

ESTIMATING THE MATRIX $p \rightarrow q$ NORM

LARRY GUTH, DOMINIQUE MALDAGUE, AND JOHN URSCHEL

ABSTRACT. The matrix $p \rightarrow q$ norm is a fundamental quantity appearing in a variety of areas of mathematics. This quantity is known to be efficiently computable in only a few special cases. The best known algorithms for approximately computing this quantity with theoretical guarantees essentially consist of computing the $p \rightarrow q$ norm for p, q where this quantity can be computed exactly or up to a constant, and applying interpolation. We analyze the matrix $2 \rightarrow q$ norm problem and provide an improved approximation algorithm via a simple argument involving the rows of a given matrix. For example, we improve the best-known $2 \rightarrow 4$ norm approximation from $m^{1/8}$ to $m^{1/12}$. This insight for the $2 \rightarrow q$ norm improves the best known $p \rightarrow q$ approximation algorithm for the region $p \leq 2 \leq q$, and leads to an overall improvement in the best-known approximation for $p \rightarrow q$ norms from $m^{25/128}$ to $m^{3-2\sqrt{2}}$.

1. INTRODUCTION

Given a matrix $A \in \mathbb{C}^{m \times n}$, the $p \rightarrow q$ norm of A , defined by

$$\|A\|_{p \rightarrow q} := \max_{x \in \mathbb{C}^n} \frac{\|Ax\|_q}{\|x\|_p}, \quad p, q \in [1, \infty),$$

where $\|\cdot\|_p$ is the ℓ_p -norm of a vector, is a measure of how large in magnitude a vector of a given p -norm can become in q -norm from the application of A . The matrix $p \rightarrow q$ norm is a fundamental quantity, appearing in a variety of areas of mathematics and computer science. The spectral norm ($p = q = 2$) is perhaps the most used and well-known instance of a $p \rightarrow q$ norm, and is equal to the largest singular value of a matrix. More generally, matrix $p \rightarrow p$ norms (simply called p -norms) appear in relative error estimates for linear systems $Ax = b$, as the $p \rightarrow p$ condition number of a matrix governs the worst-case relative error for x in p -norm due to small perturbations to A and b ; see both [13, Chapter 2.6] and [18, Chapter 14] for details. The special case of $p = \infty$ and $q = 1$ is the well-studied Grothendieck problem [14, 23]: maximize $\langle y, Ax \rangle$ subject to $\|x\|_\infty, \|y\|_\infty = 1$, of which the max-cut problem in graph theory is a special case [12]. The case of $p = 2$ and $q = 4$ is also of special significance. For instance, distinguishing between entangled and separable states in quantum information theory is known to be related to the computation of the $2 \rightarrow 4$ norm [15]. The $2 \rightarrow q$ norm, for any even $q \geq 4$, is directly related to small-set expansion in graph theory and the unique games conjecture in computational complexity [2]. More generally, norms in the hyper-contractive region of $p < q$ are closely related to mixing of Markov chains and a variety of problems in theoretical computer science, see [25, 6] for details and further examples. The most general setting of arbitrary p and q is also broadly applicable to robust optimization, see [26, Section 1.1.2] and the examples and references contained therein.

2020 *Mathematics Subject Classification.* 15A60, 65F35, 68W25.

Key words and phrases. approximation algorithm, matrix norm.

One of our initial questions about $\|A\|_{p \rightarrow q}$ was whether techniques proving L^p to L^q boundedness from analysis could be applied to improve approximation algorithms for $\|A\|_{p \rightarrow q}$. Consider, for example, the Fourier transform $\widehat{f}(\xi) = \int_{\mathbb{R}^n} e^{2\pi i x \cdot \xi} f(x) dx$ of a function $f : \mathbb{R}^n \rightarrow \mathbb{C}$. By the classical Hausdorff-Young inequality, $C_{p,q,n} := \sup_{\|f\|_{L^p}=1} \|\widehat{f}\|_{L^q}$ is finite if and only if $1 \leq p \leq 2$ and $\frac{1}{p} + \frac{1}{q} = 1$. The optimal constant $C_{p,q,n}$ is achieved by Gaussian functions, which was proved by Babenko [1] for even q and then Beckner [3] for all $2 \leq q$. In [9], Christ sharpened the Hausdorff-Young inequality by proving that

$$\|\widehat{f}\|_{L^q} \leq [C_{p,q,n} - \text{dist}_p(f, \mathcal{G})] \|f\|_{L^p}$$

for an appropriate distance function $\text{dist}_p(f, \mathcal{G})$ of f to the set of Gaussians. One basic idea we use in our approximation of $\|A\|_{p \rightarrow q}$ involves the following fundamental maximization principle that Christ used in [9]. Suppose that $\|f\|_{L^p} = 1$ and that for a small parameter $\delta > 0$, $C_{p,q,n}(1 - \delta) \leq \|\widehat{f}\|_{L^q}$. If for another small parameter $\eta > 0$, $f = g + h$ is a disjointly supported decomposition of f with $\|h\|_{L^p} \leq \eta$, then

$$C_{p,q,n}(1 - \delta) \leq \|\widehat{f}\|_{L^q} \leq \|\widehat{g}\|_{L^q} + C_{p,q,n}\|h\|_{L^p} \leq \|\widehat{g}\|_{L^q} + C_{p,q,n}\eta,$$

so $C_{p,q,n}(1 - \delta - \eta)\|g\|_{L^p} \leq \|\widehat{g}\|_{L^q}$. The basic principle is that if f has close to maximal ratio $\|\widehat{f}\|_{L^q}/\|f\|_{L^p}$ and g is a significant portion of the L^p mass of f , then g also has close to maximal ratio $\|\widehat{g}\|_{L^q}/\|g\|_{L^p}$. We use this idea in the proof of the key lemma (Lemma 2.1) below.

