Embeddings of Schatten Norms with Applications to Data Streams*

Yi Li[1] and David P. Woodruff[2]

1 **Division of Mathematics, School of Physical & Mathematical Sciences**
Nanyang Technological University
`yili@ntu.edu.sg`

2 **IBM Almaden Research Center**
`dpwoodru@us.ibm.com`

Abstract

Given an $n \times d$ matrix A, its Schatten-p norm, $p \geq 1$, is defined as $\|A\|_p = \left(\sum_{i=1}^{\text{rank}(A)} \sigma_i(A)^p\right)^{1/p}$, where $\sigma_i(A)$ is the i-th largest singular value of A. These norms have been studied in functional analysis in the context of non-commutative ℓ_p-spaces, and recently in data stream and linear sketching models of computation. Basic questions on the relations between these norms, such as their embeddability, are still open. Specifically, given a set of matrices $A^1, \ldots, A^{\text{poly}(nd)} \in \mathbb{R}^{n \times d}$, suppose we want to construct a linear map L such that $L(A^i) \in \mathbb{R}^{n' \times d'}$ for each i, where $n' \leq n$ and $d' \leq d$, and further, $\|A^i\|_p \leq \|L(A^i)\|_q \leq D_{p,q}\|A^i\|_p$ for a given approximation factor $D_{p,q}$ and real number $q \geq 1$. Then how large do n' and d' need to be as a function of $D_{p,q}$?

We nearly resolve this question for every $p, q \geq 1$, for the case where $L(A^i)$ can be expressed as $R \cdot A^i \cdot S$, where R and S are arbitrary matrices that are allowed to depend on $A^1, \ldots, A^t$, that is, $L(A^i)$ can be implemented by left and right matrix multiplication. Namely, for every $p, q \geq 1$, we provide nearly matching upper and lower bounds on the size of n' and d' as a function of $D_{p,q}$. Importantly, our upper bounds are *oblivious*, meaning that R and S do not depend on the A^i, while our lower bounds hold even if R and S depend on the A^i. As an application of our upper bounds, we answer a recent open question of Blasiok et al. about space-approximation trade-offs for the Schatten 1-norm, showing in a data stream it is possible to estimate the Schatten-1 norm up to a factor of $D \geq 1$ using $\tilde{O}(\min(n,d)^2/D^4)$ space.

1998 ACM Subject Classification G. Mathematics of Computing

Keywords and phrases data stream algorithms, embeddings, matrix norms, sketching

Digital Object Identifier 10.4230/LIPIcs.ICALP.2017.

1 Introduction

Given an $n \times d$ matrix A, its Schatten-p norm, $p \geq 1$, is defined as $\|A\|_p = \left(\sum_{i=1}^{r(A)} \sigma_i(A)^p\right)^{\frac{1}{p}}$, where $r(A)$ is the rank of A and $\sigma_i(A)$ is the i-th largest singular value of A, i.e., the square root of the i-th largest eigenvalue of $A^T A$. The Schatten-1 norm is the nuclear norm or trace norm, the Schatten-2 norm is the Frobenius norm, and the Schatten ∞-norm, defined as the limit of the Schatten-p norm when $p \to \infty$, is the operator norm. The Schatten 1-norm has applications in non-convex optimization [5], while Schatten-2 and Schatten-∞ norms are useful in geometry and linear algebra, see, e.g., [22]. Schatten-p norms for large p also provide approximations to the Schatten-∞ norm.

* Full version available at `arXiv:1702.05626 [cs.DS]`.

44th International Colloquium on Automata, Languages, and Programming (ICALP 2017).
Editors: Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl;
Article No. ; pp. :1–:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The Schatten norms appear to be significantly harder to compute or approximate than the vector ℓ_p-norms in various models of computation, and understanding the complexity of estimating them has led to new algorithmic ideas and lower bound techniques. The main difficulty is that we do not directly have access to the spectrum of A, and naïvely it is costly in space and time to extract useful information about it. A line of work has focused on understanding the complexity of estimating such norms in the data stream model with 1-pass over the stream [13] as well as with multiple passes [4], the sketching model [2, 12, 14], statistical models [9], as well as the general RAM model [17, 19]. Dimensionality reduction in these norms also has applications in quantum computing [8, 21]. It has also been asked in places if the Schatten-1 norm admits non-trivial nearest neighbor search data structures [1].

Our Results. In this paper we study the embeddability of the Schatten-p norm into the Schatten-q norm for linear maps implementable by matrix multiplication. More concretely, we first ask for the following form of embeddability: given n and t (where $t = \Omega(\log n)$), what is the smallest value of $D_{p,q}$, which we call the *distortion*, such that there exists a distribution $\mathcal{R}$ on $\mathbb{R}^{t\times n}$ satisfying, for any given $n \times d$ matrix A,

$$\Pr_{R\sim\mathcal{R}}\left\{\|A\|_p \leq \|RA\|_q \leq D_{p,q}\|A\|_p\right\} \geq 1 - \exp(-ct)?$$

Here $c > 0$ is an absolute constant. We can assume, w.l.o.g., that $n = d$ because we can first apply a so-called *subspace embedding* matrix (see, e.g., [22] for a survey) to the left or to the right of A to preserve each of its singular values up to a constant factor. We shall show that $D_{p,q} \gtrsim \hat{D}_{p,q}$, where

$$\hat{D}_{p,q} = \begin{cases} n^{\frac{1}{p}-\frac{1}{2}}/t^{\frac{1}{q}-\frac{1}{2}}, & 1 \leq p \leq q \leq 2; \\ n^{\frac{1}{p}-\frac{1}{2}}, & 1 \leq p \leq 2 \leq q; \\ \max\{(n/t)^{\frac{1}{2}-\frac{1}{p}}, t^{\frac{1}{p}-\frac{1}{q}}\}, & 2 \leq p \leq q; \\ n^{\frac{1}{2}-\frac{1}{p}}, & 1 \leq q \leq 2 \leq p; \\ n^{\frac{1}{2}-\frac{1}{p}}/t^{\frac{1}{2}-\frac{1}{q}}, & 2 \leq q \leq p; \\ \max\{(n/t)^{\frac{1}{p}-\frac{1}{2}}, (t/\ln t)^{\frac{1}{q}-\frac{1}{p}}\}, & 1 \leq q \leq p \leq 2, \end{cases} \tag{1}$$

