On Non-Approximability for Quadratic Programs

Sanjeev Arora *
Computer Science Department
Princeton University

Eli Berger
Institute for Advanced Study
Princeton

Elad Hazan
Computer Science Department
Princeton University

Guy Kindler †
Institute for Advanced Study
Princeton

Muli Safra
Institute for Advanced Study
Princeton

Abstract

This paper studies the computational complexity of the following type of quadratic programs: given an arbitrary matrix whose diagonal elements are zero, find $x \in \{-1, 1\}^n$ *that maximizes* $x^T Mx$*. This problem recently attracted attention due to its application in various clustering settings, as well as an intriguing connection to the famous* Grothendieck inequality*. It is approximable to within a factor of* $O(\log n)$*, and known to be NP-hard to approximate within any factor better than* $13/11 - \epsilon$ *for all* $\epsilon > 0$*. We show that it is quasi-NP-hard to approximate to a factor better than* $O(\log^\gamma n)$ *for some* $\gamma > 0$*.*

The integrality gap of the natural semidefinite relaxation for this problem is known as the Grothendieck constant *of the complete graph, and known to be* $\Theta(\log n)$*. The proof of this fact was* nonconstructive*, and did not yield an explicit problem instance where this integrality gap is achieved. Our techniques yield an explicit instance for which the integrality gap is* $\Omega(\frac{\log n}{\log \log n})$*, essentially answering one of the open problems of Alon et al. [AMMN].*

1. Introduction

This paper deals with the following class of quadratic programs, henceforth denoted MAXQP. For a square matrix M over the reals where all diagonal entries are zero, the quadratic program is given by

$$\begin{array}{ll} \text{Maximize} & x^T Mx \\ \text{Subject to} & x_i \in \{-1, 1\} \qquad \forall i \in [n] \end{array}$$

This subcase of quadratic programming has attracted a lot of attention recently thanks to a surprising web of connections. To begin with, it is an attractive subcase, being a generalization of problems such as MAX-CUT, in which the constraints involve *pairs* of vertices. In addition, the obvious generalization of the seminal MAX-CUT algorithm of Goemans and Williamson fails for this problem — the negative entries of M cause problems for the GW rounding algorithm. One would hope that investigating this problem would lead to new techniques for analyzing SDP relaxations for other problems.

Also, this variant seems to capture the essential difficulty of a natural optimization problem called *correlation clustering* introduced by Bansal, Blum, and Chawla [BBC], which was the motivation for its study in Charikar and Wirth [CW04]. (It is also studied in physics in context of *spin glass models*, see [Tal03]).

Finally, the integrality gap of the obvious SDP relaxation of this problem seems related to questions studied in analysis. In particular, the famous *Grothendieck's inequality* implies an $O(1)$-approximation to the *bipartite* case of this problem where the objective is $x^T My$ and x, y are vectors in $\{-1, 1\}^n$. This was pointed out by Alon and Naor [AN04], who gave an algorithmic version of Grothendieck's inequality (in other words, a *rounding algorithm* for the obvious SDP relaxation). They used this algorithm to derive an $O(1)$-

*Supported by David and Lucile Packard Fellowship, NSF grants CCR 0205594, CCR 0098180 and CCR 0514993.

†Supported by CCR grant $\mathcal{N}$CCR-0324906 and $\mathcal{N}$DMS-0111298.

approximation to the *cut norm* of a matrix, which plays an important role in approximation algorithms for dense graph problems [FK99].

Motivated by the Goemans-Williamson work, Nesterov [Nes98], and following him Nemirovskii et al. [NRT99] and Megretski [Meg01], obtained $O(\log n)$-approximations to MAXQP. This algorithm was later rediscovered in the clustering context by Charikar and Wirth, who also pointed that the known hardness results for MAX-CUT implied that $13/11 - \epsilon$ approximation is NP-hard. They raised the obvious question, whether the approximation ratio can be improved from $\log n$ to $O(1)$.

1.1. Our results

In this paper we show that unless $P = NP$, the approximation factor of MAXQP cannot be made constant, and prove the following:

Theorem 1. *There exists a constant $\gamma > 0$ such that unless $NP \subseteq DTIME(n^{\log^3 n})$, MAXQP cannot be approximated in polynomial time up to a factor smaller then $O(\log^\gamma n)$.*

In fact, we show that the existence of sufficiently strong PCPs implies that achieving $O(\log n)$-approximation is also hard.

Independently, Khot and O'Donnell [KO] have proved that MAXQP cannot be approximated in polynomial time up to a factor smaller than $O(\log \log n)$, assuming Khot's *unique games conjecture* [Kho02]. However, their reduction generates matrices with a specialized structure, which is useful for proving lower bounds for different variants of the problem.

The second part of our work gives a better understanding of the standard SDP relaxation for the MAXQP problem, which is used both in the abovementioned $O(\log n)$-approximation, as well as in a formal study by Alon et al [AMMN] of the *Grothendieck constant* of a graph. The Grothendieck constant of an n-node graph $G = (V, E)$ is the maximum integrality gap of the above SDP among all matrices M whose entries are zero for all $\{i, j\}$ that are not edges in E. Alon et al. proved that this integrality gap, the Grothendieck constant, is $\Omega(\log n)$ for the complete graph. This improved upon Kashin and Szarek [KS03], who obtained a bound of $\Omega(\sqrt{\log n})$. However, both proofs are non-constructive, in the sense that they do not generate an explicit matrix for which the integrality gap is achieved. We essentially settle a question by Alon et al. and provide an explicit quadratic form for which the integrality gap is $\Omega(\frac{\log n}{\log \log n})$.

The rest of the paper is organized as follows. First we present a few definitions and previous results & conjectures in Section 2. In Section 3 we prove Theorem 1. Finally Section 4 contains the explicit construction of an instance that achieves integrality gap of $\Omega(\frac{\log n}{\log \log n})$.

2. Preliminaries

The MAXQP problem we consider is defined as follows

Definition 1 (MAXQP). *An instance of the MAXQP problem is a matrix $M \in \mathbb{R}^{n \times n}$ with diagonal entries equal to zero and a set of variables $\{x_1, ..., x_n\}$. The objective is to find an assignment $f : \{x_i\} \mapsto \{-1, 1\}$ that maximizes the quadratic form $x^T M x$. The objective value of an instance M under assignment f is denoted by $M(f)$. The maximal objective value of an instance M is denoted $val(M) = \max_f M(f)$.*

The natural semi-definite relaxation for MAXQP is defined as

Definition 2 (MAXQP relaxed version). *Given a matrix $M \in \mathbb{R}^{n \times n}$ with diagonal entries equal to zero, assign unit vectors (i.e. vectors of l_2 norm 1) $v_i \in R^n$ such as to maximize the expression $\sum_{ij} M_{ij} \cdot \langle v_i, v_j \rangle$.*

A common starting point for our hardness results is the Label Cover problem defined below.

