

Hardness of Approximation for $2/4$ Norms through Quantum Reductions

*Rohit Agarwal
Axel Li
Anthony Maltsev
Anirba Sarkar*



Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2023-236

<http://www2.eecs.berkeley.edu/Pubs/TechRpts/2023/EECS-2023-236.html>

November 20, 2023

Copyright © 2023, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Hardness of Approximation for $2 \rightarrow 4$ Norms through Quantum Reductions

Rohit Agarwal
rohaga@berkeley.edu

Axel Li
axel.li@berkeley.edu

Anthony Maltsev
amaltsev@berkeley.edu

Anirban Sarkar
sarkar@berkeley.edu

March 2023

Abstract

We study the computational complexity of approximating $2 \rightarrow 4$ linear operator norm, defined as $\|A\|_{2 \rightarrow 4} = \max_{f \neq 0} (\|Af\|_4 / \|f\|_2)$. We explore the problem of multiplicatively approximating to such a norm to a constant factor. We present the previous results in the area, share our attempt to give a new NP-hardness proof, and discuss applications to other problems, such as Khot's Unique Games Conjecture [7].

Contribution Statement

This was a joint project for both CS270 and CS191. Anirban, Rohit, and Anthony were on the CS191 side while Anirban, Rohit, and Axel were on the CS270 side. All four of us read most of the papers together. Anirban studied quantum complexity and wrote section 3.1, Anthony focused on the role of quantum in the classical reduction and wrote section 3.2, Axel wrote section 3.3, focusing on the classical reductions, and Rohit wrote section 4, came up with the partial directions showcased in 3.3 and the appendix, as well as edited all of the sections.

1 Introduction

We study the computational complexity of approximating $2 \rightarrow 4$ linear operator norm, defined as

$$\|A\|_{2 \rightarrow 4} = \max_{f \neq 0} \frac{\|Af\|_4}{\|f\|_2}$$

We explore the problem of multiplicatively approximating to such a norm to a constant factor, e.g. for constants $1 < c < C$ deciding between the cases $\|A\|_{2 \rightarrow 4} \geq C\sigma$ and $\|A\|_{2 \rightarrow 4} \leq c\sigma$ where σ is the minimum singular value of A ; we shall call such an approximation a *good approximation* of $\|A\|_{2 \rightarrow 4}$. As a polynomial optimization over the unit sphere, this problem is of great importance to quantum information theory and in resolving Khot's Unique Games Conjecture [7].

First, we discuss the preliminaries from Fourier Analysis and quantum information needed. Then, we will give an introduction to the quantum information notions used in hardness of approximation results for this norm, including describing the quantum complexity class QMA. Next, we follow the paper of Barak et. al. [1] and attempt to explain the following reduction from Section 9.1 in a simpler fashion than the original work.

Theorem 1 (Informal Version). *Consider a 3-SAT instance ϕ , with n variables and $O(n)$ clauses. Deciding whether ϕ is satisfiable can be reduced to finding a good approximation to $\|A\|_{2 \rightarrow 4}$ where the dimensions of A are $m \times m$ where $m \approx \exp(\tilde{O}(\sqrt{n}))$.*

Thus, assuming the Exponential Time Hypothesis (ETH), one cannot give a good approximation for $\|A\|_{2 \rightarrow 4}$ in polynomial time. Afterwards, we will attempt to prove the NP-Hardness through classical reductions, namely through the classic promise problem of smooth label cover. We explain some of the roadblocks in why our current approaches do not work and show a technique that may have promise. Lastly, we provide a discussion of current/possible future results, what they imply about the Unique Games Conjecture (UGC), and related problems.

2 Notation and Preliminaries

2.1 Linear Algebra/Fourier Analysis

We will use the functional definitions of norms unless otherwise stated, as they are more ubiquitous in the literature ([1], [3]). Fix a natural number R , which will be the number of bits needed to index into vectors (the dimension of the space will be 2^R). We define L_2 as the set of all functions $f : \{\pm 1\}^R \rightarrow \mathbb{R}$ endowed with inner product

$$\langle f, g \rangle = \mathbb{E}_{x \sim \text{Unif}\{\pm 1\}^n} [f(x)g(x)] = \frac{1}{2^n} \sum_{x \in \mathcal{U}} f(x)g(x)$$

and induced norm $\|f\|_2 = \sqrt{\langle f, f \rangle}$. More generally, we define the p -norm as $\|f\|_p = \mathbb{E}_x [f(x)^p]^{1/p}$. Then, the $p \rightarrow q$ norm is

$$\|A\|_{p \rightarrow q} = \max_{0 \neq f \in L_2} \frac{\|Af\|_q}{\|f\|_p}$$

For any $f : \{\pm 1\}^R \rightarrow \mathbb{R}$, define the function $\hat{f} : [R] \rightarrow \mathbb{R}$ as $\hat{f}(i) = \mathbb{E}_x (x_i \cdot f(x))$. This is called the (partial) Fourier transform of f , which we shall denote as $\hat{f} = F_P f$. For vectors in the Fourier space, we endow them with the counting inner product and norm

$$\langle \hat{f}, \hat{g} \rangle = \sum_{S \subseteq [R]} \hat{f}(S) \hat{g}(S) \qquad \|\hat{f}\|_p = \left(\sum_{S \subseteq [R]} \hat{f}(S)^p \right)^{1/p}$$

Its inverse, which is also its adjoint, satisfies:

$$F_P^T \hat{f} = \sum_{i=1}^R \hat{f}(i) x_i$$

We will use the following theorem.