and the notation $f \gtrsim g$ means $f \geq g/C$ for some constant $C > 0$. The constant C in the $\gtrsim$ notation above depends on p and q only. This distortion is asymptotically tight, up to logarithmic factors, as we also construct a distribution $\mathcal{R}$ on t-by-n matrices for which for any $n \times d$ matrix A,

$$\Pr_{R\sim\mathcal{R}}\left\{\|A\|_p \leq \|RA\|_q \leq \tilde{D}_{p,q}\left(\log\frac{n}{t}\right)\|A\|_p\right\} \geq 1 - \exp\left(-ct\right),$$

where $\tilde{D}_{p,q}$ differs from $D_{p,q}$ by a constant or a factor of $\log t$. Specifically,

$$\tilde{D}_{p,q} \lesssim \begin{cases} \max\{(n/t)^{\frac{1}{p}-\frac{1}{2}}, t^{\frac{1}{q}-\frac{1}{p}}\}, & 1 \leq q \leq p \leq 2; \\ \hat{D}_{p,q}, & \text{otherwise}, \end{cases} \tag{2}$$

where $\hat{D}_{p,q}$ is given in (1). Replacing t with $t/(\ln(n/t))$, we arrive at a matching failure probability and distortion, while using a logarithmic factor more number of rows in R. Namely, we construct a distribution $\mathcal{R}$ on matrices with $t\ln(n/t)$ rows for which

$$\Pr_{R\sim\mathcal{R}}\left\{\|A\|_p \leq \|RA\|_q \leq \tilde{D}_{p,q}\|A\|_p\right\} \geq 1 - \exp\left(-ct\right).$$

We can also sketch RA on the right by a subspace embedding matrix S with $\Theta(t)$ rows, which yields

$$\Pr_{R,S}\left\{\|A\|_p \leq \left\|RAS^T\right\|_q \leq \tilde{D}_{p,q}\|A\|_p\right\} \geq 1-\exp\left(-ct\right).$$

We show that this two-sided sketch is asympotically optimal for two-sided sketches in its product of number of rows of R and number of columns of S, up to logarithmic factors. Formally, we next ask: what is the smallest value of $D_{p,q}$ for which there exists a distribution $\mathcal{G}_1$ on $\mathbb{R}^{r\times n}$ and a distribution $\mathcal{G}_2$ on $\mathbb{R}^{n\times s}$ satisfying

$$\Pr_{R\sim\mathcal{G}_1, S\sim\mathcal{G}_2}\left\{\|A\|_p \leq \|RAS\|_q \leq D_{p,q}\|A\|_p\right\} \geq 1-\exp(-c\min\{r,s\})?$$

Again we can assume, w.l.o.g, that $r = s$, because otherwise we can compose R or S with a subspace embedding to preserve all singular values up to a constant factor[1]. Henceforth for the two-sided problem, we assume that $\mathcal{G}_1$ and $\mathcal{G}_2$ are distributions on $\mathbb{R}^{t\times n}$. We also prove a matching lower bound that $D_{p,q} \gtrsim \hat{D}_{p,q}$ except in the case when $1 \leq q \leq p \leq 2$, where we instead obtain a matching lower bound up to logarithmic factors, namely, $D_{p,q} \gtrsim \max\{(n/t)^{\frac{1}{p}-\frac{1}{2}}/\log^{\frac{3}{2}} t, (t/\ln t)^{\frac{1}{q}-\frac{1}{p}}\}$.

In the important case when $p = q = 1$, our results show a space-approximation tradeoff for estimating the Schatten 1-norm (or trace norm) in a data stream, answering a question posed by Blasiok et al. [3]. This application crucially uses that R and S are oblivious to A, i.e., they can be sampled and succinctly stored without looking at A. Specifically, when each entry of A fits in a word of $O(\log n)$ bits, we can choose R and S to be Gaussian random matrices with entries truncated to $O(\log n)$ bits and with entries drawn from a family of random variables with bounded independence. For time-efficiency purposes, R and S can also be chosen to be Fast Johnson Lindenstrauss Transforms or sparse embedding matrices [6, 16, 18], though they will have larger dimension, especially to satisfy the exponential probability of failure in the problem statement (and even with constant failure probability, the dimension will be slightly larger; see [22] for a survey).

Choosing R and S to be Gaussian matrices, our result provides a data stream algorithm using $(n^2/D^4)\operatorname{polylog}(n)$ bits of memory, and achieving approximation factor D (taking $t = n/D^2$). While $\|A\|_2$, the Frobenius norm of A, provides a $\sqrt{n}$-approximation to $\|A\|_1$ and can be approximated up to a constant factor in a data stream using $O(1)$ words of space, if we want an algorithm achieving a better approximation factor then all that was known was an algorithm requiring $O(n^2)$ words of space, namely, the trivial algorithm of storing A exactly and achieving $D = 1$. It was asked in [3] if there is a smooth trade-off between the case when $D = 1$ and $D = \sqrt{n}$; our $(n^2/D^4)\operatorname{polylog}(n)$ space algorithm provides the first such trade-off, and is optimal at the two extremes. Our results are the first of their kind for large approximation factors $D \gg 1$ for estimating the Schatten-p norms in a data stream.

Finally, while in our upper bounds R and S are chosen obliviously to A, for our lower bounds we would like to rule out those R and S which are even allowed to depend on A. Clearly, if there is only a single matrix A, this question is ill-posed as one can just choose R and S to have a single row and column so that $\|RAS\|_q = \|A\|_p$. Instead, we ask the question analogous to the Johnson-Lindenstrauss transform (see e.g., [10]): given

[1] That is, if $r \leq s$, we can choose a subspace embedding matrix H of dimension $n \times \Theta(r)$ such that $\|RASH\|_q = \Theta(\|RAS\|_q)$ with probability $\geq 1-\exp(-s)$, and then pad R with zero rows so that R has the same number of rows as columns of S, increasing the number of rows of R by at most a constant factor.

$A^1, \ldots, A^{\mathrm{poly}(n)}$, can we construct an R with t rows and an S with t columns for which $\|A^i\|_p \leq \|RA^iS\|_q \leq D_{p,q}\|A^i\|_p$ for all i? We show that our lower bound on the trade-off between $D_{p,q}$ and t given by (1) continues to hold even in this setting.