Definition 3. *The Label Cover problem $\mathcal{L}(V, W, E, [R], \{\sigma_{v,w}\}_{(v,w) \in E})$ is defined as follows. We are given a regular bipartite graph with left side vertices V, right side vertices W, and a set of edges E. In addition, for every edge $(v, w) \in E$ we are given a map $\sigma_{v,w} : [R] \to [R]$. A* labelling *of the instance is a function ℓ assigning one label to each vertex of the graph, namely $\ell \colon V \cup W \to [R]$. A labelling ℓ satisfies an edge (v, w) if*

$$\sigma_{v,w}(\ell(v)) = \ell(w) \ .$$

The value *of a Label Cover instance, denoted $val(\mathcal{L})$, is defined to be the maximum, over all labellings, of the fraction of edges satisfied.*

The PCP Theorem [AS98, ALM+98] combined with Raz's parallel repetition theorem [Raz98] yields the following theorem, which is used in the proof of Theorem 1

Theorem 2 (Quasi-NP-hardness). *There exists a constant $\gamma > 0$ so that for any language L in NP, any input w and any $R > 0$, one can construct a labeling instance $\mathcal{L}$, with $|w|^{O(\log R)}$ vertices, and label set of* *size R, so that: If $w \in L$, val($\mathcal{L}$) $= 1$ and otherwise val($\mathcal{L}$) $< R^{-\gamma}$. Furthermore, $\mathcal{L}$ can be constructed in time polynomial in its size.*

A better lower bound can be achieved if we assume a strengthened version of the above theorem. Specifically, the parameter γ in Theorem 2 translates directly to the γ of Theorem 1, and therefore a PCP with parameter $\gamma = 1$ would imply the optimal hardness of approximation ratio for MAXQP , namely $\Theta(\log n)$.

2.1. Notions from analysis

In this paper we consider properties of real-valued and Boolean-valued functions over Boolean variables. We consider functions $f\colon \{-1,1\}^n \mapsto \mathbb{R}$ and say a function is *Boolean-valued* if its range is $\{-1,1\}$. The domain $\{-1,1\}^n$ is implicitly equipped with the uniform probability measure, unless specified otherwise. This implies a natural inner product over the space of real-valued functions $f\colon \{-1,1\}^n \mapsto \mathbb{R}$, defined by $\langle f,g\rangle = \mathbb{E}[fg]$. The associated norm in this space is given by $\|f\|_2 = \sqrt{\mathbb{E}[f^2]}$. We also define the p-norm for every $1 \le p < \infty$, by $\|f\|_p = (E[|f|^p])^{1/p}$. In addition, let $\|f\|_\infty = \max\{|f(x)|\}$.

Fourier expansion. The character functions over $\{-1,1\}^n$ are defined as follows. For $S \subseteq [n]$, define χ_S by $\chi_S(x) = \prod_{i\in S} x_i$. It is well known that the set of all such functions (characters) forms an orthonormal basis for our inner product space and thus every function $f : \{-1,1\}^n \to \mathbb{R}$ can be uniquely expressed as

$$f = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S, \qquad (1)$$

and the coefficients above satisfy *Plancherel's identity*, namely $\langle f,g\rangle = \sum_S \hat{f}(S)\hat{g}(S)$. In particular, $\|f\|_2^2 = \sum_S \hat{f}(S)^2$. The right-hand side of (1) is called the *Fourier expansion* of f, and the coefficients $\hat{f}(S) = \langle f,\chi_S\rangle$ are called the *Fourier coefficients* of f. Note that if f is Boolean-valued then $\sum_S \hat{f}(S)^2 = 1$, and if $f : \{-1,1\}^n \to [-1,1]$ then $\sum_S \hat{f}(S)^2 \le 1$.

We refer to the Fourier coefficients $\{\hat{f}(S)$ s.t. $|S| = 1\}$ corresponding to the linear functions $\{\chi_i | i \in [n]\}$ as the *linear coefficients*.

We speak of f's squared Fourier coefficients as *weights*, and we speak of the sets S being stratified into *levels* according to $|S|$. So for example, by the *weight of f at level 1* we mean $\sum_{|S|=1} \hat{f}(S)^2$.
For a function f as above we denote its linear part by

$$f^{=1} = \sum_{S\subset[n],|S|=1} \hat{f}(S)\chi_S$$

and similarly its non-linear part by

$$f^{\neq 1} = \sum_{S\subseteq[n],|S|\neq 1} \hat{f}(S)\chi_S \ .$$

Vector functions. In the analysis of the integrality gap we consider functions $f\colon \{-1,1\}^n \mapsto \mathbb{S}^{d-1}$, i.e. functions that map into vectors of l_2 norm 1 (vectors that lie on the unit d-dimensional sphere). Such functions also have a Fourier representation

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

however the Fourier coefficients $\hat{f}(S)$ are now vectors. Consider the n "coordinate mappings" $f_i\colon \{-1,1\}^n \mapsto [-1,1]$, defined by $f_i(x) \stackrel{def}{=} (f(x))_i$ (i.e., the value of f_i at x is equal to the i'th coordinate of the vector $f(x)$). It is easy to see that the Fourier coefficients of f are vectors whose coordinates are the corresponding coefficients of the functions f_i. The coefficients of f are vectors of norm at most 1, that is, lie inside the unit d-dimensional ball $\hat{f}(S) \in \mathbb{B}^{d-1}$.

3. Hardness of QP

In this section we prove Theorem 1, showing MAXQP to be hard to approximate. The proof is by reduction from Label Cover to MAXQP, obtained by using the long code to encode assignments to a given label cover instance. An assignment over the long code variables is regarded as a Boolean function, and the objective value can easily be expressed in terms of the Fourier coefficients of these functions.

