Separating the NP-Hardness of the Grothendieck problem from the Little-Grothendieck problem

Abstract

Grothendieck's inequality [Gro53] states that there is an absolute constant $K > 1$ such that for any $n \times n$ matrix A

$$\|A\|_{\infty\to 1} \;:=\; \max_{s,t\in\{\pm 1\}^n} \sum_{i,j} A[i,j]\cdot s(i)\cdot t(j) \;\geq\; \frac{1}{K}\cdot \max_{u_i,v_j\,\in\,\mathbb{S}^{n-1}} \sum_{i,j} A[i,j]\cdot \langle u_i,\, v_j\rangle.$$

In addition to having a tremendous impact on Banach space theory, this inequality has found applications in several unrelated fields like quantum information, regularity partitioning, communication complexity, etc. Let K_G (known as Grothendieck's constant) denote the smallest constant K above. Grothendieck's inequality implies that a natural semidefinite programming relaxation obtains a constant factor approximation to $\|A\|_{\infty\to 1}$. The exact value of K_G is yet unknown with the best lower bound $(1.67\dots)$ being due to Reeds and the best upper bound $(1.78\dots)$ being due to Braverman, Makarychev, Makarychev and Naor [BMMN13]. In contrast, the little Grothendieck inequality states that under the assumption that A is PSD the constant K above can be improved to $\pi/2$ and moreover this is tight.

The inapproximability of $\|A\|_{\infty\to 1}$ has been studied in several papers culminating in a tight UGC-based hardness result due to Raghavendra and Steurer (remarkably they achieve this without knowing the value of K_G). Briet, Regev and Saket [BRS15] proved tight NP-hardness of approximating the little Grothendieck problem within $\pi/2$, based on a framework by Guruswami, Raghavendra, Saket and Wu [GRSW16] for bypassing UGC for geometric problems. This also remained the best known NP-hardness for the general Grothendieck problem due to the nature of the Guruswami et al. framework, which utilized a projection operator onto the degree-1 Fourier coefficients of long code encodings, which naturally yielded a PSD matrix A.

We show how to extend the above framework to go beyond the degree-1 Fourier coefficients, using the *global* structure of optimal solutions to the Grothendieck problem. As a result, we obtain a separation between the NP-hardness results for the two problems, obtaining an inapproximability result for the Grothendieck problem, of a factor $\pi/2 + \varepsilon_0$ for a fixed constant $\varepsilon_0 > 0$.

1 Introduction

The Grothendieck inequality [Gro53] is a fundamental result from Banach space theory, which can be viewed from an optimization lens, as saying that the $\infty \to 1$ norm of a matrix A can be approximated using a vector relaxation i.e.,

$$\|A\|_{\infty\to 1} \;:=\; \max_{s,t\in\{\pm 1\}^n} \sum_{i,j} A[i,j]\cdot s(i)\cdot t(j) \;\geq\; \frac{1}{K}\cdot \max_{u_i,v_j\,\in\,\mathbb{S}^{n-1}} \sum_{i,j} A[i,j]\cdot \langle u_i,\, v_j\rangle.$$

The inequality has had a tremendous number of applications in a variety of areas including combinatorics, optimization, complexity theory, and quantum information theory. We refer the reader to the excellent surveys by Khot and Naor [KN12] and Pisier [Pis12] and the references therein, for an account of the rich history of the inequality, its variants, and their many connections and applications.

The problem of *computing* the $\infty \to 1$ norm of a given matrix A, which is the subject of the above inequality, is referred to as the Grothendieck problem. A long line of work has focused on determining the smallest constant K_G (known as Grothendieck's constant) achievable in the above inequality, or equivalently, the best approximation ratio for the Grothendieck problem, achieved by a natural semidefinite programming (SDP) relaxation. The best upper known bound on K_G is due to Braverman, Makarychev, Makarychev, and Naor [BMMN13] who proved that a previous bound of $\frac{\pi}{2\cdot(1+\sqrt{2})} \approx 1.782\ldots$ due to Krivine [Kri77] can be improved to $\frac{\pi}{2\cdot(1+\sqrt{2})} - \varepsilon_0$ for a fixed $\varepsilon_0 > 0$. The best lower bound $K_G \geq 1.6769\ldots$ was proved independently by Davie [Dav84] and Reeds [Ree91]. However, the true value of Grothendieck's constant is unknown, and determining it is an important open problem.

Approximability. From a computational perspective, a natural question to consider is the optimal approximation ratio achievable by *any* efficient algorithm, and not just the SDP relaxation. The first inapproximability result for the Grothendieck problem was obtained by Alon and Naor [AN04] (in an influential paper that established a connection to cut-norm and several combinatorial applications) by giving an approximation preserving reduction to MAX-CUT, which yields an NP-hardness of factor 17/16 via a result of Håstad [Hås01]. Assuming the Unique Games Conjecture (UGC), the best explict bound is by Khot and O'Donnell [KO09] who proved an inapproximability result matching the Davie-Reeds lower bound. A remarkable later result by Raghavendra and Steurer [RS09] proved that assuming the UGC, the approximation ratio by the semidefinite program is optimal i.e., they prove an inapproximability result within factor K_G, without having to know the true value of K_G!

The best known NP-hardness for the problem is by Briet, Regev and Saket [BRS15] who prove inapproximability within a factor of $\pi/2$. Prior NP-hardness results for the Grothendieck problem are all actually for a special subcase (known as the little Grothendieck problem), wherein the matrix A is required to be positive semidefinite (PSD). In this case, one can easily observe that the two vectors x, y achieving $\|A\|_{\infty\to 1}$ can be equal without loss of generality, since

$$\langle s, At\rangle \;=\; \langle A^{1/2}s, A^{1/2}t\rangle \;\leq\; \|A^{1/2}s\|_2\|A^{1/2}t\|_2 \;\leq\; \max\left\{\langle s, As\rangle, \langle t, At\rangle\right\}\,.$$

Using the above observation, and taking A to be the Laplacian of a graph, shows that the problem captures MAX-CUT as a subcase (although the result of Alon and Naor [AN04] used a slightly different matrix). The result of Briet, Regev and Saket [BRS15] also shows the factor $\pi/2$ inapproximability for the little Grothendieck problem. Moreover, their result is tight for the little

Grothendieck problem, by a result of Rietz [Rie74] (see also Nesterov [Nes98]). Thus, any further improvements to the NP-hardness, will require separating it from the little Grothendieck problem.

Techniques for proving inapproximability. There is also a technical reason why current NP-hardness results do not separate the little Grothendieck problem from the Grothendieck problem. This is because results for this problem, and a variety of other geometric problems, are proved by taking the matrix A to be a *projection operator*, which is of course PSD. Many such results are based on a framework by Guruswami, Raghavendra, Saket, and Wu [GRSW16] for bypassing the UGC in obtaining hardness of geometric problems. They obtained tight inapproximability results using "smooth label cover" instead of Unique Games, for the L_p-Grothendieck problem (matching the UG-hardness result of Kindler, Naor, and Schechtman [KNS10]) and the subspace approximation problem (matching the UG-hardness result of Deshpande, Tulsiani, and Vishnoi [DTV11]). This was also the framework used by Briet, Regev and Saket [BRS15] for proving $\pi/2$ inapproximability for the (little) Grothendieck problem, matching an earlier UG-hardness result by Khot and Naor [KN09]). This framework was also used in [BGG⁺19] to obtain inapproximability results for $\|A\|_{p\to q}$ for several other values of p and q.

To understand why the GRSW framework naturally leads to projection operators, in the study of all the above problems, it is instructive to consider a "dictatorship test" gadget for the Grothendieck problem. Viewing the vectors s, t as evaluation tables of Boolean functions, we can equivalently think of the objective as $\langle f, Ag\rangle$ where $f, g : \{-1,1\}^R \to \{-1,1\}$ are Boolean functions over the domain (say) $\{-1,1\}^R$. A simple test follows from the well-known fact that the ℓ_2^2 mass of the degree-1 Fourier coefficients is at most $2/\pi + \varepsilon$ for any function far from a dictator, while it is equal to 1 for a dictator function. Taking F_1 to be the level-1 Fourier projection operator (which only keeps Fourier characters and coefficients of degree 1), we have that the (normalized) optimum value of $\langle f, F_1 g\rangle$ is 1 when maximizing over all ± 1 valued functions, and at most $2/\pi + \varepsilon$ when restricted to functions far from dictators, since $\|F_1 f\|_2^2 \le ({}^2/\pi + \varepsilon)\cdot \|f\|_2^2$.

Of course dictatorship tests are nontrivial to combine with Unique Games, and even more so with Label Cover instances. Considering an instance of Label Cover with vertex set V, and taking $(f_v : \{-1,1\}^R \to \{-1,1\})_{v\in V}$ to be the "long code" encodings for the labels, let $\mathbf{f} : V \times \{-1,1\}^R \to \{-1,1\}$ denote the combined function and let $\mathbf{F}_1$ denote the operator which projects each of the long-codes to the degree-1 space i.e., $\mathbf{F}_1 : \mathbb{R}^{V\times\{-1,1\}^R} \to \mathbb{R}^{V\times[R]}$. The GRSW framework amounts to defining a *global projection operator* $\mathbf{P}$ (which depends on the underlying Label Cover instance) on the combined level-1 Fourier space, such that $\mathbf{PF}_1\mathbf{f}$ behaves as if far from a dictator in blocks corresponding to most vertices, when starting with an unsatisfiable instance of Label Cover i.e.,, $\|\mathbf{PF}_1\mathbf{f}\|_2^2 \le ({}^2/\pi + \varepsilon)\cdot\|\mathbf{f}\|_2^2$. The final operator $\mathbf{A}$ in the result of [BRS15] can be taken to be the PSD operator $\mathbf{F}_1^*\mathbf{PF}_1$. As before, the solution optimizing $\langle \mathbf{f}, \mathbf{F}_1^*\mathbf{PF}_1\mathbf{g}\rangle$ satisfies $\mathbf{f} = \mathbf{g}$, which is a dictator in all blocks when the instance of Label Cover is satisfiable, and far from dictators in most blocks otherwise. Results for all the geometric problems above similarly rely on projection operators, and an analysis of the level-1 Fourier coefficients.

