

Proof: Given Φ, we first construct an instance Φ_g of $\mathsf{MAX\ k\text{-}CSP}(g)$ as follows. For each constraint $C \in \Phi$ of the form $f(\mathbf{x}_C + \mathbf{b}_C)$, we simply replace it by a constraint C' of the form $g(\mathbf{x}_C + \mathbf{b}_C)$. Let A_0 be any assignment which satisfies $\rho(f)+\varepsilon$ fraction of the constraints in Φ. Since we have $g \geq f$, the same assignment also satisfies at least $\rho(f)+\varepsilon \geq \rho(g) + (\varepsilon - \delta) \geq \rho(g) + \varepsilon/2$ fraction of the constraints in Φ_g.

We now use the algorithm from Lemma 5.8 to find an assignment A_g which in expectation satisfies at least $\rho(g) + \frac{c\cdot\varepsilon}{\log(1/\varepsilon)}$ fraction of the constraints in Φ_g. We use this to construct a random assignment A_f as follows. For each variable x_i, we independently set (with α to be chosen later)

$$A_f(x_i) \ := \ \begin{cases} A_g(x_i) & \text{with probability } \alpha \\ 0 & \text{with probability } \frac{1-\alpha}{2} \\ 1 & \text{with probability } \frac{1-\alpha}{2} \end{cases}$$

We will show that A_f still satisfies a good fraction of the constraints in Φ_g and that for each constraint $C' \in \Phi_g$, the probability that A_f satisfies C' but does not satisfy the corresponding $C \in \Phi$ is small. Together, these will complete the proof.

Before we analyze the fraction of constraints in Φ_g satisfied by A_f, we will need an assumption on the starting assignment A_g. For each variable $x_i \in \{0,1\}$, let y_i denote $(-1)^{x_i} \in \{-1,1\}$ and for an assignment A, let $A(y_i) = (-1)^{A(x_i)}$. Since g has Fourier degree at most 2, the fraction of constraints in Φ_g satisfied by an assignment to the variables can be written as a quadratic polynomial in the variables y_i as

$$\mathop{\mathbb{E}}_{C' \in \Phi_g}\left[g(\mathbf{x}_{C'} + \mathbf{b}_{C'})\right] \ = \ \rho(g) + \sum_i a_i \cdot y_i + \sum_{i,j} b_{ij} \cdot y_i y_j \,.$$

We will assume that A_g is such that the degree 1 part of the above expression is non-negative i.e., $\sum_i a_i \cdot A(y_i) \geq 0$. If this is not the case, we can flip all the bits in A_g to ensure this. Since the value of the degree-0 and degree-2 terms remain unchanged by this, it can only increase $\mathsf{val}_{\Phi_g}(A_g)$. We can now prove the following claim.

Claim 5.10 *Let A_g and A_f be as above. Then $\mathbb{E}_{A_f}\left[\mathsf{val}_{\Phi_g}(A_f)\right] \geq \rho(g) + \alpha^2 \cdot \frac{c \cdot \varepsilon}{\log(1/\varepsilon)}$.*

Proof: We know that $\mathsf{val}_{\Phi_g}(A_g)$ equals

$$\mathsf{val}_{\Phi_g}(A_g) \;=\; \rho(g) + \sum_i a_i \cdot A_g(y_i) + \sum_{i,j} b_{ij} \cdot A_g(y_i) \cdot A_g(y_j) \;\geq\; \rho(g) + \frac{c \cdot \varepsilon}{\log(1/\varepsilon)}\,.$$

Also, from the definition of A_f, it is easy to see that for all $i \neq j$

$$\underset{A_f}{\mathbb{E}}\left[A_f(y_i)\right] = \alpha \cdot A_g(y_i) \quad \text{and} \quad \underset{A_f}{\mathbb{E}}\left[A_f(y_i) \cdot A_f(y_j)\right] = \alpha^2 \cdot A_g(y_i) \cdot A_g(y_j)\,.$$

Thus,

$$\underset{A_f}{\mathbb{E}}\left[\mathsf{val}_{\Phi_g}(A_g)\right] \;=\; \rho(g) + \alpha \cdot \sum_i a_i \cdot A_g(y_i) + \alpha^2 \cdot \sum_{i,j} b_{ij} \cdot A_g(y_i) \cdot A_g(y_j)$$

Using the assumption $\sum_i a_i \cdot A_g(y_i) \geq 0$, we get that $\mathbb{E}_{A_f}\left[\mathsf{val}_{\Phi_g}(A_f)\right] \geq \rho(g) + \alpha^2 \cdot \frac{c \cdot \varepsilon}{\log(1/\varepsilon)}$. ■

Next we argue that the probability that A_f satisfies a constraint $C' \in \Phi_g$ but does not satisfy the corresponding constraint $C \in \Phi$ is small. Let C be of the form $f(\mathbf{x}_C + \mathbf{b}_C)$ which means C' is of the form $g(\mathbf{x}_C + \mathbf{b}_C)$. We will show that the probability that $A_f(\mathbf{x}_C) + \mathbf{b}_C \in g^{-1}(1) \cap f^{-1}(0)$ is small.