Our construction is somewhat similar to other inapproximability results, such as that for MAX-CUT [KKMO] and for SPARSEST-CUT [CKK+05, KV]. The techniques in these results, however, are limited to proving gaps of $O(\log\log n)$ (technically, this arises from the tightness of Bourgain's theorem from Fourier analysis) [Bou02]. The key for achieving polylogarithmic in-approximability factors for MAXQP is the option of using negative coefficients in instances of MAXQP. This allows for long-code tests that incur severe penalty on certain illegal codewords. Specifically, looking at long-code words as Boolean functions, we use negative entries in the MAXQP matrix to significantly reduce the value obtained by our long-code test when applied to words that have any non-linear Fourier coefficients.

3.1. The reduction

Given an instance of label cover $\mathcal{L} = \mathcal{L}(V, W, E, [R], \{\sigma_{v,w}\}_{(v,w)\in E})$, we describe a reduction which constructs an instance of MaxQP denoted $M_{\mathcal{L}}$. The diagonal entries of our initial construction will not be zero, however in Subsection 3.4 we eliminate all non-zero diagonal entries in $M_{\mathcal{L}}$.

Parameters. Let $\mathcal{L} = \mathcal{L}(V, W, E, [R], \{\sigma_{v,w}\}_{(v,w)\in E})$ be an instance of label cover, where the size of the instance is $n = |V| + |W|$, and $R < \log^{10} n$. The reduction uses three parameters, ν, b, and d, which are set by

$$\nu \stackrel{def}{=} \frac{1}{2n}, \quad \text{and} \quad b, d \stackrel{def}{=} e^{10R} + 4\nu^{-6} \ .$$

The variables. For every vertex $u \in V \cup W$ of the original instance $\mathcal{L}$, the reduction generates d sets of new variables, denoted $\left\{C_u^i\right\}_{i\in[d]}$. (The reader may prefer to think of d as 1 in first reading. We later need to make many "copies" of each variable by setting d to a larger value, in order to set the diagonal elements of the generated instance to zero). Each set C_u^i will correspond to an encoding of the assignment to u, and will contain on variable $C_u^i(x) \in C_u^i$ for every element $x \in \{-1, 1\}^R$ of the R-dimensional discrete hypercube. The MaxQP instance $M_{\mathcal{L}}$ will therefore be defined over $N \stackrel{def}{=} d(|V| + |W|)2^R$ variables.

The quadratic form. When restricted to a subset C_u^i, an assignment f to the variables of the MaxQP instance can be viewed as a Boolean function f_u^i, defined by $f_u^i(x) \stackrel{def}{=} f(C_u^i(x))$. Let $f_u \stackrel{def}{=} E_{i\in[d]}[f_u^i]$. We write our quadratic form as a convex combination of bilinear forms, defined over the functions f_u^i. We have two kinds of forms: the *internal* and *external* forms.

- **Internal Forms.** For every $u \in V \cup W$ and every $i, j \in [d]$ we write

$$T_{u,i,j}(f) \stackrel{def}{=} -b \sum_{S\subseteq[R],\ |S|\neq 1} \widehat{f_u^i}(S)\widehat{f_u^j}(S) \ .$$

In addition, let

$$\begin{aligned} T_u(f) &= \mathbb{E}_{i,j\in[d]}[T_{u,i,j}(f)] \\ &= -b \sum_{S\subseteq[R],\ |S|\neq 1} \widehat{f_u}^2(S). \end{aligned}$$

Note that because of the large value of b, even tiny non-zero non-linear coefficients of f_u make $T_u(f)$ become very negative.

- **External Forms.** For every edge $(v, w) \in E$ and every $i, j \in [d]$ we write

$$T_{v,w,i,j}(f) = \sum_{k\in[R]} \widehat{f_v^i}(\{k\})\widehat{f_w^j}(\{\sigma_{v,w}(k)\}) \ ,$$

and let

$$\begin{aligned} T_{vw}(f) &= \mathbb{E}_{i,j\in[d]}[T_{v,w,i,j}(f)] \\ &= \textstyle\sum_{k\in[R]} \widehat{f_v}(\{k\})\widehat{f_w}(\{\sigma_{v,w}(k)\}). \end{aligned}$$

Our MaxQP instance is given by the following quadratic form.

$$\begin{aligned} M_{\mathcal{L}}(f) \stackrel{def}{=} &\ \nu \mathbb{E}_{u\in V\cup W}[T_u(f)] \\ &+ (1-\nu)\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \end{aligned} \qquad (2)$$

This concludes the description of our reduction except for a small modification, discussed in Subsection 3.4, needed to obtain zero diagonal entries. In the next two subsections we proceed in proving completeness and soundness properties for the reduction (Lemma 1 and Lemma 2 respectively). We then show in Subsection 3.4 that removing diagonal entries does not change the properties of $M_{\mathcal{L}}$ significantly, and finally in Subsection 3.5 we conclude the proof of Theorem 1.

3.2. Completeness

Let $\mathcal{L}$ and $M_{\mathcal{L}}$ be as above. Recall that the value of $\mathcal{L}$ is the maximal fraction of edges that can be satisfied by a labelling, and that the value of $M_{\mathcal{L}}$, $val(M_{\mathcal{L}})$, is the maximal value that the quadratic form can obtain by a Boolean assignment. The following lemma states that the value of $\mathcal{L}$ is a lower bound for the value of $M_{\mathcal{L}}$.

Lemma 1. *If $val(\mathcal{L}) \geq 1 - \varepsilon$, then $val(M_{\mathcal{L}}) \geq (1 - \varepsilon)(1 - \nu)$.*

Proof: According to the assumption, $\mathcal{L}$ has some labelling $l : V \cup W \to [R]$ satisfying at least $1 - \varepsilon$ of its constraints. We define an assignment f for the MaxQP instance by $f_u^i(x) \stackrel{def}{=} x_{l(u)}$.

The Fourier coefficients of f_u^i are $\widehat{f_u^i}(\{l(u)\}) = 1$, and $\widehat{f_u^i}(S) = 0$ whenever $S \neq \{l(u)\}$. Hence for every $u \in V \cup W$ and $i, j \in [d]$ we have $T_{u,i,j}(f) = 0$, and therefore $T_u = 0$. Next, let $(v, w) \in E$ and $i, j \in [d]$. If the edge (v, w) is satisfied by the labelling, namely $\sigma_{vw}(l(v)) = l(w)$ (this is true for at least a $(1 - \varepsilon)$-fraction of the edges), then

$$T_{v,w,i,j}(f) = \sum_{k\in[R]} \delta_{k,l(v)}\delta_{\sigma_{vu}(k),l(w)} = 1.$$

If the edge (v,w) is not satisfied by the labelling then the expression above yields 0. Hence the overall value of the MAXQP instance is

$$\begin{aligned} M_{\mathcal{L}}(f) &= \nu\mathbb{E}_{u\in V\cup W}[T_u(f)] \\ &+ (1-\nu)\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \\ &\geq (1-\nu)(1-\varepsilon)\ . \end{aligned}$$

□

3.3. Soundness

The following lemma states the soundness property of the reduction.