Computing the matrix $p \rightarrow q$ norm consists of maximizing a convex function over a convex domain, and, unfortunately, is difficult in general. The matrix $p \rightarrow q$ norm is only efficiently computable (up to arbitrary error) in three special cases: $p = q = 2$, $p = 1$, and $q = \infty$. As previously mentioned, the spectral norm $\|A\|_{2 \rightarrow 2}$ is equal to the largest singular value $\sigma_{\max}(A)$ of A and is achieved by any right singular vector v corresponding to $\sigma_{\max}(A)$. Let r_j , $j = 1, \dots, m$, denote the conjugate transpose of the rows of A . When $q = \infty$, $\|A\|_{p \rightarrow \infty} = \max_{j \in \{1, \dots, m\}} \|r_j\|_{p^*}$, where $1/p + 1/p^* = 1$, and this value is achieved by the vector $v = (|r_j(1)|^{p^*-2} r_j(1), \dots, |r_j(n)|^{p^*-2} r_j(n))$, where $r_j = (r_j(1), \dots, r_j(n))$ is a row vector with maximal p^* norm. In fact, the case of $p = 1$ is mathematically equivalent to $q = \infty$, as

$$\|A\|_{p \rightarrow q} = \max_{\substack{\|x\|_p=1, \\ \|y\|_{q^*}=1}} \langle y, Ax \rangle = \|A^*\|_{q^* \rightarrow p^*}, \quad \text{where } \frac{1}{p} + \frac{1}{p^*} = \frac{1}{q} + \frac{1}{q^*} = 1.$$

More generally, by duality, it suffices to only consider $\|A\|_{p \rightarrow q}$ for $1/p + 1/q \leq 1$. For this reason, our results focus on the regimes $p, q \geq 2$ and $1 \leq p \leq 2 \leq q$, as $1 \leq p, q \leq 2$ is mathematically equivalent to $p, q \geq 2$. In the region $1 \leq q \leq 2 \leq p$, the matrix $p \rightarrow q$ norm can be approximated up to a constant factor of at most $\pi/2$, this is often called Nesterov's $\pi/2$ Theorem, and the region $1 \leq q \leq 2 \leq p$ is often called the Nesterov region [22]. When the underlying matrix is non-negative, the matrix $p \rightarrow q$ norm can be computed exactly when $p \geq q \geq 1$ (see [26, 4]), a setting with application to oblivious routing in graph theory [11].

1.1. Hardness Results. Outside of these special cases, the majority of results for matrix norms consists of algorithmic lower bounds. In [24], Rohn showed that computing the $\infty \rightarrow 1$ norm is NP-hard, a result later expanded upon by Steinberg, who showed

that computing the $p \rightarrow q$ norm for $p > q$ is NP-hard [26]. Later, Hendrickx and Olshesky proved that computing the $p \rightarrow p$ norm for any rational $p \neq 1, 2$ is NP-hard, and that computing the $\infty \rightarrow q$ norm to an arbitrary constant for any rational q is NP-hard [16]. Bhaskara and Vijayaraghavan proved that it is NP-hard to compute the $p \rightarrow q$ norm up to any constant factor for $2 < q \leq p$, and, under a reasonable assumption from computational complexity, show that this norm cannot be approximated within a factor of $2^{\log^{1-\epsilon}(n)}$ for any constant $\epsilon > 0$ [4]. Barak et al. treated the $2 \rightarrow q$ norm problem, and showed that the $2 \rightarrow 4$ norm is NP-hard to approximate up to inverse-polynomial (in dimension) precision, and, assuming the exponential time hypothesis, better than a factor of $2^{O(\log^{1/2-\epsilon} n)}$ [2]. Finally, Bhattiprolu et al. showed, under a reasonable assumption from computational complexity, that it is NP-hard to approximate the $p \rightarrow q$ norm for $2 < p < q$ within a factor of $2^{O(\log^{1-\epsilon} n)}$ [5].

1.2. Existing Algorithms. In terms of methods, in 1974 Boyd proposed a power method type heuristic for computing the matrix $p \rightarrow q$ norm, with theoretical analysis for the case of non-negative matrices, but without rigorous guarantees for arbitrary matrices [7]. Higham later produced a modified power method using a well-chosen initial guess that performed well in practice for the $p \rightarrow p$ norm and had the theoretical approximation guarantee of $n^{1-1/p}$ for $p \geq 2$ [17]. Steinberg, by approximating norms in the Nesterov region $1 \leq q \leq 2 \leq p$ via semidefinite programming and using interpolation, produced a $(\pi/2)^{\frac{1}{q}} m^{\frac{q-2}{q^2}}$ approximation algorithm for $p, q \geq 2$ and a $m^{\frac{2(q-2)(p-1)}{pq^2}} n^{\frac{(p-1)(2-p)}{p^2}}$ algorithm for $p \leq 2 \leq q$, leading to a worst-case approximation of $O(\max\{m, n\}^{25/128})$ for an arbitrary p and q . For the $p \rightarrow p$ norm, Steinberg's algorithm has a stronger theoretical guarantee than that of Higham (e.g., $m^{1/8}$ vs $m^{1/4}$ for $\|A\|_{4 \rightarrow 4}$), but this theoretical gap is offset by the practical difference in efficiency of their two algorithms, as Steinberg's estimate relies on the solution of a semidefinite program. See [19] for a summary of the existing literature and best-known complexity for cutting plane and interior point techniques for approximately solving a semidefinite program. As previously mentioned, the $2 \rightarrow q$ norm is of particular interest in theoretical computer science. One interesting area of research for $2 \rightarrow q$ norms is in producing algorithms with bounded additive error. Barak et al. produced a $2^{O(\log^2(n)\epsilon^{-2})}$ time algorithm that computes $\|A\|_{2 \rightarrow 4}^4$ up to additive error $\epsilon \|A\|_{2 \rightarrow 2}^2 \|A\|_{2 \rightarrow \infty}^2$ [2], a result later improved and extended to arbitrary q by Brandão and Harrow [8].

1.3. Our Results. In this work, we improve upon existing algorithms for computing $\|A\|_{p \rightarrow q}$. We do so by producing superior estimates for the matrix $2 \rightarrow q$ norm (Lemma 2.1) and extending these results using interpolation. Our simplest result for estimating $\|A\|_{2 \rightarrow q}$ is recorded in the following theorem.

Theorem 1.1. *Let $A \in \mathbb{C}^{m \times n}$, $r_1, \dots, r_m \in \mathbb{C}^n$ denote the conjugate transpose of the m rows of A , $v \in \mathbb{C}^n$ denote a right singular vector corresponding to the 2-norm of A , and $S = \{v, r_1, \dots, r_m\}$. Then*

$$\max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2} \geq (2m^{\frac{1}{q}})^{-\frac{q-2}{2(q-1)}} \|A\|_{2 \rightarrow q} \quad \text{for all } q \geq 2.$$

In §2.1, we show that Theorem 1.1 is the best approximation possible with the given data. We also give examples to demonstrate the significantly improved bounds from

Theorem 1.1 compared to Steinberg's estimate using interpolation between the $2 \rightarrow 2$ and $2 \rightarrow \infty$ norms [26]. See Figure 1 for a comparison of the approximation exponents for Steinberg's interpolation approach and for Theorem 1.2. As an example, interpolation gives an $\sqrt[3]{2} m^{1/12}$ approximation for the popular $2 \rightarrow 4$ norm, while interpolation provides only a $m^{1/8}$ approximation.