Theorem 2. *For a function $f : \{\pm 1\}^R \rightarrow \mathbb{R}$, and a function $\hat{g} : [R] \rightarrow \mathbb{R}$, $\|f\|_2 \geq \|F_P f\|_2$ and $\|F_P^T \hat{g}\|_2 = \|\hat{g}\|_2$. In addition, $F_P^T F_P f = f$ if and only if f is linear.*

2.2 Quantum Information

We follow the notions from [1]. Let $L(V)$ be the set of linear operators on a vector space V . Let $\mathbf{S}(V)$ be the set of all L_2 unit vectors on V . The set of density matrices over V is defined as:

$$\mathcal{D}(V) = \{\rho \in L(V) : \rho \succeq 0, \text{Tr } \rho = 1\} = \text{conv}\{xx^* : x \in \mathbf{S}(V)\}$$

Where $\text{conv}\{\cdot\}$ is the convex hull. We define the set of separable states over k systems then as

$$\text{Sep}^k(n) = \text{conv}\{x_1x_1^* \otimes \cdots \otimes x_kx_k^* : x_1 \in \mathbf{S}(\mathbb{C}^n), \dots, x_k \in \mathbf{S}(\mathbb{C}^n)\}$$

This corresponds to operators of the form $\rho = \sum_k p_k(\rho_{1k} \otimes \rho_{2k})$, where $\sum_k p_k = 1$. If we want to use real vectors instead, we'll denote this as $\text{Sep}^k(\mathbb{R}^n)$. Finally, we define the support function of M as

$$\mathbf{h}_K(M) = \max_{\rho \in K} |\text{Tr}(M\rho^*)|$$

One can interpret this as the maximum probability that a quantum measurement implemented by the matrix M returns true for a state in K (which will typically be Sep).

2.3 Quantum Complexity

Quantum Merlin Arthur (QMA) is a quantum complexity class that is the set of all languages L that can be solved with a quantum verifier V . $\mathbf{QMA}_{f(m)}(k)_{c,s}$ is the set of all languages L that have a quantum verifier V such that for an $x \in \{0, 1\}^n$:

- (Completeness) If $x \in L$, then there exists k proofs ψ_1, \dots, ψ_k , all $l = O(f(m))$ qubits long, such that running V on input $(x \otimes \psi_1 \otimes \cdots \otimes \psi_k)$ accepts w.p. $\geq c$.
- (Soundness) If $x \notin L$, then $\forall \psi_1, \dots, \psi_k$, all at most $l = O(f(m))$ qubits long, running V on input $(x \otimes \psi_1 \otimes \cdots \otimes \psi_k)$ accepts w.p. $\leq s$.

If c and s are not specified, they are assumed to be $\frac{2}{3}, \frac{1}{3}$ respectively.

Swap Test: The swap test is a method used to check how much two states differ. It takes two input states of equal dimension, ρ and σ , as input. It returns either 0 or 1, which corresponds to whether the two states ρ and σ are close or far in inner product respectively. The test returns 0 with probability $\frac{1}{2} + \frac{1}{2} \text{Tr}(\rho\sigma)$.

Product Test: The product test follows these steps, and detects if a state is entangled:

1. Create two copies of $\psi \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$. Let these two copies be ψ_1, ψ_2 .
2. Perform the swap test on each of the n pairs of corresponding subsystems in ψ_1, ψ_2 .
3. If all of the swap tests returned 0, corresponding to the states being the same, then accept. Else, reject.

Soleimanifar and Wright show in [11] that we can upper bound the product test, which tells us that the product test will detect the distance from our product state.

Theorem 3. For state $\psi \in \mathbb{C}^n \otimes \mathbb{C}^n$ with overlap $\omega = \max_{\phi \in \text{Sep}^2(n), \phi \text{ is pure}} |\langle \psi, \phi \rangle|^2$, the product test has acceptance probability at most:

$$PT_n(\omega) \leq \begin{cases} \omega^2 - \omega + 1 & \omega \geq \frac{1}{2} \\ \frac{1}{3}\omega^2 + \frac{2}{3} & \omega < \frac{1}{2} \end{cases}$$

2.4 Label Cover

We follow the complete graph formulation of [3]. An instance of Label Cover is given by a tuple $\mathcal{L} = (G, [R], [L], \Sigma)$ that consists of a regular connected graph $G = (V, E)$, a label set $[R]$ and a collection of maps $\Sigma = \{(\pi_{e,v}, \pi_{e,w}) : e = (v, w) \in E\}$ where $\pi_{e,v} : [R] \rightarrow [L]$ for all adjacent $v \in V, e \in E$. We say for an edge $e = (v, w)$ is satisfied by some labeling $\ell : V \rightarrow [R]$ if $\pi_{e,v}(\ell(v)) = \pi_{e,w}(\ell(w))$. $\text{OPT}(\mathcal{L})$ is the maximum proportion of edges satisfied by any labeling. It is known ([3]) that the Gap version of this problem is hard.

Theorem 4. For any $\xi > 0$ there exist positive integers R and L and a label cover instance $\mathcal{L} = (G, [R], [L], \Pi)$ such that deciding between $\text{OPT}(\mathcal{L}) = 1$ and $\text{OPT}(\mathcal{L}) \leq \xi$ is NP-Hard.

3 Main Results

3.1 QMA and the Role of 3-SAT

Because 3-SAT is in the complexity class of $\mathbf{QMA}(2)$, We are able to use the hardness of 3-SAT to prove results about the hardness of $\mathbf{QMA}_{\log}(2)$. More specifically, Harrow and Montanaro show in [6] that if $\ell : \mathbb{N} \rightarrow \mathbb{N}$ and is polynomially bounded, then $\mathbf{3-SAT} \in \mathbf{QMA}_{\ell(n)\sqrt{n}\text{polylog}(n)}(2)_{2^{-\ell(n)},1}$.

As an aside, recall the definitions of \mathbf{QMA} and the product test. Harrow and Montanaro also show that $\mathbf{QMA}_n(k) = \mathbf{QMA}_n(2)$ for $2 \leq k \leq \text{poly}(n)$ where n is the size of the proof with the following protocol:

1. Create two copies of $\psi \in \mathbb{C}^{d_1} \otimes \dots \otimes \mathbb{C}^{d_n}$. Let these two copies be ψ_1, ψ_2 .
2. Do exactly one of the following, each with probability $\frac{1}{2}$:
 - (a) Run the product test on ψ_1 and ψ_2 and accept iff the product test accepts.
 - (b) Randomly pick either ψ_1 or ψ_2 and then run the quantum verifier V on this state, and accept iff the verifier accepts.

Completeness for this protocol follows directly, as it either runs the product test, which will always return the correct value, or we run the verifier on one of the two states, and the verifier is defined to accept with a probability of at least c . Therefore completeness holds. To show soundness, we find an upper bound on the probability that the product test will accept $\psi_1 \otimes \psi_2$, which eventually yields a soundness bound of $\frac{(1-s)^2}{100}$. Thus, 3-SAT is instrumental in collapsing the quantum proof size hierarchy, and this reduction relies on the product test result.