Our Techniques. We shall focus on the case $p = q$ in this description of our technical overview. For our upper bounds, a natural idea is to take R to be a (normalized) Gaussian random matrix, and the analysis of the quantity $\|RA\|_p$, when $p \geq 2$, follows fairly directly from the so-called non-commutative Khintchine inequality as follows.

▶ **Lemma 1** (Non-commutative Khintchine Inequality [15]). *Suppose that $C_1, \ldots, C_n$ are (deterministic) matrices of the same dimension and $g_1, \ldots, g_n$ are independent $N(0,1)$ variables. It holds that*

$$\mathop{\mathbb{E}}_{g_1,\ldots,g_n} \left\| \sum_i g_i C_i \right\|_p \simeq \max\left\{ \left\| \left(\sum_i C_i C_i^T \right)^{\frac{1}{2}} \right\|_p, \left\| \left(\sum_i C_i^T C_i \right)^{\frac{1}{2}} \right\|_p \right\}, \quad p \geq 2.$$

In order to estimate $\|RA\|_p$, we can write

$$RA = \sum_{i,j} r_{ij}(e_i e_j^T A) =: \sum_{i,j} r_{ij} C_{ij}$$

and it is straightforward to compute that

$$\sum_{i,j} C_{ij} C_{ij}^T = \mathrm{tr}(AA^T) I_t = \|A\|_F^2 I_t, \quad \sum_{i,j} C_{ij}^T C_{ij} = t \cdot A^T A.$$

It follows from the non-commutative Khintchine inequality that (recall that R is a normalized Gaussian matrix with $N(0, 1/t)$ entries)

$$\mathbb{E}\|RA\|_p \simeq \max\left\{ t^{\frac{1}{2}-\frac{1}{p}} \|A\|_F, \|A\|_p \right\}, \quad p \geq 2.$$

Using a concentration inequality for Lipschitz functions on Gaussian space, one can show that $\|RA\|_p$ is concentrated around $\mathbb{E}\|RA\|_p$, and using standard the standard relationship between $\|A\|_F$ and $\|A\|_p$ then completes the argument.

When $p < 2$, the non-commutative Khintchine inequality gives a much less tractable characterization, so we need to analyze $\|RA\|_p$ in a different manner, which is potentially of independent interest. Our analysis also works for non-Gaussian matrices R whenever R satisfies certain properties, which, for instance, are satisfied by a Fast Johnson-Lindenstrauss Transform.

Upper bound. We give an overview of our upper bound now, focusing on the one-sided case, since the two-sided case follows by simply right-multiplying by a generic subspace embedding S. Here we focus on the case in which R is an $r \times n$ Gaussian matrix, where $r = t \cdot \mathrm{polylog}(n)$. By rotational invariance of Gaussian matrices, and for the purposes of computing $\|AR\|_p$, we can assume that A is diagonal. Let A_1 be the restriction of A to its top $\Theta(t \log n)$ singular values. Since R is a Gaussian matrix with at least $t \log n$ rows, it is well-known that R is also a subspace embedding on A_1 (see, e.g., [20, Corollary 5.35]), namely, $\sigma_i(RA_1) \simeq \sigma_i(A_1)$ for all i, and thus $\|RA_1\|_p \simeq \|A_1\|_p = \Omega(\|A\|_p)$ when $\|A_1\|_p = \Omega(\|A\|_p)$.

If it does not hold that $\|A_1\|_p = \Omega(\|A\|_p)$, then the singular values of A are "heavy-tailed", and we show how to find a $\sigma_i(A)$ with $i < \Theta(t \log n)$ for which $\sigma_i^2(A)$ is relatively small compared to $\sigma_i^2(A) + \sigma_{i+1}^2(A) + \cdots + \sigma_n^2(A)$. More specifically, let A_2 be the restriction of A to $\sigma_i(A), \ldots, \sigma_n(A)$. Then we have that $\|A_2\|_{op} \lesssim \|A_2\|_F/\sqrt{t}$. Since for a Gaussian matrix R it holds that $\|RA_2\|_{op} \lesssim \|A_2\|_{op} + \|A\|_F/\sqrt{r}$ (see Proposition 3), we thus have that $\|RA_2\|_{op} \lesssim \|A_2\|_F/\sqrt{t}$. On the other hand, $\|RA_2\|_F \simeq \|A_2\|_F$. This implies there exist $\Omega(t)$ singular values of RA_2 that are $\Omega(\|A_2\|_F/\sqrt{t})$, which yields that $\|RA_2\|_p \gtrsim \|A_2\|_p = \Omega(\|A\|_p)$. Therefore we have established the lower bound that $\|RA\|_p \geq \max\{\|RA_1\|_p, \|RA_2\|_p\}$ in terms of $\|A\|_p$.

To upper bound $\|RA\|_p$ in terms of $\|A\|_p$, note that $\|RA\|_p \leq \|RA_1\|_p + \|RA_2\|_p$ by the triangle inequality, where A_1, A_2 are as above. Again it follows from the subspace embedding property of R that $\|RA_1\|_p \lesssim \|A_1\|_p \leq \|A\|_p$. Regarding $\|RA_2\|_p$, we relate its Schatten-p norm to its Frobenius norm and use the fact that $\|RA_2\|_F \simeq \|A_2\|_F$. This gives an upper bound of $\|RA_2\|_p$ in terms of $\|A_2\|_p$, and using that $\|A_2\|_p \leq \|A\|_p$, it gives an upper bound in terms of $\|A\|_p$. This is sufficient to obtain an overall upper bound on $\|RA\|_p$.

Lower bound. Now we give an overview of our lower bounds for some specific cases. First consider one-sided sketches. We choose our hard distribution as follows: we choose an $n \times (10t)$ Gaussian matrix G padded with 0s to become an $n \times n$ matrix. For a sketch matrix R containing t rows, by rotational invariance of Gaussian matrices, $\|RG\|_p$ is identically distributed to $\|\Sigma_R G'\|_p$, where Σ_R is the $t \times t$ diagonal matrix consisting of the singular values of R, and where G' is a $t \times (10t)$ Gaussian matrix. It is a classical result that all singular values of G' are $\Theta(\sqrt{t})$ and thus $\|RG\|_p \simeq \sqrt{t}\|R\|_p$. This implies that

$$\sqrt{n} t^{\frac{1}{2}-\frac{1}{p}} \lesssim \sqrt{t}\|R\|_p \lesssim D_{p,p}\sqrt{n} t^{\frac{1}{2}-\frac{1}{p}}, \tag{3}$$

since all non-zero singular values of G are $\Theta(\sqrt{n})$. On the other hand, applying R to the $n \times n$ identity matrix gives that

$$n^{\frac{1}{p}} \leq \|R\|_p \leq D_{p,p} n^{\frac{1}{p}}. \tag{4}$$

Combining (3) and (4) gives that $D_{p,p} \geq \max\{(n/t)^{1/2-1/p}, (n/t)^{1/p-1/2}\}$.