By also incorporating the $1 \rightarrow q$ norm and interpolation, we use Theorem 1.1 to improve $p \rightarrow q$ norm estimates for a range of p and q exponents, which leads to the following generalization of Theorem 1.1.

Theorem 1.2. *Let $A \in \mathbb{C}^{m \times n}$, $r_1, \dots, r_m \in \mathbb{C}^n$ denote the conjugate transpose of the m rows of A , $v \in \mathbb{C}^n$ denote a right singular vector corresponding to the 2-norm of A , and $y \in \mathbb{C}^n$ achieve the $1 \rightarrow q$ norm of A , and $S = \{v, r_1, \dots, r_m\}$. Then*

$$\max_{\substack{x \in S \cup \{y\} \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_p} \geq \left[(2m^{\frac{1}{q}})^{\frac{q-2}{q-1}} n^{\frac{2-p}{p}} \right]^{-\frac{p-1}{p}} \|A\|_{p \rightarrow q}$$

for all $\{(p, q) \mid p \in [1, \sqrt{2}], q \in [\frac{2}{2-p}, \frac{p}{p-1}]\}$ and $\{(p, q) \mid p \in [\sqrt{2}, 2], q \in [\frac{p}{p-1}, \frac{2}{2-p}]\}$, and

$$\max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_p} \geq \left[(2m^{\frac{1}{q}})^{\frac{q-2}{q-1}} n^{\frac{p-2}{p}} \right]^{-\frac{1}{2}} \|A\|_{p \rightarrow q}$$

for all $p, q \in [2, \infty)$.

We note that improved estimates for the regions not directly addressed in Theorem 1.2 follow from duality, as $p, q \in [1, 2]$, $\{(p, q) \mid q \in [\frac{\sqrt{2}}{\sqrt{2}-1}, \infty), p \in [\frac{q}{q-1}, \frac{2(q-1)}{q}]\}$, and $\{(p, q) \mid q \in [2, \frac{\sqrt{2}}{\sqrt{2}-1}], p \in [\frac{2(q-1)}{q}, \frac{q}{q-1}]\}$ are dual to $p, q \in [2, \infty)$, $\{(p, q) \mid p \in [1, \sqrt{2}], q \in [\frac{2}{2-p}, \frac{p}{p-1}]\}$, and $\{(p, q) \mid p \in [\sqrt{2}, 2], q \in [\frac{p}{p-1}, \frac{2}{2-p}]\}$, respectively. For the region $p, q \in [2, \infty)$ (and, again, by duality, $p, q \in [1, 2]$), the worst-case estimates for our simple estimator outperforms semidefinite programming-based interpolation estimates for square matrices when $p < 2q^2(q-1)/(q^3 - 2q^2 + 4q - 4)$. We give a visual representation of these details, as well as a contour plot of the best approximation exponent with respect to p and q in Figure 1. Theorem 1.2 is a simplified version of our main result that makes use of the worst-case Lipschitz bounds for norms. We prove a tighter version of this result in Theorem 3.1, which has the same worst-case behavior, but improved bounds for the majority of matrices A . Overall, Theorem 1.2 implies a global $O(\max\{m, n\}^{3-2\sqrt{2}})$ approximation algorithm, as

$$\max \left\{ \frac{(p-1)(2-p)}{p^2} + \frac{(p-1)(q-2)}{pq(q-1)} \mid 1 \leq p \leq 2 \leq q, \frac{1}{p} + \frac{1}{q} \leq 1 \right\} = 3 - 2\sqrt{2},$$

achieved by $p = \sqrt{2}$ and $q = \frac{\sqrt{2}}{\sqrt{2}-1}$. We note that the approximations provided by Theorem 1.2 are not simply asymptotic theoretical bounds, but concrete estimates that can be used in practice. In Table 1, we provide the worst-case approximation ratio for the $2 \rightarrow 4$ norm and any $p \rightarrow q$ norm for matrices of dimension 10^3 to 10^9 . Theorem 1.2 gives an approximation for any million by million matrix in any $p \rightarrow q$ norm within a factor of 11.081, and within a factor of 3.985 for the $2 \rightarrow 4$ norm. This illustrates that, while for sufficiently large matrices there is no hope for a polylogarithmic approximation to the $p \rightarrow q$ norm, this asymptotic algorithmic lower bound does not apply to the sizes of matrices typically seen in practice.