Consider the random string $z = A_f(\mathbf{x}_C) + A_g(\mathbf{x}_C)$. Each bit of z is 1 with probability $(1-\alpha)/2$ and 0 with probability $(1+\alpha)/2$. Let B_C denote the set $g^{-1}(1) \cap f^{-1}(0) + \mathbf{b}_C + A_g(\mathbf{x}_C)$. Then $A_f(\mathbf{x}_C) + \mathbf{b}_C \in g^{-1}(1) \cap f^{-1}(0)$ if an only if $z \in B_C$. Also, $|B_C| \leq \delta \cdot 2^k$ by assumption. For a set B, Let $\mu_p(B)$ denote the probability that a random string z_p where each bit is independently 1 with probability p and 0 with probability $1-p$, lands inside B. Then, we know that $\mu_{1/2}(B_C) \leq \delta$ and are interested in bounding $\mu_{1/2-\alpha}(B_C)$. The following claim gives the required bound.

Claim 5.11 *Let $B \subseteq \{0,1\}^k$ be such that $\mu_{1/2}(B) \leq \delta$. Then, for $\alpha \leq 1/2k$, we have that $\mu_{1/2-\alpha}(B) \leq 3\delta$.*

Proof: For any string $w \in \{0,1\}^k$, we have that

$$\mu_{1/2-\alpha}(w) \;=\; (1/2-\alpha)^{|w|} \cdot (1/2+\alpha)^{k-|w|} \;\leq\; (1/2^k) \cdot (1+2\alpha)^k\,.$$

Since $\alpha \leq 1/2k$, we have that $(1+2\alpha)^k \leq e \leq 3$. Thus, for any w, $\mu_{1/2-\alpha}(w) \leq 3 \cdot \mu_{1/2}(w)$ which gives $\mu_{1/2-\alpha}(B) \leq 3\delta$. ■

Choosing $\alpha = 1/2k$, the expected fraction of constraints $C \in \Phi$ for which $A_f(\mathbf{x}_C) + \mathbf{b}_C \in g^{-1}(1) \cap f^{-1}(0)$ is at most 3δ. Combining this with Claim 5.10, we get that

$$\underset{A_f}{\mathbb{E}}\left[\mathsf{val}_{\Phi}(A_f)\right] \;\geq\; \underset{A_f}{\mathbb{E}}\left[\mathsf{val}_{\Phi_g}(A_f)\right] - 3\delta \;\geq\; \rho(g) + \frac{1}{4k^2} \cdot \frac{c \cdot \varepsilon}{\log(1/\varepsilon)} - 3\delta\,.$$

The above is at least $\rho(g) + \frac{1}{8k^2} \cdot \frac{c \cdot \varepsilon}{\log(1/\varepsilon)}$ if $3\delta \leq \frac{1}{8k^2} \cdot \frac{c \cdot \varepsilon}{\log(1/\varepsilon)}$, which if true when $\varepsilon \geq C \cdot k^3 \delta$ for some appropriately large constant C. ■

Acknowledgements

We thank Ryan O'Donnell for help with references.

References

[1] Michael Alekhnovich, Eli Ben-Sasson, Alexander A. Razborov, and Avi Wigderson. Pseudorandom generators in propositional proof complexity. *SIAM J. Comput.*, 34(1):67–88, 2005.

[2] Michael Alekhnovich and Alexander Razborov. Lower Bounds for Polynomial Calculus: Non-Binomial Case. In *FOCS*, pages 190–199, 2001.

[3] Per Austrin and Subhash Khot. A Characterization of Approximation Resistance for Even k-Partite CSPs. *Manuscript*, 2012.

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

[5] Eli Ben-Sasson and Russell Impagliazzo. Random CNFs are Hard for the Polynomial Calculus. *Computational Complexity*, 19(4):501–519, December 2010.

[6] Eli Ben-Sasson and Avi Wigderson. Short Proofs are Narrow – Resolution Made Simple. *J. ACM*, 48(2):149–169, March 2001.

[7] Siu On Chan. Approximation Resistance from Pairwise Independent Subgroups. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:110, 2012.