Lemma 2. *If $M_{\mathcal{L}}(f) \geq \varepsilon \geq \frac{1}{R^2}$ for an assignment f, then there exists a labelling for $\mathcal{L}$ which satisfies at least an $\Omega(\varepsilon)$-fraction of the edges.*

Proof. Consider any assignment with $M_{\mathcal{L}}(f) \geq \varepsilon$. As a first step, we show that the functions f_u induced by such an assignment are extremely close to being linear functions.

Claim 1. *For all vertices $u \in V \cup W$ it holds that $\|T_v(f)\|_2^2 \leq \frac{1}{\sqrt{b}}$.*

PROOF: Note that, being averages of Boolean functions, the functions f_u take values in $[-1,1]$. Their L_2 norm is thus bounded by 1. In particular, their Fourier coefficients are each bounded by 1 in absolute value.

According to the construction, the absolute value of every external form T_{vw} is bounded by:

$$\begin{aligned} |T_{vw}(f)| &= \left|\mathbb{E}_{i,j\in[d]}[T_{vw}(i,j)]\right| \\ &= \sum_{k=1}^{R}\left|\widehat{f}_v(\{k\})\right|\left|\widehat{f}_w(\{\{\sigma_{v,w}(k)\}\})\right| \\ &\leq R \end{aligned}$$

For an internal form T_v we have

$$T_v(f) = -b\sum_{|S|\neq 1}\widehat{f}_v(S)^2 = -b\|f_v^{\neq 1}\|_2^2\ ,$$

By equation 2 and the assumption $M_{\mathcal{L}}(f) \geq \varepsilon$ we have:

$$\begin{gathered} \varepsilon \leq M_{\mathcal{L}}(f) \\ = \nu\mathbb{E}_{u\in V\cup W}[T_u(f)] + (1-\nu)\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \\ \leq -\nu b\mathbb{E}_{u\in V\cup W}[\|f_u^{\neq 1}\|_2^2] + R \end{gathered}$$

Which implies $\mathbb{E}_{u\in V\cup W}[\|f_u^{\neq 1}\|_2^2 \leq \frac{R-\varepsilon}{\nu b} \leq \frac{R}{2\nu b}$. Now suppose that there exists a u such that $\|f_u^{\neq 1}\|_2^2 > \frac{1}{\sqrt{b}}$. This implies:

$$\begin{aligned} E_{u\in V\cup W}\left[\|f_u^{\neq 1}\|_2^2\right] &\geq \frac{1}{n}\left[1\cdot\frac{1}{\sqrt{b}} + (n-1)\cdot 0\right] \\ &= \frac{1}{n\sqrt{b}} > \frac{R}{2\nu b} \end{aligned}$$

In contradiction to the previous conclusion.
□

Claim 2. *For all vertices $u \in V \cup W$ it holds that $\sum_{k=1}^{R}|\widehat{f}_v(\{k\})| \leq 2$.*

PROOF: By the previous Lemma, $\|f_v^{\neq 1}\|_2^2 \leq \frac{1}{\sqrt{b}} \leq e^{-5R}$.

Now suppose that $\sum_{k=1}^{R}|\widehat{f}_v(\{k\})| > 2$. Since $f_v^{=1}$ is a linear function with coefficients $\{\widehat{f}_v(\{k\})|k\in[R]\}$, there exists a value $y \in \{+1,-1\}^R$ for which $f_v^{=1}(y) = \sum_{k=1}^{R}|\widehat{f}_v(\{k\})| > 2$. For this y we have $f_v^{\neq 1}(y) = f_v(y) - f_v^{=1}(y) \leq -1$. Therefore,

$$\|f_v^{\neq 1}\|_2^2 \geq 2^{-R}\ ,$$

and this is a contradiction. □

The following simple argument shows that the expected value of T_{vw} is large for the assignment f.

Claim 3. $\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \geq \frac{1}{2}\varepsilon$.

PROOF: We are assuming that $M_{\mathcal{L}}(f) = \nu\mathbb{E}_u[T_u(f)] + (1-\nu)\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \geq \varepsilon$. Note that $T_u(f) \leq 0$. Hence, $\mathbb{E}_{(v,w)\in E}[T_{vw}(f)] \geq \frac{\varepsilon}{1-\nu} \geq \frac{1}{2}\varepsilon$. □

Using the previous claims, we now define a random label assignment as follows. The assignment to every $v \in V\cup W$ is randomly and independently chosen to be k with probability $\frac{1}{2}|\widehat{f}_v(\{k\})|$ (the sum of these probabilities is at most one by Claim 2), and with probability $1-\frac{1}{2}\sum_k|\widehat{f}_v(\{k\})|$ we leave v un-assigned.

Let c_{vw} be an indicator random variable that is set to 1 if and only if the label assignment above satisfies the label-cover constraint on the edge (v,w).

The expected number of constraints satisfied by our assignment is:

$$\begin{gathered} \mathbb{E}_{(v,w)\in E(\mathcal{L})}[c_{v,w}] = \\ \mathbb{E}_{v,w}\left[\sum_{k\in[R]}\frac{1}{2}|\widehat{f}_v(\{k\})|\cdot\frac{1}{2}|\widehat{f}_w(\sigma_{vw}(k))|\right] \\ \geq \frac{1}{4}\mathbb{E}_{v,w}[\sum_{k\in[R]}\widehat{f}_v(\{k\})\widehat{f}_w(\sigma_{vw}(k))] \\ = \frac{1}{4}\mathbb{E}_{v,w}\left[T_{vw}\right] \geq \frac{1}{8}\varepsilon \end{gathered}$$

Where the last inequality is by Claim 3. This completes the proof of Lemma 2.