For our purposes, the Exponential Time Hypothesis states that $\mathbf{3-SAT} \notin \mathbf{DTIME}(\exp(\ell(n)))$ for any function $\ell : \mathbb{N} \rightarrow \mathbb{N}$ for $\ell(n) = o(n)$. Here, n is the number of clauses on the 3-SAT. Assuming ETH is true, then we can conclude the following:

1. $\mathbf{QMA}_{\log(d)}(2)_{\frac{1}{2},1} \not\subseteq \mathbf{DTIME}(d^{\log^{1-\epsilon} d})$ for arbitrary $\epsilon > 0$
2. $\mathbf{QMA}_{\log(d)}(2)_{2^{-\sqrt{\log(d)}/\text{polylog}(\log(d))},1} \not\subseteq \mathbf{DTIME}(\text{poly}(d))$

These two results allow us to make stronger hardness claims for $\mathbf{QMA}_{\log(d)}(2)$, and subsequently $\mathbf{h}_{\text{Sep}(d,d)}$. Harrow and Montanaro showed that $\mathbf{QMA}_m(2)_{c,s}$ is the set of languages that can be decided by checking whether $\mathbf{h}_{\text{Sep}(2^m,2^m)}(M)$ is $\geq c$ or if it is $\leq s$ where M is a measurement operator that is able to be constructed in polynomial time on a quantum computer. This, in turn, allows us to make stronger claims about the hardness of $\mathbf{h}_{\text{Sep}(d,d)}$, which we will discuss more in the next section.

3.2 Simplifying the Barak et. al. Proof

One key result of a paper by Barak et. al. paper [1] is that approximating the $2 \rightarrow 4$ norm for a specially constructed matrix is sufficient to approximate \mathbf{h}_{Sep} to any desired level of precision and thus resolve 3-SAT. We roughly reproduce the proof of Theorem 1 here with additional insights. The first main idea of the proof is to establish equivalences between the $2 \rightarrow 4$ norm and several other norms and \mathbf{h}_{Sep} for matrices of a specific structure.

Let us first define the injective tensor norm $\|\cdot\|_{inj}$ for a vector $T \in V_1 \otimes V_2 \otimes \dots \otimes V_r$ where the V_i are some vector spaces:

$$\|T\|_{inj} = \max_{x_1 \in \mathbf{S}(V_1) \dots x_r \in \mathbf{S}(V_r)} |\langle T, x_1 \otimes \dots \otimes x_r \rangle|$$

Let us then define the symmetric subspace $\vee^r \mathbb{F}^n$ as the subspace of $(\mathbb{F}^n)^{\otimes r}$ that is invariant under any permutation of its elements. For example, if $x \otimes y \otimes z$ is in $\vee^3 \mathbb{F}^n$, then $x \otimes y \otimes z = y \otimes z \otimes x = z \otimes y \otimes x$, or any other such permutation. Then, for a matrix $T \in \vee^r \mathbb{F}^n$, the injective tensor norm can be simplified:

$$\|T\|_{inj} = \max_{x \in \mathbf{S}(\mathbb{F})} |\langle T, x^{\otimes r} \rangle|$$

This leads to a connection to the $2 \rightarrow 4$ norm of a matrix A . If we define a new tensor A_4 based on the rows a_1, \dots, a_n of A as $A_4 = \sum_{i=1}^n a_i^{\otimes 4}$, then, by direct calculation:

$$\begin{aligned} \|A\|_{2 \rightarrow 4}^4 &= \max_{x \neq 0} \left(\frac{\|Ax\|_4}{\|x\|_2} \right)^4 = \max_{x \in \mathbf{S}(\mathbb{R}^n)} \|Ax\|_4^4 \\ &= \max_{x \in \mathbf{S}(\mathbb{R}^n)} \sum_{i=1}^n \langle a_i, x \rangle^4 = \max_{x \in \mathbf{S}(\mathbb{R}^n)} \langle A_4, x^{\otimes 4} \rangle \\ &= \|A_4\|_{inj} \end{aligned}$$

The last line follows from the fact that $A_4 \in \mathbb{V}^4 \mathbb{R}^n$. Now we wish to draw a connection from the injective tensor norm of A_4 to \mathbf{h}_{Sep} . The key property that we will exploit to make this connection is the following: a tensor $T \in (\mathbb{R}^n)^{\otimes k}$ can always be written in the form $\sum_{i=1}^n T_i \otimes e_i$ where $T_i \in (\mathbb{R}^n)^{\otimes k-1}$. Then:

$$\|T\|_{inj}^2 = \mathbf{h}_{\text{Sep}^{k-1}(\mathbb{R}^n)} \left(\sum_{i=1}^n T_i T_i^* \right)$$

This is proven in [1] through direct calculation. Instead of applying this directly to A_4 , we define an intermediate matrix $A_3 = \sum_{i=1}^n a_i \otimes a_i \otimes e_i$. It is shown in [1], again through direct calculation, that $\|A_4\|_{inj} = \|A_3\|_{inj}^2$. Note that this matrix is defined in a way that can be used directly by the aforementioned key property. Then:

$$\|A_3\|_{inj}^2 = \mathbf{h}_{\text{Sep}^2(\mathbb{R}^n)} \left(\sum_{i=1}^n (a_i \otimes a_i)(a_i^* \otimes a_i^*) \right) = \mathbf{h}_{\text{Sep}^2(\mathbb{R}^n)} \left(\sum_{i=1}^n a_i a_i^* \otimes a_i a_i^* \right)$$

Let us trace the line of equivalences we have made:

$$\|A\|_{2 \rightarrow 4}^4 = \|A_4\|_{inj} = \|A_3\|_{inj}^2 = \mathbf{h}_{\text{Sep}^2(\mathbb{R}^n)} \left(\sum_{i=1}^n a_i a_i^* \otimes a_i a_i^* \right)$$

Every step we have taken works backward as well. This has a critical consequence: $\mathbf{h}_{\text{Sep}^2(\mathbb{R}^n)} \left(\sum_{i=1}^m x_i x_i^* \otimes x_i x_i^* \right)$ for any set of vectors $x \in \mathbb{R}^n$ can be expressed as the $\|\cdot\|_{2 \rightarrow 4}$ norm of some matrix of dimension m by n that can be efficiently constructed from x . This is the starting point of the rest of the proof: we will try to find a reduction from 3-SAT to the problem of approximating $\mathbf{h}_{\text{Sep}} \left(\sum x x^* \otimes x x^* \right)$ for some set of vectors x .