[8] Moses Charikar and Anthony Wirth. Maximizing Quadratic Programs: Extending Grothendieck's Inequality. In *FOCS*, pages 54–60, 2004.

[9] Ronald de Wolf. A Brief Introduction to Fourier Analysis on the Boolean Cube. *Theory of Computing, Graduate Surveys*, 1:1–20, 2008.

[10] Ehud Friedgut. Boolean Functions With Low Average Sensitivity Depend On Few Coordinates. *Combinatorica*, 18(1):27–35, 1998.

[11] Ehud Friedgut, Gil Kalai, and Assaf Naor. Boolean Functions whose Fourier Transform is Concentrated on the First Two Levels. *Advances in Applied Mathematics*, 29(3):427–437, 2002.

[12] Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. Optimal Sherali-Adams Gaps from Pairwise Independence. In *APPROX-RANDOM*, pages 125–139, 2009.

[13] Gustav Hast. *Beating a Random Assignment*. PhD thesis, Royal Institute of Technology, Sweden, 2005.

[14] Jeff Kahn, Gil Kalai, and Nathan Linial. The Influence of Variables on Boolean Functions. In *FOCS*, pages 68–80, 1988.

[15] Rafał Latała. Estimates of moments and tails of gaussian chaoses. *The Annals of Probability*, 34(6):pp. 2315–2331, 2006.

[16] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010.

[17] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[18] Ryan O'Donnell. Lecture notes for analysis of Boolean functions. Available at http://analysisofbooleanfunctions.org, 2012.

[19] Alex Samorodnitsky and Luca Trevisan. A PCP Characterization of NP with Optimal Amortized Query Complexity. In *STOC*, pages 191–199, 2000.

[20] Grant Schoenebeck. Linear Level Lasserre Lower Bounds for Certain k-CSPs. In *FOCS*, pages 593–602, 2008.

[21] Johan Håstad. Some Optimal Inapproximability Results. In *J. of the ACM*, pages 798–859, 2001.

[22] Johan Håstad. On the Efficient Approximability of Constraint Satisfaction Problems. In *Surveys in Combinatorics*, volume 346, pages 201–222. Cambridge University Press, 2007.

[23] Madhur Tulsiani. CSP gaps and Reductions in the Lasserre Hierarchy. In *STOC*, pages 303–312, 2009.

[24] Madhur Tulsiani and Pratik Worah. LS+ Lower Bounds from Pairwise Independence. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:105, 2012.

A Charikar-Wirth SDP Rounding

Charikar and Wirth [8] gave a semidefinite programming based algorithm to approximate the value of quadratic programs of the form

$$\max_{y \in \{-1,1\}^n} \sum_{i,j} B_{ij} \cdot y_i y_j .$$

We show how to modify their algorithm to obtain an approximation algorithm for $\mathsf{MAX\ k\text{-}CSP}(g)$ for $g \in \mathcal{Q}$. For a quadratic program of the above form, they solve the standard SDP relaxation which can be written as

$$\begin{aligned} \text{maximize} \qquad & \sum_{i,j} B_{ij} \cdot \langle \mathbf{U}_i, \mathbf{U}_j \rangle \\ \text{subject to} \qquad & \|\mathbf{U}_i\|^2 = 1 \quad \forall i \in [n] \end{aligned}$$

They use the following rounding algorithm which we will need to modify appropriately for $\mathsf{MAX\ k\text{-}CSP}(g)$. We can assume that $\mathbf{U}_i \in \mathbb{R}^n \ \ \forall i$. The parameter $T > 0$ will be chosen later.

- Sample a random Gaussian vector $\mathbf{g} \sim N(0,1)^n$.
- For each i, set $z_i = (1/T) \cdot \langle \mathbf{g}, \mathbf{U}_i \rangle$ if $(1/T) \cdot \langle \mathbf{g}, \mathbf{U}_i \rangle \in [-1,1]$ and $\mathsf{sign}(\langle \mathbf{g}, \mathbf{U}_i \rangle)$ otherwise.
- Set $y_i = 1$ independently with probability $(1 + z_i)/2$ and $y_i = -1$ otherwise.

They show that under this rounding scheme, for each i, j and value B_{ij}

$$\mathbb{E}\left[B_{ij} \cdot y_i y_j\right] \;\geq\; (1/T^2) \cdot B_{ij} \cdot \langle \mathbf{U}_i, \mathbf{U}_j \rangle - 8e^{-T^2/2} \cdot |B_{ij}| \; . \tag{A.1}$$

Choosing an appropriate value for T then gives the required approximation guarantee. We will use a slight modification of their algorithm to prove the following.