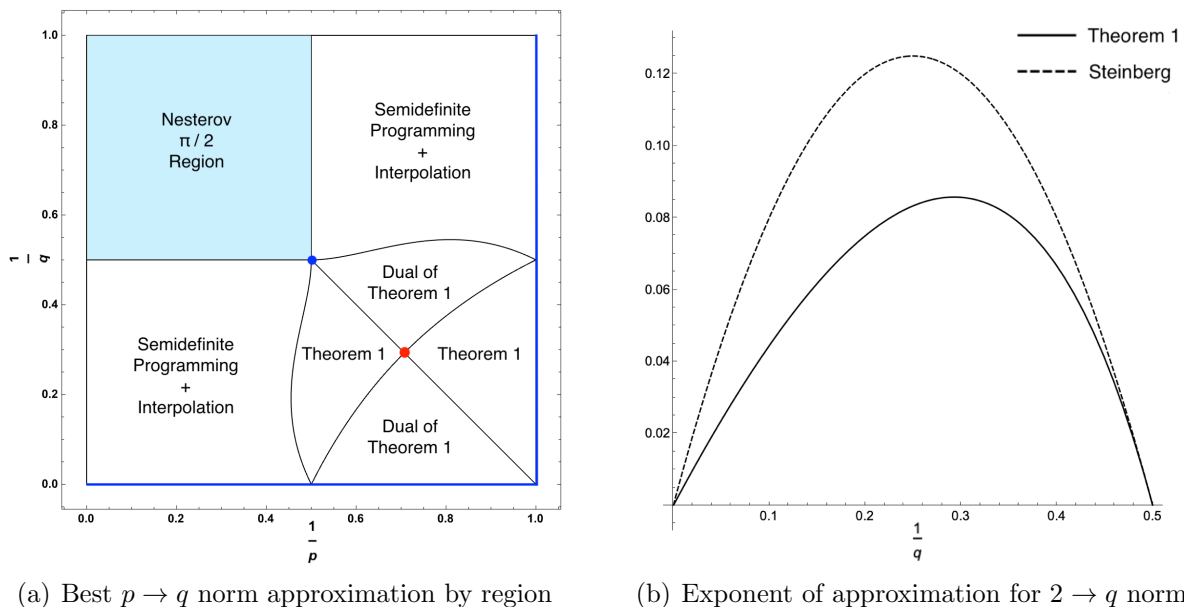


FIGURE 1. Figure (a) is a plot of the square $(1/p, 1/q) \in [0, 1]^2$. Efficiently computable p, q are in dark blue and the Nesterov region computable to a constant factor is in light blue. We note the regions of p, q for which Theorem 1.2 improves upon the best-known approximation. The worst approximation for any p, q is given by $p = \sqrt{2}, q = \sqrt{2}/(\sqrt{2} - 1)$, marked in red. Figure (b) plots the worst-case approximation exponent (e.g., the value α in a Cm^α approximation) for Steinberg’s interpolation approach and Theorem 1.2. For the popular $2 \rightarrow 4$ norm, interpolation achieves an exponent of $1/8$, while Theorem 1.2 achieves an exponent of $1/12$, a 33.3% decrease in the exponent.

$\max\{m, n\} =$	10^3	10^4	10^5	10^6	10^7	10^8	10^9
$2 \rightarrow 4$ norm	2.241	2.715	3.289	3.985	4.827	5.849	7.086
$p \rightarrow q$ norm	3.388	5.029	7.465	11.081	16.449	24.419	36.248

TABLE 1. The worst-case approximation ratio for the $2 \rightarrow 4$ norm and the $p \rightarrow q$ norm (achieved by $p = \sqrt{2}, q = \sqrt{2}/(\sqrt{2} - 1)$) for matrices of varying sizes of m and n . Despite the non-existence of an efficient polylogarithmic approximation algorithm for the matrix $p \rightarrow q$ norm problem, the exponents in the polynomial factors of Theorem 1.2 grow quite slowly. Indeed, the $p \rightarrow q$ norm of a million by million sized matrix can be quickly approximated within a factor of ≈ 11 , while $\ln(10^6) \approx 13.8$. In the case of the $2 \rightarrow 4$ norm, a billion by billion sized matrix can be approximated up to a factor of ≈ 7 , while $\ln(10^9) \approx 20.7$.

The remainder of the paper is as follows. In Section 2, we study the matrix $2 \rightarrow q$ norm problem. We produce improved results by moving beyond standard interpolation bounds and techniques and exploiting the information contained in each row of a matrix.

We illustrate the tightness of our bounds using a variety of known matrices. In Section 3, we extend our results for $2 \rightarrow q$ norms to arbitrary $p \rightarrow q$ norms, producing both a proof of Theorem 1.2 and inspiring a practical algorithm for estimating $p \rightarrow q$ norms. Finally, we perform numerical experiments illustrating the practical efficiency of the theoretical results of this work.

Acknowledgements LG is supported by a Simons Investigator award. DM is supported by the National Science Foundation under Award No. 2103249.

2. ESTIMATING THE $2 \rightarrow q$ NORM

The approximation algorithm of Steinberg approximates the $2 \rightarrow q$ norm, $q \geq 2$, of a matrix by computing the $2 \rightarrow 2$ and $2 \rightarrow \infty$ norms and using interpolation. This estimate depends on two vectors, a right singular vector corresponding to the largest singular value and the conjugate transpose of a row with maximal 2-norm. In this section, we prove that making use of all rows of the matrix leads to an improved approximation algorithm. We present the following lemma.

Lemma 2.1. *Let $A \in \mathbb{C}^{m \times n}$, $r_1, \dots, r_m \in \mathbb{C}^n$ denote the conjugate transpose of the m rows of A , $v \in \mathbb{C}^n$ denote a right singular vector corresponding to the 2-norm of A , and $S = \{v, r_1, \dots, r_m\}$. Then*

$$\max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2} \geq \left[\frac{1}{2\|A\|_{2 \rightarrow \infty}} \max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2} \right]^{\frac{q-2}{2(q-1)}} \left[\frac{\|Av\|_q}{\|Av\|_2} \right]^{\frac{1}{q-1}} \|A\|_{2 \rightarrow q} \quad \text{for all } q \geq 2.$$

Proof. Let $w \in \mathbb{C}^n$, $\|w\|_2 = 1$, satisfy $\|Aw\|_q = \|A\|_{2 \rightarrow q}$, and consider the orthogonal projection of w onto r_1, \dots, r_m . If, for some $\lambda > 0$, $|\langle w, r_i \rangle| \leq \lambda \|A\|_{2 \rightarrow \infty}$ for all $i = 1, \dots, m$, then

$$\|Aw\|_q^q = \sum_{i=1}^m |\langle w, r_i \rangle|^q \leq [\lambda \|A\|_{2 \rightarrow \infty}]^{q-2} \sum_{i=1}^m |\langle w, r_i \rangle|^2 \leq [\lambda \|A\|_{2 \rightarrow \infty}]^{q-2} \|A\|_{2 \rightarrow 2}^2,$$

$$\begin{aligned} \|A\|_{2 \rightarrow q} &\leq \lambda^{\frac{q-2}{q}} \|A\|_{2 \rightarrow \infty}^{\frac{q-2}{q}} \|A\|_{2 \rightarrow 2}^{\frac{2}{q}} \\ &= \lambda^{\frac{q-2}{q}} \left[\min_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2} \frac{\|x\|_2 \|A\|_{2 \rightarrow \infty}}{\|Ax\|_q} \right]^{\frac{q-2}{q}} \left[\frac{\|Av\|_q}{\|v\|_2} \frac{\|Av\|_2}{\|Av\|_q} \right]^{\frac{2}{q}} \\ &\leq \lambda^{\frac{q-2}{q}} \left[\frac{\|Av\|_2}{\|Av\|_q} \right]^{\frac{2}{q}} \left[\min_{\substack{x \in S \\ x \neq 0}} \frac{\|x\|_2 \|A\|_{2 \rightarrow \infty}}{\|Ax\|_q} \right]^{\frac{q-2}{q}} \max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2}. \end{aligned}$$