3.4. Removing the diagonal

The instance $M_{\mathcal{L}}$ constructed in the previous section has non-zero diagonal entries. However since we took care to have d "copies" of every set of variables, the interaction of any variable set C_u^i with itself (which occurs only in the terms T_v) is negligible. More formally, consider the MAXQP instance $B_{\mathcal{L}}$, that is obtained from $M_{\mathcal{L}}$ by removing all terms of the form $T_{u,i,i}(f)$.

Recall that $T_{u,i,j}(f) = -b\sum_{|S|\neq 1}\widehat{f_u^i}(S)\widehat{f_u^j}(S)$, and therefore

$$|T_{u,i,i}(f)| \leq b\sum_S \widehat{f_u^i}^2(S) = b\ .$$

Hence, for any specific assignment f, the difference in value of $M_{\mathcal{L}}$ and $B_{\mathcal{L}}$ is bounded by

$$\begin{aligned}|M_{\mathcal{L}}(f) - B_{\mathcal{L}}(f)| &\leq \frac{\nu}{|V\cup W|d^2}\sum_{u\in V\cup W} T_{u,i,i} \\ &\leq \frac{\nu b}{d^2} = \frac{\nu}{b} \leq e^{-10R}\end{aligned}$$

3.5. Concluding the hardness proofs

Theorem 1 now follows as simple corollary of Lemma 1 and Lemma 2.

Proof of Theorem 1. Given an instance of label cover $\mathcal{L}$ as in theorem 2, construct $M_{\mathcal{L}}$ as described above. The MAXQP instance has the following properties:

1. The size of the instance is $N = O(n^{\log R}\cdot 2^R)$.
2. By lemma 1, if there exists an assignment satisfying more than a $1-\varepsilon$ fraction of the equations of $\mathcal{L}$, then the value of the QP is at least $1-\varepsilon-o(\varepsilon)$.
3. By lemma 2, if the value of $M_{\mathcal{L}}$ is at least $\delta > \frac{1}{R^2}$, then there exists an assignment that satisfies $\Omega(\delta)$ of the constraints of $\mathcal{L}$.

Set $R = \log^2 n$. Suppose that we could approximate $val(M_{\mathcal{L}})$ in polynomial time to a factor better then $O(\log^\gamma N)$. Then if the best assignment for $\mathcal{L}$ satisfies fraction 1 of the equations, we can find a solution to the MAXQP instance of value $1/\log^\gamma(N) = \log^{-\gamma}(n^{\log R}2^R) = \Omega(\log^{-\gamma}(2^{\log^2 n})) = \Omega(R^{-\gamma})$.

On the other hand, if every assignment satisfies at most $R^{-\gamma}$ of the constraints, then any QP solution will have value at most $R^{-\gamma}$. Thus in time $poly(N) = n^{O(\log^2 n)}$ we can distinguish between the two cases of the label cover instance. By theorem 2, this implies $NP \subseteq DTIME(n^{\log^3 n})$. □

4. Explicit Integrality Gap

In this section we prove the following theorem, showing an explicit family of MAXQP instances with increasing integrality gap. Our construction was inspired by the recent embedding lower bound of Khot and Vishnoi [KV].

Theorem 3. *There exists a family of* MAXQP *instances of unbounded size, where the integrality gap of instances over n variables is $\Omega(\frac{\log n}{\log\log n})$.*

Notation. Fix any $n \in \mathbb{N}$. Let $\mathcal{F} \stackrel{def}{=} \{f | f : \{1,-1\}^n \mapsto \{1,-1\}\}$ be the set of all Boolean functions on n bits. Let $R \stackrel{def}{=} 2^n$ and $N \stackrel{\text{def}}{=} 2^R = 2^{2^n}$. For any $f\in\mathcal{F}$ and $T\subseteq[n]$, let $f\circ T\in\mathcal{F}$ denote the function defined by $f\circ T(x) = f(x\oplus T)$, where $x\oplus T$ denotes the vector obtained from x by flipping the value of x_k for every $k\in T$.

Let $f \sim_\eta f'$ the distribution on pairs of functions $f, f' \in \mathcal{F}$ where f is chosen uniformly at random and f' is obtained by flipping each value of f independently with probability η. Denote by $\rho \sim_\eta \{\pm 1\}^n$ the distribution on n-bit strings such that each entry is chosen independently to be -1 with probability η and 1 otherwise.

4.1. The construction

The construction makes use of three parameters that we fix as follows. Let $\nu = \frac{1}{R^2}$, $b = N^{10}$, $d = b^2 = N^{20}$

Variables. We generate an instance of QP, denoted M_n, over a set of variables V such that $|V| = n$. It will be more convenient for us to have more than one label for each variable. That is, we first define a quadratic form over a larger number of variables, and then identify some of them, thereby obtaining a form over a smaller number of variables each having more than one label. The initial set of variables is $V = \{\langle f,g,i\rangle | f,g\in\mathcal{F}, i\in[d]\}$. We define an equivalence relation over the variables by setting $\langle f, g\circ T, i\rangle \equiv \langle f\chi_T, g, i\rangle$ for every subset $T\subseteq[n]$, and identify all the variables that belong to the same equivalence class.

We partition the labels into disjoint sets by setting

$$V_{f,i} = \{\langle f,g,i\rangle | g\in\mathcal{F}\}.$$

Given an assignment A to the variables (whether a Boolean or a vector assignment), its restriction to $V_{f,i}$ can be viewed as a function over $\mathcal{F}$. We denote this function by A_f^i.

The quadratic form. The final quadratic form is a convex combination of bilinear forms over the functions A^i_f, which are defined in terms of their Fourier representation. As in the case of the hardness reduction, we have *internal forms* and *external forms*.