Lemma A.1 *Let $g \in \mathcal{Q}$. Then there exists a randomized polynomial time algorithm for rounding the standard SDP relaxation of MAX k-CSP(g), which given an instance Φ with SDP value $\rho(g)+\varepsilon$, outputs an assignment A satisfying at least $\rho(g) + \frac{c \cdot \varepsilon}{\log(1/\varepsilon)}$ fraction of the constraints in expectation. Here c is an absolute constant.*

Proof: Given an instance Φ of MAX k-CSP(g) in variables $x_1, \ldots, x_n$, we can write the fraction of satisfied constraints as

$$\mathop{\mathbb{E}}_{C \in \Phi_g}\left[g(\mathbf{x}_C + \mathbf{b}_C)\right] \;=\; \rho(g) + \sum_i a_i \cdot y_i + \sum_{i,j} b_{ij} \cdot y_i y_j \, ,$$

where $y_i = (-1)^{x_i}$ and we used the fact that $g \in \mathcal{Q}$. We can introduce a variable y_0 which is intended to be equal to 1 and consider the following semidefinite relaxation

$$\begin{array}{ll} \text{maximize} & \rho(g) + \sum_i a_i \cdot \langle \mathbf{U}_0, \mathbf{U}_i \rangle + \sum_{i,j} b_{ij} \cdot \langle \mathbf{U}_i, \mathbf{U}_j \rangle \\ \text{subject to} & \|\mathbf{U}_i\|^2 = 1 \quad \forall i \in \{0\} \cup [n] \end{array}$$

We (randomly) round all the variables including an extra variable corresponding to $\mathbf{U}_0$ to obtain values $w_0, w_1, \ldots, w_n \in \{-1, 1\}$ exactly as in the Charikar-Wirth algorithm. Finally, we set $y_i = w_0 \cdot w_i$ for each i. By equation A.1 we have that for each i and j

$$\begin{aligned} \mathbb{E}\left[a_i \cdot y_i\right] &= \mathbb{E}\left[a_i \cdot w_0 w_i\right] \;\geq\; (1/T^2) \cdot a_i \cdot \langle \mathbf{U}_0, \mathbf{U}_i \rangle - 8e^{-T^2/2} \cdot |a_i| \\ \mathbb{E}\left[b_{ij} \cdot y_i y_j\right] &= \mathbb{E}\left[b_{ij} \cdot w_i w_j\right] \;\geq\; (1/T^2) \cdot b_{ij} \cdot \langle \mathbf{U}_i, \mathbf{U}_j \rangle - 8e^{-T^2/2} \cdot |b_{ij}| \; . \end{aligned}$$

Let ROUND denote the fraction of constraints satisfied by the assignment given by the above algorithm and let FRAC denote the value of the SDP relaxation. Then, using the above we have

$$\mathbb{E}\left[\mathsf{ROUND} - \rho(g)\right] \;\geq\; (1/T^2) \cdot \mathbb{E}\left[\mathsf{FRAC} - \rho(g)\right] - 8e^{-T^2/2} \cdot \left(\sum_i |a_i| + \sum_{ij} |b_{ij}| \right) .$$

To obtain a bound on the second term, we note that

$$\sum_i |a_i| + \sum_{ij} |b_{ij}| \;\leq\; \mathop{\mathbb{E}}_{C \in \Phi}\left[\sum_\alpha |\hat{g}(\alpha)| \right] \;\leq\; 16 \, ,$$

Since g has Fourier degree at most 2 and can depend on at most 4 variables by Theorem 1.5. Choosing $T = \sqrt{4 \cdot \log_e(1/\varepsilon)}$ and using that $\mathsf{FRAC} - \rho(g) \geq \varepsilon$, we get that

$$\mathbb{E}\left[\mathsf{ROUND}\right] \;\geq\; \rho(g) + \frac{\varepsilon}{4 \cdot \log_e(1/\varepsilon)} - 128 \cdot \varepsilon^2$$

which is at least $\rho(g) + \frac{\varepsilon}{8 \log_e(1/\varepsilon)}$ for sufficiently small ε. ∎

B An improved version of Theorem 4.2

In this section we prove an improved version of Theorem 4.2, which is perhaps more directly comparable to the results of Friedgut et al. [11]. Theorem 4.2 shows that a a function $f : \{0,1\}^k \to \{0,1\}$ for which the Fourier mass at level r and above ($\gamma_r(f)$) is small, is close to a function g with Fourier degree at most $r-1$. There we require $\gamma_r(f)$ to be polynomially small in k ($\gamma_r(f) \leq 1/k^3$). Here, we prove that even if $\gamma_r(f)$ is a sufficiently small constant (depending on r), then f is close to a function g with Fourier degree at most $r-1$. However, the distance $\Delta(f,g)$ is no longer $O_r(\gamma_r(f))$. Formally, we prove the following.

