Electronic Colloquium on Computational Complexity, Report No. 37 (2018)

Inapproximability of Matrix $p \to q$ Norms

Vijay Bhattiprolu* Mrinalkanti Ghosh† Venkatesan Guruswami*
Euiwoong Lee‡ Madhur Tulsiani†

We study the problem of computing the $p \to q$ norm of a matrix $A \in \mathbb{R}^{m \times n}$, defined as

$$\|A\|_{p \to q} := \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_q}{\|x\|_p} .$$

This problem generalizes the spectral norm of a matrix ($p = q = 2$) and the Grothendieck problem ($p = \infty$, $q = 1$), and has been widely studied in various regimes. When $p \geq q$, the problem exhibits a dichotomy: constant factor approximation algorithms are known if $2 \in [q, p]$, and the problem is hard to approximate within almost polynomial factors when $2 \notin [q, p]$.

The regime when $p < q$, known as *hypercontractive norms*, is particularly significant for various applications but much less well understood. The case with $p = 2$ and $q > 2$ was studied by [Barak et al., STOC'12] who gave sub-exponential algorithms for a promise version of the problem (which captures small-set expansion) and also proved hardness of approximation results based on the Exponential Time Hypothesis. However, no NP-hardness of approximation is known for these problems for any $p < q$.

We study the hardness of approximating matrix norms in both the above cases. We prove the following results:

- We show that for any $1 < p < q < \infty$ with $2 \notin [p, q]$, $\|A\|_{p \to q}$ is hard to approximate within $2^{O((\log n)^{1-\varepsilon})}$ assuming $\mathsf{NP} \not\subseteq \mathsf{BPTIME}\left(2^{(\log n)^{O(1)}}\right)$. This suggests that, similar to the case of $p \geq q$, the hypercontractive setting may be qualitatively different when 2 does not lie between p and q.

- For all $p \geq q$ with $2 \in [q, p]$, we show that $\|A\|_{p \to q}$ is hard to approximate within any factor smaller than $1/(\gamma_{p^*} \cdot \gamma_q)$, where for any r, γ_r denotes the r^{th} norm of a standard normal random variable, and $p^* := p/(p-1)$ is the dual norm of p. The hardness factor is tight for the cases when p or q equals 2.

---

*Computer Science Department, Carnegie Mellon University, Pittsburgh, USA. Some of this work was done when the authors were visiting the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. Research supported in part by NSF grants CCF-1422045 and CCF-1526092. {vpb,venkatg}@cs.cmu.edu

†Toyota Technological Institute, Chicago, USA. Research supported by NSF CCF-1254044. {mkghosh,madhurt}@ttic.edu

‡Computer Science Department, New York University, New York, USA. Some of this work was done when the author was at the Simons Institute for the Theory of Computing. luw0315@gmail.com

ISSN 1433-8092

Contents

1 Introduction

We consider the problem of finding the $p \to q$ norm of a given matrix $A \in \mathbb{R}^{m \times n}$, which is defined as

$$\|A\|_{p \to q} := \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_q}{\|x\|_p}.$$

The quantity $\|A\|_{p \to q}$ is a natural generalization of the well-studied spectral norm, which corresponds to the case $p = q = 2$. For general p and q, this quantity computes the maximum distortion (stretch) of the operator A from the normed space ℓ_p^n to ℓ_q^m.

The case when $p = \infty$ and $q = 1$ is the well known Grothendieck problem [KN12, Pis12], where the goal is to maximize $\langle y, Ax \rangle$ subject to $\|x\|_\infty, \|y\|_\infty \leq 1$. In fact, via simple duality arguments (see Section 2), the general problem computing $\|A\|_{p \to q}$ can be seen to be equivalent to the following variant of the Grothendieck problem (and to $\|A^T\|_{q^* \to p^*}$)

$$\|A\|_{p \to q} = \max_{\substack{\|x\|_p \leq 1 \\ \|y\|_{q^*} \leq 1}} \langle y, Ax \rangle = \|A^T\|_{q^* \to p^*},$$

where p^*, q^* denote the dual norms of p and q, satisfying $1/p + 1/p^* = 1/q + 1/q^* = 1$.

The case when $p < q$, known as the case of *hypercontractive* norms, also has a special significance to the analysis of random walks, expansion and related problems in hardness of approximation [Bis11, BBH$^+$12]. Bounds on hypercontractive norms of operators are often used to prove small-set expansion of graphs (where the operator is related to the adjacency matrix) and thus the approximability of hypercontractive norms is closely related to the problem of determining small-set expansion of graphs. The problem of computing $\|A\|_{2 \to 4}$ is also known to be equivalent to determining the maximum acceptance probability of a quantum protocol with multiple unentangled provers, and is related to several problems in quantum information theory [HM13, BH15].

1.1 Known results

The problem of approximating norms has been studied in various contexts, and we summarize several known results in Fig. 1. While the case of $p = q = 2$ corresponds to the spectral norm, the problem is also easy when $q = \infty$ (or equivalently $p = 1$) since this corresponds to selecting the row of A with the maximum ℓ_{p^*} norm. Note that in general, Fig. 1 is symmetric about the principal diagonal. Also note that if $\|A\|_{p \to q}$ is a hypercontractive norm ($p < q$) then so is the equivalent $\|A^T\|_{q^* \to p^*}$ (the hypercontractive and non-hypercontractive case are separated by the non-principal diagonal).

The non-hypercontractive case $(p \geq q)$**.** Many results are known in the case of $p \geq q$, where the problem admits good approximations when $2 \in [q, p]$, and is hard otherwise. Determining the right constants in these approximations when $2 \in [q, p]$ has been of considerable interest in the analysis and optimization community.

For the case of $\infty \to 1$ norm, Grothendieck's theorem [Gro56] shows that the integrality gap of a semidefinite programming (SDP) relaxation is bounded by a constant, and the (unknown) optimal value is now called the Grothendieck constant K_G. Krivine [Kri77] proved an upper bound of $\pi/(2\ln(1+\sqrt{2})) = 1.782\ldots$ on K_G, and it was later shown by Braverman et al. that K_G is strictly smaller than this bound. The best known lower bound on K_G is about 1.676, due to (an unpublished manuscript of) Reeds [Ree91] (see also [KO09] for a proof).

An upper bound of K_G on the approximation factor also follows from the work of Nesterov [Nes98] for any $p \geq 2 \geq q$. A later work of Steinberg [Ste05] also gave an upper bound of $\min\{\gamma_p/\gamma_q, \gamma_{q^*}/\gamma_{p^*}\}$, where γ_p denotes p^{th} norm of a standard normal random variable (i.e. the p-th root of the p-th gaussian moment). Note that Steinberg's bound is less than K_G for some values of (p, q), in particular for all values of the form $(2, q)$ with $q \leq 2$ (and equivalently $(p, 2)$ for $p \geq 2$), where it equals $1/\gamma_q$ (and $1/\gamma_{p^*}$ for $(p, 2)$).

On the hardness side, Briët, Regev and Saket [BRS15] showed NP-hardness of $\pi/2$ for the $\infty \to 1$ norm, strengthening a hardness result of Khot and Naor based on the Unique Games Conjecture (UGC) [KN08] (for a special case of the Grothendieck problem when the matrix A is positive semidefinite). Assuming UGC, a hardness result matching Reeds' lower bound was proved by Khot and O'Donnell[KO09], and hardness of approximating within K_G was proved by Raghavendra and Steurer [RS09].

For a related problem known as the L_p-Grothendieck problem, where the goal is to maximize $\langle x, Ax\rangle$ for $\|x\|_p \leq 1$, results by Steinberg [Ste05] and Kindler, Schechtman and Naor [KNS10] give an upper bound of γ_p^2, and a matching lower bound was proved assuming UGC by [KNS10], which was strengthened to NP-hardness by Guruswami et al. [GRSW16]. However, note that this problem is quadratic and not necessarily bilinear, and is in general much harder than the Grothendieck problems considered here. In particular, the case of $p = \infty$ only admits an $\Theta(\log n)$ approximation instead of K_G for the bilinear version [AMMN06, ABH$^+$05].

For the case when $2 \notin [q, p]$, an upper bound of $O(\max\{m, n\}^{25/128})$ on the approximation ratio was proved by Steinberg [Ste05]. Bhaskara and Vijayaraghavan [BV11] showed NP-hardness of approximation whithin any constant factor, and hardness of approximation within an $O\left(2^{(\log n)^{1-\varepsilon}}\right)$ factor for arbitrary $\varepsilon > 0$ assuming $\mathsf{NP} \not\subseteq \mathsf{DTIME}\left(2^{(\log n)^{O(1)}}\right)$.

The hypercontractive case $(p < q)$**.** Relatively fewer results are known for the case when $p < q$. Steinberg's result [Ste05] also applies to this case, giving an upper bound of $O(\max\{m, n\}^{25/128})$ on the approximation factor, for all p, q. For the case of $2 \to q$ norm (for any $q > 2$), Barak et al. [BBH$^+$12] give an approximation algorithm for a promise version of the problem, running in time $\exp\left(\tilde{O}(n^{2/q})\right)$. They also provide an additive approximation algorithm for the $2 \to 4$ norm (where the error depends on $2 \to 2$ norm and $2 \to \infty$ norm of A), which was extended to the $2 \to q$ norm by Harrow and Montanaro [HM13]. Barak et al. also prove NP-hardness of approximating $\|A\|_{2\to4}$ within a factor of $1 + \tilde{O}(1/n^{o(1)})$, and hardness of approximating better than $\exp O((\log n)^{1/2-\varepsilon})$ in polynomial time, assuming the Exponential Time Hypothesis (ETH). This reduction was also used by Harrow, Natarajan and Wu [HNW16] to prove that $\tilde{O}(\log n)$ levels of the Sum-of-Squares SDP hierarchy cannot approximate $\|A\|_{2\to4}$ within any constant factor.