If there exists some r_k such that $|\langle w, r_k \rangle| > \lambda \|A\|_{2 \rightarrow \infty}$, then we can upper bound $\|A\|_{2 \rightarrow q}$ by considering the orthogonal projection of w onto r_k . We have

$$\left\| w - \frac{\langle w, r_k \rangle}{\|r_k\|_2^2} r_k \right\|_2^2 = \|w\|_2^2 - \frac{|\langle w, r_k \rangle|^2}{\|r_k\|_2^2} = 1 - \frac{|\langle w, r_k \rangle|^2}{\|r_k\|_2^2}$$

and

$$\begin{aligned}
\|Aw\|_q &\leq \left\| A \left[\frac{\langle w, r_k \rangle}{\|r_k\|_2^2} r_k \right] \right\|_q + \left\| A \left[w - \frac{\langle w, r_k \rangle}{\|r_k\|_2^2} r_k \right] \right\|_q \\
&\leq \frac{|\langle w, r_k \rangle|}{\|r_k\|_2^2} \|Ar_k\|_q + \|A\|_{2 \rightarrow q} \left\| w - \frac{\langle w, r_k \rangle}{\|r_k\|_2^2} r_k \right\|_2 \\
&= \frac{|\langle w, r_k \rangle|}{\|r_k\|_2^2} \|Ar_k\|_q + \|A\|_{2 \rightarrow q} \left[1 - \frac{|\langle w, r_k \rangle|^2}{\|r_k\|_2^2} \right]^{1/2} \\
&\leq \frac{|\langle w, r_k \rangle|}{\|r_k\|_2^2} \|Ar_k\|_q + \|A\|_{2 \rightarrow q} \left[1 - \frac{|\langle w, r_k \rangle|^2}{2\|r_k\|_2^2} \right],
\end{aligned}$$

implying that

$$\frac{\|Ar_k\|_q}{\|r_k\|_2} \geq \frac{|\langle w, r_k \rangle|}{2\|r_k\|_2} \|A\|_{2 \rightarrow q} > \frac{\lambda}{2} \frac{\|A\|_{2 \rightarrow \infty}}{\|r_k\|_2} \|A\|_{2 \rightarrow q} \geq \frac{\lambda}{2} \|A\|_{2 \rightarrow q}.$$

Altogether, we have that

$$\max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2} \geq \min \left\{ \frac{\lambda}{2}, \frac{1}{\lambda^{\frac{q-2}{q}}} \left[\frac{\|Av\|_q}{\|Av\|_2} \right]^{\frac{2}{q}} \left[\max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2 \|A\|_{2 \rightarrow \infty}} \right]^{\frac{q-2}{q}} \right\} \|A\|_{2 \rightarrow q}.$$

The quantity $\min \left\{ \frac{\lambda}{2}, \frac{1}{\lambda^{\frac{q-2}{q}}} \left[\frac{\|Av\|_q}{\|Av\|_2} \right]^{\frac{2}{q}} \left[\max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2 \|A\|_{2 \rightarrow \infty}} \right]^{\frac{q-2}{q}} \right\}$ is maximized when

$$\lambda = 2^{\frac{q}{2(q-1)}} \left[\frac{\|Av\|_q}{\|Av\|_2} \right]^{\frac{1}{q-1}} \left[\max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2 \|A\|_{2 \rightarrow \infty}} \right]^{\frac{q-2}{2(q-1)}},$$

implying our desired result. \square

We note that the bound of Lemma 2.1 has a constant term approaching $\sqrt{2}$ as q tends to infinity. This lemma is clearly not tight for $q = \infty$, giving an estimate of $\sqrt{2}$ instead of 1. However, this is not an artifact of the bounding procedure, but the technique itself, as, for any fixed $q < \infty$, this factor is unavoidable for matrices whose optimal λ is a function of m .

Theorem 1.1 follows as a corollary to Lemma 2.1. Although the estimate in Theorem 1.1 is simpler, for a particular matrix A , the bounds from Lemma 2.1 can be tighter.

Proof of Theorem 1.1. First observe that

$$\frac{1}{2\|A\|_{2 \rightarrow \infty}} \max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2} \geq \frac{\|Ar_j\|_q}{2\|Ar_j\|_\infty}.$$

It remains to note that $\frac{\|Ar_j\|_q}{2\|Ar_j\|_\infty} \geq \frac{1}{2}$ and $\frac{\|Av\|_q}{\|Av\|_2} \geq m^{-\frac{q-2}{2q}}$, which follows from Hölder's inequality. \square

2.1. Examples . We give examples comparing the bounds from Lemma 2.1 with the interpolation bounds. We show that Lemma 2.1 is essentially the best approximation possible given our inputs. We also demonstrate the dependence of Lemma 2.1 on the choice of witness v of the spectral norm.

2.1.1. *Tightness of Lemma 2.1 and comparison with interpolation.* We begin by recalling the classical interpolation estimates for $\|A\|_{2 \rightarrow q}$ which are recorded in [26]. Since $2 \leq q \leq \infty$, the matrix norm $\|A\|_{2 \rightarrow q}$ is controlled by $\|A\|_{2 \rightarrow 2}$ and $\|A\|_{2, \infty}$ by interpolation. Indeed, if $w \in \mathbb{C}^n$ satisfies $\|A\|_{2 \rightarrow q} = \|Aw\|_q$ and $\|w\|_2 = 1$, then

$$\|Aw\|_q^q = \sum_{i=1}^m |\langle w, r_i \rangle|^q \leq \max_{1 \leq k \leq m} |\langle w, r_k \rangle|^{q-2} \sum_{i=1}^m |\langle w, r_i \rangle|^2 \leq \|A\|_{2 \rightarrow \infty}^{q-2} \|A\|_{2 \rightarrow 2}^2,$$

where we use the same notation as in Lemma 2.1. Let $r_j \in \mathbb{R}^n$ be a row of A with maximal $\|r_j\|_2$ and let $v \in \mathbb{R}^n$ satisfy $\|A\|_{2 \rightarrow 2} = \frac{\|Av\|_2}{\|v\|_2}$. Then the interpolation inequality is

$$\max \left(\frac{\|Ar_j\|_q}{\|r_j\|_2}, \frac{\|Av\|_q}{\|v\|_2} \right) \leq \|A\|_{2 \rightarrow q} \leq \left[\frac{\|Ar_j\|_\infty}{\|r_j\|_2} \right]^{\frac{q-2}{q}} \left[\frac{\|Av\|_2}{\|v\|_2} \right]^{\frac{2}{q}}. \quad (1)$$

Since it is straightforward to compute r_j and to approximate v , (1) gives a polynomial time approximation of $\|A\|_{2 \rightarrow q}$ with a multiplicative error bounded by $\left[\frac{\|Ar_j\|_\infty}{\|Ar_j\|_q} \right]^{\frac{q-2}{q}} \left[\frac{\|Av\|_2}{\|Av\|_q} \right]^{\frac{2}{q}}$.