Theorem B.1 *Let $r \geq 2$. Then there exist constants c_r and C_r such that for any $f : \{0,1\}^k \to \{0,1\}$ with $\gamma_r(f) \leq c_r$, there exists $g : \{0,1\}^k \to \{0,1\}$ satisfying*

- $\Delta(f,g) \leq C_r / \log(1/\gamma_r(f))$.
- *$deg(g) \leq r-1$.*
- *g depends on at most 2^{4r} variables.*

Proof: As before, we will proceed by showing that all influences are either too small or too large, and use $I(f)$ to denote the set of coordinates with high influence. However, since we can no longer use $\gamma_r(f)$ to obtain a bound on the total influence (and hence the size of $I(f)$), we will need to work with *degree bounded influences*. For $f : \{0,1\}^k \to \{0,1\}$, the degree d bounded influence of a coordinate i is defined as

$$\mathrm{Inf}_i^{\leq d}(f) \ := \ \sum_{\substack{\alpha_i = 1 \\ |\alpha| \leq d}} \hat{f}(\alpha)^2 \,.$$

It is easy to see that the total degree d bounded influence equals $\sum_{|\alpha| \leq d} |\alpha| \cdot \hat{f}(\alpha)^2$, which is at most d. Taking $I(f)$ to be the set of coordinates with high degree $(r-1)$ influence will give the required bound on $|I(f)|$.

Again as before, we will take h to be the function obtained by dropping the terms not contained in $I(f)$ from the Fourier expansion of f, and g to be the indicator for $h \geq \frac{1}{2}$. However, to show that the Fourier mass of f is mostly contained in $I(f)$ (and hence g is close to f), it will no longer be sufficient to use a simple union bound as in the proof of Theorem 4.2. Instead, we will need to use the invariance principle of Mossel et al. [16] and tail estimates for polynomials of Gaussian variables by Latała [15].

As before, we set $\theta = \gamma_r(f)$ and begin by showing:

$$\forall i \in [k], \qquad \mathrm{Inf}_i^{\leq (r-1)}(f) \leq \theta \quad \text{or} \quad \mathrm{Inf}_i^{\leq (r-1)}(f) \geq \frac{1}{2^{3r+3}} \,. \tag{B.1}$$

Let T_ε be the noise operator with $\varepsilon = \frac{1}{\sqrt{2}}$ and let f_i be the difference function along i^{th} coordinate defined as $f_i(x) := f(x) - f(x + e_i)$. The Bonami-Beckner inequality implies

$$\begin{aligned} \left\| T_{\frac{1}{\sqrt{2}}} f_i \right\|_2^2 \ \leq \ \|f_i\|_{3/2}^2 = \ & \mathbb{E}\left[|f_i|^{3/2} \right]^{4/3} \\ = \ & 2^{8/3} \cdot (\mathrm{Inf}_i(f))^{4/3} \\ \leq \ & 2^{8/3} \cdot \left(\mathrm{Inf}_i^{\leq (r-1)}(f) + \theta \right)^{4/3} \end{aligned}$$

On the other hand (using $\sum_{|\alpha|\geq r}\hat{f}(\alpha)^2 \leq \theta$),

$$\left\|T_{\frac{1}{\sqrt{2}}} f_i\right\|_2^2 \;=\; \sum_{\alpha:\alpha_i=1} \frac{4\hat{f}(\alpha)^2}{2^{|\alpha|}} \;\geq\; \frac{4}{2^{r-1}} \sum_{\substack{\alpha:\alpha_i=1,\\ |\alpha|\leq r-1}} \hat{f}(\alpha)^2 \;=\; \frac{4}{2^{r-1}} \cdot \mathrm{Inf}_i^{\leq (r-1)}(f)\,.$$

Combining the above two inequalities we get,

$$\frac{\mathrm{Inf}_i^{\leq (r-1)}(f)}{2^{r-1/3}} \;\leq\; \left(\mathrm{Inf}_i^{\leq (r-1)}(f) + \theta\right)^{4/3}.$$

If $\mathrm{Inf}_i^{\leq (r-1)}(f) \geq \theta$, we get $\frac{\mathrm{Inf}_i^{\leq (r-1)}(f)}{2^{r-1/3}} \leq \left(2\cdot \mathrm{Inf}_i^{\leq (r-1)}(f)\right)^{4/3}$ and hence $\mathrm{Inf}_i^{\leq (r-1)}(f) \geq \frac{1}{2^{3r+3}}$.