1.2 Our contribution

We extend the hardness results of [BRS15] for the $\infty \to 1$ and $2 \to 1$ norms of a matrix to any $p \geq 2 \geq q$. The hardness factors obtained match the performance of known algorithms (due to Steinberg [Ste05]) for the cases of $2 \to q$ and $p \to 2$.

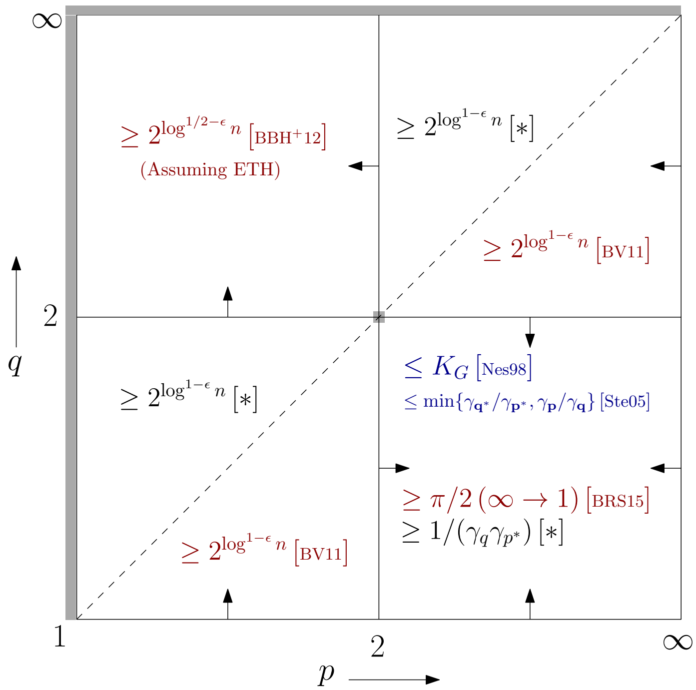

Figure 1: Upper and lower bounds for approximating $\|A\|_{p\to q}$. Arrows indicate the region to which a boundary belongs and thicker shaded regions represent exact algorithms. Our results are indicated by [∗]. We omit UGC-based hardness results in the figure.

Theorem 1.1. *For any p, q such that $\infty \geq p \geq 2 \geq q \geq 1$ and $\varepsilon > 0$, it is NP-hard to approximate the $p \to q$ norm within a factor $1/(\gamma_{p^*}\gamma_q) - \varepsilon$.*

We also prove the first strong inapproximability results for hypercontractive norms without assuming ETH. We show that it is hard to approximate $\|A\|_{p\to q}$ within almost polynomial factors unless NP is in randomized quasi-polynomial time. This is the content of the following theorem.

Theorem 1.2. *For any p, q such that $1 < p \leq q < 2$ or $2 < p \leq q < \infty$ and $\varepsilon > 0$, there is no polynomial time algorithm that approximates the $p \to q$ norm of an $n \times n$ matrix within a factor $2^{\log^{1-\varepsilon} n}$ unless NP $\subseteq$ BPTIME $\left(2^{(\log n)^{O(1)}}\right)$. When q is an even integer, the same inapproximability result holds unless NP $\subseteq$ DTIME $\left(2^{(\log n)^{O(1)}}\right)$ [1].*

We view the above theorem as providing some evidence that while hypercontractive norms have been studied as a single class so far, the case when $2 \in [p, q]$ may be qualitatively different (with respect to techniques) from the case when $2 \notin [p, q]$

As discussed below, it is indeed possible to overcome the obstructions in designing reductions for hypercontractive norms in the case when $2 \notin [p, q]$. However, the case when $2 \in [p, q]$ remains a very interesting open problem.

[1] Interestingly, Barak et al. [BBH+12] gave a subexponential algorithm for a promise version of $2 \to q$ norm, with runtime $2^{O(n^{2/q})}$ and our reduction rules out a constant factor approximation algorithm for $\|A\|_{p\to q}$ ($2 < p < q < \infty$) running in time $2^{n^{o(1/q)}}$ (assuming $NP \not\subseteq$ DTIME $\left(2^{o(n)}\right)$). That is, one cannot have a significantly better runtime than that of Barak et al. for the $p \to q$ norm problem when $2 < p < q < \infty$.

Both the above theorems are in fact consequences of our main technical theorem, which proves hardness of approximating $\|A\|_{2\to r}$ for $r < 2$ (and hence $\|A\|_{r^*\to 2}$ for $r^* > 2$) while providing additional structure in the matrix A produced by the reduction. This theorem is proved in Section 3.

1.3 Proof overview

The hardness of proving hardness for hypercontractive norms. Reductions for various geometric problems use a "smooth" version of the Label Cover problem, composed with long-code functions for the labels of the variables. In various reductions, including the ones by Guruswami et al. [GRSW16] and Briët et al. [BRS15] (which we closely follow) the solution vector x to the geometric problem consists of the Fourier coefficients of the various long-code functions, with a "block" x_v for each vertex of the label-cover instance. The relevant geometric operation (transformation by the matrix A in our case) consists of projecting to a space which enforces the consistency constraints derived from the label-cover problem, on the Fourier coefficients of the encodings.

However, this strategy presents with two problems when designing reductions for hypercontractive norms. Firstly, while projections maintain the ℓ_2 norm of encodings corresponding to consistent labelings and reduce that of inconsistent ones, their behaviour is harder to analyze for ℓ_p norms for $p \neq 2$. Secondly, the *global* objective of maximizing $\|Ax\|_q$ is required to enforce different behavior within the blocks x_v, than in the full vector x. The block vectors x_v in the solution corresponding to a satisfying assignment of label cover are intended to be highly sparse, since they correspond to "dictator functions" which have only one non-zero Fourier coefficient. This can be enforced in a test using the fact that for a vector $x_v \in \mathbb{R}^t$, $\|x_v\|_q$ is a convex function of $\|x_v\|_p$ when $p \leq q$, and is maximized for vectors with all the mass concentrated in a single coordinate. However, a global objective function which tries to maximize $\sum_v \|x_v\|_q^q$, also achieves a high value from global vectors x which concentrate all the mass on coordinates corresponding to few vertices of the label cover instance, and do not carry any meaningful information about assignments to the underlying label cover problem.

Since we can only check for a global objective which is the ℓ_q norm of some vector involving coordinates from blocks across the entire instance, it is not clear how to enforce local Fourier concentration (dictator functions for individual long codes) and global well-distribution (meaningful information regarding assignments of most vertices) using the same objective function. While the projector A also enforces a linear relation between the block vectors x_u and x_v for all edges (u, v) in the label cover instance, using this to ensure well-distribution across blocks seems to require a very high density of constraints in the label cover instance, and no hardness results are available in this regime.

Our reduction. We show that when $2 \notin [p, q]$, it is possible to bypass the above issues using hardness of $\|A\|_{2\to r}$ as an intermediate (for $r < 2$). Note that since $\|z\|_r$ is a *concave* function of $\|z\|_2$ in this case, the test favors vectors in which the mass is well-distributed and thus solves the second issue. For this, we use local tests based on the Berry-Esséen theorem (as in [GRSW16] and [BRS15]). Also, since the starting point now is the ℓ_2 norm, the effect of projections is easier to analyze. This reduction is discussed in Section 3.

By duality, we can interpret the above as a hardness result for $\|A\|_{p\to 2}$ when $p > 2$ (using $r = p^*$). We then convert this to a hardness result for $p\to q$ norm in the hypercontractive case by composing A with an "approximate isometry" B from $\ell_2 \to \ell_q$ (i.e., $\forall y \; \|By\|_q \approx \|y\|_2$) since we can replace $\|Ax\|_2$ with $\|BAx\|_q$. Milman's version of the Dvoretzky theorem [Ver17] implies random operators to a sufficiently high dimensional ($n^{O(q)}$) space satisfy this property, which then yields constant factor hardness results for the $p \to q$ norm. A similar application of Dvoretzky's theorem also appears in an independent work of Krishnan et al. [KMW18] on sketching matrix norms.

We also show that the hardness for hypercontractive norms can be amplified via tensoring. This was known previously for the $2 \to 4$ norm using an argument based on parallel repetition for QMA [HM13], and for the case of $p = q$ [BV11], but appears to have gone unnoticed for the general $p < q$ case. The amplification is then used to prove hardness of approximation within almost polynomial factors.

Nonhypercontractive norms. We also use the hardness of $\|A\|_{2\to r}$ to obtain hardness for the non-hypercontractive case of $\|A\|_{p\to q}$ with $q < 2 < p$, by using an operator that "factorizes" through ℓ_2. In particular, we obtain hardness results for $\|A\|_{p\to 2}$ and $\|A\|_{2\to q}$ (of factors $1/\gamma_{p^*}$and $1/\gamma_q$ respectively) using the reduction in Section 3. We then combine these hardness results using additional properties of the operator A obtained in the reduction, to obtain a hardness of factor $(1/\gamma_{p^*}) \cdot (1/\gamma_q)$ for the $p \to q$ norm for $p > 2 > q$. The composition, as well as the hardness results for hypercontractive norms, are presented in Section 4.

2 Preliminaries and Notation

2.1 Matrix Norms

For a vector $x \in \mathbb{R}^n$, throughout this paper we will use $x(i)$ 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 $x \in \mathbb{R}^n$,

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

Clearly $\|x\|_{\ell_p} = \|x\|_{L_p} \cdot n^{1/p}$. For $p = \infty$, we define $\|x\|_{\ell_\infty} = \|x\|_{L_\infty} := \max_{i\in[n]} |x(i)|$. We will use p^* to denote the 'dual' of p, i.e. $p^* = p/(p-1)$. Unless stated otherwise, we usually work with $\|\cdot\|_{\ell_p}$. We also define inner product $\langle x, y\rangle$ to denote the inner product under the counting measure unless stated otherwise; i.e. for two vectors $x, y \in \mathbb{R}^n$, $\langle x, y\rangle := \sum_{i\in[n]} x(i)y(i)$.