- **Internal Forms.** For every $f \in \mathcal{F}$ we let M_f be defined by

$$M_f(A) \stackrel{def}{=} \mathbb{E}_{i,j\in[d]}\left[-b\sum_{|\alpha|\neq 1}\widehat{A^i_f}(\alpha)\widehat{A^j_f}(\alpha)\right] .$$

  Note that if we define $A_f \stackrel{def}{=} \mathbb{E}_{i\in[d]}[A^i_f]$, then

$$M_f(A) = -b\sum_{|\alpha|\neq 1}\widehat{A_f}^2(\alpha) \ . \qquad (3)$$

- **External forms.** For every $f, f' \in \mathcal{F}$, let $M_{f,f'}$ be defined by

$$\begin{aligned} M_{f,f'}(A) &\stackrel{def}{=} \mathbb{E}_{i,j\in[d]}\left[\sum_{|\alpha|=1}\widehat{A^i_f}(\alpha)\widehat{A^j_{f'}}(\alpha)\right] \\ &= \sum_{|\alpha|=1}\widehat{A_f}(\alpha)\widehat{A_{f'}}(\alpha) \qquad (4)\end{aligned}$$

The final quadratic form is given by the following convex combination of the internal and external forms:

$$M_n(A) \stackrel{def}{=} \nu\cdot\mathbb{E}_{f\in\mathcal{F}}\left[M_f\right] + (1-\nu)\cdot\mathbb{E}_{f\sim_\eta f'}\left[M_{f,f'}\right]$$

We now state the main lemma of this section,

Lemma 3. *For every $0 < \eta < \frac{1}{2}$, for every large enough n, the* MaxQP *instance M_n satisfies the following properties:*

1. *For every Boolean assignment A, we have $M_n(A) \le \frac{1}{R^{\frac{\eta}{1-\eta}}}$*
2. *There exists a vector assignment A_v for which $M_n(A) \ge 1 - 2\eta - \nu$.*

Before we prove Lemma 3, let us show how it implies Theorem 3.

Proof:[Proof of Theorem 3.] The number of variables in the instance M_n is $\frac{N^2\cdot d}{R} = O(N^{22})$. According to Lemma 3, the integrality gap is $R^{\frac{\eta}{1-\eta}}\cdot(1-2\eta-\nu)$. Fix $\eta = \frac{1}{2} - \frac{1}{\log R}$, then the integrality gap becomes:

$$\begin{aligned} R^{\frac{\eta}{1-\eta}}\cdot(1-2\eta-\nu) &= \Omega(R^{1-\frac{2}{\log R}}\cdot(\frac{1}{\log R} - \frac{1}{R^2})) \\ &= \Omega(\frac{R}{\log R}) = \Omega(\frac{\log N}{\log\log N})\end{aligned}$$

□

4.2. Integral solution

In this subsection we prove the first part of Lemma 3.

Lemma 4. *For any Boolean assignment A, the value of the* MaxQP *instance M_n satisfies $M_n(A) \le \frac{1}{R^{\frac{\eta}{1-\eta}}}$.*

To prove this lemma, we start by examining a few properties of the Boolean functions $\{A_f | f \in \mathcal{F}\}$. The fact that every variable of the instance M_n has several labels implies a certain relationship between the Fourier coefficients of the functions A_f. This is formalized in the following claim.

Claim 4. *For any $T \subseteq [n]$ and any $f \in \mathcal{F}$ it holds $\forall x \subseteq [n] \ \ \widehat{A}_f(x) = \widehat{A}_{f\chi_T}(x \circ T)$.*

Proof:
Consider a certain function $f \in \mathcal{F}$ and subset $T \subseteq [n]$. Since for every function $g \in \mathcal{F}$ the vertices $\langle f, g\circ T, i\rangle \equiv \langle f\chi_T, g, i\rangle$ were identified, the assignment A must satisfy:

$$\forall g \ \ A_f(g\circ T) = A_{f\chi_T}(g)$$

Writing these equations in Fourier basis we have:

$$\forall g \ \ \sum_{x\subseteq[n]}\widehat{A}_f(x)\chi_x(g\circ T) = \sum_{x\subseteq[n]}\widehat{A}_{f\chi_T}(x)\chi_x(g)$$

If the values $\widehat{A}_f(x)$ are fixed, this system of linear equations has a possible solution $\forall x \ \ \widehat{A}_f(x) = \widehat{A}_{f\chi_T}(x\circ T)$. This is the only solution as the linear system of equations has full rank.
□

Another property of any assignment with $M_n(A) > 0$ is that each A_f is very close to being a linear function. The following two claims prove this fact for two different measures of distance - the l_1 and l_2 norms.

Claim 5. *For any assignment such that $M_n(A) > 0$ it holds that $\forall f \in \mathcal{F} \ \ \|A^{\neq 1}_f\|^2_2 \le \frac{1}{N^6}$*

Proof: By equations 3 and 4 we have

$$\begin{aligned} M_n(A) = \nu\mathbb{E}_f\left[-b\|A^{\neq 1}_f\|^2_2\right] \\ +(1-\nu)\mathbb{E}_{f\sim_\eta f'\in\mathcal{F}}\left[\textstyle\sum_{|\alpha|=1}\widehat{A}_f(\alpha)\widehat{A}_{f'}(\alpha)\right] \qquad (5)\end{aligned}$$

Assuming $M_n(A) > 0$ this translates to

$$\begin{aligned} 0 < M_n(A) &\le 1 - \nu - b\mathbb{E}_f\left[\|A^{\neq 1}_f\|^2_2\right] \\ &\le 2 - \nu b\mathbb{E}_f\left[\|A^{\neq 1}_f\|^2_2\right]\end{aligned}$$

According to the choice of parameters we obtain

$$\mathbb{E}_f\left[\|A_f^{\neq 1}\|_2^2\right] \leq \frac{2}{\nu b} < \frac{1}{N^8}$$

Now suppose that there exists an f such that $\|A_f^{\neq 1}\|_2^2 > \frac{1}{N^6}$. This implies:

$$\mathbb{E}_f\left[\|A_f^{\neq 1}\|_2^2\right] \geq \frac{1}{N}\left[1 \cdot \frac{1}{N^6} + (N-1)\cdot 0\right] > \frac{1}{N^8}$$

In contradiction to the previous conclusion.
□

Claim 6.

$$\forall f \in \mathcal{F} \quad \sum_{x \subseteq [n]} |A_f(x)| \leq 2$$

Proof: By Claim 5, for every $f \in \mathcal{F}$ we have $\|A_f^{\neq 1}\|_2^2 < \frac{1}{N^8} \leq e^{-4R}$.

The rest of the argument is the same as in claim 2.
□

We can now proceed with the proof of Lemma 4.

Proof:[Lemma 4] Consider any Boolean assignment A to the variables of M_n. Suppose that $M_n(A) > 0$ (the assignment that achieves the maximum over $M_n(A)$ definitely satisfies this). From equation 5 we have

$$M_n(A) \leq \mathbb{E}_{f \sim_\eta f' \in \mathcal{F}}[\sum_{x \subseteq [n]} \widehat{A}_f(x)\widehat{A}_{f'}(x)]$$

Using the assignment A, we proceed to define a function $\Phi = \Phi_A : \mathcal{F} \mapsto [R]$ as follows. For every set of the form $\{f\chi_S | S \subseteq [n]\}$ we pick an arbitrary representative f, and set $\Phi(f)$ to be x with probability $\frac{1}{2}|\widehat{A}_f(x)|$, and with probability $1 - \frac{1}{2}\sum_x |\widehat{A}_f(x)|$ we set it arbitrarily to zero (note that by Claim 6 the above probabilities are indeed non-negative and sum up to 1).