Let $I(f)$ denote the set of coordinates with high influence, i.e.

$$I(f) \;:=\; \left\{ i \;:\; \mathrm{Inf}_i^{\leq (r-1)}(f) \geq \frac{1}{2^{3r+3}} \right\}.$$

Since the total degree $r-1$ influence is at most $r-1$, we have that $|I(f)| \leq (r-1)\cdot 2^{3r+3} \leq 2^{4r}$.

The following lemma proves that most of the Fourier mass of f is contained in $I(f)$. The proof of the lemma is where the argument differs significantly from the one used in the proof of Theorem 4.2. Note that below we also use α to denote the set $\{i \;:\; \alpha_i = 1\}$.

Lemma B.2 *There exists a constant C_r such that for $\gamma_r(f) = \theta$ and the set $I(f)$ defined as above,*

$$\sum_{\alpha:\alpha\cap\overline{I}\neq\emptyset} \hat{f}(\alpha)^2 \;\leq\; C_r/\log(1/\theta)\,.$$

Proof: For ease of notation, we denote the set $I(f)$ simply as I. For $z \in \{0,1\}^I$, let $f_{I\to z}$ denote function obtained by fixing the inputs of f in I according to z. We will fix the bits in I so that the function $f_{I\to z}$ has variance equal to $\sum_{\alpha:\alpha\cap\overline{I(f)}\neq\emptyset}\hat{f}(\alpha)^2$ and the degree $r-1$ influence of all the unfixed variables is small. We will then bound the variance of $f_{I\to z}$ using the invariance principle.

Claim B.3 *There exists $z \in \{0,1\}^I$ such that for the function $F = f_{I\to z}$, we have*

- $\mathsf{Var}(F) \;\geq\; \sum_{\alpha:\alpha\cap\overline{I(f)}\neq\emptyset}\hat{f}(\alpha)^2$.
- *For all $i \in \overline{I}$,* $\mathrm{Inf}_i^{\leq (r-1)}(F) \;\leq\; 2^{|I|}\cdot\theta$.
- $\sum_{\substack{\beta\in\{0,1\}^{\overline{I}}\\ |\beta|\geq r}} \hat{F}(\beta)^2 \;\leq\; 2^I\cdot\theta$.

Proof: We first consider the expected variance of $f_{I\to z}$ over the choice of a random $z \in \{0,1\}^I$. For $\beta \in \{0,1\}^{\overline{I}}$ and $\beta' \in \{0,1\}^I$, we use $\beta\circ\beta'$ to denote the string in $\{0,1\}^k$ obtained by combining the two.

$$\begin{aligned}
\mathop{\mathbb{E}}_{z\in\{0,1\}^I}\left[\mathsf{Var}(f_{I\to z})\right] &= \mathop{\mathbb{E}}_{z\in\{0,1\}^I}\left[\sum_{\beta\in\{0,1\}^{\overline{I}},\beta\neq 0^{|\overline{I}|}} \hat{f}_{I\to z}(\beta)^2\right] \\
&= \mathop{\mathbb{E}}_{z\in\{0,1\}^I}\left[\sum_{\beta\in\{0,1\}^{\overline{I}},\beta\neq 0^{|\overline{I}|}} \left(\sum_{\beta'\in\{0,1\}^I} \hat{f}(\beta\circ\beta')\cdot\chi_{\beta'}(z)\right)^2\right] \\
&= \sum_{\alpha:\alpha\cap\overline{I}\neq\emptyset} \hat{f}(\alpha)^2\,.
\end{aligned}$$

Thus, there exists a $z \in \{0,1\}^I$ such that $\mathsf{Var}(f_{I\to z}) \geq \sum_{\alpha:\alpha\cap\overline{I}\neq\emptyset} \hat{f}(\alpha)^2$. We fix such a z and let F denote the function $f_{I\to z}$ on $\{0,1\}^{\overline{I}}$. By a calculation as above, it is easy to check that for any $i \in \overline{I}$, $\mathbb{E}_{z\in\{0,1\}^I}\left[\mathrm{Inf}_i^{\leq(r-1)}(f_{I\to z})\right] = \mathrm{Inf}_i^{\leq(r-1)}(f)$. Since $\mathrm{Inf}_i^{\leq(r-1)}(f) \leq \theta$ for all $i \in \overline{I}$, we must have that $\mathrm{Inf}_i^{\leq(r-1)}(F) \leq 2^{|I|}\cdot\theta$ for each $i \in \overline{I}$. Similarly, we also have that

$$\begin{aligned}\mathop{\mathbb{E}}_{z\in\{0,1\}^I}\left[\sum_{|\beta|\geq r}\hat{f}_{I\to z}(\beta)^2\right] &= \mathop{\mathbb{E}}_{z\in\{0,1\}^I}\left[\sum_{|\beta|\geq r}\left(\sum_{\beta'\in\{0,1\}^I}\hat{f}(\beta\circ\beta')\cdot\chi_{\beta'}(z)\right)^2\right]\\ &= \sum_{|\alpha\cap\overline{I}|\geq r}\hat{f}(\alpha)^2 \;\leq\; \theta\,.\end{aligned}$$