We next record a well-known fact about p-norms that is used in establishing many duality statements.

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

We next define the primary problems of interest in this paper.

Definition 2.2. *For $p, q \in [1, \infty]$, the $p \to q$ norm problem is to maximize*

$$\frac{\|Ax\|_{\ell_q}}{\|x\|_{\ell_p}}$$

given an $m \times n$ matrix A.

Definition 2.3. *For $p, q \in [1, \infty]$, we define a generalization of the Grothendieck problem, namely (p, q)-Grothendieck, as the problem of computing*

$$\sup_{\|y\|_{\ell_p}=1} \sup_{\|x\|_{\ell_q}=1} \langle y, Ax \rangle$$

given an $m \times n$ matrix A.

The original Grothendieck problem is precisely (∞, ∞)-Grothendieck. We next state the well known equivalence of $p \to q$ norm, (q^*, p)-Grothendieck, and $q^* \to p^*$ norm.

Observation 2.4. *For any $p, q \in [1, \infty]$ and any matrix A,*

$$\|A\|_{\ell_p \to \ell_q} = \sup_{\|y\|_{\ell_{q^*}}=1} \sup_{\|x\|_{\ell_p}=1} \langle y, Ax \rangle = \|A^T\|_{\ell_{q^*} \to \ell_{p^*}}.$$

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

$$\begin{aligned} \|A\|_{\ell_p \to \ell_q} &= \sup_{\|x\|_{\ell_p}=1} \|Ax\|_{\ell_q} = \sup_{\|x\|_{\ell_p}=1} \sup_{\|y\|_{\ell_{q^*}}=1} \langle y, Ax \rangle = \sup_{\|y\|_{\ell_{q^*}}=1} \sup_{\|x\|_{\ell_p}=1} \langle y, Ax \rangle \\ &= \sup_{\|x\|_{\ell_p}=1} \sup_{\|y\|_{\ell_{q^*}}=1} \langle x, A^T y \rangle = \sup_{\|y\|_{\ell_{q^*}}=1} \|A^T y\|_{\ell_{p^*}} = \|A^T\|_{\ell_{q^*} \to \ell_{p^*}}. \end{aligned}$$

∎

The following observation will be useful for composing hardness maps for $p \to 2$ norm and $2 \to q$ norm to get $p \to q$ norm hardness for when $p > q$ and $p \geq 2 \geq q$.

Observation 2.5. *For any $p, q, r \in [1, \infty]$ and any matrices B, C,*

$$\|BC\|_{\ell_p \to \ell_q} = \sup_x \frac{\|BCx\|_{\ell_q}}{\|x\|_{\ell_p}} \leq \sup_x \frac{\|B\|_{\ell_r \to \ell_q} \|Cx\|_{\ell_r}}{\|x\|_{\ell_p}} \leq \|B\|_{\ell_r \to \ell_q} \|C\|_{\ell_p \to \ell_r}.$$

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

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

The *Fourier transform* refers to a linear operator F that maps f to $\widehat{f}$ as defined as (2). We interpret $\widehat{f}$ as a 2^R-dimensional vector whose coordinates are indexed by $S \subseteq [R]$. Endow the expectation norm and the expectation norm to f and $\widehat{f}$ respectively; i.e.,

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

as well as the corresponding inner products $\langle f, g\rangle$ and $\langle \widehat{f}, \widehat{g}\rangle$ consistent with their 2-norms. We also define the *inverse Fourier transform* F^T to be a linear operator that maps a given $\widehat{f} : 2^R \to \mathbb{R}$ to $f : \{\pm 1\}^R \to \mathbb{R}$ defined as in (1). We state the following well-known facts from Fourier analysis.

Observation 2.6 (Parseval's Theorem)**.** *For any $f : \{\pm 1\}^R \to \mathbb{R}$, $\|f\|_{L_2} = \|Ff\|_{\ell_2}$.*

Observation 2.7. *F and F^T form an adjoint pair; i.e., for any $f : \{\pm 1\}^R \to \mathbb{R}$ and $\widehat{g} : 2^R \to \mathbb{R}$,*

$$\langle \widehat{g}, Ff\rangle = \langle F^T \widehat{g}, f\rangle.$$

Observation 2.8. *$F^T F$ is the identity operator.*

In Section 3, we also consider a *partial* Fourier transform F_P that maps a given function $f : \{\pm 1\}^R \to \mathbb{R}$ to a vector $\widehat{f} : [R] \to \mathbb{R}$ defined as $\widehat{f}(i) = \mathbb{E}_{x \in \{\pm 1\}^R}[f(x) \cdot x_i]$ for all $i \in [R]$. It is the original Fourier transform where $\widehat{f}$ is further projected to R coordinates corresponding to linear coefficients. The partial inverse Fourier transform F_P^T is a transformation that maps a vector $\widehat{f} : [R] \to \mathbb{R}$ to a function $f : \{\pm 1\}^R \to \mathbb{R}$ as in (1) restricted to $S = \{i\}$ for some $i \in [R]$. These partial transforms satisfy similar observations as above: (1) $\|f\|_{L_2} \geq \|F_P f\|_{\ell_2}$, (2) $\|F_P^T \widehat{f}\|_{L_2} = \|\widehat{f}\|_{\ell_2}$, (3) F_P and F_P^T form an adjoint pair, and (4) $(F_P^T F_P) f = f$ if and only if f is a linear function.

2.3 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 the original construction due to [Kho02]. The theorem also describes the various structural properties, including smoothness, that are identified by the hard instances.

Theorem 2.9. *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 Label Cover instance $(G, [R], [L], \Sigma)$ as above such that*

- *(Hardness): It is NP-hard to distinguish between the following two cases:*
  - *(Completeness): $\mathsf{OPT}(\mathcal{L}) = 1$.*
  - *(Soundness): $\mathsf{OPT}(\mathcal{L}) \leq \xi$.*
- *(Structural 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 $|\pi_{e,v}^{-1}(i)| \leq D$; that is at most D elements in $[R]$ are mapped to the same element in $[L]$.*
  - *(Weak Expansion): For any $\delta > 0$ and vertex set $V' \subseteq V$ such that $|V'| = \delta \cdot |V|$, the number of edges among the vertices in $|V'|$ is at least $(\delta^2/2)|E|$.*

3 Hardness of $2 \to r$ norm with $r < 2$

This section proves the following theorem that serves as a starting point of our hardness results. The theorem is stated for the expectation norm for consistency with the current literature, but the same statement holds for the counting norm, since if A is an $n \times n$ matrix, $\|A\|_{\ell_2 \to \ell_r} = n^{1/r-1/2} \cdot \|A\|_{L_2 \to L_r}$. Note that the matrix A used in the reduction below does not depend on r

Theorem 3.1. *For any $\varepsilon > 0$, there is a polynomial time reduction that takes a 3-CNF formula φ and produces a symmetric matrix $A \in \mathbb{R}^{n \times n}$ with $n = \mathrm{poly}(|\varphi|)$ such that*

- *(Completeness) If φ is satisfiable, there exists $x \in \mathbb{R}^n$ with $|x(i)| = 1$ for all $i \in [n]$ and $Ax = x$. In particular, $\|A\|_{L_2 \to L_r} \geq 1$ for all $1 \leq r \leq \infty$.*
- *(Soundness) $\|A\|_{L_2 \to L_r} \leq \gamma_r + \varepsilon^{2-r}$ for all $1 \leq r < 2$.*

We adapt the proof by Briët, Regev and Saket for the hardness of $2 \to 1$ and $\infty \to 1$ norms to prove the above theorem. A small difference is that, unlike their construction which starts with a Fourier encoding of the long-code functions, we start with an evaluation table (to ensure that the resulting matrices are symmetric). We also analyze their dictatorship tests for the case of fractional r.

3.1 Reduction and Completeness

Let $\mathcal{L} = (G, [R], [L], \Sigma)$ be an instance of Label Cover with $G = (V, E)$. In the rest of this section, $n = |V|$ and our reduction will construct a self-adjoint linear operator $\mathbf{A} : \mathbb{R}^N \to \mathbb{R}^N$ with $N = |V| \cdot 2^R$, which yields a symmetric $N \times N$ matrix representing $\mathbf{A}$ in the standard basis. This section concerns the following four Hilbert spaces based on the standard Fourier analysis composed with $\mathcal{L}$.

1. Evaluation space $\mathbb{R}^{2^R}$. Each function in this space is denoted by $f : \{\pm 1\}^R \to \mathbb{R}$. The inner product is defined as $\langle f, g \rangle := \mathbb{E}_{x \in \{\pm 1\}^R}[f(x)g(x)]$, which induces $\|f\|_2 := \|f\|_{L_2}$. We also define $\|f\|_{L_p} := \mathbb{E}_x[|f(x)|^p]^{1/p}$ in this space.
2. Fourier space $\mathbb{R}^R$. Each function in this space is denoted by $\widehat{f} : [R] \to \mathbb{R}$. The inner product is defined as $\langle \widehat{f}, \widehat{g} \rangle := \sum_{i \in [R]} \widehat{f}(i)\widehat{g}(i)$, which induces $\|\widehat{f}\|_2 := \|\widehat{f}\|_{\ell_2}$.
3. Combined evaluation space $\mathbb{R}^{V \times 2^R}$. Each function in this space is denoted by $\mathbf{f} : V \times \{\pm 1\}^R \to \mathbb{R}$. The inner product is defined as $\langle \mathbf{f}, \mathbf{g} \rangle := \mathbb{E}_{v \in V}[\mathbb{E}_{x \in \{\pm 1\}^R}[\mathbf{f}(v, x)\mathbf{g}(v, x)]]$, which induces $\|\mathbf{f}\|_{L_2} := \|\mathbf{f}\|_{L_2}$. We also define $\|\mathbf{f}\|_p := \mathbb{E}_{v,x}[|\mathbf{f}(v, x)|^p]^{1/p}$ in this space.
4. Combined Fourier space $\mathbb{R}^{V \times R}$. Each function in this space is denoted by $\widehat{\mathbf{f}} : V \times [R] \to \mathbb{R}$. The inner product is defined as $\langle \widehat{\mathbf{f}}, \widehat{\mathbf{g}} \rangle := \mathbb{E}_{v \in V}[\sum_{i \in [R]} \widehat{\mathbf{f}}(v, i)\widehat{\mathbf{g}}(v, i)]$, which induces $\|\widehat{\mathbf{f}}\|_2$, which is neither a counting nor an expectation norm.