For the two-sided sketch, we change the hard distribution to (i) $n \times n$ Gaussian random matrix F and (ii) the distribution of GH^T, where G and H are $n \times \Theta(t)$ Gaussian random matrices. The proof then relies on the analysis for $\left\|RFS^T\right\|_p$ and $\left\|RGH^T S^T\right\|_p$. When $p \geq 2$, non-commutative Khintchine inequality gives immediately that

$$\left\|RGH^T S^T\right\|_p \simeq \sqrt{t}\left\|RFS^T\right\|_p \simeq \sqrt{t}\max\{\|R\|_p\|S\|_{op}, \|R\|_{op}\|S\|_p\}, \quad p \geq 2. \tag{5}$$

When $p < 2$, a different approach is followed. We divide the singular values of R and S into bands, where each band contains singular values within a factor of 2 from each other. We shall consider the first $\Theta(\log t)$ bands only because the remaining singular values are $1/\operatorname{poly}(t)$ and negligible. Now, if all singular values of R' and S' are within a factor of 2 from each other, then $\left\|R'F(S')^T\right\|_p \simeq \|R'\|_{op}\|S'\|_{op}\|F\|_p$ and $\left\|R'GH^T(S')^T\right\|_p \simeq \|R'\|_{op}\|S'\|_{op}\left\|GH^T\right\|_p$. It is not difficult to see that

$$\|GH^T\|_p \simeq \sqrt{t}\|F\|_p \tag{6}$$

Since R' and S' consist of one of the $\Theta(\log t)$ bands of R and S, respectively, it follows that

$$\left\|RGH^T S^T\right\|_p \simeq \sqrt{t}/\operatorname{polylog}(t) \cdot \left\|RFS^T\right\|_p, \quad p < 2. \tag{7}$$

A lower bound of $D_{p,p}$ then follows from combining (6), (5) (or (7)) with

$$\|F\|_p \leq \left\|RFS^T\right\|_p \leq D_{p,p}\|F\|_p, \quad \text{and} \quad \left\|GH^T\right\|_p \leq \left\|RGH^T S^T\right\|_p \leq D_{p,p}\left\|GH^T\right\|_p.$$

To strengthen the lower bound for the sketches that even depend on the input matrix, we follow the approach in [10]. We first work with random hard instances, and then sample input matrices $A^1, \ldots, A^{\mathrm{poly}(n)}$ from the hard distribution, and apply a net argument on sketching matrices R and S to obtain a deterministic statement, which states that for any fixed R and S such that the distortion guarantee is satisfied with all samples $A^1, \ldots, A^{\mathrm{poly}(n)}$, the distortion lower bound remains to hold.

2 Preliminaries

Notation. Throughout the paper, we use $f \lesssim g$ to denote $f \leq Cg$ for some constant C, $f \gtrsim g$ to denote $f \geq Cg$ for some constant C, and $f \simeq g$ to denote $C_1 g \leq f \leq C_2 g$ for some constants C_1 and C_2.

Bands of Singular Values. Given a matrix A, we split the singular values of A, $\sigma_1(A) \geq \sigma_2(A) \geq \cdots$, into bands such that the singular values in each band are within a factor of 2 from each other. Formally, define the i-th singular value band of A to be

$$\mathcal{B}_i(A) = \left\{ k : \frac{\|A\|_{op}}{2^{i+1}} < \sigma_k(A) \leq \frac{\|A\|_{op}}{2^i} \right\}, \quad i \geq 0,$$

and let $N_i(A) = |\mathcal{B}_i(A)|$, the cardinality of the i-th band.

Extreme Singular Values of Gaussian Matrices. We shall repeatedly use the following results on Gaussian matrices.

▶ Proposition 2 ([20, Corollary 5.35]). Let G be an $r \times n$ ($r < n$) Gaussian random matrix of i.i.d. entries $N(0,1)$. With probability at least $1-2\exp(-u^2/2)$, it holds that $\sqrt{n} - \sqrt{r} - u \leq s_{\min}(G) \leq s_{\max}(G) \leq \sqrt{n} + \sqrt{r} + u$.

Combining [11, Corollary 3.21] and the concentration bound in Gauss space [20, Proposition 5.34], we also have

▶ Proposition 3. Let A be a deterministic $n \times n$ matrix and G be an $r \times n$ ($r < n$) Gaussian random matrix of i.i.d. entries $N(0,1)$. Then for any K, it holds that $\|GA\|_{op} \leq K(\|A\|_{op}\sqrt{r} + \|A\|_F)$ with probability at least $1 - \exp(-c\sqrt{K}r)$, where $c > 0$ is an absolute constant.

Nets on Matrices. The following fact concerns nets of matrices and was used in [10]. We shall use it in our lower bound arguments.

▶ Proposition 4 ([10, Lemma 2]). There exists a net $\mathcal{R} \subset \bigcup_{t=1}^{t_0} \mathbb{R}^{t\times n}$ of size $\exp(O(t_0 n \ln(Dn/\eta)))$ such that for any $R \in \mathbb{R}^{t\times n}$ ($1 \leq t \leq t_0$) with column norms in $[1, D]$, we can find $R' \in \mathcal{R}$ such that $\|R - R'\|_{op} \leq \eta$.

3 Lower Bounds

In this section we show the full proof of the $(n/t)^{1/2-1/p}$ lower bound for one-sided sketches (Theorem 5) and the $(n/t)^{1/p-1/2}/\log^{3/2} t$ bound for two-sided sketches (Theorem 6), which demonstrates our techniques. Other cases can be found in the full version.