A reduction from 3-SAT to the problem of approximating $\mathbf{h}_{\text{Sep}}(M)$ for a p.s.d. matrix $0 \preceq M \preceq I$ is already known; this follows from the quantum complexity of 3-SAT. In particular, since it is in $\mathbf{QMA}_{\tilde{O}(\sqrt{n})}(2)$, then there exists a verifier V which accepts two small-sized proofs, i.e. a product state over two vectors in $\mathbb{C}^{2^{\tilde{O}(\sqrt{n})}}$. Recall that this means that $\mathbf{h}_{\text{Sep}}(V)$ represents the maximum probability a proof can be accepted by V . Then, the remaining work is to show that we can transform M into some M' of the desired form. More formally, if we are given that $\mathbf{h}_{\text{Sep}}(M)$ is either 1 or below $(1 - \delta)$ (a so-called promise problem), we wish to find a related matrix M' such that $\mathbf{h}_{\text{Sep}}(M')$ is also either 1 or below $(1 - \delta)$.

The main idea of the transformation to M' is best understood through a quantum lens: it is based on the product test studied by Harrow and Montanaro in [6]. We desire our resulting matrix to be in the form of a quantum state tensored with itself. The product test would accept such a state with probability 1, and it would reject states not in that form with some constant probability. We can implement this as a projection matrix, where things that are close to products are mostly preserved, and things far from products are attenuated. The projector P onto the subspace symmetric under permutation of the 1st and 3rd element and 2nd and 4th elements (denoted $P_n(1,3)$ and $P_n(2,4)$) can be thought of as applying the product test to a 2-subsystem system $(a \otimes b) \hat{\otimes} (c \otimes d)$. The notation $\hat{\otimes}$ denotes the cut across which the product test is performed. Formally, the projector matrix is:

$$P = \frac{I + P_n(1,3)}{2} \cdot \frac{I + P_n(2,4)}{2} = \mathbb{E}_{a,b}(aa^* \otimes bb^* \hat{\otimes} aa^* \otimes bb^*)$$

The second equality comes from an application of Isserlis'/Wick's theorem. The expectation is over complex gaussian distributed vectors a and b with $\mathbb{E}_a \|a\|_2^2 = \mathbb{E}_b \|b\|_2^2 = \frac{n}{\sqrt{2}}$. This form is easier to work

with, so we will use the expectation formulation of P . In this form, we see that we are close to the form that we desired for our matrix. We want to use this projector to relate M to some M_1 . We will do this as follows. Define M_1 as:

$$M_1 = (\sqrt{M} \hat{\otimes} \sqrt{M}) P (\sqrt{M} \hat{\otimes} \sqrt{M})$$

Here \sqrt{M} denotes the unique square root of positive matrix M (it is a POVM): $\sqrt{M} = U \sqrt{\Lambda} U^T$ where $M = U \Lambda U^T$ is the spectral decomposition of M . Plugging in the expectation form for P :

$$\begin{aligned} M_1 &= (\sqrt{M} \hat{\otimes} \sqrt{M}) \mathbb{E}_{a,b} (aa^* \otimes bb^* \hat{\otimes} aa^* \otimes bb^*) (\sqrt{M} \hat{\otimes} \sqrt{M}) \\ &= \mathbb{E}_{a,b} ((\sqrt{M} aa^* \otimes bb^* \sqrt{M}) \hat{\otimes} (\sqrt{M} aa^* \otimes bb^* \sqrt{M})) \\ &= \mathbb{E}_{a,b} (((\sqrt{M} a \otimes b)(a^* \otimes b^* \sqrt{M})) \hat{\otimes} ((\sqrt{M} a \otimes b)(a^* \otimes b^* \sqrt{M}))) \end{aligned}$$

Let us define $v_{a,b} = \sqrt{M}(a \otimes b)$. Then, we can see that,

$$M_1 = \mathbb{E}_{a,b} (v_{a,b} v_{a,b}^* \hat{\otimes} v_{a,b} v_{a,b}^*)$$

Recall that we want some matrix that can be written in the form $\sum xx^* \hat{\otimes} xx^*$. Since the expectation is a weighted sum of elements, this matrix M_1 is in the desired form. However, if we were to reconstruct a matrix A such that $\|A\|_{2 \rightarrow 4} = \mathbf{h}_{\text{Sep}}(M_1)$, this matrix A would have infinite dimension because the expectation is a sum over an infinite amount of tensors. This is insufficient for any hardness claims so we wish to find some finite sized equivalent matrix. For this purpose, we can apply Caratheodory's Theorem to M_1 :

$$M_1 = \mathbb{E}_{a,b} (v_{a,b} v_{a,b}^* \hat{\otimes} v_{a,b} v_{a,b}^*) = \sum_{i=1}^{n^2} z_i z_i^* \hat{\otimes} z_i z_i^*$$

Caratheodory's theorem states that for any point inside the convex hull of some set can be represented as the convex combination of a small subset of the points that define that convex hull. The size of this subset is defined by the dimension of the space containing that convex hull. It is clear that $\mathbb{E}_{a,b} (v_{a,b} v_{a,b}^* \hat{\otimes} v_{a,b} v_{a,b}^*)$ is a point inside the convex hull of points of the form $v_{a,b} v_{a,b}^* \hat{\otimes} v_{a,b} v_{a,b}^*$. Then, applying the theorem, we get the finite summation representation of $M_1 = \sum_{i=1}^{n^2} z_i z_i^* \hat{\otimes} z_i z_i^*$ as above. There are n^2 vectors needed because the dimension of $v_{a,b}$ is n^2 , as each of a and b are of dimension n . Note that we could have just found a gadget summation that captured this instead of resorting to Wick's theorem. However, finding such a gadget is tricky, as the number of terms explodes quite substantially even for small n .