Note that $\mathbf{f} \in \mathbb{R}^{V \times 2^R}$ and a vertex $v \in V$ induces $f_v \in \mathbb{R}^{2^R}$ defined by $f_v(x) := \mathbf{f}(v, x)$, and similarly $\widehat{\mathbf{f}} \in \mathbb{R}^{V \times R}$ and a vertex $v \in V$ induces $\widehat{f_v} \in \mathbb{R}^R$ defined by $\widehat{f_v}(x) := \widehat{\mathbf{f}}(v, x)$.

As defined in Section 2.2, we use the standard following (partial) *Fourier transform F* that maps $f \in \mathbb{R}^{2^R}$ to $\widehat{f} \in \mathbb{R}^R$ as follows. [2]

$$\widehat{f}(i) = (Ff)(i) := \mathop{\mathbb{E}}_{x \in \{\pm 1\}^R} [x_i f(x)]. \tag{3}$$

The (partial) *inverse Fourier transform F^T* that maps $\widehat{f} \in \mathbb{R}^R$ to $f \in \mathbb{R}^{2^R}$ is defined by

$$f(x) = (F^T \widehat{f})(x) := \sum_{i \in [R]} x_i \widehat{f}(i). \tag{4}$$

This Fourier transform can be naturally extended to combined spaces by defining $\mathbf{F} : \mathbf{f} \mapsto \widehat{\mathbf{f}}$ as $f_v \mapsto \widehat{f_v}$ for all $v \in V$. Then $\mathbf{F}^T$ maps $\widehat{\mathbf{f}}$ to $\mathbf{f}$ as $\widehat{f_v} \mapsto f_v$ for all $v \in V$.

Finally, let $\widehat{\mathbf{P}} : \mathbb{R}^{V \times R} \to \mathbb{R}^{V \times R}$ be the orthogonal projector to the following subspace of the combined Fourier space:

$$\widehat{\mathbf{L}} := \left\{ \widehat{\mathbf{f}} \in \mathbb{R}^{V \times R} : \sum_{j \in \pi_{e,u}^{-1}(i)} \widehat{f}_u(i) = \sum_{j \in \pi_{e,v}^{-1}(i)} \widehat{f}_v(j) \text{ for all } (u,v) \in E \text{ and } i \in [L] \right\}. \tag{5}$$

Our transformation $\mathbf{A} : \mathbb{R}^{V \times 2^R} \to \mathbb{R}^{V \times 2^R}$ is defined by

$$\mathbf{A} := (\mathbf{F}^T) \widehat{\mathbf{P}} \mathbf{F}. \tag{6}$$

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 $\widehat{\mathbf{P}}$ is a projector, $\mathbf{A}$ is self-adjoint by design.

We also note that a similar reduction that produces $(\mathbf{F}^T)\widehat{\mathbf{P}}$ was used in Guruswami et al. [GRSW16] and Briët et al. [BRS15] for subspace approximation and Grothendieck-type problems, and indeed this reduction suffices for Theorem 3.1 except the self-adjointness and additional properties in the completeness case.

Completeness. We prove the following lemma for the completeness case. A simple intuition is that if $\mathcal{L}$ admits a good labeling, we can construct a $\mathbf{f}$ such that each f_v is a linear function and $\widehat{\mathbf{f}}$ is already in the subspace $\widehat{\mathbf{L}}$. Therefore, each of Fourier transform, projection to $\widehat{\mathbf{L}}$, and inverse Fourier transform does not really change $\mathbf{f}$.

Lemma 3.2 (Completeness)**.** *Let $\ell : V \to [R]$ be a labeling that satisfies every edge of $\mathcal{L}$. There exists a function $\mathbf{f} \in \mathbb{R}^{V \times 2^R}$ such that $\mathbf{f}(v,x)$ is either $+1$ or -1 for all $v \in V, x \in \{\pm 1\}^R$ and $\mathbf{Af} = \mathbf{f}$.*

Proof. Let $\mathbf{f}(v,x) := x_{\ell(v)}$ for every $v \in V, x \in \{\pm 1\}^R$. Consider $\widehat{\mathbf{f}} = \mathbf{Ff}$. For each vertex $v \in V$, $\widehat{\mathbf{f}}(v,i) = \widehat{f_v}(i) = 1$ if $i = \ell(v)$ and 0 otherwise. Since ℓ satisfies every edge of $\mathcal{L}$, $\widehat{\mathbf{f}} \in \widehat{\mathbf{L}}$ and $\widehat{\mathbf{P}}\widehat{\mathbf{f}} = \widehat{\mathbf{f}}$. Finally, since each f_v is a linear function, the partial inverse Fourier transform F^T satisfies $(F^T)\widehat{f_v} = f_v$, which implies that $(\mathbf{F}^T)\widehat{\mathbf{f}} = \mathbf{f}$. Therefore, $\mathbf{Af} = (\mathbf{F}^T \widehat{\mathbf{P}} \mathbf{F})\mathbf{f} = \mathbf{f}$. ■

[2] We use only *linear Fourier coefficients* in this work. F was defined as F_P in Section 2.2.

3.2 Soundness

We prove the following soundness lemma. This finishes the proof of Theorem 3.1 since Theorem 2.9 guarantees NP-hardness of Label Cover for arbitrarily small $\xi > 0$ and arbitrarily large $J \in \mathbb{N}$.

Lemma 3.3 (Soundness). *For every $\varepsilon > 0$, there exist $\xi > 0$ (that determines $D = D(\xi)$ as in Theorem 2.9) and $J \in \mathbb{N}$ such that if $\mathsf{OPT}(\mathcal{L}) \leq \xi$, $\mathcal{L}$ is D-to-1, and $\mathcal{L}$ is J-smooth, $\|\mathbf{A}\|_{L_2 \to L_r} \leq \gamma_r + 4\varepsilon^{2-r}$ for every $1 \leq r < 2$.*

Proof. Let $\mathbf{f} \in \mathbb{R}^{V \times 2^R}$ be an arbitrary vector such that $\|\mathbf{f}\|_{L_2} = 1$. Let $\widehat{\mathbf{f}} = \mathbf{F}\mathbf{f}$, $\widehat{\mathbf{g}} = \widehat{\mathbf{L}}\widehat{\mathbf{f}}$, and $\mathbf{g} = \mathbf{F}^T\widehat{\mathbf{g}}$ so that $\mathbf{g} = (\mathbf{F}^T\widehat{\mathbf{L}}\mathbf{F})\mathbf{f} = \mathbf{A}\mathbf{f}$. By Parseval's theorem, $\|\widehat{f}_v\|_{\ell_2} \leq \|f_v\|_{L_2}$ for all $v \in V$ and $\|\widehat{\mathbf{f}}\|_2 \leq \|\mathbf{f}\|_{L_2} \leq 1$. Since $\widehat{\mathbf{L}}$ is an orthogonal projection, $\|\widehat{\mathbf{g}}\|_2 \leq \|\widehat{\mathbf{f}}\|_2 \leq 1$. Fix $1 \leq r < 2$ and suppose

$$\|\mathbf{g}\|_{L_r}^r = \mathop{\mathbb{E}}_{v \in V}\left[\|g_v\|_{L_r}^r\right] \geq \gamma_r^r + 4\varepsilon^{2-r}. \tag{7}$$

Use Lemma A.2 to obtain $\delta = \delta(\varepsilon)$ such that $\|g_v\|_{L_p}^p > (\gamma_p^p + \varepsilon)\|\widehat{g}_v\|_{\ell_2}^p$ implies $\|\widehat{g}\|_{\ell_4} > \delta\|\widehat{g}\|_{\ell_2}$ for all $1 \leq p < 2$ (so that δ does not depend on r), and consider

$$V_0 := \{v \in V : \|\widehat{g}_v\|_{\ell_4} > \delta\varepsilon \text{ and } \|\widehat{g}_v\|_{\ell_2} \leq 1/\varepsilon\}. \tag{8}$$

We prove the following lemma that lower bounds the size of V_0.