The bounds in Lemma 2.1 involve more calculations: we require $\|Av\|_q/\|v\|_2$ and $\|Ar_i\|_q/\|r_i\|_2$ for each $i = 1, \dots, m$. Using Lemma 2.1, we approximate $\|A\|_{2, q}$ by

$$\max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2} \leq \|A\|_{2 \rightarrow q} \leq \left[\max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{2\|A\|_{2 \rightarrow \infty}\|x\|_2} \right]^{-\frac{q-2}{2(q-1)}} \left[\frac{\|Av\|_q}{\|Av\|_2} \right]^{-\frac{1}{q-1}} \max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2}. \quad (2)$$

We will give examples of $m \times m$ matrices B and D which have identical data as inputs for both (1) and (2). For these examples, the range for the $2 \rightarrow q$ norms given by interpolation is a factor of $m^{\frac{(q-2)^2}{2q^2(q-1)}}$ larger than the range given by Lemma 2.1. Furthermore, the $2 \rightarrow q$ norms of B and D take values in the opposite ends of the range given by Lemma 2.1, which demonstrates that the approximation is the best possible with the given data. In this section, the notation \sim and \lesssim means within a factor of 10.

Let $n = m+2$. Let $e_i \in \mathbb{R}^n$ be the standard basis vectors with a 1 in the i th coordinate and 0 otherwise. Define the rows of the $m \times n$ matrix B by $r_i^B = (1 - m^{-\frac{2}{q}})^{\frac{1}{2}} e_i + m^{-\frac{1}{q}} e_n$ for $i = 1, \dots, m$. Thus B has the form

$$\left[\begin{array}{c|cc} & 0 & m^{-\frac{1}{q}} \\ [1 - m^{-\frac{2}{q}}]^{1/2} I_m & \vdots & \vdots \\ & 0 & m^{-\frac{1}{q}} \end{array} \right]$$

in which I_m is the $m \times m$ identity.

Define the rows of the $m \times n$ matrix D by $r_i^D = m^{-\frac{1}{q}} e_1 + m^{-\frac{q-2}{2q(q-1)}} e_2 + [1 - m^{-\frac{2}{q}} - m^{-\frac{q-2}{q(q-1)}}]^{1/2} e_{i+2}$ if $1 \leq i \leq \lfloor m^{\frac{q-2}{q-1}} \rfloor$ and $r_i^D = m^{-\frac{1}{q}} e_1 + [1 - m^{-\frac{2}{q}}]^{1/2} e_{i+2}$ if $\lfloor m^{\frac{q-2}{q-1}} \rfloor + 1 \leq$

vector corresponding to the 2-norm of A , $y \in \mathbb{C}^n$ achieve the $1 \rightarrow q$ norm of A , and $S = \{v, r_1, \dots, r_m\}$. Then

$$\max_{\substack{x \in S \cup \{y\} \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_p} \geq \left[\frac{\|y\|_1}{\|y\|_p} \right]^{\frac{2-p}{p}} \left(\left[\frac{\|Ar_j\|_q}{2\|Ar_j\|_\infty} \right]^{\frac{q-2}{2(q-1)}} \left[\frac{\|Av\|_q}{\|Av\|_2} \right]^{\frac{1}{q-1}} \min_{\substack{x \in S \\ x \neq 0}} \frac{\|x\|_2}{\|x\|_p} \right)^{\frac{2(p-1)}{p}} \|A\|_{p \rightarrow q}$$

for all $1 \leq p \leq 2 \leq q$, and

$$\max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_p} \geq n^{\frac{2-p}{2p}} \left(\left[\frac{\|Ar_j\|_q}{2\|Ar_j\|_\infty} \right]^{\frac{q-2}{2(q-1)}} \left[\frac{\|Av\|_q}{\|Av\|_2} \right]^{\frac{1}{q-1}} \min_{\substack{x \in S \\ x \neq 0}} \frac{\|x\|_2}{\|x\|_p} \right) \|A\|_{p \rightarrow q}$$

for all $p, q \geq 2$.

Proof. Using the bound $\|A\|_{p \rightarrow q} \leq \|A\|_{1 \rightarrow q}^{\frac{2-p}{p}} \|A\|_{2 \rightarrow q}^{\frac{2(p-1)}{p}}$ and the results of Lemma 2.1, our estimates for $\|A\|_{p \rightarrow q}$, $1 \leq p \leq 2 \leq q$, follow relatively quickly:

$$\begin{aligned} \|A\|_{p \rightarrow q} &\leq \|A\|_{1 \rightarrow q}^{\frac{2-p}{p}} \|A\|_{2 \rightarrow q}^{\frac{2(p-1)}{p}} \\ &\leq \left[\frac{\|Ay\|_q}{\|y\|_1} \right]^{\frac{2-p}{p}} \left(\left[\frac{2\|Ar_j\|_\infty}{\|Ar_j\|_q} \right]^{\frac{q-2}{2(q-1)}} \left[\frac{\|Av\|_2}{\|Av\|_q} \right]^{\frac{1}{q-1}} \max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2} \right)^{\frac{2(p-1)}{p}} \\ &= \left[\frac{\|y\|_p}{\|y\|_1} \frac{\|Ay\|_q}{\|y\|_p} \right]^{\frac{2-p}{p}} \left(\left[\frac{2\|Ar_j\|_\infty}{\|Ar_j\|_q} \right]^{\frac{q-2}{2(q-1)}} \left[\frac{\|Av\|_2}{\|Av\|_q} \right]^{\frac{1}{q-1}} \max_{\substack{x \in S \\ x \neq 0}} \frac{\|x\|_p}{\|x\|_2} \frac{\|Ax\|_q}{\|x\|_p} \right)^{\frac{2(p-1)}{p}} \\ &\leq \left[\frac{\|y\|_p}{\|y\|_1} \right]^{\frac{2-p}{p}} \left(\left[\frac{2\|Ar_j\|_\infty}{\|Ar_j\|_q} \right]^{\frac{q-2}{2(q-1)}} \left[\frac{\|Av\|_2}{\|Av\|_q} \right]^{\frac{1}{q-1}} \max_{\substack{x \in S \\ x \neq 0}} \frac{\|x\|_p}{\|x\|_2} \right)^{\frac{2(p-1)}{p}} \max_{\substack{x \in S \cup \{y\} \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_p}. \end{aligned}$$