▶ **Theorem 5** (One-sided sketch). *Let $p > 2$ and $p > q$. There exist a set $T \subset \mathbb{R}^{n\times n}$ with $|T| = \mathrm{poly}(n)$ and an absolute constant $c \in (0,1)$ such that, if it holds for some matrix*

$R \in \mathbb{R}^{t\times n}$ with $t \leq cn$ and for all $A \in T$ that $\|A\|_p \leq \|RA\|_q \leq D_{p,q}\|A\|_p$, then it must hold that $D_{p,q} \gtrsim (n/t)^{\frac{1}{2}-\frac{1}{p}}$.

Instead of proving this theorem, we prove the following rephrased version.

▶ **Theorem 5' (rephrased).** Let $p > 2$ and $p > q$. There exists an absolute constant D_0 and a set $T \subset \mathbb{R}^{n\times n}$ with $|T| = O(n \ln(Dn))$ such that, if $D \geq D_0$ and it holds for some matrix $R \in \mathbb{R}^{t\times n}$ and for all $A \in T$ that

$$\|A\|_p \leq \|RA\|_q \leq D^{\frac{1}{2}-\frac{1}{p}}\|A\|_p, \tag{8}$$

then it must hold that $t \gtrsim n/D$.

Proof. Let $r = n/(\rho^2 D)$ and $t_0 = \theta r$ for some constants $\rho > 1$ and $\theta \in (0,1)$ to be determined. We shall show that if $t \leq t_0$, it will not happen that R satisfies (8) for all A in a carefully chosen set T.

Let $\mathcal{D}$ be the distribution of Gaussian random matrices of dimension $n \times r$ with i.i.d. entries $N(0, 1/r)$. Let $R = U\Sigma V^T$ be the singular value decomposition of R and $A \sim \mathcal{D}$. Then by rotational invariance of the Schatten norm and Gaussian random matrices, we know that $\|RA\|_q$ is identically distributed to $\|\Sigma A\|_q = \|B^T\Sigma'\|_q$, where Σ' is the left $t \times t$ block of Σ and B is formed by the first t rows of A.

It follows from Proposition 2 that with probability $\geq 1 - \exp(-c_1c_2r)$, $s_{\max}(B) \leq 1 + 2c_1\sqrt{t/r} \leq 1 + 2\sqrt{\theta}c_1$, and thus

$$\|B^T\Sigma'\|_q \leq s_{\max}(B)\|\Sigma'\|_q \leq (1+2\sqrt{\theta}c_1)\|\Sigma'\|_q = (1+\sqrt{\theta}c_1)\|R\|_q \leq (1+2\sqrt{\theta}c_1)D^{\frac{1}{2}-\frac{1}{p}}n^{\frac{1}{p}},$$

that is, with probability $\geq 1 - \exp(-c_1c_2r)$,

$$\|RA\|_q \leq (1+2\sqrt{\theta}c_1)D^{\frac{1}{2}-\frac{1}{p}}n^{\frac{1}{p}}.$$

On the other hand, with probability $\geq 1 - \exp(-c_1c_2r)$, all singular values of A are at least $\sqrt{n/r} - 2c_1 = \rho\sqrt{D} - 2c_1 \geq (1-\epsilon)\rho\sqrt{D}$ if we choose $D_0 \geq 4c_1^2/\epsilon^2$. Then

$$\|RA\|_q \geq \|A\|_p \geq (1-\epsilon)sr^{\frac{1}{p}}\sqrt{D} = (1-\epsilon)\rho^{1-\frac{2}{p}}n^{\frac{1}{p}}D^{\frac{1}{2}-\frac{1}{p}}.$$

Also, with probability $\geq 1 - \exp(-c_1c_2r)$, all singular values of A are at most $\sqrt{n/r} + 2c_1 = \rho\sqrt{D} + 2c_1 \leq (1+\epsilon)\rho\sqrt{D}$ and thus

$$\|A\|_p \leq r^{\frac{1}{p}}(1+\epsilon)s\sqrt{D} = (1+\epsilon)\rho^{1-\frac{2}{p}}n^{\frac{1}{p}}D^{\frac{1}{2}-\frac{1}{p}}.$$

This motivates the following definitions of constraints for $R \in \mathbb{R}^{t\times n}$ and $A \in \mathbb{R}^{n\times n}$:

$$\mathrm{P}_1(R,A): \|RA\|_q \leq (1+2\sqrt{\theta}c_1)D^{\frac{1}{2}-\frac{1}{p}}n^{\frac{1}{p}} \quad \mathrm{P}_2(R,A) : \|RA\|_q \geq (1-\epsilon)\rho^{1-\frac{2}{p}}n^{\frac{1}{p}}D^{\frac{1}{2}-\frac{1}{p}}$$

$$\mathrm{P}_3(A): \|A\|_p \leq (1+\epsilon)\rho^{1-\frac{2}{p}}n^{\frac{1}{p}}D^{\frac{1}{2}-\frac{1}{p}}.$$

Now, for m samples $A_1, \ldots, A_m$ drawn from $\mathcal{D}$, it holds for any fixed R that

$$\Pr_{A_1,\ldots,A_m}\{\exists i \text{ s.t. } \mathrm{P}_1(R,A) \text{ and } \mathrm{P}_2(R,A) \text{ and } \mathrm{P}_3(A) \text{ hold}\} \geq 1 - e^{-c_1c_2mr}. \tag{9}$$

Since $1 \leq \|Ge_ie_i^T\|_q \leq D$ and $\|Ge_ie_i^T\|_q = \|R_i\|_2$, we can restrict the matrix R to matrices with column norm in $[1, D]$. Thus we can find a net $\mathcal{R} \subset \bigcup_{t=1}^{t_0}\mathbb{R}^{t\times n}$ of size $\exp(O(t_0 n \ln(Dn/\eta))$ such that for any R with column norms in $[1, D]$, we can find $R' \in \mathcal{R}$ such that $\|R - R'\|_{op} \leq \eta$.

Now it follows from (9) that

$$\Pr_{A_1,\ldots,A_m}\{\forall R\in\mathcal{R},\exists i,\ \mathrm{P}_1(R,A)\text{ and }\mathrm{P}_2(R,A)\text{ and }\mathrm{P}_3(R,A)\text{ hold}\}$$
$$\geq 1-\exp\left(O\left(t_0 n\ln\frac{Dn}{\eta}\right)\right)\exp\left(-\frac{c_1c_2}{D}mn\right)>0,$$

provided that $m=\Theta(n\ln(Dn))$. Fix $A_1,\ldots,A_m$ such that for each $R\in\mathcal{R}$ there exists an i such that $\mathrm{P}_1(R,A_i)$ and $\mathrm{P}_2(R,A_i)$ and $\mathrm{P}_3(A_i)$ all hold.