While improved dictatorship tests are indeed known for the Grothendieck problem, this requires going beyond the level-1 Fourier coefficients. Indeed the dictatorship test used in the UG-hardness result of Khot and O'Donnell [KO09] uses the operator $F_1 - \lambda \cdot \mathrm{Id}$ where Id denotes the identity operator. They call this the Davie-Reeds operator, since it is based on the lower bound constructions of Davie and Reeds, which can be viewed as integrality gap instances for the SDP relaxation of the Grothendieck problem. Raghavendra and Steurer [RS09] obtain their result using operators of the form $\sum_{i\ge 0} \lambda_i \cdot F_i$, where F_i is the level-i Fourier projection, and the coefficients $\lambda_i \in \mathbb{R}$ can be chosen using any solution to the SDP relaxation that exhibits an integrality gap.

However, it is not clear how to combine these tests with the Label Cover based projection operator $\mathbf{P}$ defined by GRSW, since it only acts on the level-1 Fourier coefficients. Moreover, the analysis in the case of the PSD operator $\mathbf{F}_1\mathbf{P}\mathbf{F}_1$ can be local, since we can write $\langle \mathbf{f}, \mathbf{F}_1^*\mathbf{P}\mathbf{F}_1\rangle$ as $\|\mathbf{P}\mathbf{F}_1\|_2^2$, which can be analyzed by understanding the level-1 Fourier mass of the projected function $\mathbf{f}$ separately in each block corresponding to some vertex v. Since the symmetry of of the optimal solution $\mathbf{f} = \mathbf{g}$ and the interpretation of the objective as an ℓ_2^2 norm is not available when the operator is not PSD, results based on the GRSW framework have been limited to projection operators.

Our techniques. We consider an operator $\mathbf{A}$ based on the Davie-Reeds operator. In particular, we take

$$\mathbf{A} \;=\; \mathbf{F}_1^*\mathbf{P}\mathbf{F}_1 - \lambda \cdot \mathrm{Id}\,,$$

where Id is the identity operator in the global space, and $\lambda > 0$ is a small constant. The optimizers of $\langle \mathbf{f}, \mathbf{A}\mathbf{g}\rangle$ no longer enjoy the symmetry $\mathbf{f} = \mathbf{g}$ that holds in the PSD case, but let us still suppose this is the case for a moment. This suffices to finish the proof since

$$\langle f, \mathbf{A}\mathbf{g}\rangle \;=\; \|\mathbf{P}\mathbf{F}_1\mathbf{f}\|_2^2 - \lambda \cdot \|\mathbf{f}\|_2^2 \;\leq\; \left(\frac{2}{\pi} + \varepsilon\right) \cdot \|\mathbf{f}\|_2^2 - \lambda \cdot \|\mathbf{f}\|_2^2 \;=\; \left(\frac{2}{\pi} + \varepsilon - \lambda\right)\,,$$

using the norm-reducing property of the GRSW projection operator, when starting from an unsatisfiable instance of label cover. One can check that for satisfiable instances, the optimal value is $1 - \lambda$, leading to a ratio strictly larger than $\pi/2$ when $\lambda > 0$.

The problem then reduces to still showing an approximate symmetry in the solution, namely that $\|\mathbf{f} - \mathbf{g}\|$ is small. We now rely on the *global structure* of the solution instead of the PSD nature of the operator to conclude this. A simple (but crucial) observation in our analysis is that the optimal solutions $\mathbf{f}$ and $\mathbf{g}$ must be close to linear threshold functions (LTFs). Indeed we must have for all $v \in V$, that $g_v(x) = \mathrm{sgn}(\langle (\mathbf{P}\mathbf{F}_1\mathbf{f})_v, x\rangle - \lambda \cdot f_v(x))$ (whenever $\langle (\mathbf{P}\mathbf{F}_1\mathbf{f})_v, x\rangle - \lambda \cdot f_v(x)$ is non-zero) and vice-versa for $f_v(x)$. For an LTF $\mathrm{sgn}(\langle a, x\rangle)$ we will refer to a as the *linear weights* associated to the LTF. By stability results for regular LTFs, we can then reduce the problem to showing that $\mathbf{P}\mathbf{F}_1\mathbf{f}$ is close to $\mathbf{P}\mathbf{F}_1\mathbf{g}$ i.e., regular LTFs are close, if their associated linear weights are close. The most technical part of the result is actually showing the regularity of the LTFs to apply this argument. Finally, the optimality of the solutions $\mathbf{f}$ and $\mathbf{g}$ can be used to show the closeness of the linear weights, since the term

$$\langle \mathbf{f}, \mathbf{F}_1^*\mathbf{P}\mathbf{F}_1\mathbf{g}\rangle \;=\; \langle \mathbf{P}\mathbf{F}_1\mathbf{f}, \mathbf{P}\mathbf{F}_1\mathbf{g}\rangle\,,$$

which is part of the objective, and can be viewed as a measure of the correlation of the linear weights for the above LTFs.

Note that the above is a departure from the usual analysis of long codes, which considers a global function $\mathbf{f}$ and decomposes it into block functions f_v which are analyzed individually in the evaluation or Fourier space. Instead, we need the global LTF structure of the solutions. We then decompose the global functions into local blocks in the "linear weights space". We hope such an analysis relying not only on local Fourier analysis, but also on global properties of the optimal solution, will be helpful in further strengthening the results for other geometric problems of interest.

2 Preliminaries and Notation

2.1 p-Norms

For a vector $s \in \mathbb{R}^n$, throughout this paper we will use $s(i)$ to denote its i-th coordinate. For $p \in [1,\infty)$, we define $\|\cdot\|_{\ell_p}$ to denote the counting p-norm and $\|\cdot\|_{L_p}$ to denote the expectation p-norm; i.e., for a vector $s \in \mathbb{R}^n$,

$$\|s\|_{\ell_p} := \left(\sum_{i\in[n]} |s(i)|^p\right)^{1/p} \quad \text{and} \quad \|s\|_{L_p} := \mathop{\mathbb{E}}_{i\sim[n]} \left[|s(i)|^p\right]^{1/p} = \left(\frac{1}{n}\cdot\sum_{i\in[n]} |s(i)|^p\right)^{1/p}.$$

Clearly $\|s\|_{\ell_p} = \|s\|_{L_p} \cdot n^{1/p}$. For $p = \infty$, we define $\|s\|_{\ell_\infty} = \|s\|_{L_\infty} := \max_{i\in[n]} |s(i)|$. We also use $\langle s,t\rangle_c$ to explicitly denote the inner product under the counting measure, i.e., for two vectors $s,t \in \mathbb{R}^n$, $\langle s,t\rangle_c := \sum_{i\in[n]} s(i)t(i)$. Later in the paper we will work with four different inner product spaces and will always use $\langle\cdot,\cdot\rangle$ to denote the associated inner product.

We will use p^* to denote the 'dual' of p, i.e. $p^* = p/(p-1)$. We also use the convention that $1^* = \infty$ and $\infty^* = 1$. We next record a well-known fact about p-norms; namely that the dual norm of the p-norm is the p^* norm.

Observation 2.1. *For any $p \in [1,\infty]$, $\|s\|_{\ell_p} = \sup_{\|t\|_{\ell_{p^*}}=1} \langle t,s\rangle_c$.*

We next define the operator norm between ℓ_p^n spaces.

Definition 2.2. *For $p,q \in [1,\infty]$, and a linear operator $A : \ell_p^n \to \ell_q^m$ the operator norm is defined as*

$$\|A\|_{\ell_p\to\ell_q} := \max_{s\in\mathbb{R}^n} \frac{\|As\|_{\ell_q}}{\|s\|_{\ell_p}}$$

We say *Grothendieck optimization problem* to refer to the important special case $\|A\|_{\ell_\infty\to\ell_1}$. We next state the well known fact that the $\infty \to 1$ operator norm is equivalent to bilinear maximization over the hypercube.

Fact 2.3. *For an $m \times n$ matrix A,*

$$\|A\|_{\ell_\infty\to\ell_1} = \sup_{s\in\{\pm1\}^n} \sup_{t\in\{\pm1\}^m} \langle t, As\rangle_c = \|A^*\|_{\ell_\infty\to\ell_1}.$$

Proof. Using $\langle y, Ax\rangle = \langle x, A^*y\rangle_c$,

$$\|A\|_{\ell_\infty\to\ell_1} = \sup_{\|s\|_{\ell_\infty}\le 1} \|As\|_{\ell_1} = \sup_{\|s\|_{\ell_\infty},\|t\|_{\ell_\infty}\le 1} \langle t, As\rangle_c = \sup_{s\in\{\pm1\}^n} \sup_{t\in\{\pm1\}^m} \langle t, As\rangle_c$$

where the final equality follows since if any $s(i)$ is in the interval $(-1,1)$ then setting $s(i) := \mathrm{sgn}(\sum_j A[i,j]\cdot t(j))$ cannot decrease the value. Similarly for any $t(j) \in (-1,1)$, setting $t(j) := \mathrm{sgn}(\sum_i A[i,j]\cdot s(i))$ cannot decrease the value. ∎

2.2 Fourier Analysis

We introduce some basic facts about Fourier analysis of Boolean functions. Let $R \in \mathbb{N}$ be a positive integer, and consider a function $f : \{\pm 1\}^R \to \mathbb{R}$. For any subset $S \subseteq [R]$ let $\chi_S := \prod_{i \in S} x_i$. Then we can represent f as

$$f(x_1, \ldots, x_R) = \sum_{S \subseteq [R]} \widehat{f}(S) \cdot \chi_S(x_1, \ldots x_R), \tag{1}$$

where

$$\widehat{f}(S) = \mathbb{E}_{x \in \{\pm 1\}^R}[f(x) \cdot \chi_S(x)] \text{ for all } S \subseteq [R]. \tag{2}$$

We interpret $\widehat{f}$ as a vector in $\mathbb{R}^{2^{[R]}}$ whose coordinates are indexed by $S \subseteq [R]$. We will always use the expectation norms for f and counting norms for $\widehat{f}$; i.e.,

$$\|f\|_{L_p} = \left(\mathop{\mathbb{E}}_{x \in \{\pm 1\}^R} [|f(x)|^p] \right)^{1/p} \quad \text{and} \quad \|\widehat{f}\|_{\ell_p} = \left(\sum_{S \subseteq [R]} |\widehat{f}(S)|^p \right)^{1/p}.$$

Similarly the we use the expectation inner product for $\langle f, g \rangle$ and we use the counting inner product for $\langle \widehat{f}, \widehat{g} \rangle$.