Lemma 3.4. *For $V_0 \subseteq V$ defined as in (8), we have $|V_0| \geq \varepsilon^2|V|$.*

Proof. The proof closely follows the proof of Lemma 3.4 of [BRS15]. Define the sets

$$\begin{aligned}
V_1 &= \{v \in V : \|\widehat{g}_v\|_{\ell_4} \leq \delta\varepsilon \text{ and } \|\widehat{g}_v\|_{\ell_2} < \varepsilon\}, \\
V_2 &= \{v \in V : \|\widehat{g}_v\|_{\ell_4} \leq \delta\varepsilon \text{ and } \|\widehat{g}_v\|_{\ell_2} \geq \varepsilon\}, \\
V_3 &= \{v \in V : \|\widehat{g}_v\|_{\ell_2} > 1/\varepsilon\}.
\end{aligned}$$

From (7), we have

$$\sum_{v \in V_0} \|g_v\|_{L_r}^r + \sum_{v \in V_1} \|g_v\|_{L_r}^r + \sum_{v \in V_2} \|g_v\|_{L_r}^r + \sum_{v \in V_3} \|g_v\|_{L_r}^r \geq (\gamma_r^r + 4\varepsilon^{2-r})|V|. \tag{9}$$

We bound the four sums on the left side of (9) individually. Parseval's theorem and the fact that $r < 2$ implies $\|g_v\|_{L_r} \leq \|g_v\|_{L_2} = \|\widehat{g}_v\|_{\ell_2}$, and since $\|\widehat{g}_v\|_{\ell_2} \leq 1/\varepsilon$ for every $v \in V_0$, the first sum in (9) can be bounded by

$$\sum_{v \in V_0} \|g_v\|_{L_r}^r \leq |V_0|/\varepsilon^r. \tag{10}$$

Similarly, using the definition of V_1 the second sum in (9) is at most $\varepsilon^r|V|$. By Lemma A.2, for each $v \in V_2$, we have $\|g_v\|_{L_r}^r \leq (\gamma_r^r + \varepsilon)\|\widehat{g}_v\|_{\ell_2}^r$. Therefore, the third sum in (9) is bounded as

$$
\begin{aligned}
\sum_{v \in V_2} \|g_v\|_{L_r}^r &\le (\gamma_r^r + \varepsilon) \sum_{v \in V_2} \|\widehat{g}_v\|_{\ell_2}^r \\
&= (\gamma_r^r + \varepsilon)|V_2| \mathbb{E}_{v \in V_2}[\|\widehat{g}_v\|_{\ell_2}^r] \\
&\le (\gamma_r^r + \varepsilon)|V_2| \mathbb{E}_{v \in V_2}[\|\widehat{g}_v\|_{\ell_2}^2]^{r/2} && \text{(By Jensen using } r < 2) \\
&= (\gamma_r^r + \varepsilon)|V_2| \left(\frac{\sum_{v \in V_2} \|\widehat{g}_v\|_{\ell_2}^2}{|V_2|} \right)^{r/2} \\
&\le (\gamma_r^r + \varepsilon)|V_2|^{1-r/2}|V|^{r/2} && \Big(\sum_{v \in V_2} \|\widehat{g}_v\|_{\ell_2}^2 \le \sum_{v \in V} \|\widehat{g}_v\|_{\ell_2}^2 \le |V|\Big) \\
&\le (\gamma_r^r + \varepsilon)|V|. && (11)
\end{aligned}
$$

Finally, the fourth sum in (9) is bounded by

$$
\begin{aligned}
\sum_{v \in V_3} \|g_v\|_{L_r}^r &\le \sum_{v \in V_3} \|g_v\|_{L_2}^r && \text{(Since } r < 2) \\
&= \sum_{v \in V_3} \|\widehat{g}_v\|_{\ell_2}^r && \text{(By Parseval's theorem)} \\
&= \sum_{v \in V_3} \|\widehat{g}_v\|_{\ell_2}^{r-2} \|\widehat{g}_v\|_{\ell_2}^2 \\
&< \sum_{v \in V_3} \varepsilon^{2-r} \|\widehat{g}_v\|_{\ell_2}^2 && (\|\widehat{g}_v\|_{\ell_2} > 1/\varepsilon \text{ for } v \in V_3, \text{ and } r < 2) \\
&= \varepsilon^{2-r} \sum_{v \in V_3} \|\widehat{g}_v\|_{\ell_2}^2 \le \varepsilon^{2-r}|V|. && (12)
\end{aligned}
$$

Combining the above with (9) yields

$$
\begin{aligned}
|V_0| &\ge \varepsilon^r \sum_{v \in V_0} \|g_v\|_{L_r}^r \\
&\ge \varepsilon^r \Big((\gamma_r^r + 4\varepsilon^{2-r})|V| - \varepsilon^r|V| - (\gamma_r^r + \varepsilon)|V| - \varepsilon^{2-r}|V| \Big) \\
&\ge \varepsilon^r \varepsilon^{2-r}|V| = \varepsilon^2|V|, && (13)
\end{aligned}
$$

where the last inequality uses the fact that $\varepsilon^{2-r} \ge \varepsilon \ge \varepsilon^r$. ■

Therefore, $|V_0| \ge \varepsilon^2|V|$ and every vertex of v satisfies $\|\widehat{g}_v\|_{\ell_4} > \delta\varepsilon$ and $\|\widehat{g}_v\|_{\ell_2} \le 1/\varepsilon$. Using only these two facts together with $\widehat{\mathbf{g}} \in \widehat{\mathbf{L}}$, Briët et al. [BRS15] proved that if the smoothness parameter J is large enough given other parameters, $\mathcal{L}$ admits a labeling that satisfies a significant fraction of edges.

Lemma 3.5 (Lemma 3.6 of [BRS15])**.** *Let $\beta := \delta^2\varepsilon^3$. There exists an absolute constant $c' > 0$ such that if $\mathcal{L}$ is T-to-1 and $T/(c'\varepsilon^8\beta^4)$-smooth for some $T \in \mathbb{N}$, there is a labeling that satisfies at least $\varepsilon^8\beta^4/1024$ fraction of E.*

This finishes the proof of Lemma 3.3 by setting $\zeta := \varepsilon^8\beta^4/1024$ and $J := D(\zeta)/(c'\varepsilon^8\beta^4)$ with $D(\zeta)$ defined in Theorem 2.9. ■

4 Hardness of $p \to q$ norm

In this section, we prove our main results. We prove Theorem 1.1 on hardness of approximating $p \to q$ norm when $p \geq 2 \geq q$, and Theorem 1.2 on hardness of approximating $p \to q$ norm when $2 < p < q$. By duality, the same hardness is implied for the case of $p < q < 2$.

Our result for $p \geq 2 \geq q$ in Section 4.1 follows from Theorem 3.1 using additional properties in the completeness case. For hypercontractive norms, we start by showing constant factor hardness via reduction from $p \to 2$ norm (see Section 4.2), and then amplify the hardness factor by using the fact that all hypercontractive norms productivize under Kronecker product, which we prove in Section 4.4.

4.1 Hardness for $p \geq 2 \geq q$

We use Theorem 3.1 to prove hardness of $p \to q$ norm for $p \geq 2 \geq q$, which proves Theorem 1.1.

Proof of Theorem 1.1: Fix p, q, and $\delta > 0$ such that $\infty \geq p \geq 2 \geq q$ and $p > q$. Our goal is to prove that $p \to q$ norm is NP-hard to approximate within a factor $1/(\gamma_{p^*}\gamma_q + \delta)$. For $2 \to q$ norm for $1 \leq q < 2$, Theorem 3.1 (with $\varepsilon \leftarrow \delta^{1/(2-q)}$) directly proves a hardness ratio of $1/(\gamma_q + \varepsilon^{2-q}) = 1/(\gamma_q + \delta)$. By duality, it also gives an $1/(\gamma_{p^*} + \delta)$ hardness for $p \to 2$ norm for $p > 2$.

For $p \to q$ norm for $p > 2 > q$, apply Theorem 3.1 with $\varepsilon = (\delta/3)^{\max(1/(2-p^*), 1/(2-q))}$. It gives a polynomial time reduction that produces a symmetric matrix $A \in \mathbb{R}^{n \times n}$ given a 3-SAT formula φ. Our instance for $p \to q$ norm is $AA^T = A^2$.

- (Completeness) If φ is satisfiable, there exists $x \in \mathbb{R}^n$ such that $|x(i)| = 1$ for all $i \in [N]$ and $Ax = x$. Therefore, $A^2 x = x$ and $\|A^2\|_{L_p \to L_q} \geq 1$.
- (Soundness) If φ is not satisfiable,

$$\begin{aligned} \|A\|_{L_p \to L_2} &= \|A\|_{L_2 \to L_{p^*}} \leq \gamma_{p^*} + \varepsilon^{2-p^*} \leq \gamma_{p^*} + \delta/3, \text{ and} \\ \|A\|_{L_2 \to L_q} &\leq \gamma_q + \varepsilon^{2-q} \leq \gamma_q + \delta/3. \end{aligned}$$

  This implies that

$$\|A^2\|_{L_p \to L_q} \leq \|A\|_{L_p \to L_2} \|A\|_{L_2 \to L_q} \leq (\gamma_{p^*} + \delta/3)(\gamma_q + \delta/3) \leq \gamma_{p^*}\gamma_q + \delta\,.$$

This creates a gap of $1/(\gamma_{p^*}\gamma_q + \delta)$ between the completeness and the soundness case. The same gap holds for the counting norm since $\|A^2\|_{\ell_p \to \ell_q} = n^{1/q - 1/p} \cdot \|A^2\|_{L_p \to L_q}$. ∎

4.2 Reduction from $p \to 2$ norm via Approximate Isometries

Let $A \in \mathbb{R}^{n \times n}$ be a hard instance of $p \to 2$ norm. For any $q \geq 1$, if a matrix $B \in \mathbb{R}^{m \times n}$ satisfies $\|Bx\|_{\ell_q} = (1 \pm o(1))\|x\|_{\ell_2}$ for all $x \in \mathbb{R}^n$, then $\|BA\|_{p \to q} = (1 \pm o(1))\|A\|_{p \to 2}$. Thus BA will serve as a hard instance for $p \to q$ norm if one can compute such a matrix B efficiently. In fact, a consequence of the Dvoretzky-Milman theorem is that a sufficiently tall random matrix B satisfies the aforementioned property with high probability. In other words, for $m = m(q,n)$ sufficiently large, a random linear operator from ℓ_2^n to ℓ_q^m is an approximate isometry.

To restate this from a geometric perspective, for $m(q,n)$ sufficiently larger than n, a random section of the unit ball in ℓ_q^m is approximately isometric to the unit ball in ℓ_2^n. In the interest of simplicity, we will instead state and use a corollary of the following matrix deviation inequality due to Schechtman (see [Sch06], Chapter 11 in [Ver17]).