Take $T=\{I_n,e_1e_1^T,\ldots,e_ne_n^T,A_1,\ldots,A_m\}$. We know that if R satisfies (8) for all $A\in T$, then there exists R' such that $\|R'-R\|_{op}\leq\eta$, and there exists $1\leq i\leq m$ such that $\mathrm{P}_1(R',A_i)$, $\mathrm{P}_2(R',A_i)$ and $\mathrm{P}_3(A_i)$ all hold. It follows that

$$\begin{aligned}\|RA_i\|_q &\leq \|R'A_i\|_q+\|(R-R')A_i\|_q\leq\|R'A_i\|_q+\|R-R'\|_{op}\|A_i\|_p\\ &\leq\left(1+2\sqrt{\theta}c_1+(1+\epsilon)\rho^{1-\frac{2}{p}}\eta\right)D^{\frac12-\frac1p}n^{\frac1p}\end{aligned}$$

and

$$\begin{aligned}\|RA_i\|_q &\geq \|RA_i\|_q-\|(R-R')A_i\|_q\geq\|R'A_i\|_q-\|R-R'\|_{op}\|A_i\|_p\\ &\geq\left((1-\epsilon)-(1+\epsilon)\eta\right)\rho^{1-\frac{2}{p}}D^{\frac12-\frac1p}n^{\frac1p}\end{aligned}$$

We meet a contradiction when θ, ϵ and η are all sufficiently small and ρ is sufficiently large, for instance, when $\eta=\Theta(\epsilon)$, $\theta=\Theta(\epsilon^2/c_2^2)$ and $\rho=\Theta(1+p\epsilon/(p-2))$. ◀

▶ **Theorem 6** (Two-sided sketch). *Let $p<2$. There exist a set $T\subset\mathbb{R}^{n\times n}$ with $|T|=\mathrm{poly}(n)$ and an absolute constant $c\in(0,1)$ such that, if it holds for some matrices $R,S\in\mathbb{R}^{t\times n}$ with $t\leq cn$ and for all $A\in T$ that $\|A\|_p\leq\left\|RAS^T\right\|_q\leq D_{p,q}\|A\|_p$, it must hold that $D_{p,q}\gtrsim(n/t)^{\frac1p-\frac12}/\log^{\frac32}t$.*

Instead of proving this theorem, we prove the following rephrased version.

▶ Theorem 6' (rephrased). Let $p<2$, $p>q$ and $D\geq D_0$ for some an absolute constant D_0. There exists a set $T\subset\mathbb{R}^{n\times n}$ with $|T|=O(n\ln(Dn))$ such that it holds for some matrices $R,S\in\mathbb{R}^{t\times n}$ and for all $A\in T$ that

$$\|A\|_p\leq\left\|RAS^T\right\|_q\leq D^{\frac1p-\frac12}\|A\|_p,\tag{10}$$

then it must hold that $t\gtrsim n/(D\log^{3p/(2-p)}t)$.

We need two auxiliary lemmata, whose proofs are omitted owing to space limitations.

▶ **Lemma 7.** *Let A and B be deterministic $n\times n$ matrices and G be a Gaussian random matrix of i.i.d. $N(0,1)$ entries. It holds with probability $1-O(1)$ that*

$$\|AGB\|_p\lesssim(\log^{\frac52}n)(\log\log n)\|A\|_{op}\|B\|_{op}E_p(A,B),$$

where

$$E_p(A,B)=\max_{0\leq i,j\leq 3\log n}\frac{1}{2^{i+j}}\cdot\min\left\{N_i(A),N_j(B)\right\}^{\frac1p}\cdot\max\left\{\sqrt{N_i(A)},\sqrt{N_j(B)}\right\}.\tag{11}$$

▶ **Lemma 8.** *Let A and B be deterministic $n\times N$ matrices and G,H be $N\times r$ Gaussian random matrices of i.i.d. $N(0,1)$ entries. Suppose that $n\leq cr$ for some absolute constant $c\in(0,1)$. It holds with probability $1-O(1)$ that $\|AGH^TB^T\|_p\gtrsim\sqrt{r}\|A\|_{op}\|B\|_{op}E_p(A,B)$, where $E_p(A,B)$ is as defined in (11).*

Now we are ready to show Theorem 6'.

Proof of Theorem 6'. Without loss of generality, we can assume that the maximum column norm of R and that of S are the same; otherwise we can rescale R and S.

Let $r = n/(\rho^2 D)$ and $t_0 = \theta r$ for some $\rho = \Theta(\log^{3p/(2-p)} t)$ and $\theta \in (0,1)$ to be determined. We shall show that if $t \leq t_0$, it will not happen that R and S satisfy (10) for all $A \in T$.

Let $\mathcal{D}$ be the distribution of Gaussian random matrices of dimension $n \times r$ with i.i.d. entries $N(0,1)$ and let $G, H \sim \mathcal{D}$ be independent. It follows from Lemma 8 that with probability $\geq 1 - O(1)$,

$$\|\Sigma_R G H^T \Sigma_S^T\|_q \gtrsim \sqrt{r} E_q(R,S). \tag{12}$$

On the other hand, it follows from (10) that with probability $\geq 1 - \exp(-c_1 n)$,

$$\|\Sigma_R G H^T \Sigma_S^T\|_q \leq D^{\frac{1}{2}-\frac{1}{p}} \|GH^T\|_p \lesssim D^{\frac{1}{2}-\frac{1}{p}} n r^{\frac{1}{p}}. \tag{13}$$

Now, let $\mathcal{F}$ be the distribution of an $n \times n$ Gaussian matrix of i.i.d. entries $N(0,1)$ and let F be drawn from $\mathcal{F}$. Then $\|RFS\|_q$ is identically distributed as $\Sigma_R F' \Sigma_S$, where F' is a random $t \times t$ Gaussian matrix of i.i.d. entries $N(0,1)$. It follows from Lemma 7 that with probability $\geq 1 - O(1)$,

$$\|\Sigma_R F' \Sigma_S^T\|_q \lesssim (\log^{\frac{5}{2}} t)(\log\log t) E_q(R,S) \leq (\log^3 t) E_q(R,S) \tag{14}$$