The *Fourier transform* refers to the linear operator F that maps f to $\widehat{f}$ as defined in (2). The *inverse Fourier transform* is the linear operator that maps $\widehat{f} : 2^{[R]} \to \mathbb{R}$ to $f : \{\pm 1\}^R \to \mathbb{R}$ defined as in (1). The inverse Fourier transform is simply the adjoint F^* of the Fourier transform.

Fact 2.4. *$F^* F$ is the identity operator.*

We refer to $\dot{\widehat{f}} := (\widehat{f}(\{1\}), \ldots, \widehat{f}(\{R\}))$ as the linear Fourier coefficients of f (indeed f is a linear function if and only if $\widehat{f}$ is supported completely inside $\dot{\widehat{f}}$). We define the *linear Fourier transform* denoted by F_1 as the (non-invertible) linear operator mapping f to $\dot{\widehat{f}}$. The adjoint F_1^* maps $\dot{\widehat{f}}$ to the boolean linear function $x \mapsto \langle \dot{\widehat{f}}, x \rangle$. We define the level-1 weight of f as $\mathrm{W}_1(f) := \|\dot{\widehat{f}}\|_{\ell_2}^2$. Similarly the level-$k$ weight of f is defined as $\mathrm{W}_k(f) := \sum_{S \in \binom{[R]}{k}} \widehat{f}(S)^2$. So we have $\|\widehat{f}\|_{\ell_2}^2 = \sum_{k \in \{0, \ldots R\}} \mathrm{W}_k(f)$.

The following well-known fact from Fourier analysis states that the expectation 2-norm on f coincides with the counting 2-norm on $\widehat{f}$.

Fact 2.5 (Parseval)**.** *For any $f : \{\pm 1\}^R \to \mathbb{R}$, $\|f\|_{L_2} = \|\widehat{f}\|_{\ell_2}$.*

In particular we conclude from this that $\mathrm{W}_1(f) \le \|f\|_{L_2}^2$.

2.3 Hilbert Spaces

Recall a Hilbert space is a vector space endowed with an inner product which we denote by $\langle \cdot, \cdot \rangle$. The inner product induces a Hilbert norm which we will denote by $\|h\|_H := \sqrt{\langle h, h \rangle}$. In this paper we work predominantly with four finite dimensional real Hilbert spaces defined below. In what follows V denotes the index set of the vertices of a graph. For the remainder of this paper, we assume $|V| = n$.

1. *Boolean function evaluation space H_E* over the vector space $\mathbb{R}^{\{\pm 1\}^R}$ whose elements are denoted throughout by lower case letters (e.g., $f, g, f_v, g_v, \ldots$). H_E is endowed with the expectation inner product $\langle f, g \rangle := \mathbb{E}_{x \in \{\pm 1\}^R}[f(x) g(x)]$, which induces the Hilbert norm $\|f\|_H = \|f\|_{L_2}$.

2. *Linear Fourier coefficient space* H_F^1 over the vector space $\mathbb{R}^{[R]}$ whose elements are denoted throughout by lower case hatted letters with a dot (e.g., $\hat{\dot{f}}, \hat{\dot{g}}, \hat{\dot{f}}_v, \hat{\dot{g}}_v, \ldots$). H_F^1 is endowed with the usual counting inner product $\langle \hat{\dot{f}}, \hat{\dot{g}} \rangle := \sum_{i \in [R]} \widehat{f}(\{i\})\widehat{g}(\{i\})$, which induces the Hilbert norm $\|\hat{\dot{g}}\|_H = \|\hat{\dot{g}}\|_{\ell_2}$.

3. *Concatenated evaluation space* $\mathbf{H}_E = H_E^{\oplus V}$ over the vector space $\mathbb{R}^{V \times \{\pm 1\}^R}$ whose elements are tuples of boolean functions denoted as $\mathbf{f} = (f_v)_{v \in V}$. Elements of $\mathbf{H}_E$ are denoted throughout by bold lower case letters (e.g., $\mathbf{f}, \mathbf{g}, \ldots$). $\mathbf{H}_E$ is endowed with the expectation inner product $\langle \mathbf{f}, \mathbf{g} \rangle := \mathbb{E}_{v \in V}[\langle f_v, g_v \rangle] = \mathbb{E}_{v \in V}[\mathbb{E}_{x \in \{\pm 1\}^R}[f_v(x) g_v(x)]]$ which induces the Hilbert norm $\|\mathbf{f}\|_H = \|\mathbf{f}\|_{L_2}$.

4. *Concatenated linear Fourier coefficient space* $\mathbf{H}_F^1 = (H_F^1)^{\oplus V}$ over the vector space $\mathbb{R}^{V \times [R]}$ whose elements are tuples of linear Fourier coefficient vectors denoted as $\hat{\dot{\mathbf{f}}} = (\hat{\dot{f}}_v)_{v \in V}$. Elements of $\mathbf{H}_F^1$ are denoted throughout by hatted bold lower case letters with a dot (e.g., $\hat{\dot{\mathbf{f}}}, \hat{\dot{\mathbf{g}}}, \ldots$). $\mathbf{H}_F^1$ is endowed with the inner product $\langle \hat{\dot{\mathbf{f}}}, \hat{\dot{\mathbf{g}}} \rangle := \mathbb{E}_{v \in V}[\langle \hat{\dot{f}}_v, \hat{\dot{g}}_v \rangle]$. Note that the induced Hilbert norm $\|\hat{\dot{\mathbf{f}}}\|_H = \|\hat{\dot{\mathbf{f}}}\|_{\ell_2} / \sqrt{n}$ is neither a counting nor an expectation norm.

The linear Fourier transform can be naturally extended to the concatenated space by defining $\mathbf{F}_1 : \mathbf{f} \mapsto \hat{\dot{\mathbf{f}}}$ which represents the vertex-wise map $f_v \mapsto \hat{\dot{f}}_v$ for all $v \in V$. The adjoint $\mathbf{F}_1^*$ maps $\hat{\dot{\mathbf{f}}} = (\hat{\dot{f}}_v)_{v \in V}$ to the tuple of boolean linear functions $(x \mapsto \langle \hat{\dot{f}}_v, x \rangle)_{v \in V}$.

2.4 Smooth Label Cover

An instance of Label Cover is given by a quadruple $\mathcal{L} = (G, [R], [L], \Sigma)$ that consists of a regular connected graph $G = (V, E)$, a label set $[R]$ for some positive integer n, and a collection $\Sigma = ((\pi_{e,v}, \pi_{e,w}) : e = (v, w) \in E)$ of pairs of maps both from $[R]$ to $[L]$ associated with the endpoints of the edges in E. Given a *labeling* $\ell : V \to [R]$, we say that an edge $e = (v, w) \in E$ is *satisfied* if $\pi_{e,v}(\ell(v)) = \pi_{e,w}(\ell(w))$. Let $\mathsf{OPT}(\mathcal{L})$ be the maximum fraction of satisfied edges by any labeling.

The following hardness result for Label Cover, given in [GRSW16], is a slight variant of a construction originally due to [Kho02]. The theorem also describes several structural properties, including smoothness, that are satisfied by the Label Cover instances.

Theorem 2.6. *For any $\xi > 0$ and $J \in \mathbb{N}$, there exist positive integers $R = R(\xi, J), L = L(\xi, J)$ and $D = D(\xi)$, and a polynomial time reduction $\varphi \mapsto \mathcal{L}(\varphi)$ from a 3-CNF instance φ to a Label Cover instance $\mathcal{L}(\varphi) = (G, [R], [L], \Sigma)$ such that*

- *(Hardness):*
  - *(Completeness): If φ is satisfiable, then $\mathsf{OPT}(\mathcal{L}(\varphi)) = 1$.*
  - *(Soundness): If φ is unsatisfiable, then $\mathsf{OPT}(\mathcal{L}(\varphi)) \leq \xi$.*
- *(Structural Properties): For any φ, $\mathcal{L}(\varphi)$ has the following properties*
  - *(J-Smoothness): For every vertex $v \in V$ and distinct $i, j \in [R]$, we have*

$$\mathop{\mathbb{P}}_{e: v \in e}[\pi_{e,v}(i) = \pi_{e,v}(j)] \leq 1/J.$$

  - *(D-to-1): For every vertex $v \in V$, edge $e \in E$ incident on v, and $i \in [L]$, we have $|\pi_{e,v}^{-1}(i)| \leq D$; i.e., at most D elements in $[R]$ are mapped to the same element in $[L]$.*

– *(Weak Expansion): For any $\delta > 0$ and any subset of vertices $V' \subseteq V$ such that $|V'| = \delta \cdot |V|$, the number of edges induced by the vertices in $|V'|$ is at least $(\delta^2/2)|E|$.*

2.5 Label Cover Consistency Subspace for Linear Fourier Coefficients

Let $\mathcal{L} = (G, [R], [L], \Sigma)$ be an instance of Label Cover with $G = (V, E)$ and let $\mathbf{P} : \mathbb{R}^{V\times[R]} \to \mathbb{R}^{V\times[R]}$ be the orthogonal projector to the subspace $\widehat{\mathbf{L}}$ of H_F^1 which is defined as:

$$\widehat{\mathbf{L}} := \left\{ \dot{\widehat{\mathbf{f}}} \in H_F^1 : \sum_{j\in\pi_{e,u}^{-1}(i)} \dot{\widehat{f}}_u(j) = \sum_{j\in\pi_{e,v}^{-1}(i)} \dot{\widehat{f}}_v(j) \text{ for all } (u,v) \in E \text{ and } i \in [L] \right\}. \tag{3}$$