The regime of $p, q \geq 2$ is also straightforward, and follows from the bound $\|A\|_{p \rightarrow q} \leq n^{1/2-1/p} \|A\|_{2 \rightarrow q}$:

$$\begin{aligned} \|A\|_{p \rightarrow q} &\leq n^{\frac{p-2}{2p}} \|A\|_{2 \rightarrow q} \\ &\leq n^{\frac{p-2}{2p}} \left[\frac{2\|Ar_j\|_\infty}{\|Ar_j\|_q} \right]^{\frac{q-2}{2(q-1)}} \left[\frac{\|Av\|_2}{\|Av\|_q} \right]^{\frac{1}{q-1}} \max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_2} \\ &\leq n^{\frac{p-2}{2p}} \left[\frac{2\|Ar_j\|_\infty}{\|Ar_j\|_q} \right]^{\frac{q-2}{2(q-1)}} \left[\frac{\|Av\|_2}{\|Av\|_q} \right]^{\frac{1}{q-1}} \max_{\substack{x \in S \\ x \neq 0}} \frac{\|x\|_p}{\|x\|_2} \max_{\substack{x \in S \\ x \neq 0}} \frac{\|Ax\|_q}{\|x\|_p}. \end{aligned}$$

□

Theorem 1.2 follows immediately from the above theorem and the basic p -norm inequality $\|x\|_q \leq \|x\|_p \leq n^{1/p-1/q} \|x\|_q$ for all $x \in \mathbb{C}^n$ and $p \leq q$.

3.2. From Theory to Practice. To this point, our analysis has consisted entirely of the worst-case theoretical behavior. Theorem 3.1 gives a theoretical improvement over previous techniques. Here, we focus on how the theoretical results of this work influence the practical implementation of algorithms in practice. In terms of complexity, computing the largest singular value up to a relative accuracy of order ϵ with high probability requires at most $O(\log(n/\epsilon)/\epsilon^{1/2})$ matrix vector products using Krylov subspace methods (see [21, 27, 20] and references therein for details). The ratio of $\|Ar_j\|_q/\|r_j\|_p$ for all

graph	n	deg_{avg}	Adj. Mat. A			Norm. Adj. \mathcal{A}		
			α_{apx}	α_{imp}	M_{imp}	α_{apx}	α_{imp}	M_{imp}
144	144,649	14.86	2.11	0.5% [†]	2.8%	3.30	22.1%	0.0%
598a	110,971	13.37	2.60	12.3%	0.0%	3.20	21.0%	0.0%
auto	448,695	14.77	2.23	5.2%	0.0%	3.40	25.1%	5.1%
ca-AstroPh	18,772	21.10	1.91	0.0%	0.0%	1.84	0.0%	0.0%
ca-HepPh	12,008	19.74	1.43	0.0%	0.0%	1.92	0.0%	0.0%
coPapersCiteseer	434,102	73.88	1.33	13.1%	0.0%	2.94	17.6%	0.0%
coPapersDBLP	540,486	56.41	1.88	9.9%	0.0%	3.51	24.5%	0.0%
data	2,851	10.59	2.10	3.8%	2.9%	2.33	8.9%	3.2%
fe_rotor	99,617	13.30	1.65	0.0%	0.0%	3.13	21.0%	2.2%
fe_tooth	78,136	11.58	2.58	11.6%	0.0%	3.10	19.7%	0.0%
m14b	214,765	15.64	2.43	9.4%	0.0%	3.26	21.7%	0.0%
mycielskian16	49,151	679.18	2.38	11.3%	6.8%	2.47	10.4%	0.6%
wing_nodal	10,937	13.80	2.42	9.1%	0.0%	2.65	14.1%	2.4%

TABLE 2. We consider the matrix $2 \rightarrow 4$ norm for a variety of graphs taken from application. This table contains experiments for every undirected graph in the SuiteSparse Matrix Collection [10] with between 1,000 and 1,000,000 vertices and average degree deg_{avg} at least 10. For either the adjacency matrix A or the normalized adjacency matrix \mathcal{A} , we define a number of parameters. The quantity α_{apx} is the approximation produced by Lemma 2.1, e.g., Lemma 2.1 produces a vector x and a number α_{apx} such that $\|Ax\|_4/\|x\|_2 \geq \alpha_{\text{apx}}\|A\|_{2 \rightarrow 4}$. The quantity α_{imp} is the percent decrease in the approximation between Lemma 2.1 and standard interpolation bounds applied to the vectors v and r achieving the $2 \rightarrow 2$ and $2 \rightarrow \infty$ norms, respectively (e.g., the quantity $[\|Ar\|_\infty\|Av\|_2/(\|Ar\|_4\|Av\|_4)]^{1/2}$). The quantity M_{imp} is the percent increase in largest computed $\|Ax\|_4/\|x\|_2$ between Lemma 2.1 and the test vectors v and r produced by interpolation. The additional computation of $\|Ax\|_4/\|x\|_2$ for the rows of A led to an improvement in the best-known vector approximation over 30% of the time. However, even when the computation did not produce an improved vector approximation, it provided an improved guarantee (e.g., approximation ratio) of the quality of the existing vector approximation over 80% of the time.

$j = 1, \dots, m$ can be computed in matrix multiplication time, and for a matrix $A \in \mathbb{C}^{m \times n}$ with at most k non-zeros per row, in $O(k^2n)$ time, making our results particularly attractive for sparse matrices. In addition, we note that the vector with largest $p \rightarrow q$ norm produced by Theorem 3.1 is a strong initial guess for the matrix $p \rightarrow q$ norm, but would certainly benefit from the application of a modified power method in the spirit of Boyd and Higham [7, 17], or a general non-linear solver.