Theorem 4.1 (Schechtman [Sch06])**.** *Let B be an $m \times n$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries. Let $f : \mathbb{R}^m \to \mathbb{R}$ be a positive-homogeneous and subadditive function, and let b be such that $f(y) \le b\|y\|_{\ell_2}$ for all $y \in \mathbb{R}^m$. Then for any $T \subset \mathbb{R}^n$,*

$$\sup_{x \in T} |f(Bx) - \mathbb{E}[f(Bx)]| = O(b \cdot \gamma(T) + t \cdot \mathrm{rad}(T))$$

with probability at least $1 - e^{-t^2}$, where $\mathrm{rad}(T)$ is the radius of T, and $\gamma(T)$ is the gaussian complexity of T defined as

$$\gamma(T) := \mathop{\mathbb{E}}_{g \sim \mathcal{N}(0,I_n)} \left[\sup_{t \in T} |\langle g, t \rangle| \right]$$

The above theorem is established by proving that the random process given by $X_x := f(Bx) - \mathbb{E}[f(Bx)]$ has sub-gaussian increments with respect to L_2 and subsequently appealing to Talagrand's Comparison tail bound.

We will apply this theorem with $f(\cdot) = \|\cdot\|_{\ell_q}$, $b = 1$ and T being the unit ball under $\|\cdot\|_{\ell_2}$. We first state a known estimate of $\mathbb{E}[f(Bx)] = \mathbb{E}[\|Bx\|_{\ell_q}]$ for any fixed x satisfying $\|x\|_{\ell_2} = 1$. Note that when $\|x\|_{\ell_2} = 1$, Bx has the same distribution as an m-dimensional random vector with i.i.d. $\mathcal{N}(0,1)$ coordinates.

Theorem 4.2 (Biau and Mason [BM15])**.** *Let $X \in \mathbb{R}^m$ be a random vector with i.i.d. $\mathcal{N}(0,1)$ coordinates. Then for any $q \ge 2$,*

$$\mathbb{E}\left[\|X\|_{\ell_q}\right] = m^{1/q} \cdot \gamma_q + O(m^{(1/q)-1}).$$

We are now equipped to see that a tall random gaussian matrix is an approximate isometry (as a linear map from ℓ_2^n to ℓ_q^m) with high probability.

Corollary 4.3. *Let B be an $m \times n$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries where $m = \omega(n^{q/2})$. Then with probability at least $1 - e^{-n}$, every vector $x \in \mathbb{R}^n$ satisfies,*

$$\|Bx\|_{\ell_q} = (1 \pm o(1)) \cdot m^{1/q} \cdot \gamma_q \cdot \|x\|_{\ell_2}.$$

Proof. We apply Theorem 4.1 with function f being the ℓ_q norm, $b = 1$, and $t = \sqrt{n}$. Further we set T to be the ℓ_2 unit sphere, which yields $\gamma(T) = \Theta(\sqrt{n})$ and $\mathrm{rad}(T) = 1$. Applying Theorem 4.2 yields that with probability at least $1 - e^{t^2} = 1 - e^{-n}$, for all x with $\|x\|_{\ell_2} = 1$, we have

$$\begin{aligned}
\left| \|Bx\|_{\ell_q} - m^{1/q} \cdot \gamma_q \right| &\le \left| \|Bx\|_{\ell_q} - \mathbb{E}\left[\|X\|_{\ell_q}\right] \right| + \left| \mathbb{E}\left[\|X\|_{\ell_q}\right] - m^{1/q} \cdot \gamma_q \right| \\
&\le O(b \cdot \gamma(T) + t \cdot \mathrm{rad}(T) + m^{(1/q)-1}) \\
&\le O(\sqrt{n} + \sqrt{n} + m^{(1/q)-1}) \\
&\le o(m^{1/q}).
\end{aligned}$$

∎

We thus obtain the desired constant factor hardness:

Proposition 4.4. *For any $p > 2$, $2 \le q < \infty$ and any $\varepsilon > 0$, there is no polynomial time algorithm that approximates $p{\to}q$ norm (and consequently $q^*{\to}p^*$ norm) within a factor of $1/\gamma_{p^*} - \varepsilon$ unless $\mathsf{NP} \not\subseteq \mathsf{BPP}$.*

Proof. By Corollary 4.3, for every $n \times n$ matrix A and a random $m \times n$ matrix B with i.i.d. $\mathcal{N}(0,1)$ entries ($m = \omega(n^{q/2})$), with probability at least $1 - e^{-n}$, we have

$$\|BA\|_{\ell_p \to \ell_q} = (1 \pm o(1)) \cdot \gamma_q \cdot m^{1/q} \cdot \|A\|_{\ell_p \to \ell_2}.$$

Thus the reduction $A \to BA$ combined with $p{\to}2$ norm hardness implied by Theorem 3.1, yields the claim. ∎

The generality of the concentration of measure phenomenon underlying the proof of the Dvoretzky-Milman theorem allows us to generalize Proposition 4.4, to obtain constant factor hardness of maximizing various norms over the ℓ_p ball ($p > 2$). In this more general version, the strength of our hardness assumption is dependent on the gaussian width of the dual of the norm being maximized. Its proof is identical to that of Proposition 4.4.

Theorem 4.5. *Consider any $p > 2, \varepsilon > 0$, and any family $(f_m)_{m \in \mathbb{N}}$ of positive-homogeneous and subadditive functions where $f_m : \mathbb{R}^m \to \mathbb{R}$. Let $(b_m)_{m \in \mathbb{N}}$ be such that $f_m(y) \le b_m \cdot \|y\|_{\ell_2}$ for all y and let $N = N(n)$ be such that $\gamma_*(f_N) = \omega(b_N \cdot \sqrt{n})$, where*

$$\gamma^*(f_N) := \mathop{\mathbb{E}}_{g \sim \mathcal{N}(0, I_N)} [f_N(g)] \,.$$

Then unless $\mathsf{NP} \not\subseteq \mathsf{BPTIME}(N(n))$, there is no polynomial time $(1/\gamma_{p^*} - \varepsilon)$-approximation algorithm for the problem of computing $\sup_{\|x\|_p = 1} f_m(Ax)$, given an $m \times n$ matrix A.

4.3 Derandomized Reduction

In this section, we show how to derandomize the reduction in Proposition 4.4 to obtain NP-hardness when $q \ge 2$ is an even integer and $p > 2$. Similarly to Section 4.2, given $A \in \mathbb{R}^{n \times n}$ as a hard instance of $p{\to}2$ norm, our strategy is to construct a matrix $B \in \mathbb{R}^{m \times n}$ and output BA as a hard instance of $p{\to}q$ norm.

Instead of requiring B to satisfy $\|Bx\|_{\ell_q} = (1 \pm o(1))\|x\|_{\ell_2}$ for all $x \in \mathbb{R}^n$, we show that $\|Bx\|_{\ell_q} \le (1 + o(1))\|x\|_{\ell_2}$ for all $x \in \mathbb{R}^n$ and $\|Bx\|_{\ell_q} \ge (1 - o(1))\|x\|_{\ell_2}$ when every coordinate of x has the same absolute value. Since Theorem 3.1 ensures that $\|A\|_{\ell_p \to \ell_2}$ is achieved by $x = Ax$ for such a well-spread x in the completeness case, BA serves as a hard instance for $p{\to}q$ norm.

We use the following construction of q-wise independent sets to construct such a B deterministically.

Theorem 4.6 (Alon, Babai, and Itai [ABI86])**.** *For any $k \in \mathbb{N}$, one can compute a set S of vectors in $\{\pm 1\}^n$ of size $O(n^{k/2})$, in time $n^{O(k)}$, such that the vector random variable Y obtained by sampling uniformly from S satisfies that for any $I \in \binom{[n]}{k}$, the marginal distribution $Y|_I$ is the uniform distribution over $\{\pm 1\}^k$.*

For a matrix B as above, a randomly chosen row behaves similarly to an n-dimensional Rademacher random vector with respect to $\|\cdot\|_{\ell_q}$.

Corollary 4.7. *Let $R \in \mathbb{R}^n$ be a vector random variable with i.i.d. Rademacher (± 1) coordinates. For any even integer $q \geq 2$, there is an $m \times n$ matrix B with $m = O(n^{q/2})$, computable in $n^{O(q)}$ time, such that for all $x \in \mathbb{R}^n$, we have*

$$\|Bx\|_{\ell_q} = m^{1/q} \cdot \mathop{\mathbb{E}}_{R} \left[\langle R, x \rangle^q \right]^{1/q}.$$

Proof. Let B be a matrix, the set of whose rows is precisely S. By Theorem 4.6,

$$\|Bx\|_{\ell_q}^q = \sum_{Y \in S} \langle Y, x \rangle^q = m \cdot \mathop{\mathbb{E}}_{R} \left[\langle R, x \rangle^q \right] . \qquad \blacksquare$$

We use the following two results that will bound $\|BA\|_{\ell_p \to \ell_q}$ for the completeness case and the soundness case respectively.

Theorem 4.8 (Stechkin [Ste61])**.** *Let $R \in \mathbb{R}^n$ be a vector random variable with i.i.d. Rademacher coordinates. Then for any $q \geq 2$ and any $x \in \mathbb{R}^n$ whose coordinates have the same absolute value,*

$$\mathbb{E} \left[\langle R, x \rangle \right]^{1/q} = (1 - o(1)) \cdot \gamma_q \|x\|_{\ell_2}.$$

Theorem 4.9 (Khintchine inequality [Haa81])**.** *Let $R \in \mathbb{R}^n$ be a vector random variable with i.i.d. Rademacher coordinates. Then for any $q \geq 2$ and any $x \in \mathbb{R}^n$,*

$$\mathbb{E} \left[\langle R, x \rangle^q \right]^{1/q} \leq \gamma_q \cdot \|x\|_{\ell_2}.$$

We finally prove the derandomimzed version of Proposition 4.4 for even $q \geq 2$.

Proposition 4.10. *For any $p > 2, \varepsilon > 0$, and any even integer $q \geq 2$, it is NP-hard to approximate $p \to q$ norm within a factor of $1/\gamma_{p^*} - \varepsilon$.*

Proof. Apply Theorem 3.1 with $r_1 \leftarrow p^*$ and $\varepsilon \leftarrow \varepsilon$. Given an instance φ of 3-SAT, Theorem 3.1 produces a symmetric matrix $A \in \mathbb{R}^{n \times n}$ in polynomial time as a hard instance of $p \to 2$ norm. Our instance for $p \to q$ norm is BA where B is the $m \times n$ matrix given by Corollary 4.7 with $m = O(n^{q/2})$.

