Approximability of $p \to q$ Matrix Norms: Generalized Krivine Rounding and Hypercontractive Hardness

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

Abstract

We study the problem of computing the $p \to q$ operator norm of a matrix A in $\mathbb{R}^{m\times n}$, defined as $||A||_{p\to q} := \sup_{x\in\mathbb{R}^n\setminus\{0\}} ||Ax||_q/||x||_p$. 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 is in $[q, p]$, and the problem is hard to approximate within almost polynomial factors when 2 is not in [q,p]. For the case when 2 is in $[q, p]$ we prove almost matching approximation and NP-hardness results.

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 prove the first NP-hardness result for approximating hypercontractive norms. We show that for any $1 < p < q < \infty$ with 2 not in $[p, q]$, $||A||_{p\to q}$ is hard to approximate within $2^{O(\log^{1-\varepsilon} n)}$ assuming NP is not contained in BPTIME($2^{\log^{O(1)} n}$).

*Carnegie Mellon University. Supported by NSF CCF-1422045 and CCF-1526092. Part of the work was done while visiting UC Berkeley and CMSA, Harvard. `vpb@cs.cmu.edu`

†Toyota Technological Institute Chicago. Supported by NSF CCF-1254044 `mkghosh@ttic.edu`

‡Carnegie Mellon University. Supported in part by NSF grant CCF-1526092. Part of the work was done while visiting CMSA, Harvard. `guruswami@cmu.edu`

§New York University. Supported by the Simons Institute for the Theory of Computing. `euiwoong@cims.nyu.edu`

¶Toyota Technological Institute Chicago. Supported by NSF grants CCF-1254044 and CCF-1816372 `madhurt@ttic.edu`

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, 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 $1/p + 1/p^* = 1/q + 1/q^* = 1$.

1.1 Hypercontractive norms.

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]. 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].

Bounds on hypercontractive norms of operators are also used to prove expansion of small sets in graphs. Indeed, if f is the indicator function of set S of measure δ in a graph with adjacency matrix A, then we have that for any $p \leq q$,

$$\Phi(S) \;=\; 1 - \frac{\langle f, Af\rangle}{\|f\|_2^2} \;\geq\; 1 - \frac{\|f\|_{q^*} \cdot \|Af\|_q}{\delta} \;\geq\; 1 - \|A\|_{p\to q} \cdot \delta^{1/p-1/q}\,.$$

It was proved by Barak et al. [BBH+12] that the above connection to small-set expansion can in fact be made two-sided for a special case of the $2\to q$ norm. They proved that to resolve the promise version of the small-set expansion (SSE) problem, it suffices to distinguish the cases $\|A\|_{2\to q} \leq c\cdot\sigma_{\min}$ and $\|A\|_{2\to q} \geq C\cdot\sigma_{\min}$, where $\sigma_{\min}$ is the least non-zero singular value of A and $C > c > 1$ are appropriately chosen constants based on the parameters of the SSE problem. Thus, the approximability of $2\to q$ norm is closely related to the small-set expansion problem. In particular, proving NP-hardness of approximating $2\to q$ norm is (necessarily) an intermediate goal towards proving the Small-Set Expansion Hypothesis of Raghavendra and Steurer [RS10].

However, relatively few results algorithmic and hardness results are known for approximating hypercontractive norms. A result of Steinberg's [Ste05] gives 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 the promise version of the problem described above, running in time $\exp\left(\tilde{O}(n^{2/q})\right)$. They also provide an additive approximation algorithm for $2\to 4$ norm (where the error depends on the $2\to 2$ 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 $\|A\|_{2\to 4}$ 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 quasi-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 $\|A\|_{2\to 4}$ within any constant factor.

It is natural to ask if the bottleneck in proving (constant factor) hardness of approximation for $2\to q$ norm arises from the fact from the nature of the domain (the ℓ_2 ball) or from hypercontractive nature of the objective. As discussed in Section 4, *all* hypercontractive norms present a barrier for gadget reductions, since if a "true" solution x is meant to encode the assignment to a (say) label cover problem with consistency checked via local gadgets, then (for $q > p$), a "cheating solution" may make the value of $\|Ax\|_q$ very large by using a sparse x which does not carry any meaningful information about the underlying label cover problem.

We show that (somewhat surprisingly, at least for the authors) it is indeed possible to overcome the barrier for gadget reductions for hypercontractive norms, for any $2 < p < q$ (and by duality, for any $p < q < 2$). This gives the first NP-hardness result for hypercontractive norms (under randomized reductions). Assuming ETH, this also rules out a constant factor approximation algorithm that runs in 2^{n^δ} for some $\delta := \delta(p,q)$.

Theorem 1.1. *For any p,q such that $1 < p \leq q < 2$ or $2 < p \leq q < \infty$ and a constant $c > 1$, unless NP $\in$ BPP, no polytime algorithm approximates $p\to q$ norm within a factor of c. The reduction runs in time $n^{B_p q}$ for $2 < p < q$, where $B_p = \mathrm{poly}(1/(1-\gamma_{p^*}))$.*

We show that the above hardness can be strengthened to any constant factor via a simple tensoring argument. In fact, this also shows 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$\left(2^{(\log n)^{O(1)}}\right)$*

We also note that the operator A arising in our reduction in Theorem 1.1 satisfies $\sigma_{\min}(A) \approx 1$ (and is in fact a product of a carefully chosen projection and a scaled random Gaussian matrix). For such an A, we prove the hardness of distinguishing $\|A\|_{p\to q} \leq c$ and $\|A\|_{p\to q} \geq C$, for constants $C > c > 1$. For the corresponding problem in the case of $2\to q$ norm, Barak et al. [BBH+12] gave a subexponential algorithm running in time $\exp\left(O(n^{2/q})\right)$ (which works for every $C > c > 1$). On the other hand, since the running time of our reduction is $n^{O(q)}$, we get that assuming ETH, no algorithm can distinguish the above cases for $p\to q$ norm in time $\exp\left(n^{o(1/q)}\right)$, for any $p \leq q$ when $2 \notin [p,q]$.