We have constructed a matrix M_1 related to M that is in the desired form to connect to the $2 \rightarrow 4$ norm, but does this matrix M_1 serve as a good approximation of M ? We claim that it does. Formally:

$$\mathbf{h}_{\text{Sep}^2(n^2)}(M_1) \begin{cases} = 1 & \mathbf{h}_{\text{Sep}^2(n)}(M) = 1 \text{ (Case Y)} \\ \leq 1 - \delta/2 & \mathbf{h}_{\text{Sep}^2(n)}(M) \leq 1 - \delta \text{ (Case N)} \end{cases}$$

A proof that includes mechanical details of this is provided at the end of section 9.2 of [1]. We will discuss the intuition of the proof instead.

In Case Y, there must exist some separable state $x \otimes y$ with x, y of dimension n and magnitude 1 that satisfy $\text{Tr}(M(xx^* \otimes yy^*)) = 1$. Then, $\langle x \otimes y, Mx \otimes y \rangle = 1$. This means that $x \otimes y$ is an eigenvector of M with eigenvalue 1. Then, we can define $z = (x \otimes y) \hat{\otimes} (x \otimes y)$ such that $\text{Tr}(M_1(zz^*)) = 1$. This can be broken up into $z^* (\sqrt{M} \hat{\otimes} \sqrt{M}) P (\sqrt{M} \hat{\otimes} \sqrt{M}) z$. z is an eigenvector of $(\sqrt{M} \hat{\otimes} \sqrt{M})$ with eigenvalue 1 because $x \otimes y$ was an eigenvector of M and \sqrt{M} with eigenvalue 1. Similarly, z is an eigenvector of P with eigenvalue 1 by construction. Then, $\text{Tr}(M_1 z z^*) = z^* z = 1$.

The idea of the proof for Case N is similar: either $(\sqrt{M} \hat{\otimes} \sqrt{M})$ or P will shrink any vector $z = x \otimes y$ where x, y are of dimension n^2 and norm 1. If x and y are tensor products of the form $a \otimes b$ where a, b are of dimension n , then $(\sqrt{M} \hat{\otimes} \sqrt{M})$ must shrink x and y since $\mathbf{h}_{\text{Sep}^2(n)}(M) \leq 1 - \delta$. Otherwise, if x and y are not shrunk significantly by $(\sqrt{M} \hat{\otimes} \sqrt{M})$, then they must be shrunk by the projection matrix P . This last argument is analogous to the fact that the product test has a small probability of success for non product states.

Thus, we have a matrix M_1 of the desired form that achieves about the same separation between cases Y and N as M does. However, we would like a matrix that achieves the same or stronger separation between cases. Therefore, we want to achieve some completeness-soundness gap amplification. This is achieved by constructing a matrix M_2 from M_1 that amplifies the gap while maintaining the desired

form. Recall that $M_1 = \mathbb{E}_{a,b}(v_{a,b}v_{a,b}^* \hat{\otimes} v_{a,b}v_{a,b}^*) = \sum_{i=1}^{n^2} z_i z_i^* \hat{\otimes} z_i z_i^*$. We will define M_2 as tensoring M_1 with itself k times. Specifically, this tensoring is not across the product cut.

$$\begin{aligned} M_2 &= M_1^{\otimes k} = \mathbb{E}_{a_1 \dots a_k, b_1 \dots b_k}((v_{a_1, b_1} v_{a_1, b_1}^* \otimes \dots \otimes v_{a_k, b_k} v_{a_k, b_k}^*) \hat{\otimes} (v_{a_1, b_1} v_{a_1, b_1}^* \otimes \dots \otimes v_{a_k, b_k} v_{a_k, b_k}^*)) \\ &= \sum_{i_1, i_2, \dots, i_k \in [n^2]} (z_{i_1} z_{i_1}^* \otimes \dots \otimes z_{i_k} z_{i_k}^*) \hat{\otimes} (z_{i_1} z_{i_1}^* \otimes \dots \otimes z_{i_k} z_{i_k}^*) \end{aligned}$$

We now claim that $\mathbf{h}_{\text{Sep}^2(n^{2k})}(M_2) = (\mathbf{h}_{\text{Sep}^2(n^2)}(M_1))^k$ by Lemma 10 of [6], but once again this fact can be seen in a quantum information way: if the probability of accepting the test M_1 is at most p , then the probability of accepting k parallel copies of the test M_1 is at most p^k . Thus, we can amplify the soundness-completeness gap however much we want at the cost of growing the resulting matrix.

This concludes the proof that 3-SAT reduces to approximating $\|A\|_{2 \rightarrow 4}$ for some matrix A . Recall that n is the number of states in the corresponding proof provided to the QMA protocol for 3-SAT. The proofs provided are $\tilde{O}(\sqrt{n'})$ qubits long, so $n = \exp(\tilde{O}(\sqrt{n'}))$, where n' denotes the number of variables in the 3-SAT instance ϕ that we are reducing from. The conclusion, then, is that determining whether there exists a satisfying assignment of variables for a 3-SAT instance ϕ reduces to approximating $\|A\|_{2 \rightarrow 4}$ for a related A where A has sub-exponential dimensions $\exp(\tilde{O}(\sqrt{n'}))$. This implies, assuming the Exponential Time Hypothesis, that approximating $\|A\|_{2 \rightarrow 4}$ is NP-Hard. This presentation of the proof excludes some details for the sake of clarity, but we include all of the main ideas necessary to understand the logic of the proof. One such detail is that the norm of the matrix A is defined to be achieved with real vectors, but quantum proofs can be complex vectors, so an additional step of converting complex vectors to real vectors is needed (this is done by multiplication by a constant gadget matrix).

3.3 Attempts at Showing Full NP-Hardness

To show full NP-hardness, we move away from the language of Quantum Information and instead present a reduction from a classically hard problem to approximate, the problem of Gap Label Cover. We attempt to follow the proof of [3], who are able to prove hardness for approximating $2 \rightarrow r$ norms for $r < 2$, though we were ultimately unsuccessful along this path.

At a high-level, what the NP-Hardness proof in [3] does is replace every vertex label of the label cover problem with a Boolean hypercube function. A covering (proof) of perfect completeness can be represented as hypercube functions that are all dictator functions $f_v(x) = x_{\ell(v)}$. A projection matrix is used to verify the consistency of the label cover constraints; e.g. whether they (approximately) fit the constraints. A codeword test, i.e. a local test that ensures that each function is actually a dictator function, exists implicitly in the construction of the matrix and causes unsatisfactory functions \mathbf{f} to have small norm when transformed, resulting in a small $2 \rightarrow r$ norm if no good label cover can be found.