On the other hand, it follows from (10) that with probability $\geq 1 - \exp(-c_2 n)$,

$$\|RFS^T\|_q \geq \|F\|_p \gtrsim n^{1/p}\sqrt{n}. \tag{15}$$

Define events $\mathrm{P}_1(G,H,R,S)$ and $\mathrm{P}_2(F,R,S)$ to be (12) and (14) respectively. Further define

$$\mathrm{P}_3(G,H): \ \left\|GH^T\right\|_p \lesssim n r^{1/p} \qquad \text{and} \qquad \mathrm{P}_4(F): \ \|F\|_p \lesssim n^{1/p}\sqrt{n}.$$

Both $\mathrm{P}_3(G,H)$ and $\mathrm{P}_4(F)$ hold with probability $\geq 1 - e^{-c_3 n}$ when $G, H \sim \mathcal{D}$ and $F \sim \mathcal{F}$.

Now, for $2m$ samples $G_1, \dots, G_m, H_1, \dots, H_m$ independently drawn from $\mathcal{D}$, and m samples $F_1, \dots, F_m$ independently drawn from $\mathcal{F}$, it holds for any fixed R and S that

$$\Pr_{G_i,H_i,F_i}\left\{\exists i,\ \mathrm{P}_1(G_i,H_i,R,S), \mathrm{P}_2(F_i,R,S), \mathrm{P}_3(G_i,H_i), \mathrm{P}_4(F_i) \text{ all hold}\right\} \geq 1 - e^{-c_4 m}. \tag{16}$$

Since $1 \leq \|R e_i e_j^T S^T\|_q = \|R_i\|_2 \|S_j\|_2 \leq D$, we can restrict the matrix R and S to matrices with column norm in $[1, \sqrt{D}]$. Thus we can find a net $\mathcal{M} \subset \bigcup_{t=1}^{t_0} \mathbb{R}^{t \times n}$ of size $\exp(O(t_0 n \ln(Dn/\eta))$ such that for any M with column norms in $[1, \sqrt{D}]$, there exists $M' \in \mathcal{G}$ such that $\|M - M'\|_{op} \leq \eta$.

Now it follows from (16) that

$$\Pr_{G_i,H_i,F_i}\left\{\forall R, S \in \mathcal{M}, \exists i,\ \mathrm{P}_1(G_i,H_i,R,S), \mathrm{P}_2(F_i,R,S), \mathrm{P}_3(G_i,H_i), \mathrm{P}_4(F_i) \text{ all hold}\right\}$$
$$\geq 1 - \exp\left(O\left(t_0 n \ln \frac{Dn}{\eta}\right)\right) \exp\left(-c_4 m\right) > 0,$$

provided that $m = \Theta(n \ln(Dn))$. Fix $\{G_i, H_i, F_i\}_i$ such that for each pair $R', S' \in \mathcal{M}$ there exists i such that $\mathrm{P}_1(G_i, H_i, R', S')$ and $\mathrm{P}_2(F_i, R', S')$ and $\mathrm{P}_3(G_i, H_i)$ and $\mathrm{P}_4(F_i)$ all hold.

Take $T = \{I_n\} \cup \{e_i e_j^T\}_{1 \leq i,j \leq n} \cup \{G_i H_i^T\}_{1 \leq i \leq m} \cup \{F_i\}_{1 \leq i \leq m}$. We know that if (R,S) satisfies (10) for all $A \in T$, then there exists R' and S' such that $\|R' - R\|_{op} \leq \eta$ and

$\|S' - S\|_{op} \leq \eta$, and there exists $1 \leq i \leq m$ such that $\mathrm{P}_1(G_i, H_i, R', S')$ and $\mathrm{P}_2(F_i, R', S')$ and $\mathrm{P}_3(G_i, H_i)$ and $\mathrm{P}_4(F_i)$ all hold. One can then show that (12), (15) hold with slightly smaller constants and (13), (14) with slightly larger constants for R and S. It follows that

$$\frac{n^{\frac{1}{p}}\sqrt{n}}{\log^3 t} \lesssim D^{\frac{1}{p}-\frac{1}{2}}\sqrt{r}nr^{\frac{1}{p}}, \quad \text{or,} \quad \frac{1}{\log^3 t} \lesssim \left(\frac{rD}{n}\right)^{\frac{1}{p}-\frac{1}{2}} = \frac{1}{\rho^{\frac{2}{p}-1}},$$

which contradicts our choice of ρ (the hidden constant in $\lesssim$ above depends only on D_0, θ and η, and then we can choose the hidden constant in the Θ-notation for ρ). ◀

4 Upper Bounds

We shall only show the upper bounds for $1 \leq p \leq q \leq 2$ in this section. Other cases can be found in the full version.

Let $G \in \mathbb{R}^{r\times n}$ ($r \geq Ct$) be a random matrix and $c, c', \eta > 0$ be absolute constants which satisfy the following properties:

(a) (subspace embedding) For a fixed t-dimensional subspace $X \subseteq \mathbb{R}^n$ it holds with probability $\geq 1 - \exp(-c't)$ that

$$(1-\eta)\|x\|_2 \leq \|Gx\|_2 \leq (1+\eta)\|x\|_2, \quad \forall x \in X;$$

(b) For a fixed $A \in \mathbb{R}^{n\times n}$ it holds with probability $\geq 1 - \exp(-c'r)$ that

$$\|GA\|_{op} \leq c\left(\|A\|_{op} + \frac{1}{\sqrt{r}}\|A\|_F\right);$$

(c) For a fixed $A \in \mathbb{R}^{n\times n}$ it holds with probability $\geq 1 - \exp(-c'r)$ that

$$(1-\eta)\|A\|_F \leq \|GA\|_F \leq (1+\eta)\|A\|_F.$$

Consider the singular value decomposition $A = U\Sigma V^T$, where U and V are orthogonal matrices, $\Sigma = \mathrm{diag}\{\sigma_1, \ldots, \sigma_n\}$ with $\sigma_1 \geq \sigma_2 \geq \cdots$. For an index set $I \subseteq [n]$, define $A_I = U\Sigma_I V^T$, where Σ_I is Σ restricted to the diagonal elements with indices inside I (the diagonal entries with indices outside I are replaced with 0).