While the above results give some possible reductions for working with hypercontractive norms, it remains an interesting problem to understand the role of the domain as a barrier to proving hardness results for the $2\to q$ norm problems. In fact, no hardness results are available even for the more general problem of polynomial optimization over the ℓ_2 ball. 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

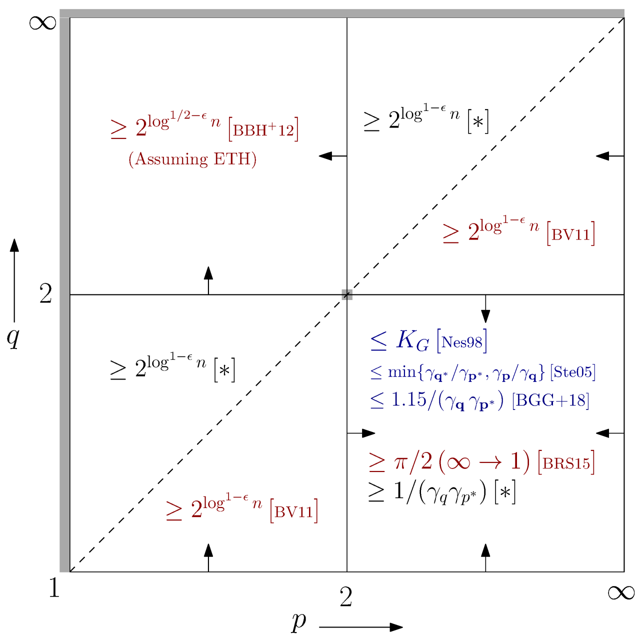

Figure 1: Upper and lower bounds for approximating $\|A\|_{p\to q}$. Arrows indicate the region to which a boundary belongs and thicker shaded regions represent exact algorithms. Our results are indicated by [*]. We omit UGC-based hardness results in the figure.

$2 \notin [p, q]$. This is indeed known to be true in the *non-hypercontractive case* with $p \geq q$. In fact, our results are obtained via new hardness results for the case $p \geq q$, as described below.

1.2 The non-hypercontractive case. Several results are known in the case when $p \geq q$, and we summarize known results for matrix norms in Figure 1, for the both the hypercontractive and non-hypercontractive cases. 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, Figure 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).

As is apparent from the figure, the problem of approximating $\|A\|_{p\to q}$ for $p \geq q$ admits good approximations when $2 \in [q, p]$, and is hard otherwise. 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 within any constant factor, and hardness of approximation within an $O\left(2^{(\log n)^{1-\varepsilon}}\right)$ factor for arbitrary $\varepsilon > 0$ assuming $\mathsf{NP} \not\subseteq \mathsf{DTIME}\left(2^{(\log n)^{O(1)}}\right)$.

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) [KN09] (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].

Hardness. 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 $p \to 2$, and moreover almost matches the algorithmic results mentioned above.

THEOREM 1.3. *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 $1/(\gamma_{p^*}\gamma_q) - \varepsilon$.*

Both Theorem 1.1 and Theorem 1.3 are consequences of a more 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. We also show our methods can be used to provide a simple proof (albeit via randomized reductions) of the $2^{\Omega((\log n)^{1-\varepsilon})}$ hardness for the non-hypercontractive case when $2 \notin [q, p]$, which was proved by [BV11].

The Search For Optimal Constants and Optimal Algorithms. The goal of determining the right approximation ratio for these problems is closely related to the question of finding the optimal (rounding) algorithms. For the Grothendieck problem, the goal is to find $y \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ with $\|y\|_\infty, \|x\|_\infty \leq 1$, and one considers the following semidefinite relaxation:

$$\begin{array}{lll} \text{maximize} & \displaystyle\sum_{i,j} A_{i,j} \cdot \langle u^i, v^j \rangle \quad \text{s.t.} & \\ \text{subject to} & \|u^i\|_2 \leq 1, \|v^j\|_2 \leq 1 & \forall i \in [m], j \in [n] \\ & u^i, v^j \in \mathbb{R}^{m+n} & \forall i \in [m], j \in [n] \end{array}$$

By the bilinear nature of the problem above, it is clear that the optimal x, y can be taken to have entries in $\{-1, 1\}$. A bound on the approximation ratio[1] of the above program is then obtained by designing a good "rounding" algorithm which maps the vectors u^i, v^j to values in $\{-1, 1\}$. Krivine's analysis [Kri77] corresponds to a rounding algorithm which considers a random vector $\mathbf{g} \sim \mathcal{N}(0, I_{m+n})$ and rounds to x, y defined as

$$y_i := \operatorname{sgn}\left(\langle \varphi(u^i), \mathbf{g}\rangle\right) \quad \text{and} \quad x_j := \operatorname{sgn}\left(\langle \psi(v^j), \mathbf{g}\rangle\right),$$

for some appropriately chosen transformations φ and ψ. This gives the following upper bound on the approximation ratio of the above relaxation, and hence on the value of the Grothendieck constant K_G:

$$K_G \leq \frac{1}{\sinh^{-1}(1)} \cdot \frac{\pi}{2} = \frac{1}{\ln(1+\sqrt{2})} \cdot \frac{\pi}{2}.$$

[1]Since we will be dealing with problems where the optimal solution may not be integral, we will use the term "approximation ratio" instead of "integrality gap".

Braverman et al. [BMMN13] show that the above bound can be strictly improved (by a very small amount) using a two dimensional analogue of the above algorithm, where the value y_i is taken to be a function of the two dimensional projection $(\langle \varphi(u^i), \mathbf{g}_1\rangle, \langle \varphi(u^i), \mathbf{g}_2\rangle)$ for independent Gaussian vectors $\mathbf{g}_1, \mathbf{g}_2 \in \mathbb{R}^{m+n}$ (and similarly for x). Naor and Regev [NR14] show that such schemes are optimal in the sense that it is possible to achieve an approximation ratio arbitrarily close to the true (but unknown) value of K_G by using k-dimensional projections for a large (constant) k. A similar existential result was also proved by Raghavendra and Steurer [RS09] who proved that the there exists a (slightly different) rounding algorithm which can achieve the (unknown) approximation ratio K_G.