The following lemma shown in [BRS15] (informally speaking) states that if $\mathcal{L}$ is far from satisfiable then for any element of $\dot{\widehat{\mathbf{f}}} \in \widehat{\mathbf{L}}$ there cannot be many vertices with influential coordinates (otherwise one can decode an assignment to $\mathcal{L}$ contradicting unsatisfiability). In other words, projection to $\widehat{\mathbf{L}}$ acts as a test of Label Cover consistency. For technical ease of use we state the lemma in terms of projections of concatenated boolean functions $\mathbf{f}$:

Lemma 2.7 (Corollary of Lemma 3.6 of [BRS15])**.** *There exists an absolute constant $C > 1$ such that if $\mathcal{L}$ is a T-to-1 label cover instance for some $T \in \mathbb{N}$ with soundness ξ, smoothness $C \cdot T/\xi$ and weak expansion, then for any $\mathbf{f} \in \{\pm 1\}^{V\times 2^R}$, we have*

$$|\{v \in V \mid \|(\mathbf{P}\dot{\widehat{\mathbf{f}}})_v\|_{\ell_\infty} > \xi^{1/14},\ \|(\mathbf{P}\dot{\widehat{\mathbf{f}}})_v\|_{\ell_2} \le 1/\xi^{1/28}\}| < O(\xi^{1/14} \cdot n)$$

Combining the preceding lemma with an appropriate dictatorship test (namely the bilinear form $(f, g) \mapsto \langle \dot{\widehat{f}}, \dot{\widehat{g}} \rangle$), [BRS15] showed the following soundness claim en route to their hardness result for little Grothendieck.

Theorem 2.8 (Implicit in proof of Theorem 1.3 of [BRS15])**.** *There exist absolute constants $C > 1$ and $c \in (0, 1)$ such that if $\mathcal{L}$ is a T-to-1 label cover instance for some $T \in \mathbb{N}$ with soundness ξ, smoothness $C \cdot T/\xi$ and weak expansion, then for any $\mathbf{f} \in \{\pm 1\}^{V\times\{\pm 1\}^R}$, we have $\|\mathbf{P}\dot{\widehat{\mathbf{f}}}\|_2 \le \sqrt{2/\pi} + \xi^c$.*

2.6 Central Limit Phenomena and Linear Threshold Functions

Recall the classical Berry-Esseen central limit theorem states

Theorem 2.9 (Berry-Esseen Central Limit Theorem)**.** *Let $S := X_1 + \cdots + X_R$ where $X_1, \ldots, X_R$ are independent centered random variables with $\mathbb{E}[X_i^2] = a_i^2$ and $\mathbb{E}[|X_i|^3] = b_i^3$. Let Ψ and Φ respectively denote the CDFs of S and of a centered Gaussian distribution with variance $\|a\|_{\ell_2}^2$. Then*

$$\sup_{x\in\mathbb{R}} |\Psi(x) - \Phi(x)| \le 10 \cdot \frac{\|b\|_{\ell_3}^3}{\|a\|_{\ell_2}^3}$$

An *unbiased linear threshold function* (henceforth LTF) is a boolean function of the form $\mathrm{sgn}(\langle a, x\rangle)$ for some vector $a \in \mathbb{R}^R$. We will refer to the entries of a as the *linear weights* of the LTF. Due to the nature of our reduction, we will frequently deal with LTFs and perturbed LTFs in the analysis. When $\|a\|_{\ell_2} = 1$ and $\|a\|_{\ell_\infty} \le \varepsilon$, the LTF is called *regular*. In this section we collect and derive some facts about central limit phenomena exhibited by regular LTFs (intuitively this is because for a random ± 1 vector x, $\langle a, x\rangle$ exhibits similar behaviour to a Gaussian random variable). The following result stated as Theorem 5.17 in [O'D14] is a corollary of a multidimensional version of the Berry-Esseen Central limit theorem due to [Ben05]. It states that the noise stability (and level-1 weight) of an LTF behaves like that of the function $\operatorname{sgn}(x_1)$ in gaussian space.

Theorem 2.10 (Noise Stability and Level-1 Weight of LTFs)**.**
Let $f(x) = \operatorname{sgn}(\langle a, x\rangle)$ be an unbiased LTF where $\|a\|_{\ell_2} = 1$ and $\|a\|_{\ell_\infty} \le \varepsilon$. Then for any $\rho \in (-1, 1)$,

$$\sum_{k \ge 1} \mathrm{W}_k(f) \cdot \rho^k \le \frac{2}{\pi} \cdot \arcsin \rho + O\Big(\frac{\varepsilon}{\sqrt{1-\rho^2}}\Big)\,.$$

Since $\mathrm{W}_k(f) \ge 0$ and $\arcsin \rho \le \rho + 10 \cdot \rho^3$, setting $\rho := \sqrt{\varepsilon}$ above implies the level-1 bound

$$\mathrm{W}_1(f) \le 2/\pi + O(\sqrt{\varepsilon})\,.$$

We require a version of Theorem 2.10 for perturbed LTFs:

Lemma 2.11 (Level-1 Weight of λ-Perturbed LTFs)**.**
Let $a \in \mathbb{R}^R$ and $K > 1$ be such that $\|a\|_{\ell_2} \ge \frac{1}{4\pi}$, $\|a\|_{\ell_\infty} \le \varepsilon$, and let $f, g : \{\pm 1\}^n \to \{\pm 1\}$ be boolean functions satisfying $g(x) = \operatorname{sgn}(\langle a, x\rangle - \lambda \cdot f(x))$ whenever x is such that $\langle a, x\rangle - \lambda \cdot f(x) \neq 0$ (where $\lambda \in (0, 1)$). Then the fraction of inputs on which $g(x)$ and $\operatorname{sgn}(\langle a, x\rangle - \lambda \cdot f(x))$ disagree is at most $4\sqrt{2\pi} \cdot \lambda + O(\varepsilon)$, and moreover $\mathrm{W}_1(g) \le 2/\pi + 2^{5/4}\pi^{1/4} \cdot \sqrt{\lambda} + O(\sqrt{\varepsilon})$.

Proof. Observe that the fraction of inputs on which $g(x)$ and $\operatorname{sgn}(\langle a, x\rangle - \lambda \cdot f(x))$ disagree is at most

$$\begin{aligned}\mathop{\mathbb{P}}_{x \sim \{\pm 1\}^R}[(g(x) \neq \operatorname{sgn}(\langle a, x\rangle)) \vee (\lambda \cdot f(x) = \operatorname{sgn}(\langle a, x\rangle))] &\le \mathop{\mathbb{P}}_{x \sim \{\pm 1\}^R}[|\langle a, x\rangle| \le \lambda] \\ &= 2 \cdot \mathop{\mathbb{P}}_{x \sim \{\pm 1\}^R}[\langle a, x\rangle \le \lambda] - 1\,.\end{aligned}$$

Let $\Phi : \mathbb{R} \to \mathbb{R}$ denote the CDF of a Gaussian random variable with mean 0 and variance $\|a\|_{\ell_2}^2$. Since $\|a\|_{\ell_\infty} \le \varepsilon$, we have $\|a\|_{\ell_3}^3 / \|a\|_{\ell_2}^3 \le \varepsilon / \|a\|_{\ell_2} \le 4\pi\varepsilon$. Thus by central limit theorem (Theorem 2.9) we conclude that

$$\begin{aligned}&2 \cdot \mathop{\mathbb{P}}_{x \sim \{\pm 1\}^R}[\langle a, x\rangle \le \lambda] - 1 \;\le\; 2 \cdot \Phi(\lambda) - 1 + O(\varepsilon) \\ &\le\; O(\varepsilon) + \sqrt{\tfrac{2}{\pi}} \cdot \tfrac{1}{\|a\|_{\ell_2}} \cdot \textstyle\int_0^\lambda e^{-t^2/(2\|a\|_{\ell_2}^2)}\, dt \;\le\; O(\varepsilon) + \int_0^\lambda \sqrt{\tfrac{2}{\pi}} \cdot \tfrac{dt}{\|a\|_{\ell_2}} \;\le\; O(\varepsilon) + 4\sqrt{2\pi}\lambda\end{aligned}$$

as desired.

For the second claim, we have

$$\begin{aligned}&\mathrm{W}_1(g)^{1/2} \le \mathrm{W}_1(\operatorname{sgn}(\langle a, x\rangle))^{1/2} + \mathrm{W}_1(g - \operatorname{sgn}(\langle a, x\rangle))^{1/2} \\ &\le \sqrt{2/\pi} + O(\sqrt{\varepsilon}) + \mathrm{W}_1(g - \operatorname{sgn}(\langle a, x\rangle))^{1/2} \\ &\le \sqrt{2/\pi} + O(\sqrt{\varepsilon}) + \|g - \operatorname{sgn}(\langle a, x\rangle)\|_2 \le \sqrt{2/\pi} + 2^{5/4}\pi^{1/4}\sqrt{\lambda} + O(\sqrt{\varepsilon})\end{aligned}$$

where the first inequality follows from triangle inequality over the space ℓ_2^R, the second inequality follows from Theorem 2.10, and the fourth inequality follows from the $1 - 4\sqrt{2\pi}\lambda - O(\varepsilon)$ agreement that we proved above. ∎

We also need the following lemma which informally states that two regular LTFs are close whenever their linear weights are close. Again we proceed by first passing to an appropriate analogue over Gaussians.