Once $\Phi(f)$ has been set, we let $\Phi(f\chi_T) \stackrel{def}{=} \Phi(f) \oplus T$ for every $T \subseteq [n]$. Note that the resulting function Φ must be balanced, that is, it must satisfy

$$\forall x \subseteq [n] \ . \ \Pr_{f \in \mathcal{F}}[\Phi(f) = x] = \frac{1}{R} \tag{6}$$

This implies the following bound on the stability of Φ_A, attributed to O'Donnell [O'D04] and communicated to us by Khot [Kho] (see proof in the appendix).

Lemma 5 (O'Donnell). *For any function* $\Psi : \{\pm 1\}^t \mapsto [R]$ *such that* $\forall i \in [R]$. $\Pr_{x \in \{\pm 1\}^t}[\Psi(x) = i] = \frac{1}{R}$ *it holds that:*

$$\Pr_{x \sim_\eta x' \in \{\pm 1\}^t}[\Psi(x) = \Psi(x')] \leq \frac{1}{R^{\frac{\eta}{1-\eta}}} \ .$$

In the following claim we show how Φ_A can be used to bound $M_n(A)$.

Claim 7.

$$M_n(A) \leq 4 \Pr_{f \sim_\eta f' \in \mathcal{F}}[\Phi(f) = \Phi(f')] + \nu$$

Proof: Let $N(f) = \{f\chi_T | T \subseteq [x]\}$, and $\Pr[f, f'] = \Pr_{\rho \sim_\eta \{\pm 1\}^R}[f' = f\rho]$ be the probability of obtaining f' from f under η noise. Let $I(f, f')$ be the indicator random variable that is 1 if and only if $\Phi(f) = \Phi(f')$. By definition of Φ we have:

$$\begin{aligned}
&\Pr_{f \sim_\eta f' \in \mathcal{F}}[\Phi(f) = \Phi(f')] \\
&= \frac{1}{|F|^2} \sum_{f, f' \in \mathcal{F}} \Pr[f, f'] \cdot I(f, f') \\
&= \frac{1}{|F|^2} \sum_{f \in \mathcal{F}} \Bigg(\sum_{f' \notin N(f)} \Pr[f, f'] I(f, f') \\
&+ \sum_{T \subseteq [x]} \Pr[f, f\chi_T] I(f, f') \Bigg) \\
&\geq \frac{1}{|F|^2} \sum_{f \in \mathcal{F}} \sum_{f' \notin N(f)} \Pr[f, f'] I(f, f')
\end{aligned}$$

Now since for $f \notin N(f)$ the values of $\Phi(f)$ and $\Phi(f')$ are independent

$$\begin{aligned}
&= \frac{1}{|F|^2} \sum_{f \in \mathcal{F}} \sum_{f' \notin N(f)} \Pr[f, f'] \sum_{x \subseteq [n]} \frac{1}{4} |\widehat{A}_f(x)||\widehat{A}_{f'}(x)| \\
&= \frac{1}{4|F|^2} \Bigg(\sum_{f, f' \in \mathcal{F}} \Pr[f, f'] \sum_{x \subseteq [n]} |\widehat{A}_f(x)||\widehat{A}_{f'}(x)| \\
&- \sum_{f \in \mathcal{F}} \sum_{T \subseteq [n]} \Pr[f, f\chi_T] \sum_{x \subseteq [n]} |\widehat{A}_f(x)||\widehat{A}_{f\chi_T}(x)| \Bigg) \\
&\geq \frac{1}{4} M_n(A) - \frac{2R}{N} \geq \frac{1}{4} M_n(A) - \frac{1}{4}\nu
\end{aligned}$$

□

This lemma together with the bound of Lemma 5 yields

$$M_n(A) \leq \frac{4}{R^{\frac{\eta}{1-\eta}}} + \nu = O(\frac{1}{R^{\frac{\eta}{1-\eta}}}) \ , \tag{7}$$

Proving Lemma (4).
□

4.3. Vector solution

Consider the vector assignment, given by the Fourier coefficients:

$$\forall \alpha, i \quad \widehat{A}^i_f(\alpha) = \begin{cases} \frac{1}{R} f\chi_T & |\alpha| = 1, \alpha = T \subseteq [n] \\ 0 & o/w \end{cases}$$

As for the integral solution, let $A_f \stackrel{def}{=} \mathbb{E}_{i \in [d]}[A^i_f]$. Since $\forall i, j \,.\, \widehat{A}^i_f = \widehat{A}^j_f$, we have $\forall i \,.\, \widehat{A}_f = \widehat{A}^i_f$. Notice that the vectors $\widehat{A}_f(\alpha)$ are orthogonal, and their norms satisfy

$$\forall \alpha \quad \|\widehat{A}_f(\alpha)\|_2 = \begin{cases} \frac{1}{R} & |\alpha| = 1 \\ 0 & o/w \end{cases} \tag{8}$$

In the standard basis, these vectors can be written as:

$$A_f(g) = A^i_f(g) = \frac{1}{R} \sum_{T \subseteq [n]} g(1 \circ T) \cdot f\chi_T$$

Khot and Vishnoi [KV] observed that the above vector assignment assigns the same vector to all vertices in the equivalence classes $\{f\chi_T | T \subseteq [n]\}$.

Lemma 6. *For the vector solution above we have* $M_n(A) \geq 1 - 2\eta - \nu$.

PROOF: Recall by equation 5

$$\begin{aligned} M_n(A) &= \nu \mathbb{E}_f \left[-b \|A^{\neq 1}_f\|_2^2 \right] \\ &+ (1-\nu)\mathbb{E}_{f \sim_\eta f' \in \mathcal{F}} \left[\sum_\alpha \widehat{A}_f(\alpha) \widehat{A}_{f'}(\alpha) \right] \\ &\quad \text{by equation 8} \\ &= (1-\nu)\mathbb{E}_{f \sim_\eta f' \in \mathcal{F}} \left[\sum_\alpha \widehat{A}_f(\alpha) \widehat{A}_{f'}(\alpha) \right] \\ &= (1-\nu)\mathbb{E}_{f \sim_\eta f' \in \mathcal{F}} \left[\sum_{T \subseteq [n]} \frac{1}{R} f\chi_T \cdot \frac{1}{R} f'\chi_T \right] \\ &= (1-\nu)\mathbb{E}_{f \sim_\eta f' \in \mathcal{F}} \left[\frac{1}{R} \langle f, f' \rangle \right] \\ &= (1-\nu)(1-2\eta) \geq 1 - 2\eta - \nu \end{aligned}$$

□

Lemma 3 now follows from Lemmas 4 and 6.