More specifically, for a particular label covering problem $\mathcal{L} = (G, [R], [L], \Sigma)$ where $G = (V, E)$, the square matrix $A = F_P^\top \hat{P} F_P$ of size $|V|2^R$ is defined, where F_P is the partial Fourier transform defined previously and \hat{P} is an orthogonal projection onto the subspace containing all valid dictatorship encodings of the labels for a valid label cover instance (but it has other vectors, too). Effectively, this projection checks for label cover consistency in the Fourier domain, shrinking the norm for vectors that don't correspond to a correct label cover solution.

For instances that have a satisfying assignment of labels, there is a corresponding vector \mathbf{f} which has f_v be a dictator function for the corresponding label value of vertex v , for every vertex v . This results in $A\mathbf{f} = \mathbf{f}$, so $\|A\|_{2 \rightarrow r} \geq 1$. On the other hand, if there is no good solution to the label cover instance, then the matrix A will have $2 \rightarrow r$ norm bounded by a constant strictly less than 1 for fixed r via lemma A.2 of [3], which acts as a codeword test.

This method works for $r < 2$ due to the convexity of the norms. For a fixed value of the 2 norm (an L_2 ball), maximizing the r norm means distributing the mass equally amongst its components. Thus, we cannot generalize directly as the opposite behavior occurs in the $r > 2$ domain, leading to the possibility of ‘‘cheating’’ solutions that have concentrated mass and nearly all elements around 0 (and thus are not valid dictators/labelings) but still have a high objective value.

We attempted to modify this approach by defining the 1, 4 mixed vector norm as follows: $\|g\|_{1,4} = \mathbb{E}_v[\mathbb{E}_x[|g_v(x)|^4]^{1/4}]$. Consider the matrix C such that $C\mathbf{g} = \frac{1}{|V|} \sum_{v \in V} g_v$, and define $B = CA$. By the triangle inequality, $\|B\|_{2 \rightarrow 4} \leq \|A\|_{2 \rightarrow 1,4}$. In fact, if we're in the case where $\text{OPT}(\mathcal{L}) \leq \xi(\varepsilon)$, then $\|A\|_{2 \rightarrow 1,4} \leq 3\varepsilon 2^{R/4} + \gamma_4 + \varepsilon$ by Theorem 5 in the Appendix. This will be our soundness. On the other

hand, if the label cover is satisfiable there exists some \mathbf{f} such that $A\mathbf{f} = \mathbf{f}$, so $B\mathbf{f} = \frac{1}{|V|} \sum_{v \in V} f_v$. With a possible degradation of parameters (R becomes $|V|R$), we can state that the labels are distinct and thus the dictators f_v 's are linearly independent, so there exists at least one nonzero entry of $B\mathbf{f}$. Since f_v consists of ± 1 for all v , $\|B\mathbf{f}\|_4 \geq \frac{1}{|V|^{2R}}$.

Due to our inexperience, this approach ended up being problematic in multiple ways, however, and we were not able to fix the flaws. To start off with, the $\gamma_4 > 1$ term automatically disqualifies this method; we had hoped that a stronger alternative to lemma A.2 of [3] could be applied, but there is the bigger issue of parameters. The degradation of parameters to ensure our claim of linear independents creates a possibly exponential dependence on $|V|$ in R , when we had previously taken R to be fixed. Finally, we were working with the belief that ε could be set arbitrarily small to create a gap between $\|B\|_{2 \rightarrow 4}$ between the completeness and soundness cases. However, in this setting, $R = \text{poly}(1/\varepsilon)$, meaning that no such gap could be proven. All of these troubles led to this method being abandoned ultimately. However, we think that considering other, more exotic norms is a relatively unexplored and promising technique for this problem.

4 Discussion - Applications to Unique Games

The previous results only show an ETH-based hardness, but a full NP-hardness would be a large result for the reasons we discuss below.

4.1 Unique Games Conjecture

Let's reframe the Label Cover problem slightly differently, inspired by [8]. Instead of having a complete graph G , imagine having a bipartite directed graph $G' = (V \cup W, E)$, and instead of a function for each edge-vertex incidence, we have a projection for each edge $\pi_e : [R] \rightarrow [L]$ and satisfying an edge (v, w) means $\pi_e(\ell(v)) = \ell(w)$ (where $\ell(V) \subseteq [R]$ and $\ell(W) \subseteq [L]$). For a sketch of why these two settings are equivalent, consider making a vertex in V for each vertex in G and for each edge make a vertex in W . Finally connect every vertex in V with its incident edges in G with the projections $\pi_{(v,e)} = \pi_{e,v}$. This exactly doubles the number of edges, so this exactly reduces the covering by a constant factor; the graph doesn't get much bigger either. One can think of this as a probabilistic game—a challenger names an edge to two provers and the provers must return labels for each of the vertices across the edge. The proof is accepted if the projection condition for that edge is satisfied. In this sense, the problem is sometimes called the 2-Prover-1-Round Game.

There is a slightly related problem; relax the graph to be any directed graph, not necessarily bipartite, have only one label set, and use projections that are bijections. Concretely,

Definition 1. *An instance of the Unique Game Problem is $U(G(V, E), [n], \{\pi_e \mid e \in E\})$. The goal is to assign to each vertex a label from the set $[n]$. The constraint on an edge $e = (v, w) \in E$ is described by a bijection $\pi_e : [n] \rightarrow [n]$. A labeling $\ell : V \rightarrow [n]$ satisfies edge $e = (v, w)$ if and only if $\pi_e(\ell(v)) = \ell(w)$. Let $OPT(U)$ denote the maximum fraction of constraints that can be satisfied by any labeling:*

$$OPT(U) = \max_{\ell} \frac{1}{|E|} |\{e \in E : \ell \text{ satisfies } e\}|$$

This trivializes the completeness case; since the constraint functions are bijections, one can just try every possibility for one of the vertices and fill in the rest in its connected component efficiently (e.g. Breadth-first search). The Unique Games Conjecture (UGC) in [7] is then that gap version of this problem is NP-hard. In particular, it says calling the optimal fraction of edges that can be satisfied OPT , that it is NP-hard to distinguish the cases where $OPT \geq 1 - \epsilon$ and $OPT \leq \delta$ for every $\delta, \epsilon > 0$ (the problem size is a constant $n = n(\delta, \epsilon)$). We briefly mention that this problem has known sub-exponential algorithms (unlike say, 3-SAT), the most famous of which are based on Semidefinite Programming (SDP) and Rounding [2]. In fact, many these algorithms actually solve a more general class of problems, called constraint-satisfaction problems (CSPs).