For the case of arbitrary $p \geq 2 \geq q$, Nesterov [Nes98] considered the convex program in Figure 2, denoted as $\mathsf{CP}(A)$, generalizing the one above. Note

$$\begin{array}{rl} \text{maximize} & \displaystyle\sum_{i,j} A_{i,j} \cdot \langle u^i, v^j \rangle = \left\langle A, UV^T \right\rangle \\ \text{subject to} & \displaystyle\sum_{i \in [m]} \|u^i\|_2^{q^*} \leq 1 \\ & \displaystyle\sum_{j \in [n]} \|v^j\|_2^{p} \leq 1 \\ & u^i, v^j \in \mathbb{R}^{m+n} \end{array}$$

u^i (resp. v^j) is the i-th (resp. j-th) row of U (resp. V)

Figure 2: The relaxation $\mathsf{CP}(A)$ for approximating $p{\to}q$ norm of a matrix $A \in \mathbb{R}^{m\times n}$.

that since $q^* \geq 2$ and $p \geq 2$, the above program is convex in the entries of the Gram matrix of the vectors $\{u^i\}_{i\in[m]} \cup \{v^j\}_{j\in[n]}$. Although the stated bound in [Nes98] is slightly weaker (as it is proved for a larger class of problems), the approximation ratio of the above relaxation can be shown to be bounded by K_G. By using the Krivine rounding scheme of considering the sign of a random Gaussian projection (aka random hyperplane rounding) one can show that Krivine's upper bound on K_G still applies to the above problem.

Motivated by applications to robust optimization, Steinberg [Ste05] considered the dual of (a variant of) the above relaxation, and obtained an upper bound of $\min\{\gamma_p/\gamma_q, \gamma_{q^*}/\gamma_{p^*}\}$ on the approximation factor. Note that while Steinberg's bound is better (approaches 1) as p and q approach 2, it is unbounded when $p, q^* \to \infty$ (as in the Grothendieck problem).

Based on the inapproximability result of factor $1/(\gamma_{p^*} \cdot \gamma_q)$ obtained in this work, it is natural to ask if this is the "right form" of the approximation ratio.

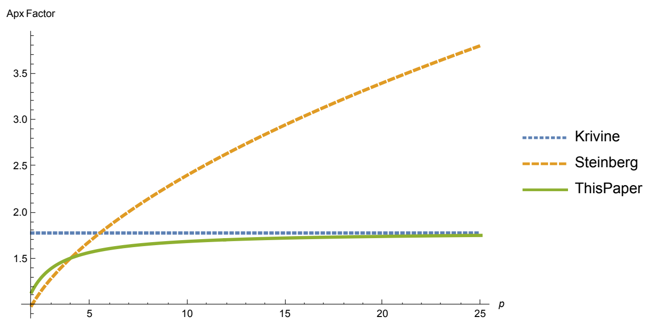

Figure 3: A comparison of the bounds for approximating $p \to p^*$ norm obtained from Krivine's rounding for K_G, Steinberg's analysis, and our bound. While our analysis yields an improved bound for $4 \le p \le 66$, we believe that the rounding algorithm achieves an improved bound for all p.

Indeed, this ratio is $\pi/2$ when $p^* = q = 1$, which is the ratio obtained by Krivine's rounding scheme, up to a factor of $\ln(1+\sqrt{2})$. We extend Krivine's result to all $p \ge 2 \ge q$ as below.

THEOREM 1.4. *There exists a fixed constant $\varepsilon_0 \le 0.00863$ such that for all $p \ge 2 \ge q$, the approximation ratio of the convex relaxation $\mathsf{CP}(A)$ is upper bounded by*

$$\frac{1+\varepsilon_0}{\sinh^{-1}(1)} \cdot \frac{1}{\gamma_{p^*} \cdot \gamma_q} \;=\; \frac{1+\varepsilon_0}{\ln(1+\sqrt{2})} \cdot \frac{1}{\gamma_{p^*} \cdot \gamma_q} .$$

Perhaps more interestingly, the above theorem is proved via a generalization of hyperplane rounding, which we believe may be of independent interest. Indeed, for a given collection of vectors $w^1, \ldots, w^m$ considered as rows of a matrix W, Gaussian hyperplane rounding corresponds to taking the "rounded" solution y to be the

$$y \;:=\; \underset{\|y'\|_\infty \le 1}{\operatorname{argmax}} \langle y', W\mathbf{g} \rangle \;=\; \left(\operatorname{sgn}\left(\langle w^i, \mathbf{g} \rangle\right)\right)_{i \in [m]} .$$

We consider the natural generalization to (say) ℓ_r norms, given by

$$y \;:=\; \underset{\|y'\|_r \le 1}{\operatorname{argmax}} \langle y', W\mathbf{g} \rangle \;=\; \left(\frac{\operatorname{sgn}\left(\langle w^i, \mathbf{g} \rangle\right) \cdot \left|\langle w^i, \mathbf{g} \rangle\right|^{r^*-1}}{\|W\mathbf{g}\|_{r^*}^{r^*-1}} \right)_{i \in [m]} .$$

We refer to y as the "Hölder dual" of $W\mathbf{g}$, since the above rounding can be obtained by viewing $W\mathbf{g}$ as lying in the dual (ℓ_{r^*}) ball, and finding the y for which Hölder's inequality is tight. Indeed, in the above language, Nesterov's rounding corresponds to considering the ℓ_∞ ball (hyperplane rounding). While Steinberg used a somewhat different relaxation, the rounding there can be obtained by viewing $W\mathbf{g}$ as lying in the primal (ℓ_r) ball instead of the dual one. In case of hyperplane rounding, the analysis is motivated by the identity that for two unit vectors u and v, we have

$$\underset{\mathbf{g}}{\mathbb{E}}\left[\operatorname{sgn}(\langle \mathbf{g}, u \rangle) \cdot \operatorname{sgn}(\langle \mathbf{g}, v \rangle)\right] \;=\; \frac{2}{\pi} \cdot \sin^{-1}(\langle u, v \rangle) .$$

We prove the appropriate extension of this identity to ℓ_r balls (and analyze the functions arising there) which may also be of interest for other optimization problems over ℓ_r balls.