Hence, $\sum_{|\beta|\geq r}\hat{F}(\beta)^2 \leq 2^{|I|}\cdot\theta$. ■

By the invariance principle of Mossel et al. [16], we will get that F is close to a function $\tilde{F}$ on Gaussian variables obtained by replacing each $(-1)^{x_i}$ by a standard normal variable G_i in the Fourier expansion of F. Since F is Boolean, it takes values only in the interval $[0,1]$. By the invariance principle, $\tilde{F}$ must also take values only in this interval, except for very small probability. By the results of Latała [15], this will lead to a bound on $\mathsf{Var}(\tilde{F}) = \mathsf{Var}(F)$. The actual proof is a little more complicated since we will actually apply the invariance principle to the "low-degree" part of F, which is not Boolean (but will take large values only with a small probability).

We now consider the low-degree part of F. Let ν denote the quantity $\sum_{\substack{\beta\in\{0,1\}^I\\ 0<|\beta|\leq r-1}}\hat{F}(\beta)^2$. Note that it is sufficient to upper bound ν since we already know that $\sum_{\substack{\beta\in\{0,1\}^{\overline{I}}\\ |\beta|\geq r}}\hat{F}(\beta)^2 \leq 2^I\cdot\theta$. We define the following (normalized) polynomial $\mathcal{P}$ on variables $Z_1,\ldots,Z_{|\overline{I}|}$, corresponding to the low-degree part in the Fourier expansion of F.

$$\mathcal{P} \;:=\; \frac{1}{\sqrt{\nu}}\cdot\sum_{\substack{\beta\in\{0,1\}^{\overline{I}}\\ |\beta|\leq r-1}}\hat{F}(\beta)\cdot\prod_{i\in\beta}Z_i \;=\; \sum_{\substack{\beta\in\{0,1\}^{\overline{I}}\\ |\beta|\leq r-1}}c_\beta\cdot\prod_{i\in\beta}Z_i\,,$$

where $c_\beta = \hat{F}/\sqrt{\nu}$. By definition of $\mathcal{P}$, we have that

$$\sum_{\beta\neq 0^{|\overline{I}|}} c_\beta^2 = 1 \qquad\text{and}\qquad \forall i\in\overline{I},\;\; \sum_{\beta\ni i}c_\beta^2 = \frac{1}{\nu}\cdot\mathrm{Inf}_i^{\leq(r-1)}(F) \leq \frac{2^{|\overline{I}|}\cdot\theta}{\nu}\,.$$

Let $\tau = \frac{2^{|\overline{I}|}\cdot\theta}{\nu}$. Then by the invariance principle (Theorem 2.1) of [16], we get that

$$\sup_t\left|\mathbb{P}\left[\mathcal{P}(X_1,\ldots,X_{|\overline{I}|})\geq t\right] - \mathbb{P}\left[\mathcal{P}(G_1,\ldots,G_{|\overline{I}|})\geq t\right]\right| \;\leq\; O(r\cdot\tau^{1/(8(r-1))})\,, \tag{B.2}$$

where $X_1,\ldots,X_{\overline{I}}$ are independent random variables taking values in $\{-1,1\}$ and $G_1,\ldots,G_{\overline{I}}$ are independent random variables with each $G_i\sim N(0,1)$. By estimates on the tails of polynomials in Gaussian variables Latała [15], we get that there exists a constant $C(r)$ such that for any $t\geq 0$ and polynomial $\mathcal{P}$ of degree r as above

$$\mathbb{P}\left[\mathcal{P}(G_1,\ldots,G_{|\overline{I}|})\geq t\right] \;\geq\; \frac{1}{C(r)}\exp\left(-C(r)\cdot t^2\right) \tag{B.3}$$

We now to upper bound the probability that $\mathcal{P}(X_1, \ldots, X_{|\overline{I}|}) > t$. In the calculation below, for $x \in \{0,1\}^{\overline{I}}$, we abuse the notation $\mathcal{P}(x)$ to denote $\mathcal{P}((-1)^{x_1}, \ldots, (-1)^{x_{|\overline{I}|}})$. We then want to bound the probability over $x \in \{0,1\}^{\overline{I}}$ that $\mathcal{P}(x) \geq t$.