4.2 SDPs for Unique Games

To give a flavor of what these SDPs look like, consider the analysis first studied by Feige and Lovasz in [5]. We first formulate an integer quadratic program. Let the variables $x_{v,i} \in \{0, 1\}$ represent the labeling and be 1 if and only if $\ell(v) = i$. Then the natural constraints are for all $v \in V$, $\sum_{i=1}^n x_{v,i}^2 = 1$,

$x_{v,i}x_{v,j} = 0$ for $i \neq j$ (every vertex must have a unique label). In addition, $x_{v,i}x_{w,j} \geq 0$ for all $v, w \in V$ and $i, j \in [n]$ (that way we do not get x 's of different signs). The vector programming relaxation replaces all these products with vector dot products. We can state it as:

$$\begin{aligned} & \text{maximize} && \frac{1}{|E|} \sum_{e=(v,w)} \sum_{i=1}^n \langle \mathbf{x}_{v,i}, \mathbf{x}_{w,\pi_e(i)} \rangle \\ & \text{subject to} && \\ & \forall v \in V && \sum_{i=1}^n \|\mathbf{x}_{v,i}\|^2 = 1 \\ & \forall v \in V, i, j \in [n], i \neq j && \langle \mathbf{x}_{v,i}, \mathbf{x}_{v,j} \rangle = 0 \\ & \forall v, w \in V, i, j \in [n] && \langle \mathbf{x}_{v,i}, \mathbf{x}_{w,j} \rangle \geq 0 \end{aligned}$$

Ragheendra [9] proved that if the UGC is true, then no algorithm can approximate CSPs efficiently to a better factor than this SDP and associated rounding scheme (which we do not describe here). In particular applying the Lasserre and Sherali-Adams Hierarchies cannot help.

4.3 Small-Set Expansion and Subexponential Algorithms

There is another problem called the Small-Set Expansion Problem, which is the following problem. For a d -regular graph $G = (V, E)$, recall the conductance of a cut $S \subset V$ is

$$\Phi_G(S) = \frac{|E(S, V \setminus S)|}{d|S|}$$

Then we define its expansion profile for $\delta \in [0, 1/2]$ as

$$\Phi_G(\delta) = \min_{|S| \leq \delta|V|} \Phi_G(S)$$

We consider the gap problem, i.e. for constants $\xi, \delta > 0$, deciding between $\Phi_G(\delta) \geq 1 - \xi$ or $\Phi_G(\delta) \leq \xi$. Intuitively, we are deciding between the cases where all small sets have many edges leaving them and where there exists a small set that has few edges leaving it. Spectral graph theory seems to fail spectacularly for this problem, so many posit it is hard, but not too hard. In fact, Ragheendra and Stuerer [10] showed a reduction from the Small-Set Expansion Problem to the Unique Games Problem. So, it also inherits the subexponential runtime of Unique Games SDPs. The two also give rise to the *Small-Set Expansion Hypothesis* (SSEH), which conjectures that the other direction is true, too, that Unique Games reduces to Small-Set Expansion.

Further, in [1], Barak et. al. also show a subexponential algorithm for approximating $\|A\|_{2 \rightarrow 4}$. It is also an SDP algorithm called Tensor-SDP (Section 5 of [1]). It uses the Sum-of-Squares hierarchy to construct a natural relaxation of the optimization problem, whose solution approaches the true optimum (at level r of the hierarchy, the maximizing vector f is restricted to being a polynomial of degree at most $O(r)$, which takes runtime $n^{O(r)}$). They are then able to use this algorithm to solve the gap Small-Set Expansion problem. The connection is through a spectral lens. Let A be a graph G 's normalized adjacency matrix and call $\{(\lambda_i, v_i)\}$ its eigenvalue eigenvector pairs. $P_{\geq \lambda}(G)$ be the projector onto $\text{span}\{v_i : \lambda_i \geq \lambda\}$; then the $2 \rightarrow 4$ norm corresponds to the expansion of the graph (Theorem 2.4 in [1]).

Barak et. al. also show that $\|A\|_{2 \rightarrow 4}$ is Small-Set Expansion-Hard. Therefore, $\|A\|_{2 \rightarrow 4}$ is harder than Unique Games assuming SSEH, but it's not that much harder, as it admits a subexponential algorithm. In a sense, they are similarly complex problems. However, there are not even ETH-level Hardness results known for Unique Games. Proving an NP-hardness result for Unique Games would resolve UGC, while proving an NP-hardness result for $\|A\|_{2 \rightarrow 4}$ would give strong evidence for UGC being true. Thus, any progress in this direction has the capacity to be applicable to the study of Khot's conjecture.

Acknowledgements

We'd like to thank Venkatesan Guruswami for suggesting this problem to us and giving us some guidance on where we were going wrong, as well as Vijay Bhattiprolu from showing us the matrix product state results. Finally, we'd like to thank the course staffs of the Spring 2023 iterations of CS270 and CS191 for their helpful feedback and the chance to do this project.

References

- [1] Boaz Barak, Fernando G. S. L. Brandão, Aram W. Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proc. STOC 2012*, pp. 307–326, 2012.
- [2] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. *Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS*, 04 2011.
- [3] Vijay Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Inapproximability of matrix $p \rightarrow q$ norms, 2018.
- [4] Jop Briët, Oded Regev, and Rishi Saket. Tight hardness of the non-commutative grothendieck problem. *Theory of Computing*, 13(15):1–24, 2017.
- [5] Uriel Feige and László Lovász. Two-prover one-round proof systems: Their power and their problems (extended abstract). In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '92, page 733–744, New York, NY, USA, 1992. Association for Computing Machinery.
- [6] Aram Harrow and Ashley Montanaro. Testing product states, quantum merlin-arthur games and tensor optimization. *Journal of the ACM*, 60, 01 2010.
- [7] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing*, STOC '02, page 767–775, New York, NY, USA, 2002. Association for Computing Machinery.
- [8] Subhash Khot. On the unique games conjecture (invited survey). In *2010 IEEE 25th Annual Conference on Computational Complexity*, pages 99–121, 2010.
- [9] Prasad Raghavendra. Optimal algorithms and inapproximability results for every csp? In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, page 245–254, New York, NY, USA, 2008. Association for Computing Machinery.
- [10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, page 755–764, New York, NY, USA, 2010. Association for Computing Machinery.
- [11] Mehdi Soleimanifar and John Wright. Testing matrix product states, 2022.