Relation to Factorization Theory. Let X, Y be Banach spaces, and let $A : X \to Y$ be a continuous linear operator. As before, the norm $\|A\|_{X \to Y}$ is defined as

$$\|A\|_{X \to Y} \;:=\; \sup_{x \in X \setminus \{0\}} \frac{\|Ax\|_Y}{\|x\|_X} .$$

The operator A is said to be factorize through Hilbert space if the factorization constant of A defined as

$$\Phi(A) \;:=\; \inf_H \inf_{BC=A} \frac{\|C\|_{X \to H} \cdot \|B\|_{H \to Y}}{\|A\|_{X \to Y}}$$

is bounded, where the infimum is taken over all Hilbert spaces H and all operators $B : H \to Y$ and $C : X \to H$. The factorization gap for spaces X and Y is then defined as $\Phi(X, Y) := \sup_A \Phi(A)$ where the supremum runs over all continuous operators $A : X \to Y$.

The theory of factorization of linear operators is a cornerstone of modern functional analysis and has also found many applications outside the field (see [Pis86, AK06] for more information). An application to theoretical computer science was found by Tropp [Tro09] who used the Grothendieck factorization [Gro56] to give an algorithmic version of a celebrated column subset selection result of Bourgain and Tzafriri [BT87].

As an almost immediate consequence of convex programming duality, our new algorithmic results also imply some improved factorization results for ℓ_p^n, ℓ_q^m (a similar observation was already made by Tropp [Tro09] in the special case of ℓ_∞^n, ℓ_1^m and for a slightly different relaxation). We first state some classical factorization results, for which we will use $T_2(X)$ and $C_2(X)$ to respectively denote the Type-2 and Cotype-2 constants of X. We refer the interested reader to the full version [BGG+18b] for a more detailed description of factorization theory as well as the relevant functional analysis preliminaries.

The Kwapień-Maurey [Kwa72, Mau74] theorem states that for any pair of Banach spaces X and Y

$$\Phi(X,Y) \;\leq\; T_2(X)\cdot C_2(Y)\,.$$

However, Grothendieck's result [Gro56] shows that a much better bound is possible in a case where $T_2(X)$ is unbounded. In particular,

$$\Phi(\ell_\infty^n,\ell_1^m) \;\leq\; K_G\,,$$

for all $m,n \in \mathbb{N}$. Pisier [Pis80] showed that if X or Y satisfies the approximation property (which is always satisfied by finite-dimensional spaces), then

$$\Phi(X,Y) \;\leq\; \left(2\cdot C_2(X^*)\cdot C_2(Y)\right)^{3/2}\,.$$

We show that the approximation ratio of Nesterov's relaxation is in fact an upper bound on the factorization gap for the spaces ℓ_p^n and ℓ_q^m. Combined with our upper bound on the integrality gap, we show an improved bound on the factorization constant, i.e., for any $p \geq 2 \geq q$ and $m,n \in \mathbb{N}$, we have that for $X = \ell_p^n$, $Y = \ell_q^m$

$$\Phi(X,Y) \;\leq\; \frac{1+\varepsilon_0}{\sinh^{-1}(1)}\cdot\left(C_2(X^*)\cdot C_2(Y)\right)\,,$$

where $\varepsilon_0 \leq 0.00863$ as before. This improves on Pisier's bound for all $p \geq 2 \geq q$, and for certain ranges of (p,q) it also improves upon K_G and the bound of Kwapień-Maurey.

Approximability and Factorizability. Let (X_n) and (Y_m) be sequences of Banach spaces such that X_n is over the vector space $\mathbb{R}^n$ and Y_m is over the vector space $\mathbb{R}^m$. We shall say a pair of sequences $((X_n),(Y_m))$ factorize if $\Phi(X_n,Y_m)$ is bounded by a constant independent of m and n. Similarly, we shall say a pair of families $((X_n),(Y_m))$ are computationally approximable if there exists a polynomial $R(m,n)$, such that for every $m,n \in \mathbb{N}$, there is an algorithm with runtime $R(m,n)$ approximating $\|A\|_{X_n\to Y_m}$ within a constant independent of m and n (given an oracle for computing the norms of vectors and a separation oracle for the unit balls of the norms). We consider the natural question of characterizing the families of norms that are approximable and their connection to factorizability and Cotype.

The pairs (p,q) for which (ℓ_p^n,ℓ_q^m) is known (resp. not known) to factorize, are precisely those pairs (p,q) which are known to be computationally approximable (resp. inapproximable assuming hardness conjectures like $\mathsf{P} \neq \mathsf{NP}$ and ETH). Moreover the Hilbertian case which trivially satisfies factorizability, is also known to be computationally approximable (with approximation factor 1).

It is tempting to ask whether the set of computationally approximable pairs coincides with the set of factorizable pairs or the pairs for which X_n^*, Y_m have bounded (independent of m,n) Cotype-2 constant. Further yet, is there a connection between the approximation factor and the factorization constant, or approximation factor and Cotype-2 constants (of X_n^* and Y_m)? Our work gives some modest additional evidence towards such conjectures. Such a result would give credibility to the appealing intuitive idea of the approximation factor being dependent on the "distance" to a Hilbert space.

2 Full Version

For the proofs of all results discussed above, we refer the interested reader to the two-part full version: [BGG+18a] (Hardness) and [BGG+18b] (Algorithm).

3 Layout

The subsequent sections are laid out in order as follows: we give an overview of the hardness results (hypercontractive followed by non-hypercontractive), followed by an overview of the algorithmic results, followed by a detailed description of the generalized Krivine rounding procedure.

4 Hardness 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.

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]. We give a simple argument based on convexity, which proves this for all $p \leq q$, but appears to have gone unnoticed previously. The amplification is then used to prove hardness of approximation within almost polynomial factors.

Non-hypercontractive 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 discussed above. 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 $p > 2 > q$.

We also obtain a simple proof of the $2^{\Omega((\log n)^{1-\varepsilon})}$ hardness for the non-hypercontractive case when $2 \notin [q, p]$ (already proved by Bhaskara and Vijayaraghavan [BV11]) via an approximate isometry argument as used in the hypercontractive case. In the hypercontractive case, we started from a constant factor hardness of the $p\to 2$ norm and the same factor for $p\to q$ norm using the fact that for a random Gaussian matrix B of appropriate dimensions, we have $\|Bx\|_q \approx \|x\|_2$ for all x. We then amplify the hardness via tensoring. In the non-hypercontractive case, we start with a hardness for $p\to p$ norm (obtained via the above isometry), which we *first* amplify via tensoring. We then apply another approximate isometry result due to Schechtman [Sch87], which gives a samplable distribution $\mathcal{D}$ over random matrices B such that with high probability over B, we have $\|Bx\|_q \approx \|x\|_p$ for all x.