Here we consider the practical effectiveness of Lemma 2.1 for estimating the $2 \rightarrow 4$ norm of the adjacency matrix of sparse graphs. This is a natural choice, given the connection between the $2 \rightarrow 4$ norm and small set expansion in graph theory [2]. Let $G = (V, E)$, $n = |V|$, be an undirected graph. We denote by $\text{deg}(i)$ the degree of vertex i ,

i.e., the number of edges containing vertex i . The adjacency matrix $A \in \mathbb{R}^{n \times n}$ of a graph G has entries $A(i, j) = 1$ if there is an edge between vertex i and vertex j , and $A(i, j) = 0$ otherwise. The normalized adjacency matrix \mathcal{A} has entries $\mathcal{A}(i, j) = 1/[\deg(i)\deg(j)]^{1/2}$ if there is an edge between vertex i and vertex j , and $\mathcal{A}(i, j) = 0$ otherwise. A graph is said to be d -regular if $\deg(i) = d$ for all vertices i . In this setting, we note that $\mathcal{A} = A/d$. We compare the performance of our results (Lemma 2.1) to interpolation in two different settings.

First, we consider d -regular graphs, a setting for which Lemma 2.1 will clearly outperform interpolation techniques. Because each row contains exactly d entries equal to one and all other entries zero, the row norm contains no information save for the degree d . In addition, by the Perron-Frobenius Theorem, the top singular value is always d and the constant vector is always a corresponding right singular vector. Therefore, interpolation techniques cannot reliably discern any structure from a d -regular graph, save for the fact that it is d -regular. In contrast, using the rows as test vectors gives further information.

Next, we consider large real-world graphs with non-uniform degree structure, taken from the SuiteSparse matrix collection (formerly referred to as the University of Florida Sparse Matrix Collection) [10]. This collection contains graphs from a variety of applications, including structural finite element meshes, social networks, collaboration networks, road networks, and many other real-world examples. We consider every undirected graph from this collection with between 1,000 and 1,000,000 vertices and average degree at least 10. We analyze both the adjacency and normalized adjacency matrices of these graphs, and report the performance of Lemma 2.1. In a number of cases, a row without maximum 2-norm provides the best approximation. More generally, experimentally we note that even if the testing of the rows of a matrix do not improve the estimate of $\|A\|_{2 \rightarrow 4}$ itself, it often still improves the approximation guarantee – an equally important parameter that verifies the quality of the estimate. We provide the details of these experiments in Table 2.

REFERENCES

- [1] K. I. Babenko. An inequality in the theory of Fourier integrals. *Izv. Akad. Nauk SSSR Ser. Mat.*, 25:531–542, 1961.
- [2] 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, 2012.
- [3] William Beckner. Inequalities in Fourier analysis. *Ann. of Math. (2)*, 102(1):159–182, 1975.
- [4] 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.
- [5] Vijay Bhattiprolu, Mrinal Kanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Inapproximability of matrix norms. *SIAM Journal on Computing*, 52(1):132–155, 2023.
- [6] Punyashloka Biswal. Hypercontractivity and its applications. *arXiv preprint arXiv:1101.2913*, 2011.
- [7] David W Boyd. The power method for l_p norms. *Linear Algebra and its Applications*, 9:95–101, 1974.
- [8] Fernando GSL Brandao and Aram W Harrow. Estimating operator norms using covering nets. *arXiv preprint arXiv:1509.05065*, 2015.
- [9] Michael Christ. A sharpened hausdorff-young inequality, 2014.
- [10] Timothy A Davis and Yifan Hu. The university of florida sparse matrix collection. *ACM Transactions on Mathematical Software (TOMS)*, 38(1):1–25, 2011.

- [11] Matthias Englert and Harald Räcke. Oblivious routing for the lp-norm. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 32–40. IEEE, 2009.
- [12] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.
- [13] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.
- [14] Alexander Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. Sao Paulo*, 8:1–79, 1953.
- [15] Aram W Harrow, Anand Natarajan, and Xiaodi Wu. Limitations of semidefinite programs for separable states and entangled games. *Communications in Mathematical Physics*, 366:423–468, 2019.
- [16] Julien M Hendrickx and Alex Olshevsky. Matrix p-norms are np-hard to approximate if $p \neq 1, 2, \infty$. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2802–2812, 2010.
- [17] Nicholas J Higham. Estimating the matrix p-norm. *Numerische Mathematik*, 62:539–555, 1992.
- [18] Nicholas J Higham. *Accuracy and stability of numerical algorithms*. SIAM, 2002.
- [19] Haotian Jiang, Tarun Kathuria, Yin Tat Lee, Swati Padmanabhan, and Zhao Song. A faster interior point method for semidefinite programming. In *2020 IEEE 61st annual symposium on foundations of computer science (FOCS)*, pages 910–918. IEEE, 2020.
- [20] Raphael A Meyer, Cameron Musco, and Christopher Musco. On the unreasonable effectiveness of single vector krylov methods for low-rank approximation. *arXiv preprint arXiv:2305.02535*, 2023.
- [21] Cameron Musco and Christopher Musco. Randomized block krylov methods for stronger and faster approximate singular value decomposition. *Advances in neural information processing systems*, 28, 2015.
- [22] Yu Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization methods and software*, 9(1-3):141–160, 1998.
- [23] Gilles Pisier. Grothendieck’s theorem, past and present. *Bulletin of the American Mathematical Society*, 49(2):237–323, 2012.
- [24] Jiri Rohn. Computing the norm $\|A\|_{\infty,1}$ is NP-hard. *Linear and Multilinear Algebra*, 47(3):195–204, 2000.
- [25] Laurent Saloff-Coste. Lectures on finite markov chains. *Lectures on probability theory and statistics: École d’été de Probabilités de Saint-Flour XXVI-1996*, pages 301–413, 1997.
- [26] Daureen Steinberg. Computation of matrix norms with applications to robust optimization. *Research thesis, Technion-Israel University of Technology*, 2, 2005.
- [27] John C Urschel. Uniform error estimates for the lanczos method. *SIAM Journal on Matrix Analysis and Applications*, 42(3):1423–1450, 2021.

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MA 02139, USA

Email address: lguth@math.mit.edu

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MA 02139, USA

Email address: dmal@mit.edu

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MA 02139, USA

Email address: urschel@mit.edu