$$\begin{aligned}
\mathbb{P}\left[\mathcal{P}(x) \geq t\right] \;=\; \mathbb{P}\left[\sqrt{\nu} \cdot \mathcal{P}(x) \geq t \cdot \sqrt{\nu}\right] \leq \;& \mathbb{P}\left[\left|F(x) - \sqrt{\nu} \cdot \mathcal{P}(x)\right| \geq t\sqrt{\nu} - 1\right] \\
\leq \;& \frac{\|F - \nu \cdot \mathcal{P}\|_2^2}{(t \cdot \sqrt{\nu} - 1)^2} \\
= \;& \frac{\sum_{|\beta| \geq r} \hat{F}(\beta)^2}{(t \cdot \sqrt{\nu} - 1)^2} \;\leq\; \frac{2^{|I|} \cdot \theta}{(t \cdot \sqrt{\nu} - 1)^2} \qquad \text{(B.4)}
\end{aligned}$$

Combining the bounds from equations (B.2), (B.3) and (B.4), and substituting $t = 2/\sqrt{\nu}$ gives

$$\frac{1}{C(r)} \cdot \exp\left(-2 \cdot C(r)/\nu\right) - 2^{|I|} \cdot \theta \;\leq\; O\left(r \cdot \left(2^{|I|} \cdot \theta/\nu\right)^{1/(8(r-1))}\right).$$

If $\nu \leq \sqrt{2^{|I|} \cdot \theta}$, then we are done. Else, the above inequality gives

$$\frac{1}{C(r)} \cdot \exp\left(-2 \cdot C(r)/\nu\right) \;\leq\; O\left(r \cdot \left(2^{|I|} \cdot \theta\right)^{1/(16(r-1))}\right).$$

Since $|I| \leq 2^{4r}$, this implies $\nu \leq C'(r)/\log(1/\theta)$ for some constant $C'(r)$.

Finally, we get

$$\sum_{\alpha:\alpha \cap \overline{I} \neq \emptyset} \hat{f}(\alpha)^2 \;\leq\; \mathsf{Var}(F) \;=\; \sum_{0<|\beta|\leq r-1} \hat{F}(\beta)^2 + \sum_{|\beta| \geq r} \hat{F}(\beta)^2 \;\leq\; \frac{C'(r)}{\log(1/\theta)} + 2^{|I|} \cdot \theta \;\leq\; \frac{C_r}{\log(1/\theta)},$$

for some constant C_r. ■

Given the above lemma, we again take

$$h \;:=\; \sum_{\alpha \subseteq I(f)} \hat{f}(\alpha)\chi_\alpha \quad \text{and} \quad g \;:=\; \mathbb{1}_{\{h \geq 1/2\}}.$$

Again, g depends only on the variables in $I(f)$. Also, from Lemma B.2, we have that

$$\|f - h\|_2^2 \;=\; \sum_{\alpha:\alpha \cap \overline{I(f)} \neq \emptyset} \hat{f}(\alpha)^2 \;\leq\; C_r/\log(1/\theta).$$

Also, for each x, $|f(x) - g(x)| \leq 2 \cdot |f(x) - h(x)|$ and hence

$$\Delta(f,g) \;=\; \mathbb{P}\left[f \neq g\right] \;=\; \|f - g\|_2^2 \;\leq\; 4 \cdot C_r/\log(1/\theta).$$

Finally, we show $deg(g) \leq r - 1$ as before. If this is not the case then we must have a Fourier coefficient in g with degree at least r and magnitude at least $1/2^{|I(f)|}$. Since $\|f - g\|_2^2 \leq 4 \cdot C_r/\log(1/\theta)$, this Fourier coefficient in f must have magnitude at least $1/2^{|I(f)|} - \sqrt{4 \cdot C_r/\log(1/\theta)}$. Since the Fourier mass above level $r - 1$ is θ, we must have that

$$\frac{1}{2^{|I(f)|}} - \sqrt{\frac{4 \cdot C_r}{\log(1/\theta)}} \;\leq\; \sqrt{\theta}.$$

If θ is sufficiently small so that $\sqrt{4 \cdot C_r/\log(1/\theta)} \leq (1/2) \cdot (1/2^{2^{4r}}) \leq (1/2) \cdot (1/2^{|I(f)|})$ and $\sqrt{\theta} < (1/2) \cdot (1/2^{2^{4r}}) \leq (1/2) \cdot (1/2^{|I(f)|})$, then this leads to a contradiction. ■

ECCC ISSN 1433-8092
http://eccc.hpi-web.de