A Proof of $2 \rightarrow 1, 4$ Bound

Theorem 5. *For every $\varepsilon > 0$, there exists a $\xi > 0$ and a label cover instance $\mathcal{L} = (G, [L], [R], \Sigma)$ such that if $\text{OPT}(\mathcal{L}) \leq \xi$, \mathcal{L} is D -to-1, and \mathcal{L} is J -smooth, then $\|A\|_{2 \rightarrow 1,4} \leq \gamma_4 + \varepsilon + 3\varepsilon 2^{R/4}$.*

Proof. We follow closely the proof of Theorem 3.3 from [3]. Let $\mathbf{f} \in \mathbb{R}^{V \times 2^R}$ be an arbitrary L_2 unit vector. Call $\hat{\mathbf{f}} = F_P \mathbf{f}$, $\hat{\mathbf{g}} = \hat{\mathbf{P}} \mathbf{f}$, and $\mathbf{g} = F_P^T \hat{\mathbf{g}}$. By Parseval's and the fact that $\hat{\mathbf{P}}$ is an orthogonal projection, $\|\hat{\mathbf{g}}\|_2 \leq \|\hat{\mathbf{f}}\|_2 \leq 1$. Assume for the sake of contrapositive that

$$\|\mathbf{g}\|_{1,4} > \gamma_4 + \varepsilon + 3\varepsilon 2^{R/4}$$

Lemma A.2 from [3] claims it works it only works for $r < 2$, but in reality works for all r with a slight degradation in parameters. In particular, we can use it to obtain a $\delta = \delta(\varepsilon)$ such that $\|g_v\|_4 > (\gamma_4 + \varepsilon)\|g_v\|_2$ implies $\|\hat{g}_v\|_4 > \delta\|\hat{g}_v\|_2$. We separate the vertices into four mutually disjoint sets

$$\begin{aligned} V_0 &= \{v \in V : \|\hat{g}_v\|_4 > \delta\varepsilon \text{ and } \|\hat{g}_v\|_2\} \\ V_1 &= \{v \in V : \|\hat{g}_v\|_4 \leq \delta\varepsilon \text{ and } \|\hat{g}_v\| < \varepsilon\} \\ V_2 &= \{v \in V : \|\hat{g}_v\|_4 \leq \delta\varepsilon \text{ and } \|\hat{g}_v\| \geq \varepsilon\} \\ V_3 &= \{v \in V : \|\hat{g}_v\|_2 > 1/\varepsilon\} \end{aligned}$$

and attempt to bound them separately. In fact, we will show $|V_0| \geq \varepsilon^2|V|$. By the arguments made by Lemma 3.5 of [3] and Lemma 3.6 of [4] before them this means that there exists a labeling satisfying more than $\text{poly}(\varepsilon)$ edges; setting ξ to this quantity means we're done.

Bounding sums yields:

$$\begin{aligned}
\sum_{v \in V_0} \|g_v\|_4 &\leq \sum_{v \in V_0} 2^{R/4} \|g_v\|_2 \\
&= 2^{R/4} \sum_{v \in V_0} \frac{1}{\varepsilon} \\
&= \frac{2^{R/4}}{\varepsilon} |V_0| \\
\sum_{v \in V_1} \|g_v\|_4 &\leq \sum_{v \in V_1} 2^{R/4} \|g_v\|_2 \\
&= 2^{R/4} \sum_{v \in V_1} \|\hat{g}_v\|_2 \\
&= 2^{R/4} \sum_{v \in V_1} \varepsilon \\
&\leq \varepsilon 2^{R/4} |V_1| \\
&\leq \varepsilon 2^{R/4} |V| \\
\sum_{v \in V_2} \|g_v\|_4 &\leq (\gamma_4 + \varepsilon) \sum_{v \in V_2} \|g_v\|_2 \\
&\leq (\gamma_4 + \varepsilon) \sqrt{|V_2| \left(\sum_{v \in V_2} \|g_v\|_2^2 \right)} \\
&\leq (\gamma_4 + \varepsilon) \sqrt{|V_2| |V|} \\
&\leq (\gamma_4 + \varepsilon) |V| \\
\sum_{v \in V_3} \|g_v\|_4 &\leq \sum_{v \in V_3} 2^{R/4} \|g_v\|_2 \\
&= 2^{R/4} \sum_{v \in V_3} \|\hat{g}_v\|_2 \\
&= 2^{R/4} \sum_{v \in V_3} \frac{\|\hat{g}_v\|_2^2}{\|\hat{g}_v\|_2} \\
&\leq 2^{R/4} \sum_{v \in V_3} \varepsilon \|\hat{g}_v\|_2^2 \\
&\leq \varepsilon 2^{R/4} |V|
\end{aligned}$$

Putting everything together gives us

$$\begin{aligned}
|V_0| &\geq \frac{\varepsilon}{2^{R/4}} \sum_{v \in V_0} \|g_v\|_4 \\
|V_0| &\geq \frac{\varepsilon}{2^{R/4}} |V| \left(\|\mathbf{g}\|_{1,4} |V| - \sum_{v \in V_1} \|g_v\|_4 - \sum_{v \in V_2} \|g_v\|_4 - \sum_{v \in V_3} \|g_v\|_4 \right) \\
|V_0| &\geq \frac{\varepsilon}{2^{R/4}} |V| \left((\gamma_4 + \varepsilon + 3\varepsilon 2^{R/4}) - \varepsilon 2^{R/4} - (\gamma_4 + \varepsilon) - \varepsilon 2^{R/4} \right) \\
|V_0| &\geq \varepsilon^2 |V|
\end{aligned}$$

□