4.4. Removing the diagonal

As the parameter d is much larger b, we can apply a modification very similar to the corresponding modification in the hardness of approximation result (Subsection 3.4), to obtain a matrix with zero diagonal entries. The details are omitted for brevity.

References

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and intractability of approximation problems. *Journal of the ACM*, 45:501–555, 1998.

[AMMN] Noga Alon, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. Quadratic forms on graphs. *To appear STOC 2005.*

[AN04] Noga Alon and Assaf Naor. Approximating the cut-norm via grothendieck's inequality. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 72–80, New York, NY, USA, 2004. ACM Press.

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45:70–122, 1998.

[BBC] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. *Mach. Learn.*, 56(1-3):89–113.

[Bou02] J. Bourgain. On the distribution of the fourier spectrum of boolean functions. *Israel Journal of Mathematics*, 131:269–276, 2002.

[CKK+05] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. In *manuscript*, 2005.

[CW04] Moses Charikar and Anthony Wirth. Maximizing quadratic programs: Extending grothendieck's inequality. In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04)*, pages 54–60, Washington, DC, USA, 2004. IEEE Computer Society.

[FK99] A. M. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19:175–200, 1999.

[Kho] Subhash Khot. personal communications, march 2005.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775, New York, NY, USA, 2002. ACM Press.

[KKMO] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan ODonnell. Optimal inapproximability results for max-cut and other 2-variable csps? In *FOCS 2004*.

[KO] S. Khot and R. O'Donnell. personal communications, march 2005.

[KS03] B. S. Kashin and S. J. Szarek. On the gram matrices of systems of uniformly bounded functions. *Proceedings of the Steklov Institute of Mathematics*, 243:227–233, 2003.

[KV] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. *to appear FOCS 2005.*

[Meg01] A. Megretski. Relaxation of quadratic programs in operator theory and system analysis. *Systems, Approximation, Singular Integral Operators, and Related Topics (Bordeaux, 2000)*, (3):365–392, 2001.

[Nes98] Y. Nesterov. Global quadratic optimization via conic relaxation. *Working paper CORE*, 1998.

[NRT99] A. Nemirovski, C. Roos, and T. Terlaky. On maximization of quadratic form over intersection of ellipsoids with common center. *Mathematical Programming*, 86(3):463–473, 1999.

[O'D04] R. O'Donnell. P.hd thesis. In *MIT*, 2004.

[Raz98] R. Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998.

[Tal03] Michel Talagrand. *Spin Glasses: a Challenge to Mathematicians*, volume 46 of *Ergbnisse der Mathematik und ihrer Grenzgebiete*. New York, 2003.

A. Stability of balanced multi-valued functions

For completeness, we provide the proof O'Donnell's Lemma [O'D04]:

PROOF:[Lemma 5] Given Ψ, define $\forall j \in [R] \quad \Phi_j(x) : \{\pm 1\}^t \mapsto \{0,1\}$ as follows:

$$\Phi_j(x) \stackrel{def}{=} \begin{cases} 1 & \Psi(x) = j \\ 0 & o/w \end{cases}$$

Then:

$$\begin{aligned}
& \Pr_{x \sim_\eta x' \in \{\pm 1\}^t}[\Psi(x) = \Psi(x') = j] \\
= & \Pr_{x \sim_\eta x' \in D}[\Phi_j(x) = \Phi_j(x') = 1] \\
= & \mathbb{E}_{x \sim_\eta x' \in \{\pm 1\}^t}[\Phi_j(x)\Phi_j(x')] \\
= & \mathbb{E}_{x,\rho}[(\sum_{\alpha \subseteq [t]} \widehat{\Phi_j}(\alpha)\chi_\alpha(x))(\sum_{\beta \subseteq [t]} \widehat{\Phi_j}(\beta)\chi_\beta(x \oplus \rho))] \\
= & \sum_{\alpha \subseteq [t]} \widehat{\Phi_j}^2(\alpha)\mathbb{E}_{\rho \sim_\eta \{\pm\}^t}[\chi_\alpha(\rho)))] \\
= & \sum_{\alpha \subseteq [t]} \widehat{\Phi_j}^2(\alpha)(1-2\eta)^{|\alpha|} \\
= & \|T_{\sqrt{1-2\eta}}[\Phi_j]\|_2^2
\end{aligned}$$

Where $T_\delta[f]$ is the Beckner operator:

$$T_\delta[f] = \sum_S \delta^{|S|}\widehat{f}(S)\chi_S$$

Now using the Beckner inequality (which states $\|T_\delta[f]\|_p \leq \|f\|_r$ for $r \leq p$, $\delta \leq \sqrt{\frac{r-1}{p-1}}$):

$$\begin{aligned}
& \Pr_{x \sim_\eta x' \in \{\pm 1\}^t}[\Psi(x) = \Psi(x') = j] \\
& = \|T_{\sqrt{1-2\eta}}[\Phi_j]\|_2^2 \\
& \leq \|\Phi_j\|_{2-2\eta}^2 && \text{using Beckner} \\
& = \mathbb{E}_{x \in \{\pm 1\}^t}[\Phi_j(x)^{2-2\eta}]^{2/2-2\eta} \\
& = (\tfrac{1}{R})^{\frac{1}{1-\eta}} && \text{by properties of } \Phi_j
\end{aligned}$$

Therefore:

$$\begin{aligned}
& \Pr_{x \sim_\eta x' \in \{\pm 1\}^t}[\Psi(x) = \Psi(x')] \\
= & \sum_{j \in [R]} \Pr_{x \sim_\eta x' \in \{\pm 1\}^t}[\Psi(x) = \Psi(x') = j] \\
\leq & R \cdot \frac{1}{R^{1/1-\eta}} = \frac{1}{R^{\frac{\eta}{1-\eta}}}
\end{aligned}$$

□