- (Completeness) If φ is satisfiable, there exists a vector $x \in \{\pm \frac{1}{\sqrt{n}}\}^n$ such that $Ax = x$. So we have $\|BAx\|_{\ell_q} = \|Bx\|_{\ell_q} = (1 - o(1)) \cdot m^{1/q} \cdot \gamma_q$, where the last equality uses Corollary 4.7 and Theorem 4.8. Thus $\|BA\|_{\ell_p \to \ell_q} \geq (1 - o(1)) \cdot m^{1/q} \cdot \gamma_q$.
- (Soundness) If φ is not satisfiable, then for any x with $\|x\|_{\ell_p} = 1$,

$$\begin{aligned} \|BAx\|_{\ell_q} &= m^{1/q} \cdot \mathop{\mathbb{E}}_{R} \left[\langle R, Ax \rangle^q \right]^{1/q} \leq m^{1/q} \cdot \gamma_q \cdot \|Ax\|_{\ell_2} \\ &\leq \; m^{1/q} \cdot \gamma_q \cdot \|A\|_{\ell_p \to \ell_2} \leq \; m^{1/q} \cdot \gamma_q \cdot (\gamma_{p^*} - \varepsilon) \end{aligned}$$

  where the first inequality is a direct application of Theorem 4.9. $\blacksquare$

4.4 Hypercontractive Norms Productivize

We will next amplify our hardness results using the fact that hypercontractive norms productivize under the natural operation of Kronecker or tensor product. Bhaskara and Vijayraghavan [BV11] showed this for the special case of $p = q$ and the Harrow and Montanaro [HM13] showed this for 2→4 norm (via parallel repetition for QMA(2)). In this section we prove this claim whenever $p \leq q$.

Theorem 4.11. *Let A and B be $m_1 \times n_1$ and $m_2 \times n_2$ matrices respectively. Then for any $1 \leq p \leq q < \infty$, $\|A \otimes B\|_{\ell_p \to \ell_q} \leq \|A\|_{\ell_p \to \ell_q} \cdot \|B\|_{\ell_p \to \ell_q}$.*

Proof. We will begin with some notation. Let a_i, b_j respectively denote the i-th and j-th rows of A and B. Consider any $z \in \mathbb{R}^{[n_1] \times [n_2]}$ satisfying $\|z\|_{\ell_p} = 1$. For $k \in [n_1]$, let $z_k \in \mathbb{R}^{n_2}$ denote the vector given by $z_k(\ell) := z(k, \ell)$. For $j \in [m_2]$, let $\bar{z}_j \in \mathbb{R}^{n_1}$ denote the vector given by $\bar{z}_j(k) := \langle b_j, z_k \rangle$. Finally, for $k \in [n_1]$, let $\lambda_k := \|z_k\|_{\ell_p}^p$ and let $v_k \in \mathbb{R}^{m_2}$ be the vector given by $v_k(j) := |\bar{z}_j(k)|^p / \lambda_k$.

We begin by 'peeling off' A:

$$\begin{aligned}
\|(A \otimes B)z\|_{\ell_q}^q &= \sum_{i,j} |\langle a_i \otimes b_j, z \rangle|^q = \sum_j \sum_i |\langle a_i, \bar{z}_j \rangle|^q \\
&= \sum_j \|A\bar{z}_j\|_{\ell_q}^q \\
&\leq \|A\|_{\ell_p \to \ell_q}^q \cdot \sum_j \|\bar{z}_j\|_{\ell_p}^q \\
&= \|A\|_{\ell_p \to \ell_q}^q \cdot \sum_j \left(\|\bar{z}_j\|_{\ell_p}^p\right)^{q/p}
\end{aligned}$$

In the special case of $p = q$, the proof ends here since the expression is a sum of terms of the form $\|By\|_{\ell_p}^p$ and can thus be upper bounded term-wise by $\|B\|_{\ell_p \to \ell_p}^p \cdot \|z_k\|_{\ell_p}^p$ which sums to $\|B\|_{\ell_q \to \ell_p}^p$. To handle the case of $q > p$, we will use a convexity argument:

$$\begin{aligned}
&\|A\|_{\ell_p \to \ell_q}^q \cdot \sum_j \left(\|\bar{z}_j\|_{\ell_p}^p\right)^{q/p} \\
= \ &\|A\|_{\ell_p \to \ell_q}^q \cdot \sum_j \left(\sum_k |\bar{z}_j(k)|^p\right)^{q/p} \\
= \ &\|A\|_{\ell_p \to \ell_q}^q \cdot \|\sum_k \lambda_k \cdot v_k\|_{\ell_{q/p}}^{q/p} && (|\bar{z}_j(k)|^p = \lambda_k v_k(j)) \\
\leq \ &\|A\|_{\ell_p \to \ell_q}^q \cdot \sum_k \lambda_k \cdot \|v_k\|_{\ell_{q/p}}^{q/p} && (\text{by convexity of } \|\cdot\|_{q/p}^{q/p} \text{ when } q \geq p) \\
\leq \ &\|A\|_{\ell_p \to \ell_q}^q \cdot \max_k \|v_k\|_{\ell_{q/p}}^{q/p}
\end{aligned}$$

It remains to show that $\|v_k\|_{\ell_{q/p}}^{q/p}$ is precisely $\|Bz_k\|_{\ell_q}^q / \|z_k\|_{\ell_p}^q$.

$$\begin{aligned}
\|A\|_{\ell_p\to\ell_q}^q \cdot \max_k \|v_k\|_{\ell_{q/p}}^{q/p} &= \|A\|_{\ell_p\to\ell_q}^q \cdot \max_k \frac{1}{\|z_k\|_{\ell_p}^q} \cdot \sum_j |\overline{z}_j(k)|^q \\
&= \|A\|_{\ell_p\to\ell_q}^q \cdot \max_k \frac{1}{\|z_k\|_{\ell_p}^q} \cdot \sum_j |\langle b_j, z_k\rangle|^q \\
&= \|A\|_{\ell_p\to\ell_q}^q \cdot \max_k \frac{\|Bz_k\|_{\ell_q}^q}{\|z_k\|_{\ell_p}^q} \\
&\le \|A\|_{\ell_p\to\ell_q}^q \cdot \|B\|_{\ell_p\to\ell_q}^q
\end{aligned}$$

Thus we have established $\|A \otimes B\|_{\ell_p\to\ell_q} \le \|A\|_{\ell_p\to\ell_q} \cdot \|B\|_{\ell_p\to\ell_q}$. Lastly, the claim follows by observing that the statement is equivalent to the statement obtained by replacing the counting norms with expectation norms. ■

We finally establish super constant NP-Hardness of approximating $p{\to}q$ norm, proving Theorem 1.2.

Proof of Theorem 1.2: Fix $2 < p \le q < \infty$. Proposition 4.4 states that there exists $c = c(p,q) > 1$ such that any polynomial time algorithm approximating the $p{\to}q$ norm of an $n \times n$-matrix A within a factor of c will imply $\mathsf{NP} \subseteq \mathsf{BPP}$. Using Theorem 4.11, for any integer $k \in \mathbb{N}$ and $N = n^k$, any polynomial time algorithm approximating the $p{\to}q$ norm of an $N \times N$-matrix $A^{\otimes k}$ within a factor of c^k implies that NP admits a randomized algorithm running in time $\mathrm{poly}(N) = n^{O(k)}$. Under $\mathsf{NP} \not\subseteq \mathsf{BPP}$, any constant factor approximation algorithm is ruled out by setting k to be a sufficiently large constant. For any $\varepsilon > 0$, setting $k = \log^{1/\varepsilon} n$ rules out an approximation factor of $c^k = 2^{O(\log^{1-\varepsilon} N)}$ unless $\mathsf{NP} \subseteq \mathsf{BPTIME}\left(2^{\log^{O(1)} n}\right)$.

By duality, the same statements hold for $1 < p \le q < 2$. When $2 < p \le q$ and q is an even integer, all reductions become deterministic due to Proposition 4.10. ■

A Dictatorship Test

First we prove an implication of Berry-Esséen estimate for fractional moments (similar to Lemma 3.3 of [GRSW16], see also [KNS10]).

Lemma A.1. *There exist universal constants $c > 0$ and $\delta_0 > 0$ such that the following statement is true. If $X_1, \cdots, X_n$ are bounded independent random variables with $|X_i| \le 1$, $\mathbb{E}[X_i] = 0$ for $i \in [n]$, and $\sum_{i\in[n]} \mathbb{E}[X_i^2] = 1$, $\sum_{i\in[n]} \mathbb{E}[|X_i|^3] \le \delta$ for some $0 < \delta < \delta_0$, then for every $p \ge 1$:*

$$\left(\mathbb{E}\left[\left|\sum_{j=1}^n X_j\right|^p\right]\right)^{\frac{1}{p}} \le \gamma_p \cdot \left(1 + c\delta\left(\log\left(1/\delta\right)\right)^{\frac{p}{2}}\right).$$

Now we state and prove the main lemma of this section:

Lemma A.2. *Let $f : \{\pm 1\}^R \to \mathbb{R}$ be a linear function for some positive integer $R \in \mathbb{N}$ and $\widehat{f} : [R] \to \mathbb{R}$ be its linear Fourier coefficients defined by*

$$\widehat{f}(i) := \mathop{\mathbb{E}}_{x \in \{\pm 1\}^R} [x_i f(x)] .$$

For all $\varepsilon > 0$, there exists $\delta > 0$ such that if $\|f\|_{L_r} > (\gamma_r + \varepsilon)\|\widehat{f}\|_{\ell_2}$ then $\|\widehat{f}\|_{\ell_4} > \delta\|\widehat{f}\|_{\ell_2}$ for all $1 \le r < 2$.