We thus view the above results as showing that combined with a basic hardness for $p\to 2$ norm, the basic ideas of duality, tensoring, and embedding (which builds on powerful results from functional analysis) can be combined in powerful ways to prove strong results in both the hypercontractive and non-hypercontractive regimes.

5 Algorithm Proof overview

As discussed above, we consider Nesterov's convex relaxation and generalize the hyperplane rounding scheme using "Hölder duals" of the Gaussian projections, instead of taking the sign. As in the Krivine rounding scheme, this rounding is applied to transformations of the SDP solutions. The nature of these transformations depends on how the rounding procedure changes the correlation between two vectors. Let $u, v \in \mathbb{R}^N$ be two unit vectors with $\langle u, v\rangle = \rho$. Then, for $\mathbf{g} \sim \mathcal{N}(0, I_N)$, $\langle \mathbf{g}, u\rangle$ and $\langle \mathbf{g}, v\rangle$ are ρ-correlated Gaussian random variables. Hyperplane rounding then gives ± 1 valued random variables whose correlation is given by

$$\mathop{\mathbb{E}}_{\mathbf{g}_1 \sim_\rho \mathbf{g}_2} [\mathrm{sgn}(\mathbf{g}_1) \cdot \mathrm{sgn}(\mathbf{g}_2)] \;=\; \frac{2}{\pi} \cdot \sin^{-1}(\rho)\,.$$

The transformations φ and ψ (to be applied to the vectors u and v) in Krivine's scheme are then chosen depending on the Taylor series for the sin function, which is the inverse of function computed on the correlation. For the case of Hölder-dual rounding, we prove the following generalization of the above identity

$$\begin{aligned}&\mathop{\mathbb{E}}_{\mathbf{g}_1 \sim_\rho \mathbf{g}_2} \left[\mathrm{sgn}(\mathbf{g}_1)\,|\mathbf{g}_1|^{q-1} \cdot \mathrm{sgn}(\mathbf{g}_2)\,|\mathbf{g}_2|^{p^*-1}\right] \\ &= \; \gamma_q^q \cdot \gamma_{p^*}^{p^*} \cdot \rho \cdot {}_2F_1\!\left(1-\frac{q}{2}, 1-\frac{p^*}{2}\,;\,\frac{3}{2}\,;\,\rho^2\right),\end{aligned}$$

where ${}_2F_1$ denotes a hypergeometric function with the specified parameters. The proof of the above identity combines simple tools from Hermite analysis with known integral representations from the theory of special functions, and may be useful in other applications of the rounding procedure.

Note that in the Grothendieck case, we have $\gamma_{p^*}^{p^*} = \gamma_q^q = \sqrt{2/\pi}$, and the remaining part is simply the $\sin^{-1}$ function. In the Krivine rounding scheme, the transformations φ and ψ are chosen to satisfy $(2/\pi) \cdot \sin^{-1}(\langle \varphi(u), \psi(v)\rangle) = c \cdot \langle u, v\rangle$, where the constant c then governs the approximation ratio. The transformations $\varphi(u)$ and $\psi(v)$ taken to be of the form $\varphi(u) = \oplus_{i=1}^{\infty} a_i \cdot u^{\otimes i}$ such that

$$\langle \varphi(u), \psi(v)\rangle = c' \sin(\langle u, v\rangle) \text{ and } \|\varphi(u)\|_2 = \|\psi(v)\| = 1$$

If f represents (a normalized version of) the function of ρ occurring in the identity above (which is $\sin^{-1}$ for hyperplane rounding), then the approximation ratio is governed by the function h obtained by replacing every Taylor coefficient of f^{-1} by its absolute value. While f^{-1} is simply the sin function (and thus h is the sinh function) in the Grothendieck problem, no closed-form expressions are available for general p and q.

The task of understanding the approximation ratio thus reduces to the analytic task of understanding the *family* of the functions h obtained for different values of p and q. Concretely, the approximation ratio is given by the value $1/(h^{-1}(1) \cdot \gamma_q\,\gamma_{p^*})$. At a high level, we prove bounds on $h^{-1}(1)$ by establishing properties of the Taylor coefficients of the family of functions f^{-1}, i.e., the family given by

$$\left\{f^{-1} \mid f(\rho) = \rho \cdot {}_2F_1\big(a_1, b_1\,;\,3/2\,;\,\rho^2\big)\,, a_1, b_1 \in [0, 1/2]\right\}$$

While in the cases considered earlier, the functions h are easy to determine in terms of f^{-1} via succinct formulae [Kri77, Haa81, AN04] or can be truncated after the cubic term [NR14], neither of these are true for the family of functions we consider. Hypergeometric functions are a rich and expressive class of functions, capturing many of the special functions appearing in Mathematical Physics and various ensembles of orthogonal polynomials. Due to this expressive power, the set of inverses is not well understood. In particular, while the coefficients of f are monotone in p and q, this is not true for f^{-1}. Moreover, the rates of decay of the coefficients may range from inverse polynomial to super-exponential. We analyze the coefficients of f^{-1} using complex-analytic methods inspired by (but quite different from) the work of Haagerup [Haa81] on bounding the complex Grothendieck constant. The key technical challenge in our work is in *arguing systematically about a family of inverse hypergeometric functions* which we address by developing methods to estimate the values of a family of contour integrals.

While our methods only gives a bound of the form $h^{-1}(1) \geq \sinh^{-1}(1)/(1+\varepsilon_0)$, we believe this is an artifact of the analysis and the true bound should indeed be $h^{-1}(1) \geq \sinh^{-1}(1)$.

6 Detailed Description of Algorithm

6.1 Notation For a non-negative real number r, we define the r-th Gaussian norm of a standard gaussian g as $\gamma_r := (\mathbb{E}_{g\sim\mathcal{N}(0,1)}[|g|^r])^{1/r}$.