Lemma 2.12 (Agreement of Regular LTFs with Correlated Linear Coefficients). *Let $a, b \in \mathbb{R}^R$ be such that $\|a\|_{\ell_2}, \|b\|_{\ell_2} \geq 1/(4\pi)$, and $\|a\|_{\ell_\infty}, \|b\|_{\ell_\infty} \leq \varepsilon$. Then*

$$\mathop{\mathbb{P}}_{x}[\mathrm{sgn}(\langle a, x\rangle) \neq \mathrm{sgn}(\langle b, x\rangle)] \leq 4\sqrt{2}\|a-b\|_{\ell_2} + O(\varepsilon^{1/6})$$

Proof. Let $u := a/\|a\|_{\ell_2}$, $v := b/\|b\|_{\ell_2}$, $\rho := \langle u, v\rangle$. Note that $\|u\|_{\ell_\infty}, \|v\|_{\ell_\infty} \leq 4\pi \cdot \varepsilon$ and further that

$$\begin{aligned}\|a-b\|_{\ell_2}^2 &= \|a\|_{\ell_2}^2 + \|b\|_{\ell_2}^2 - 2\rho\|a\|_{\ell_2}\|b\|_{\ell_2} \underset{\text{AM-GM}}{\geq} 2\|a\|_{\ell_2}\|b\|_{\ell_2} - 2\rho\|a\|_{\ell_2}\|b\|_{\ell_2} \\ &= \|a\|_{\ell_2}\|b\|_{\ell_2}\|u-v\|_{\ell_2}^2 \geq (1/16\pi^2)\cdot\|u-v\|_{\ell_2}^2 .\end{aligned}$$

Thus it suffices to show a bound of $(\sqrt{2}/\pi)\|u-v\|_{\ell_2} + O(\varepsilon^{1/6})$.

To this end let $K := \{y \in \mathbb{R}^R \mid \langle u, y\rangle \geq 0,\ \langle v, y\rangle \leq 0\}$ be the intersection of two (regular) halfspaces, let x be a uniformly random vector in $\{\pm 1\}^R$ and let $\gamma \in \mathbb{R}^R$ be a vector with independent standard Gaussian coordinates. It suffices to show $\mathbb{P}[x \in K] \leq \|u-v\|_{\ell_2}/(\sqrt{2}\pi) + O(\varepsilon^{1/6})$ since we have $\mathbb{P}_x[\mathrm{sgn}(\langle a, x\rangle) \neq \mathrm{sgn}(\langle b, x\rangle)] \leq \mathbb{P}_x[x \in K] + \mathbb{P}_x[x \in -K] = 2 \cdot \mathbb{P}_x[x \in K]$. By Invariance principle for the intersection of regular halfspaces (e.g. Theorem 3.1 in [HKM13]) we have

$$\mathop{\mathbb{P}}_{x}[x \in K] \leq \mathop{\mathbb{P}}_{\gamma}[\gamma \in K] + O(\varepsilon^{1/6}) = \frac{\cos^{-1}\rho}{2\pi} + O(\varepsilon^{1/6})$$

where the final equality follows since the probability of a random hyperplane lying between two vectors u, v is precisely $\cos^{-1}\rho/\pi$ (sometimes referred to as the Grothendieck identity). By Taylor expansion, we have $\rho := \cos\theta \leq 1 - \theta^2/4$. Therefore, $\cos^{-1}\rho \leq \sqrt{4-4\rho} = \sqrt{2}\|u-v\|_{\ell_2}$, and we obtain $\mathbb{P}_x[x \in K] \leq \|u-v\|_{\ell_2}/(\sqrt{2}\pi) + O(\varepsilon^{1/6})$ as desired. ∎

3 $(\frac{\pi}{2} + \varepsilon_0)$ NP-Hardness of $\|\cdot\|_{\ell_\infty^n \to \ell_1^n}$

3.1 Reduction from Smooth Label Cover

Here we describe a polynomial time reduction taking as input a Label Cover instance $\mathcal{L}$ and producing a self-adjoint linear operator $\mathbf{A} : \mathbb{R}^{V\times\{\pm 1\}^R} \to \mathbb{R}^{V\times\{\pm 1\}^R}$. Let $\lambda \in (0,1)$ be a constant whose value will be fixed later. $\mathbf{A}$ is defined as follows

$$\mathbf{A} := \mathbf{F}_1^* \, \mathbf{P} \, \mathbf{F}_1 - \lambda \cdot \mathrm{Id} . \tag{4}$$

Equivalently the corresponding bilinear form is given by

$$\langle \mathbf{f}, \mathbf{A}\mathbf{g}\rangle = \langle \dot{\widehat{\mathbf{f}}}, \mathbf{P}\dot{\widehat{\mathbf{g}}}\rangle - \lambda \cdot \langle \mathbf{f}, \mathbf{g}\rangle . \tag{5}$$

In other words, given $\mathbf{f}$, we apply the Fourier transform for each $v \in V$, project the combined Fourier coefficients to $\widehat{\mathbf{L}}$ that checks the Label Cover consistency, and apply the inverse Fourier transform. Since $\mathbf{P}$ is a projector, $\mathbf{A}$ is self-adjoint by design.

Remark 3.1. *Our reduction is inspired both by the reduction $\mathcal{L} \mapsto (\mathbf{F}_1^*)\mathbf{P}\mathbf{F}_1$ used in [GRSW16], [BRS15] and by the dictatorship test $\mathbb{F}_1^*\mathbb{F}_1 - \lambda \cdot \mathrm{Id}$ used in [KO09] (which was based on a gap instance due independently to Davie and Reeds).*

3.2 Proof Sketch

It is easily seen that in the completeness case assigning each $f_v(x) = g_v(x) = x_{\ell(v)}$ to be dictator functions (where ℓ is some satisfying label cover assignment) yields a value of $1 - \lambda$.

So to obtain a gap of $\pi/2 + \varepsilon_0$ it suffices to show that for a sufficiently small constant λ soundness is upper bounded by $2/\pi - k\lambda$ for any constant $k > 2/\pi$. We will do this by showing the stronger bound of $2/\pi - \lambda + O(\lambda^{3/2} + \xi^{c'})$ and taking λ, ξ sufficiently small (here ξ is label cover soundness and can be taken to be an arbitrarily small constant independent of λ). By Theorem 2.8 we already have $\langle \dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}} \rangle = \langle \mathbf{P}\dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}} \rangle \leq 2/\pi + 3\xi^c$. Thus it suffices to show that if $\mathbf{f}, \mathbf{g}$ are optimal then they must satisfy $\langle \mathbf{f}, \mathbf{g} \rangle \geq 1 - O(\sqrt{\lambda})$ (since the remaining $-\lambda\langle f, g\rangle$ term would then subtract the necessary amount from $2/\pi$ to yield our desired soundness).

Closeness of $\mathbf{f}, \mathbf{g}$. We begin with the crucial observation that optimal $\mathbf{f}, \mathbf{g}$ are λ-perturbed LTFs. Indeed it must be that whenever $\langle (\mathbf{P}\dot{\hat{\mathbf{f}}})_v, x\rangle - \lambda \cdot f_v(x) \neq 0$, we have $g_v(x) = \operatorname{sgn}(\langle (\mathbf{P}\dot{\hat{\mathbf{f}}})_v, x\rangle - \lambda \cdot f_v(x))$ (otherwise $\mathbf{f}, \mathbf{g}$ are not optimal as the value can be improved). Using this structure (as well as the central limit phenomenon) we show in Lemma 3.2 which is the most technical part of the proof, that most vertices (at least $1 - O(\sqrt{\lambda})$ fraction) satisfy that $\|(\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_2}, \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} \geq 1/(4\pi)$ (i.e., are not too small in norm). Further by Lemma 2.7 most vertices ($1 - O(\xi^{c'})$ fraction) satisfy that $(\mathbf{P}\dot{\hat{\mathbf{f}}})_v, (\mathbf{P}\dot{\hat{\mathbf{g}}})_v$ do not have any large coordinates. Thus for most vertices we may leverage the central limit phenomenon by applying Lemma 2.11 to conclude that $\mathbf{f}$ is close to $(\operatorname{sgn}(\langle (\mathbf{P}\dot{\hat{\mathbf{g}}})_v, x\rangle))_{v\in V}$ and $\mathbf{g}$ is close to $(\operatorname{sgn}(\langle (\mathbf{P}\dot{\hat{\mathbf{f}}})_v, x\rangle))_{v\in V}$. Finally we will conclude $\mathbf{f}$ is close to $\mathbf{g}$ by showing that $\mathbf{P}\dot{\hat{\mathbf{f}}}$ is close to $\mathbf{P}\dot{\hat{\mathbf{g}}}$.

Closeness of $\mathbf{P}\dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}}$. Note that $\langle \mathbf{f}, \mathbf{g}\rangle \in [-1, 1]$ and so $\langle \mathbf{f}, \mathbf{A}\mathbf{g}\rangle \leq \langle \dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}}\rangle + \lambda$. So we may assume that $\langle \dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}}\rangle \geq 2/\pi - 2\lambda$ since otherwise we have already proved soundness of $2/\pi - \lambda$. Thus $\langle \mathbf{P}\dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}}\rangle = \langle \dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}}\rangle \geq 2/\pi - 2\lambda$. On the other hand, Theorem 2.8 states that $\|\mathbf{P}\dot{\hat{\mathbf{f}}}\|_H, \|\mathbf{P}\dot{\hat{\mathbf{g}}}\|_H \leq \sqrt{2/\pi} + \xi^c$. Thus $\mathbf{P}\dot{\hat{\mathbf{f}}}$ is close to $\mathbf{P}\dot{\hat{\mathbf{g}}}$ allowing us to conclude $\mathbf{f}$ is close to $\mathbf{g}$ using Lemma 2.12.

3.3 Analysis

Let $\mathbf{f}, \mathbf{g} \in \{\pm 1\}^{V \times \{\pm 1\}^R}$ be maximizers of $\langle \mathbf{f}, \mathbf{A}\mathbf{g}\rangle$. We begin the proof by defining various subsets of vertices for which $(\mathbf{P}\dot{\hat{\mathbf{f}}})_v, (\mathbf{P}\dot{\hat{\mathbf{g}}})_v$ have anomalous behaviour. In Lemma 3.2 we will show that all of these anomalous sets are small.