Proof. We will prove this lemma by the method of contradiction. Let us assume $\|\widehat{f}\|_{\ell_4} \le \delta\|\widehat{f}\|_{\ell_2}$, for δ to be fixed later.

Let us define $y_i := \frac{\widehat{f}(i)}{\|\widehat{f}\|_{\ell_2}}$. Then, for all $x \in \{-1,1\}^R$,

$$g(x) := \sum_{i \in [n]} x_i \cdot y_i = \frac{f(x)}{\|\widehat{f}\|_{\ell_2}} .$$

Let $Z_i = x_i \cdot y_i$ be the random variable when x_i is independently uniformly randomly chosen from $\{-1,1\}$. Now

$$\sum_{i \in [n]} \mathbb{E}\left[Z_i^2\right] = \sum_{i \in [n]} \frac{\widehat{f}(i)^2}{\|\widehat{f}\|_{\ell_2}^2} = 1 .$$

and

$$\sum_{i \in [n]} \mathbb{E}\left[|Z_i|^3\right] = \sum_{i \in [n]} \frac{\left|\widehat{f}(i)\right|^3}{\|\widehat{f}\|_{\ell_2}^3} = \sum_{i \in [n]} \frac{\left|\widehat{f}(i)\right|^2}{\|\widehat{f}\|_{\ell_2}^2} \cdot \frac{\left|\widehat{f}(i)\right|}{\|\widehat{f}\|_{\ell_2}} \le \frac{\|\widehat{f}\|_{\ell_4}^2}{\|\widehat{f}\|_{\ell_2}^2} \le \delta^2 ,$$

where the penultimate inequality follows from Cauchy-Schwarz ineqality.

Hence, by applying Lemma A.1 on the random variables $Z_1, \cdots, Z_n$, we get:

$$\begin{aligned} \frac{\|f\|_{L_r}}{\|\widehat{f}\|_{\ell_2}} = \|g\|_{L_r} &= \left(\mathop{\mathbb{E}}_{x \in \{-1,1\}^n} \left[|g(x)|^r\right]\right)^{\frac{1}{r}} \\ &= \left(\mathop{\mathbb{E}}_{x \in \{-1,1\}^n} \left[\left|\sum_{i \in [n]} Z_i\right|^r\right]\right)^{\frac{1}{r}} \\ &\le \gamma_r \left(1 + c\delta^2 \left(\log \frac{1}{\delta}\right)^r\right) \end{aligned}$$

We choose $\delta > 0$ small enough (since $1 \le r < 2$, setting $\delta < \frac{\sqrt{\varepsilon}}{\min(\delta_0, \sqrt{\gamma_2} \log \frac{c\gamma_2}{\varepsilon})} = \frac{\sqrt{\varepsilon}}{\min(\delta_0, \log \frac{c}{\varepsilon})}$ suffices) so that $\delta^2(\log \frac{1}{\delta})^r < \frac{\varepsilon}{c\gamma_r}$. For this choise of δ, we get: $\|f\|_{L_r} \le (\gamma_r + \varepsilon)\|\widehat{f}\|_{\ell_2}$ – a contradiction. And hence the proof follows. ■

Finally we prove Lemma A.1:

Proof of Lemma A.1: The proof is almost similar to that of Lemma 2.1 of [KNS10]. From Berry-Esséen theorem (see [vB72] for the constant), we get that:

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} X_i\right| \geq u\right] \leq \mathbb{P}\left[|g| \geq u\right] + 2\sum_{i=1}^{n}\mathbb{E}\left[|X_i|^3\right] \leq \mathbb{P}\left[|g| \geq u\right] + 2\delta\,,$$

for every $u > 0$ and where $g \sim \mathcal{N}(0,1)$. By Hoeffding's lemma,

$$\mathbb{P}\left[\left|\sum_{i\in[n]} X_i\right| \geq t\right] < 2\mathrm{e}^{-2t^2}$$

for every $t > 0$. Combining the above observations, we get:

$$\begin{aligned}
\mathbb{E}\left[\left|\sum_{i=1}^{n} X_i\right|^p\right] &= \int_0^\infty pu^{p-1}\mathbb{P}\left[\left|\sum_{i=1}^{n} X_i\right| \geq u\right] du \\
&\leq \int_0^a pu^{p-1}\mathbb{P}\left[|g| > u\right] du + 2\delta a^p + 2\int_a^\infty pu^{p-1}\mathrm{e}^{-2u^2} du \\
&= \sqrt{\frac{2}{\pi}}\int_0^a u^p \mathrm{e}^{-u^2/2} du + 2\delta a^p + \frac{2p}{2^{\frac{p-1}{2}}}\int_{2a^2}^\infty z^{\frac{p+1}{2}-1}\mathrm{e}^{-z} dz \\
&= \gamma_p^p - \sqrt{\frac{2}{\pi}}\int_a^\infty u^p \mathrm{e}^{-u^2/2} du + 2\delta a^p + \Gamma\left(\frac{p+1}{2}, 2a^2\right)\,,
\end{aligned}$$

where $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function and a is a large constant determined later depending on δ and p. The second term is bounded as

$$\int_a^\infty u^p \mathrm{e}^{-u^2/2} du = a^{p-1}\mathrm{e}^{-a^2/2} + (p-1)\int_a^\infty u^{p-2}\mathrm{e}^{-u^2/2} du \leq a^{p-1}\mathrm{e}^{-a^2/2} + \frac{p-1}{a^2}\int_a^\infty u^p \mathrm{e}^{-u^2/2} du\,.$$

Hence $\int_a^\infty u^p \mathrm{e}^{-u^2/2} du \leq \frac{a^{p+1}e^{-a^2/2}}{1+a^2-p}$.

We know, $\Gamma(p+1/2, x) \to x^{\frac{p-1}{2}}\mathrm{e}^{-x}$ as $x \to \infty$. We choose $a = \gamma_p\sqrt{\log\frac{1}{\delta}}$. Hence there exists δ_0 so that for all small enough $\delta < \delta_0$, we have $\Gamma(p+1/2, 2a^2) \sim 2^{\frac{p-1}{2}}a^{p-1}\delta^{2\gamma_p^2} \ll \delta a^p$ where the last inequality follows from the fact that $2\gamma_p^2 > 1$ (as $p > 1$). Putting all this together, we get:

$$2\delta a^p + \Gamma\left(\frac{p+1}{2}, 2a^2\right) - \sqrt{\frac{2}{\pi}}\int_a^\infty u^p \mathrm{e}^{-u^2/2} du \ll 3\delta a^p - \sqrt{\frac{2}{\pi}}\frac{a^{p+1}e^{-a^2/2}}{1+a^2-p} \leq c\gamma_p^p\delta\left(\log\frac{1}{\delta}\right)^{p/2}\,,$$

where c is an absolute constant independent of a and p. This completes the proof of the lemma. ■

References

[ABH+05] Sanjeev Arora, Eli Berger, Elad Hazan, Guy Kindler, and Muli Safra. On non-approximability for quadratic programs. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 206–215. IEEE, 2005. 2

[ABI86] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of algorithms*, 7(4):567–583, 1986. 14

[AMMN06] Noga Alon, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. Quadratic forms on graphs. *Inventiones mathematicae*, 163(3):499–522, 2006. Conference version in STOC '05. 2

[BBH+12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326. ACM, 2012. 1, 2, 3

[BH15] Fernando GSL Brandao and Aram W Harrow. Estimating operator norms using covering nets. *arXiv preprint arXiv:1509.05065*, 2015. 1

[Bis11] Punyashloka Biswal. Hypercontractivity and its applications. *arXiv preprint arXiv:1101.2913*, 2011. 1

[BM15] Gérard Biau and David M Mason. High-dimensional p-norms. In *Mathematical Statistics and Limit Theorems*, pages 21–40. Springer, 2015. 13

[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. 2, 4, 9, 10, 11

[BV11] Aditya Bhaskara and Aravindan Vijayaraghavan. Approximating matrix p-norms. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 497–511. SIAM, 2011. 2, 5, 16

[Gro56] Alexandre Grothendieck. *Résumé de la théorie métrique des produits tensoriels topologiques*. Soc. de Matemática de São Paulo, 1956. 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. 2, 4, 7, 9, 17

[Haa81] Uffe Haagerup. The best constants in the khintchine inequality. *Studia Mathematica*, 70(3):231–283, 1981. 15

[HM13] Aram W Harrow and Ashley Montanaro. Testing product states, quantum Merlin-Arthur games and tensor optimization. *Journal of the ACM (JACM)*, 60(1):3, 2013. 1, 2, 5, 16

[HNW16] Aram W Harrow, Anand Natarajan, and Xiaodi Wu. Limitations of semidefinite programs for separable states and entangled games. *arXiv preprint arXiv:1612.09306*, 2016. 2

[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. 7

[KMW18] Aditya Krishnan, Sidhanth Mohanty, and David P. Woodruff. On sketching q to p norms. Manuscript, 2018. 5

[KN08] Subhash Khot and Assaf Naor. Linear equations modulo 2 and the l_1 diameter of convex bodies. *SIAM Journal on Computing*, 38(4):1448–1463, 2008. 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, 17, 19

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

[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

[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. 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. 2

[Sch06] Gideon Schechtman. Two observations regarding embedding subsets of euclidean spaces in normed spaces. *Advances in Mathematics*, 200(1):125–135, 2006. 13

[Ste61] Sergei Borisovich Stechkin. On best lacunary systems of functions. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 25(3):357–366, 1961. 15

[Ste05] Daureen Steinberg. Computation of matrix norms with applications to robust optimization. *Research thesis, Technion-Israel University of Technology*, 2005. 2

[vB72] Paul van Beek. An application of Fourier methods to the problem of sharpening the Berry-Esseen inequality. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 23(3):187–196, 1972. 19

[Ver17] Roman Vershynin. *High Dimensional Probability*. 2017. 5, 13

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il