Given a vector x, we define the r-norm as $\|x\|_r^r = \sum_i |x_i|^r$ for all $r \geq 1$. For any $r \geq 0$, we denote the dual norm by r^*, which satisfies the equality: $\frac{1}{r} + \frac{1}{r^*} = 1$.

For $p \geq 2 \geq q \geq 1$, we will use the following notation: $a := p^* - 1$ and $b := q - 1$. We note that $a, b \in [0, 1]$.

For a $m \times n$ matrix M (or vector, when $n = 1$). For an unitary function f, we define $f[M]$ to be the matrix M with entries defined as $(f[M])_{i,j} = f(M_{i,j})$ for $i \in [m], j \in [n]$. For vectors $u, v \in \mathbb{R}^\ell$, we denote by $u \circ v \in \mathbb{R}^\ell$ the entry-wise/Hadamard product of u and v. We denote the concatenation of two vectors u and v by $u \oplus v$. For a vector u, we use D_u to denote the diagonal matrix with the entries of u forming the diagonal, and for a matrix M we use $\mathrm{diag}(M)$ to denote the vector of diagonal entries.

For a function $f(\tau) = \sum_{k\geq 0} f_k \cdot \tau^k$ defined as a power series, we denote the function $\mathrm{abs}(f)(\tau) := \sum_{k\geq 0} |f_k| \cdot \tau^k$.

6.2 Krivine's Rounding Procedure

Krivine's procedure centers around the classical random hyperplane rounding. In this context, we define the random hyperplane rounding procedure on an input pair of matrices $U \in \mathbb{R}^{m\times\ell}$, $V \in \mathbb{R}^{n\times\ell}$ as outputting the vectors $\mathrm{sgn}[U\mathbf{g}]$ and $\mathrm{sgn}[V\mathbf{g}]$ where $\mathbf{g} \in \mathbb{R}^{\ell}$ is a vector with i.i.d. standard Gaussian coordinates ($f[v]$ denotes entry-wise application of a scalar function f to a vector v. We use the same convention for matrices.). The so-called Grothendieck identity states that for vectors $u, v \in \mathbb{R}^{\ell}$,

$$\mathbb{E}\left[\mathrm{sgn}\langle \mathbf{g}\,, u\rangle \cdot \mathrm{sgn}\langle \mathbf{g}\,, v\rangle\right] = \frac{\sin^{-1}\langle \widehat{u}\,, \widehat{v}\rangle}{\pi/2}$$

where $\widehat{u}$ denotes $u/\|u\|_2$. This implies the following equality which we will call the hyperplane rounding identity:

$$\mathbb{E}\left[\mathrm{sgn}[U\mathbf{g}](\mathrm{sgn}[V\mathbf{g}])^T\right] = \frac{\sin^{-1}[\widehat{U}\widehat{V}^T]}{\pi/2}\,. \tag{6.1}$$

where for a matrix U, we use $\widehat{U}$ to denote the matrix obtained by replacing the rows of U by the corresponding unit (in ℓ_2 norm) vectors. Krivine's main observation is that for any matrices U, V, there exist matrices $\varphi(\widehat{U}), \psi(\widehat{V})$ with unit vectors as rows, such that

$$\varphi(\widehat{U})\,\psi(\widehat{V})^T = \sin[(\pi/2)\cdot c\cdot \widehat{U}\widehat{V}^T]$$

where $c = \sinh^{-1}(1)\cdot 2/\pi$. Taking $\widehat{U}, \widehat{V}$ to be the optimal solution to $\mathsf{CP}(A)$, it follows that

$$\begin{aligned}\|A\|_{\infty\to 1} &\geq \left\langle A\,, \mathbb{E}\left[\mathrm{sgn}[\varphi(\widehat{U})\,\mathbf{g}]\ (\mathrm{sgn}[\psi(\widehat{V})\,\mathbf{g}])^T\right]\right\rangle \\ &= \langle A\,, c\cdot \widehat{U}\widehat{V}^T\rangle = c\cdot \mathsf{CP}(A)\,.\end{aligned}$$

The proof of Krivine's observation follows from simulating the Taylor series of a scalar function using inner products. We will now describe this more concretely.

Lemma 6.1. (Krivine) *Let $f : [-1,1] \to \mathbb{R}$ be a scalar function satisfying $f(\rho) = \sum_{k\geq 1} f_k\,\rho^k$ for an absolutely convergent series (f_k). Let $\mathrm{abs}\,(f)\,(\rho) := \sum_{k\geq 1} |f_k|\,\rho^k$ and further for vectors $u, v \in \mathbb{R}^{\ell}$ of ℓ_2-length at most 1, let*

$$S_L(f,u) := (\mathrm{sgn}(f_1)\sqrt{f_1}\cdot u) \oplus (\mathrm{sgn}(f_2)\sqrt{f_2}\cdot u^{\otimes 2}) \oplus \cdots$$
$$S_R(f,v) := (\sqrt{f_1}\cdot v) \oplus (\sqrt{f_2}\cdot v^{\otimes 2}) \oplus \cdots$$

Then for any $U \in \mathbb{R}^{m\times\ell}$, $V \in \mathbb{R}^{n\times\ell}$, $S_L(f, \sqrt{c_f}\cdot \widehat{U})$ and $S_R(f, \sqrt{c_f}\cdot\widehat{V})$ have ℓ_2-unit vectors as rows, and

$$S_L(f, \sqrt{c_f}\cdot \widehat{U})\ S_R(f, \sqrt{c_f}\cdot \widehat{V})^T = f\,[c_f\cdot \widehat{U}\widehat{V}^T]$$

where $S_L(f, W)$ for a matrix W, is applied to row-wise and $c_f := (\mathrm{abs}\,(f)^{-1})(1)$.

Proof. Using the facts $\langle y^1\otimes y^2\,, y^3\otimes y^4\rangle = \langle y^1\,, y^3\rangle\cdot\langle y^2\,, y^4\rangle$ and
$\langle y^1\oplus y^2\,, y^3\oplus y^4\rangle = \langle y^1\,, y^3\rangle + \langle y^2\,, y^4\rangle$, we have

- $\langle S_L(f,u)\,, S_R(f,v)\rangle = f(\langle u\,, v\rangle)$
- $\|S_L(f,u)\|_2 = \sqrt{\mathrm{abs}\,(f)\,(\|u\|_2^2)}$
- $\|S_R(f,v)\|_2 = \sqrt{\mathrm{abs}\,(f)\,(\|v\|_2^2)}$

The claim follows.