Vertices with excessively large norm

$$V_0^f := \{v \in V \mid \|(\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_2} > 1/\xi^{1/28}\}$$
$$V_0^g := \{v \in V \mid \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} > 1/\xi^{1/28}\}$$
$$V_0 := V_0^f \cup V_0^g$$
$$\overline{V}_0 := V \setminus V_0 = \{v \in V \mid \|(\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_2} \leq 1/\xi^{1/28} \wedge \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} \leq 1/\xi^{1/28}\}$$

Vertices with an influential coordinate

$$V_1 := \{v \in \overline{V}_0 \mid \|(\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_\infty} > \xi^{1/14} \vee \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_\infty} > \xi^{1/14}\}$$

$\overline{V}_1 := \overline{V}_0 \setminus V_1 = \{v \in \overline{V}_0 \mid \|(\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_\infty} \le \xi^{1/14} \ \wedge \ \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_\infty} \le \xi^{1/14}\}$

Vertices with excessively small norm after projecting g

$V_2 := \{v \in \overline{V}_1 \mid \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} < 1/(2\pi)\}$

$\overline{V}_2 := \overline{V}_1 \setminus V_2 = \{v \in \overline{V}_1 \mid \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} \ge 1/(2\pi)\}$

Vertices with excessively small norm after projecting f

$V_3 := \{v \in \overline{V}_2 \mid \|(\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_2} < 1/(4\pi)\}$

$\overline{V}_3 := \overline{V}_2 \setminus V_3 = \{v \in \overline{V}_1 \mid \|(\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_2} \ge 1/(4\pi) \ \wedge \ \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} \ge 1/(2\pi)\}\,.$

$\overline{V}_3$ is the set of vertices on which we may use the central limit phenomenon (Lemma 2.11) for showing closeness of $\mathbf{f}, \mathbf{g}$. We next show that $\overline{V}_3$ forms the vast majority of the vertices.

Lemma 3.2 (Most Vertices have Well Behaved Projections)**.**
There exist absolute constants $C > 1$ and $c, c_1 \in (0,1)$ such that if $\mathcal{L}$ is a T-to-1 label cover instance for some $T \in \mathbb{N}$ with soundness ξ, smoothness $C \cdot T/\xi$ and weak expansion, and $\mathbf{f}, \mathbf{g} \in \{\pm 1\}^{V \times \{\pm 1\}^R}$ are maximizers of $\langle \mathbf{f}, \mathbf{A}\mathbf{g} \rangle$, then we have $|V \setminus \overline{V}_3| \le (1005 \cdot \delta + (2\pi)^{5/4} \cdot \sqrt{\lambda} + \xi^{c_1}) \cdot n$, where $\delta := 2/\pi + 3\xi^c - \langle \dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}} \rangle$.

Proof. We begin by showing that $\mathbf{P}\dot{\hat{\mathbf{f}}}$ and $\mathbf{P}\dot{\hat{\mathbf{g}}}$ are very close (as a function of δ, λ). Indeed by Theorem 2.8 we know

$$\|\mathbf{P}\dot{\hat{\mathbf{f}}}\|_H\,, \|\mathbf{P}\dot{\hat{\mathbf{g}}}\|_H \le \sqrt{\tfrac{2}{\pi}} + \xi^c\,. \tag{6}$$

Thus $\delta \ge 0$ and further since $\mathbf{P}^2 = \mathbf{P}$, we have $\langle \mathbf{P}\dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}} \rangle = \langle \dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}} \rangle = 2/\pi + 3\xi^c - \delta$. Combining this with (6) implies

$$\|\mathbf{P}\dot{\hat{\mathbf{f}}} - \mathbf{P}\dot{\hat{\mathbf{g}}}\|_H^2 \le 2 \cdot \delta + O(\xi^c)\,. \tag{7}$$

By Cauchy-Schwarz we have $\|\mathbf{P}\dot{\hat{\mathbf{f}}}\|_H \cdot \|\mathbf{P}\dot{\hat{\mathbf{g}}}\|_H \ge \langle \mathbf{P}\dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}} \rangle = 2/\pi + 3\xi^c - \delta$. Combining this with (6) yields

$$\|\mathbf{P}\dot{\hat{\mathbf{f}}}\|_H\,, \|\mathbf{P}\dot{\hat{\mathbf{g}}}\|_H \ge \sqrt{\tfrac{2}{\pi}} - \sqrt{\tfrac{\pi}{2}} \cdot \delta - O(\xi^c)\,. \tag{8}$$

We next use the fact that $\mathbf{P}\dot{\hat{\mathbf{f}}}$ is close to $\mathbf{P}\dot{\hat{\mathbf{g}}}$ to conclude that there aren't many vertices where $\|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2}$ is sufficiently large but $\|(\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_2}$ is very small. We have

$$\begin{aligned}
\|\mathbf{P}\dot{\hat{\mathbf{f}}} - \mathbf{P}\dot{\hat{\mathbf{g}}}\|_H^2 \ge \frac{1}{n} \cdot \sum_{v \in \overline{V}_2} \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v - (\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_2}^2 &\ge \frac{1}{n} \cdot \sum_{v \in \overline{V}_2} (\|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} - \|(\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_2})^2 \\
&\ge \frac{1}{n} \cdot \sum_{v \in V_3} (\|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} - \|(\mathbf{P}\dot{\hat{\mathbf{f}}})_v\|_{\ell_2})^2 \\
&\ge |V_3|/(16\pi^2 \cdot n)
\end{aligned}$$

Thus by (7) we conclude that

$$|V_3| \le (32\pi^2 \cdot \delta + O(\xi^c)) \cdot n\,. \tag{9}$$

Since $\|\mathbf{P}\dot{\hat{\mathbf{f}}}\|_H \le \|\mathbf{f}\|_H \le 1$ (and the same for $\mathbf{P}\dot{\hat{\mathbf{g}}}$), we also have

$$|V_0| \le 2 \cdot \xi^{1/14}\,. \tag{10}$$

Thus since V_0, V_3 and V_1 (using Lemma 2.7) are small, we have established that most of the vertices in V lie inside $\overline{V}_3 \cup V_2$. The rest of the proof is organized as follows. We will argue V_2 is small by showing that if V_2 were large then this would contradict the fact that $\mathbf{P}\dot{\hat{\mathbf{g}}}$ is the projection of $\hat{\mathbf{g}}$. To do this we will show that the inner product $\langle \hat{\mathbf{g}}, \mathbf{P}\dot{\hat{\mathbf{g}}} \rangle$ is too small (i.e., bounded away from $2/\pi$) since the inner product terms involving the vertices in V_2 are small by definition of V_2 and the inner product terms involving vertices in $\overline{V}_3$ aren't larger than $2/\pi$ due to the central limit phenomenon (Lemma 2.11). This forces V_2 to be small.

We proceed with formalizing the aforementioned sketch. Since $\langle \hat{\mathbf{g}}, \mathbf{P}\dot{\hat{\mathbf{g}}} \rangle = \|\mathbf{P}\dot{\hat{\mathbf{g}}}\|_2^2$, we have by (8) that

$$\langle \hat{\mathbf{g}}, \mathbf{P}\dot{\hat{\mathbf{g}}} \rangle \ge 2/\pi - 2 \cdot \delta - O(\xi^c)\,. \tag{11}$$

On the other hand we have the upper bound

$$\begin{aligned}
\langle \hat{\mathbf{g}}, \mathbf{P}\dot{\hat{\mathbf{g}}} \rangle &= \frac{1}{n} \cdot \sum_{v \in V} \langle \dot{\hat{g}}_v, (\mathbf{P}\dot{\hat{\mathbf{g}}})_v \rangle \le \frac{1}{n} \cdot \sum_{v \in V} \mathrm{W}_1(g_v)^{1/2} \cdot \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} && \text{(Cauchy-Schwarz)} \\
&\le \frac{1}{n} \cdot \sum_{v \in \overline{V}_1} \mathrm{W}_1(g_v)^{1/2} \cdot \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} + \frac{1}{n} \cdot \sum_{v \in V_0 \cup V_1} \mathrm{W}_1(g_v)^{1/2} \cdot \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} \\
&\le \frac{1}{n} \cdot \sum_{v \in \overline{V}_1} \mathrm{W}_1(g_v)^{1/2} \cdot \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} + \frac{1}{n} \cdot \sum_{v \in V_0 \cup V_1} \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} && (\mathrm{W}_1(g_v) \le 1) \qquad (12)
\end{aligned}$$

We bound the main term and the error term separately. We begin with the error term:

$$\begin{aligned}
\sum_{v \in V_0 \cup V_1} \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} \;&\le\; \frac{|V_1| + |V_0^f \setminus V_0^g|}{\xi^{1/28}} + \sum_{v \in V_0^g} \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} \\
&\underset{(10),\ \textit{Lemma 2.7}}{\le} O(\xi^{1/28} \cdot n) + \sum_{v \in V_0^g} \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} \;\le\; O(\xi^{1/28} \cdot n) + \sum_{v \in V_0^g} \xi^{1/28} \cdot \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2}^2 \\
&\le\; O(\xi^{1/28} \cdot n) + \xi^{1/28} \cdot \|\mathbf{P}\dot{\hat{\mathbf{g}}}\|_{\ell_2}^2 \;\le\; O(\xi^{1/28} \cdot n) + \xi^{1/28} \cdot n \cdot \|\mathbf{g}\|_H^2 \;\le\; O(\xi^{1/28} \cdot n) \qquad (13)
\end{aligned}$$