Before stating our full rounding procedure, we first discuss a natural generalization of random hyperplane rounding, and much like in Krivine's case this will guide the final procedure.

6.3 Generalizing Random Hyperplane Rounding – Hölder Dual Rounding

Fix any convex bodies $B_1 \subset \mathbb{R}^m$ and $B_2 \subset \mathbb{R}^k$. Suppose that we would like a strategy that for given vectors $y \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, outputs $\overline{y} \in B_1$, $\overline{x} \in B_2$ so that $y^T A\,x = \langle A\,, y\,x^T\rangle$ is close to $\langle A\,, \overline{y}\,\overline{x}^T\rangle$ for all A. A natural strategy is to take

$$\begin{aligned}(\overline{y}, \overline{x}) &:= \underset{(\widetilde{y},\widetilde{x})\in B_1\times B_2}{\mathrm{argmax}}\ \left\langle \widetilde{y}\,\widetilde{x}^T\,, y\,x^T\right\rangle \\ &= \left(\underset{\widetilde{y}\in B_1}{\mathrm{argmax}}\,\langle\widetilde{y}\,, y\rangle\ ,\ \underset{\widetilde{x}\in B_2}{\mathrm{argmax}}\,\langle\widetilde{x}\,, x\rangle\right)\end{aligned}$$

In the special case where B is the unit ℓ_p ball, there is a closed form for an optimal solution to $\max_{\widetilde{x}\in B}\langle\widetilde{x}\,, x\rangle$, given by $\Psi_{p^*}(x)/\|x\|_{p^*}^{p^*-1}$, where $\Psi_{p^*}(x) := \mathrm{sgn}[x]\circ|[x]|^{p^*-1}$. Note that for $p = \infty$, this strategy recovers the random hyperplane rounding procedure. We shall call this procedure, *Gaussian Hölder Dual Rounding* or *Hölder Dual Rounding* for short.

Just like earlier, we will first understand the effect of *Hölder Dual Rounding* on a solution pair U, V. For $\rho \in [-1,1]$, let $\mathbf{g}_1 \sim_\rho \mathbf{g}_2$ denote ρ-correlated standard Gaussians, i.e., $\mathbf{g}_1 = \rho\,\mathbf{g}_2 + \sqrt{1-\rho^2}\,\mathbf{g}_3$ where $(\mathbf{g}_2, \mathbf{g}_3) \sim \mathcal{N}(0, \mathrm{I}_2)$, and let

$$\widetilde{f}_{a,\,b}(\rho) := \underset{\mathbf{g}_1\sim_\rho\mathbf{g}_2}{\mathbb{E}}\left[\mathrm{sgn}(\mathbf{g}_1)|\mathbf{g}_1|^b\,\mathrm{sgn}(\mathbf{g}_2)|\mathbf{g}_1|^a\right]$$

We will work towards a better understanding of $\widetilde{f}_{a,\,b}(\cdot)$ in later sections. For now note that we have for vectors $u, v \in \mathbb{R}^{\ell}$,

$$\begin{aligned}&\mathbb{E}\left[\mathrm{sgn}\langle\mathbf{g}\,, u\rangle\,|\langle\mathbf{g}\,, u\rangle|^b\cdot\mathrm{sgn}\langle\mathbf{g}\,, v\rangle\,|\langle\mathbf{g}\,, v\rangle|^a\right] \\ &= \|u\|_2^b\cdot\|v\|_2^a\cdot\widetilde{f}_{a,\,b}(\langle\widehat{u}\,, \widehat{v}\rangle)\,.\end{aligned}$$

Thus given matrices U, V, we obtain the following generalization of the hyperplane rounding identity for *Hölder Dual Rounding* :

$$\begin{aligned} &\mathbb{E}\left[\Psi_q([U\mathbf{g}])\,\Psi_{p^*}([V\mathbf{g}])^T\right] \\ (6.2)\qquad &= D_{(\|u^i\|_2^b)_{i\in[m]}} \cdot \widetilde{f}_{a,b}([\widehat{U}\widehat{V}^T]) \cdot D_{(\|v^j\|_2^a)_{j\in[n]}}\,. \end{aligned}$$

6.4 Generalized Krivine Transformation and the Full Rounding Procedure

We are finally ready to state the generalized version of Krivine's algorithm. At a high level the algorithm simply applies *Hölder Dual Rounding* to a transformed version of the optimal convex program solution pair U, V. Analogous to Krivine's algorithm, the transformation is a type of "inverse" of Eq. 6.2.

(Inversion 1) Let (U, V) be the optimal solution to $\mathsf{CP}(A)$, and let $(u^i)_{i\in[m]}$ and $(v^j)_{j\in[n]}$ respectively denote the rows of U and V.

(Inversion 2) Let $c_{a,b} := \left(\mathrm{abs}\left(\widetilde{f}_{a,b}^{-1}\right)\right)^{-1}(1)$ and let

$$\varphi(U) := D_{(\|u^i\|_2^{1/b})_{i\in[m]}}\; S_L(\widetilde{f}_{a,b}^{-1}, \sqrt{c_{a,b}}\cdot\widehat{U})\,,$$
$$\psi(V) := D_{(\|v^j\|_2^{1/a})_{j\in[n]}}\; S_R(\widetilde{f}_{a,b}^{-1}, \sqrt{c_{a,b}}\cdot\widehat{V})\,.$$

(Hölder-Dual 1) Let $\mathbf{g} \sim \mathcal{N}(0, \mathrm{I})$ be an infinite dimensional i.i.d. Gaussian vector.

(Hölder-Dual 2) Return $y := \Psi_q(\varphi(U)\,\mathbf{g})/\|\varphi(U)\,\mathbf{g}\|_q^b$ and $x := \Psi_{p^*}(\psi(V)\,\mathbf{g})/\|\psi(V)\,\mathbf{g}\|_{p^*}^a$.