Let $c_1 := \min\{1/28, c\}$. We now bound the main (first) term in (12):

$$\begin{aligned}
&\tfrac{1}{n} \cdot \textstyle\sum_{v \in \overline{V}_1} \mathrm{W}_1(g_v)^{1/2} \cdot \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} \\
&< \tfrac{1}{n} \cdot \textstyle\sum_{v \in \overline{V}_2} \mathrm{W}_1(g_v)^{1/2} \cdot \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} + \tfrac{|V_2|}{2\pi \cdot n} && \left(\forall v \in V_2,\ \|(\mathbf{P}\dot{\hat{\mathbf{g}}})_v\|_{\ell_2} < \tfrac{1}{2\pi}\right) \\
&\le \tfrac{1}{n} \cdot \sqrt{\textstyle\sum_{v \in \overline{V}_2} \mathrm{W}_1(g_v)} \cdot \|\mathbf{P}\dot{\hat{\mathbf{g}}}\|_{\ell_2} + \tfrac{|V_2|}{2\pi \cdot n} && \text{(Cauchy-Schwarz)} \\
&\le \tfrac{1}{\sqrt{n}} \cdot \sqrt{\textstyle\sum_{v \in \overline{V}_2} \mathrm{W}_1(g_v)} \cdot \left(\sqrt{\tfrac{2}{\pi}} + \xi^c\right) + \tfrac{|V_2|}{2\pi \cdot n} && \text{(by (6))} \\
&\le \sqrt{32\pi^2 \cdot \delta + O(\xi^c) + \textstyle\sum_{v \in \overline{V}_3} \mathrm{W}_1(g_v)/n} \cdot \left(\sqrt{\tfrac{2}{\pi}} + \xi^c\right) + \tfrac{|V_2|}{2\pi \cdot n} && \text{(by (9))}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sqrt{32\pi^2 \cdot \delta + O(\xi^{c_1}) + \tfrac{|\overline{V}_3|}{n}\left(\tfrac{2}{\pi} + 2^{5/4}\pi^{1/4}\sqrt{\lambda}\right)} \cdot \left(\sqrt{\tfrac{2}{\pi}} + \xi^c\right) + \tfrac{|V_2|}{2\pi \cdot n} && \text{(Lemma 2.11 with } \varepsilon := \xi^{\frac{1}{14}}) \\
&\leq \sqrt{32\pi^2 \cdot \delta + O(\xi^{c_1}) + \tfrac{2|\overline{V}_3|}{\pi n} + 2^{5/4}\pi^{1/4}\sqrt{\lambda}} \cdot \left(\sqrt{\tfrac{2}{\pi}} + \xi^c\right) + \tfrac{|V_2|}{2\pi \cdot n} && (|\overline{V}_3| \leq n) \\
&\leq \sqrt{\tfrac{2}{\pi} - \tfrac{2|V_2|}{\pi n} + 32\pi^2 \cdot \delta + 2^{5/4}\pi^{1/4}\sqrt{\lambda} + O(\xi^{c_1})} \cdot \left(\sqrt{\tfrac{2}{\pi}} + \xi^c\right) + \tfrac{|V_2|}{2\pi \cdot n} && (|\overline{V}_3| \leq n - |V_2|) \\
&= \sqrt{1 - \tfrac{|V_2|}{n} + 16\pi^3 \cdot \delta + 2^{1/4}\pi^{5/4} \cdot \sqrt{\lambda} + O(\xi^{c_1})} \cdot \left(\tfrac{2}{\pi} + O(\xi^c)\right) + \tfrac{|V_2|}{2\pi \cdot n} && \text{(factoring out } \sqrt{2/\pi}) \\
&\leq \frac{2}{\pi}\left(1 - \frac{|V_2|}{2n} + 8\pi^3 \cdot \delta + 2^{-3/4}\pi^{5/4} \cdot \sqrt{\lambda} + O(\xi^{c_1})\right) + \frac{|V_2|}{2\pi \cdot n} && \left(\forall x \in (-1,1),\ \sqrt{1+x} \leq 1 + \tfrac{x}{2}\right) \\
&\leq \frac{2}{\pi} - \frac{|V_2|}{2\pi \cdot n} + 16\pi^2 \cdot \delta + 2^{1/4}\pi^{1/4}\sqrt{\lambda} + O(\xi^{c_1})\,. && (14)
\end{aligned}
$$

The application of Lemma 2.11 above is valid since by assumption of optimality of $\mathbf{f}, \mathbf{g}$, whenever $\langle (\mathbf{P}\dot{\hat{\mathbf{f}}})_v, x\rangle - \lambda \cdot f_v(x) \neq 0$, we have $g_v(x) = \operatorname{sgn}(\langle (\mathbf{P}\dot{\hat{\mathbf{f}}})_v, x\rangle - \lambda \cdot f_v(x))$ (otherwise f, g are not optimal as the value can be improved). Combining (11), (12), (13) and (14) with the fact that $2\pi(16\pi^2 + 2) \leq 1005$ yields $|V_2|/n \leq 1005 \cdot \delta + (2\pi)^{5/4} \cdot \sqrt{\lambda} + O(\xi^{c_1})$. Combining this with (10), (9), Lemma 2.7 and the fact that $\overline{V}_3 = V \setminus (V_0 \cup V_1 \cup V_2 \cup V_3)$ yields $|V \setminus \overline{V}_3|/n \leq 1005 \cdot \delta + (2\pi)^{5/4} \cdot \sqrt{\lambda} + O(\xi^{c_1})$ as desired. ∎

Lemma 3.2 allows us to leverage the central limit phenomenon (Lemma 2.11) on most vertices thereby obtaining that $\mathbf{f}$ is close to $(\operatorname{sgn}(\langle (\mathbf{P}\dot{\hat{\mathbf{g}}})_v, x\rangle))_{v \in V}$ and $\mathbf{g}$ is close to $(\operatorname{sgn}(\langle (\mathbf{P}\dot{\hat{\mathbf{f}}})_v, x\rangle))_{v \in V}$. Finally the proximity of $\mathbf{P}\dot{\hat{\mathbf{f}}}$ and $\mathbf{P}\dot{\hat{\mathbf{g}}}$ allows us to conclude that $\mathbf{f}$ is close to $\mathbf{g}$ using Lemma 2.12.

Lemma 3.3 (Closeness of $\mathbf{f}, \mathbf{g}$).
There are absolute constants $C > 1$ and $c, c_2 \in (0,1)$ such that if $\mathcal{L}$ is a T-to-1 label cover instance for some $T \in \mathbb{N}$ with soundness ξ, smoothness $C \cdot T/\xi$ and weak expansion, and $\mathbf{f}, \mathbf{g} \in \{\pm 1\}^{V \times \{\pm 1\}^R}$ are maximizers of $\langle \mathbf{f}, \mathbf{A}\mathbf{g}\rangle$, then setting $\delta := 2/\pi + 3\xi^c - \langle \dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}}\rangle$ we have

$$\langle \mathbf{f}, \mathbf{g}\rangle \geq 1 - 2^{9/4}\pi^{5/4}\sqrt{\lambda} - 8\sqrt{\delta} - 2010 \cdot \delta + \xi^{c_2}\,.$$

Proof. By assumption of optimality we must have $g_v(x) = \operatorname{sgn}(\langle (\mathbf{P}\dot{\hat{\mathbf{f}}})_v, x\rangle - \lambda \cdot f_v(x))$ whenever $\langle (\mathbf{P}\dot{\hat{\mathbf{f}}})_v, x\rangle - \lambda \cdot f_v(x) \neq 0$ and $f_v(x) = \operatorname{sgn}(\langle (\mathbf{P}\dot{\hat{\mathbf{g}}})_v, x\rangle - \lambda \cdot g_v(x))$ whenever $\langle (\mathbf{P}\dot{\hat{\mathbf{g}}})_v, x\rangle - \lambda \cdot g_v(x) \neq 0$ (otherwise f, g are not optimal as the value can be improved). So we have

$$
\begin{aligned}
&\langle \mathbf{f}, \mathbf{g}\rangle \\
&= \sum_{v \in \overline{V}_3} \langle f_v, g_v\rangle / n + \sum_{v \in V \setminus \overline{V}_3} \langle f_v, g_v\rangle / n \\
&\geq \Big(\sum_{v \in \overline{V}_3} \langle f_v, g_v\rangle / n\Big) - 1005 \cdot \delta - 2^{5/4}\pi^{5/4} \cdot \sqrt{\lambda} - O(\xi^{c_1}) && \text{(Lemma 3.2)} \\
&= \sum_{v \in \overline{V}_3} (1 - \mathbb{P}_x[f_v(x) \neq g_v(x)])/n - 1005 \cdot \delta - 2^{5/4}\pi^{5/4} \cdot \sqrt{\lambda} - O(\xi^{c_1}) \\
&\geq |\overline{V}_3|/n - 1005 \cdot \delta - 2^{5/4}\pi^{5/4} \cdot \sqrt{\lambda} - O(\xi^{c_1}) - \sum_{v \in \overline{V}_3} \mathbb{P}_x[f_v(x) \neq g_v(x)]/n \\
&\geq 1 - 2010 \cdot \delta - 2^{9/4}\pi^{5/4} \cdot \sqrt{\lambda} - O(\xi^{c_1}) - \sum_{v \in \overline{V}_3} \mathbb{P}_x[f_v(x) \neq g_v(x)]/n && \text{(Lemma 3.2)} \quad (15)
\end{aligned}
$$

Further we have

$$\begin{aligned}
&\sum_{v\in\overline{V}_3} \mathop{\mathbb{P}}_x \left[f_v(x) \neq g_v(x)\right]/n \\
&\leq \sum_{v\in\overline{V}_3} \mathop{\mathbb{P}}_x \left[f_v(x) \neq \mathrm{sgn}(\langle(\mathbf{P}\dot{\widehat{\mathbf{g}}})_v, x\rangle)\right]/n \;+ \sum_{v\in\overline{V}_3} \mathop{\mathbb{P}}_x \left[g_v(x) \neq \mathrm{sgn}(\langle(\mathbf{P}\dot{\widehat{\mathbf{f}}})_v, x\rangle)\right]/n \;+ \\
&\quad \sum_{v\in\overline{V}_3} \mathop{\mathbb{P}}_x \left[\mathrm{sgn}(\langle(\mathbf{P}\dot{\widehat{\mathbf{g}}})_v, x\rangle) \neq \mathrm{sgn}(\langle(\mathbf{P}\dot{\widehat{\mathbf{f}}})_v, x\rangle)\right]/n \\
&\leq \sum_{v\in\overline{V}_3} \mathop{\mathbb{P}}_x \left[\mathrm{sgn}(\langle(\mathbf{P}\dot{\widehat{\mathbf{g}}})_v, x\rangle) \neq \mathrm{sgn}(\langle(\mathbf{P}\dot{\widehat{\mathbf{f}}})_v, x\rangle)\right]/n + 4\sqrt{2\pi}\cdot\lambda + O(\xi^{1/28}) \qquad (16)
\end{aligned}$$

(applying Lemma 2.11 to first two sums with $\varepsilon := \xi^{1/14}$)

Thus we have

$$\begin{aligned}
&\sum_{v\in\overline{V}_3} \mathop{\mathbb{P}}_x \left[\mathrm{sgn}(\langle(\mathbf{P}\dot{\widehat{\mathbf{g}}})_v, x\rangle) \neq \mathrm{sgn}(\langle(\mathbf{P}\dot{\widehat{\mathbf{f}}})_v, x\rangle)\right]/n \\
&\leq O(\xi^{1/84}) + \sum_{v\in\overline{V}_3} 4\sqrt{2}\cdot \|(\mathbf{P}\dot{\widehat{\mathbf{g}}})_v - (\mathbf{P}\dot{\widehat{\mathbf{f}}})_v\|_{\ell_2}/n && \text{(by Lemma 2.12)} \\
&\leq O(\xi^{1/84}) + 4\sqrt{2}\cdot \Big(\sum_{v\in\overline{V}_3} \|(\mathbf{P}\dot{\widehat{\mathbf{g}}})_v - (\mathbf{P}\dot{\widehat{\mathbf{f}}})_v\|_{\ell_2}^2/n\Big)^{1/2} && \text{(Cauchy-Schwarz)} \\
&\leq 8\sqrt{\delta} + O(\xi^{1/84}) && \text{(by (7))}
\end{aligned}$$

Plugging this back in (15) and setting $c_2 := \min\{c_1, 1/84\}$ yields the claim. ■

We are equipped to prove our main result.

Theorem 3.4 ($\pi/2+\varepsilon_0$ NP-Hardness of Grothendieck Optimization Problem)**.** *There exists a constant $\varepsilon_0 \in (0,1)$ such that it is NP-Hard to approximate $\|\cdot\|_{\ell_\infty\to\ell_1}$ within a factor of $\pi/2+\varepsilon_0$.*

Proof. We proceed by showing that the reduction from Smooth Label Cover in Section 3.1 satisfies the following

- (Completeness) If $\mathcal{L}$ is satisfiable, there exists $\mathbf{f},\mathbf{g} \in \{\pm1\}^{V\times\{\pm1\}^R}$ such that $\langle f, \mathbf{A}\mathbf{g}\rangle \geq 1-\lambda$.
- (Soundness) There are absolute constants $C > 1$ and $c_2 \in (0,1)$ such that if $\mathcal{L}$ is a T-to-1 label cover instance for some $T \in \mathbb{N}$ with soundness ξ, smoothness $C\cdot T/\xi$ and weak expansion, then for any $\mathbf{f},\mathbf{g} \in \{\pm1\}^{V\times 2^R}$ we have $\langle \mathbf{f}, \mathbf{A}\mathbf{g}\rangle \leq 2/\pi - \lambda + 32\cdot\lambda^{3/2} + 4020\cdot\lambda^2 + O(\xi^{c_2})$.

Completeness follows from assigning dictators to all vertices via the substitution $f_v(x) := g_v(x) := x_{\ell(v)}$ where ℓ is any assignment completely satisfying $\mathcal{L}$. We then have $\langle \mathbf{f}, \mathbf{A}\mathbf{g}\rangle = \langle \dot{\widehat{\mathbf{f}}}, \dot{\widehat{\mathbf{g}}}\rangle - \lambda\langle \mathbf{f}, \mathbf{g}\rangle = 1-\lambda$ since $\dot{\widehat{\mathbf{f}}}, \dot{\widehat{\mathbf{g}}}$ already lie inside the subspace $\widehat{\mathbf{L}}$ and therefore $\mathbf{P}$ acts as the identity map on $\dot{\widehat{\mathbf{f}}}$ and $\dot{\widehat{\mathbf{g}}}$.

For soundness consider any $\mathbf{f},\mathbf{g} \in \{\pm1\}^{V\times 2^R}$ that are maximizers of $\langle \mathbf{f}, \mathbf{A}\mathbf{g}\rangle$. Let $\delta := 2/\pi + 3\xi^c - \langle \dot{\widehat{\mathbf{f}}}, \mathbf{P}\dot{\widehat{\mathbf{g}}}\rangle$. Since $\langle \mathbf{f}, \mathbf{g}\rangle \leq 1$, we know $\langle \mathbf{f}, \mathbf{A}\mathbf{g}\rangle \leq \langle \dot{\widehat{\mathbf{f}}}, \mathbf{P}\dot{\widehat{\mathbf{g}}}\rangle + \lambda$. Thus we may assume without loss of generality that $\langle \dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}}\rangle > 2/\pi - 2\lambda$ (i.e., $\delta < 2\lambda + 3\xi^c$) since otherwise the soundness claim is already true. We then have

$$
\begin{aligned}
\langle \mathbf{f}, \mathbf{A}\mathbf{g}\rangle &= \langle \dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}}\rangle - \lambda \cdot \langle \mathbf{f}, \mathbf{g}\rangle \\
&= \langle \mathbf{P}\dot{\hat{\mathbf{f}}}, \mathbf{P}\dot{\hat{\mathbf{g}}}\rangle - \lambda \cdot \langle \mathbf{f}, \mathbf{g}\rangle && (\mathbf{P}^2 = \mathbf{P}) \\
&= \|\mathbf{P}\dot{\hat{\mathbf{f}}}\|_2 \cdot \|\mathbf{P}\dot{\hat{\mathbf{g}}}\|_2 - \lambda \cdot \langle \mathbf{f}, \mathbf{g}\rangle && \text{(Cauchy-Schwarz)} \\
&\leq \frac{2}{\pi} + O(\xi^c) - \lambda \cdot \langle \mathbf{f}, \mathbf{g}\rangle && \text{(by Theorem 2.8)} \\
&\leq \frac{2}{\pi} + O(\xi^c) - \lambda \cdot (1 - 2^{9/4}\pi^{5/4}\sqrt{\lambda} - 8\sqrt{\delta} - 2010 \cdot \delta - O(\xi^{c_2})) && \text{(by Lemma 3.3)} \\
&\leq \frac{2}{\pi} - \lambda + 32 \cdot \lambda^{3/2} + 4020 \cdot \lambda^2 + O(\xi^{c_2}) && (\delta \leq 2\lambda + 3\xi^c)
\end{aligned}
$$

This completes the proof of soundness.

By Theorem 2.6 (smooth label cover hardness) we may take ξ to be an arbitrary small constant independent of λ. Thus setting $\lambda := 1/30000$ we obtain an inapproximability factor of at least $\frac{\pi}{2} + 3 \cdot 10^{-6}$ as desired. ■

References

[AN04] Noga Alon and Assaf Naor. Approximating the cut-norm via grothendieck's inequality. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 72–80, 2004. 1

[Ben05] Vidmantas Bentkus. A lyapunov-type bound in R^d. *Theory of Probability & Its Applications*, 49(2):311–323, 2005. 8

[BGG+19] Vijay Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Approximability of pâĘŠ q matrix norms: generalized krivine rounding and hypercontractive hardness. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1358–1368. SIAM, 2019. 2

[BMMN13] Mark Braverman, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. The Grothendieck constant is strictly smaller than KrivineâĂŹs bound. In *Forum of Mathematics, Pi*, volume 1. Cambridge University Press, 2013. Conference version in FOCS '11. 0, 1

[BRS15] Jop Briët, Oded Regev, and Rishi Saket. Tight hardness of the non-commutative Grothendieck problem. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 1108–1122. IEEE, 2015. 0, 1, 2, 7, 9

[Dav84] Alexander Davie. A lower bound for k_g. Manuscript, 1984. 1

[DTV11] Amit Deshpande, Madhur Tulsiani, and Nisheeth K Vishnoi. Algorithms and hardness for subspace approximation. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 482–496. SIAM, 2011. 2

[Gro53] Alexandre Grothendieck. *Résumé de la théorie métrique des produits tensoriels topologiques*. Soc. de Matemática de São Paulo, 1953. 0, 1

[GRSW16] Venkatesan Guruswami, Prasad Raghavendra, Rishi Saket, and Yi Wu. Bypassing UGC from some optimal geometric inapproximability results. *ACM Transactions on Algorithms (TALG)*, 12(1):6, 2016. Conference version in SODA '12. 0, 2, 6, 9

[Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, 2001. 1

[HKM13] Prahladh Harsha, Adam Klivans, and Raghu Meka. An invariance principle for polytopes. *Journal of the ACM (JACM)*, 59(6):1–25, 2013. 9

[Kho02] Subhash Khot. Hardness results for coloring 3-colorable 3-uniform hypergraphs. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 23–32. IEEE, 2002. 6

[KN09] Subhash Khot and Assaf Naor. Approximate kernel clustering. *Mathematika*, 55(1-2):129–165, 2009. 2

[KN12] Subhash Khot and Assaf Naor. Grothendieck-type inequalities in combinatorial optimization. *Communications on Pure and Applied Mathematics*, 65(7):992–1035, 2012. 1

[KNS10] Guy Kindler, Assaf Naor, and Gideon Schechtman. The UGC hardness threshold of the Lp Grothendieck problem. *Mathematics of Operations Research*, 35(2):267–283, 2010. Conference version in SODA '08. 2

[KO09] Subhash Khot and Ryan O'Donnell. SDP gaps and UGC-hardness for Max-Cut-Gain. *Theory OF Computing*, 5:83–117, 2009. 1, 2, 9

[Kri77] Jean-Louis Krivine. Sur la constante de Grothendieck. *CR Acad. Sci. Paris Ser. AB*, 284(8):A445–A446, 1977. 1

[Nes98] Yurii Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization methods and software*, 9(1-3):141–160, 1998. 2

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014. 8

[Pis12] Gilles Pisier. Grothendieck's theorem, past and present. *Bulletin of the American Mathematical Society*, 49(2):237–323, 2012. 1

[Ree91] JA Reeds. A new lower bound on the real Grothendieck constant. Manuscript, 1991. 1

[Rie74] Ronald E Rietz. A proof of the grothendieck inequality. *Israel journal of mathematics*, 19(3):271–276, 1974. 2

[RS09] Prasad Raghavendra and David Steurer. Towards computing the Grothendieck constant. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 525–534. Society for Industrial and Applied Mathematics, 2009. 1, 2