Remark 1. Note that $\|\Psi_r(\overline{x})\|_{r^*} = \|\overline{x}\|_r^{r-1}$ and so the returned solution pair always lie on the unit ℓ_{q^*} and ℓ_p spheres respectively.

Remark 2. Like in [AN04] the procedure above can be made algorithmic by observing that there always exist $\varphi'(U) \in \mathbb{R}^{m\times(m+n)}$ and $\psi'(V) \in \mathbb{R}^{m\times(m+n)}$, whose rows have the exact same lengths and pairwise inner products as those of $\varphi(U)$ and $\psi(V)$ above. Moreover they can be computed without explicitly computing $\varphi(U)$ and $\psi(V)$ by obtaining the Gram decomposition of

$$M := \left[\begin{array}{cc} \mathrm{abs}\left(\widetilde{f}_{a,b}^{-1}\right)[c_{a,b}\cdot\widehat{V}\widehat{V}^T] & \widetilde{f}_{a,b}^{-1}([c_{a,b}\cdot\widehat{U}\widehat{V}^T]) \\ \widetilde{f}_{a,b}^{-1}([c_{a,b}\cdot\widehat{V}\widehat{U}^T]) & \mathrm{abs}\left(\widetilde{f}_{a,b}^{-1}\right)[c_{a,b}\cdot\widehat{V}\widehat{V}^T] \end{array}\right],$$

and normalizing the rows of the decomposition according to the definition of $\varphi(\cdot)$ and $\psi(\cdot)$ above. The entries of M can be computed in polynomial time with exponentially (in m and n) good accuracy by implementing the Taylor series of $\widetilde{f}_{a,b}^{-1}$ upto $\mathrm{poly}(m, n)$ terms (Taylor series inversion can be done upto k terms in time $\mathrm{poly}(k)$).

Remark 3. Note that the 2-norm of the i-th row (resp. j-th row) of $\varphi(U)$ (resp. $\psi(V)$) is $\|u^i\|_2^{1/b}$ (resp. $\|v^j\|_2^{1/a}$).

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. 3

[AK06] Fernando Albiac and Nigel J Kalton. *Topics in Banach space theory, volume 233 of Graduate Texts in Mathematics*. Springer, New York, 2006. 5

[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. 3

[AN04] Noga Alon and Assaf Naor. Approximating the cut-norm via Grothendieck's inequality. In *Proceedings of the 36th annual ACM symposium on Theory of computing*, pages 72–80. ACM, 2004. 8, 10

[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

[BGG+18a] V. Bhattiprolu, M. Ghosh, V. Guruswami, E. Lee, and M. Tulsiani. Inapproximability of Matrix $p \to q$ Norms. *ArXiv e-prints*, February 2018. 6

[BGG+18b] Vijay Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Approximating operator norms via generalized krivine rounding. *CoRR*, abs/1804.03644, 2018. 5, 6

[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

[BMMN13] Mark Braverman, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. The Grothendieck constant is strictly smaller than Krivine's bound. In *Forum of Mathematics, Pi*, volume 1. Cambridge University Press, 2013. Conference version in FOCS '11. 4

[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. 3, 6, 7

[BT87] Jean Bourgain and Lior Tzafriri. Invertibility of large submatrices with applications to the geometry of banach spaces and harmonic analysis. *Israel journal of mathematics*, 57(2):137–224, 1987. 5

[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. 3, 4, 7

[Gro56] Alexandre Grothendieck. *Résumé de la théorie métrique des produits tensoriels topologiques*. Soc. de Matemática de São Paulo, 1956. 3, 5, 6

[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. 3, 6, 7

[Haa81] Uffe Haagerup. The best constants in the khintchine inequality. *Studia Mathematica*, 70(3):231–283, 1981. 8

[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, 7

[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

[KMW18] Aditya Krishnan, Sidhanth Mohanty, and David P. Woodruff. On sketching q to p norms. Manuscript, 2018. 7

[KN09] Subhash Khot and Assaf Naor. Approximate kernel clustering. *Mathematika*, 55(1-2):129–165, 2009. 3

[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. 3

[KO09] Subhash Khot and Ryan O'Donnell. SDP gaps and UGC-hardness for Max-Cut-Gain. *Theory OF Computing*, 5:83–117, 2009. 3

[Kri77] Jean-Louis Krivine. Sur la constante de Grothendieck. *CR Acad. Sci. Paris Ser. AB*, 284(8):A445–A446, 1977. 3, 4, 8

[Kwa72] Stanislaw Kwapień. Isomorphic characterizations of inner product spaces by orthogonal series with vector valued coefficients. *Stud. Math.*, 44:583–595, 1972. 6

[Mau74] Bernard Maurey. Théorèmes de factorisation pour les opérateurs à valeurs dans un espace l_p. *Séminaire Analyse fonctionnelle (dit)*, pages 1–5, 1974. 6

[Nes98] Yurii Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization methods and software*, 9(1-3):141–160, 1998. 3, 4

[NR14] Assaf Naor and Oded Regev. Krivine schemes are optimal. *Proceedings of the American Mathematical Society*, 142(12):4315–4320, 2014. 4, 8

[Pis80] Gilles Pisier. Un theoreme sur les operateurs lineaires entre espaces de banach qui se factorisent par un espace de hilbert. In *Annales scientifiques de lEcole Normale Superieure*, volume 13, pages 23–43. Elsevier, 1980. 6

[Pis86] Gilles Pisier. *Factorization of linear operators and geometry of Banach spaces*. Number 60. American Mathematical Soc., 1986. 5

[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. 3

[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. 3, 4

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd annual ACM Symposium on Theory of Computing*, pages 755–764. ACM, 2010. 2

[Sch87] Gideon Schechtman. More on embedding subspaces of\(l_p\) in\(l^ n_r\). *Compositio Math*, 61(2):159–169, 1987. 7

[Ste05] Daureen Steinberg. Computation of matrix norms with applications to robust optimization. *Research thesis, Technion-Israel University of Technology*, 2005. 2, 3, 4

[Tro09] Joel A Tropp. Column subset selection, matrix factorization, and eigenvalue optimization. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 978–986. Society for Industrial and Applied Mathematics, 2009. 5

[Ver17] Roman Vershynin. *High Dimensional Probability*. 2017. 7