▶ **Theorem 9.** *Let $1 \leq p \leq q \leq 2$. There exist constants $\theta = \theta(p,q) < 1$ small enough and $C = C(p,q)$ large enough such that for $t \leq \theta n$ and matrix G satisfying the aforementioned properties, it holds for any (fixed) $A \in \mathbb{R}^{n\times n}$ with probabilty $1 - \exp(-c''t)$ that*

$$\frac{t^{\frac{1}{q}-\frac{1}{2}}}{n^{\frac{1}{p}-\frac{1}{2}}\log\frac{n}{t}}\|A\|_p \lesssim \|GA\|_q \lesssim \|A\|_p.$$

Note that for $t = \Omega(\log n)$ and $r = Ct$ for some large constant C, a Gaussian random matrix of i.i.d. entries $N(0, 1/r)$, or a randomized Hadamard Transform matrix of $r = \Theta(t\,\mathrm{polylog}(t))$ rows, satisfies the three conditions on G [7]. The following corollary of Theorem 9 is immediate.

▶ **Corollary 10.** *Suppose that $1 \leq p \leq q$ and $c\log n \leq t \leq \theta n$ for some absolute constants $\theta \in (0,1)$ and $c \geq 1$. There exists (random) $G \in \mathbb{R}^{r\times m}$ with $r \gtrsim t$ such that with probability $\geq 1 - \exp(-c''t)$,*

$$\|A\|_p \leq \|GA\|_q \lesssim \frac{n^{\frac{1}{p}-\frac{1}{2}}}{t^{\frac{1}{q}-\frac{1}{2}}}\left(\log\frac{n}{t}\right)\|A\|_p.$$

In particular when $p = q$,

$$\|A\|_p \le \|GA\|_p \lesssim \left(\frac{n}{t}\right)^{\frac{1}{p}-\frac{1}{2}} \left(\log\frac{n}{t}\right).$$

Now we prove Theorem 9. We first need an auxiliary lemma.

▶ **Lemma 11.** *Let θ, t, C and G be as defined in Theorem 9 and $b = \Theta(\log(n/t))$. At least one of the following conditions will hold:*

$$\sum_{i=1}^{bt} \sigma_i^p \ge \frac{1}{2}\sum_{i=1}^{n}\sigma_i^p \tag{17}$$

and

$$\sigma_s^2 \le \frac{2}{t}\sum_{i=s}^{n}\sigma_i^2 \quad \textit{for some} \quad s \le bt. \tag{18}$$

To prove the preceding lemma, consider the first b blocks of singular values of A each of size t, that is, $I_1 = \{\sigma_1, \dots, \sigma_t\}$, …, $I_b = \{\sigma_{(b-1)t+1}, \dots, \sigma_{bt}\}$.

▶ **Lemma 12.** *If (18) does not hold for any $s \le bt$, it must hold for all $2 \le j \le b$ that $\sigma_{jt} \le \frac{1}{2}\sigma_{(j-1)t}$.*

Proof. If this is not true for some j then $\sum_{i=(j-1)t+1}^{jt}\sigma_i^2 \ge t\sigma_{jt}^2 > \frac{t}{2}\sigma_{(j-1)t}^2$, which contradicts (18) with $s = (j-1)t \le bt$. ◀

Proof of Lemma 11. Suppose that (17) does not hold and we need to show that (18) holds for some $s \le bt$. Otherwise, it follows from Lemma 12 that $\sigma_{bt+1} \le \frac{\sigma_1}{2^b} \le \left(\frac{t}{n}\right)^2\sigma_1$ and thus

$$\sum_{i=bt+1}^{n} \sigma_i^p < n\sigma_{bt+1}^p \le \frac{t^{2p}}{n^{2p-1}}\sigma_1^p \le t\theta^{2p-1}\sigma_1^p, \tag{19}$$

On the other hand,

$$\sum_{i=1}^{bt}\sigma_i^p \ge t\sigma_1^p\left(\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^b}\right) = \left(1-\frac{1}{2^b}\right)t\sigma_1^p = (1-\theta^2)t\sigma_1^p. \tag{20}$$

Using the assumption on θ, we see that the rightmost side of (20) is bigger than the rightmost side of (19), which contradicts the assumption that (17) does not hold. ◀

▶ **Lemma 13.** *Let $1 \le p \le q \le 2$, and t, b and G be defined as in Lemma 11. Suppose that s satisfies (18) and let $J = \{s, s+1, \dots, n\}$. Then*

$$\|GA_J\|_q \gtrsim \frac{t^{\frac{1}{q}-\frac{1}{2}}}{n^{\frac{1}{p}-\frac{1}{2}}}\|A_J\|_p.$$

Proof. Combining Property (b) of G with (18) yields that

$$\|GA_J\|_{op} \le \frac{c}{\sqrt{t}}\left(\sqrt{2}+\sqrt{\frac{1}{C}}\right)\|A_J\|_F =: \frac{K}{\sqrt{t}}\|A_J\|_F$$

On the other hand, Property (c) states that $\|GA_J\|_F \ge \frac{1}{2}\|A_J\|_F$. This implies that at least αr singular values of GA_J are at least $\frac{\gamma}{\sqrt{t}}\|A_J\|_F$, provided that $C\left((1-\alpha)\gamma^2+\alpha K^2\right) < \frac{1}{4}$, which is satisfied if we choose $\gamma \simeq 1/\sqrt{C}$ and $\alpha \simeq 1/K^{2/q}$. It follows that

$$\|GA_J\|_q \ge (\alpha r)^{\frac{1}{q}}\frac{\gamma}{\sqrt{t}}\|A_J\|_F \ge (\alpha C)^{\frac{1}{q}}\gamma\cdot\frac{t^{\frac{1}{q}-\frac{1}{2}}}{n^{\frac{1}{p}-\frac{1}{2}}}\|A_J\|_p.$$ ◀

▶ **Lemma 14.** *Let $1 \le p \le q \le 2$, and t, b and G be defined as in Lemma 11. Suppose that s satisfies* (18) *and let $J = \{s, s+1, \ldots, n\}$. Then*

$$\|GA_J\|_q \lesssim \frac{1}{t^{\frac{1}{p}-\frac{1}{q}}}\|A_J\|_p$$

Proof. When $p \le 2$, it holds that $\|A_J\|_F^2 \le \|A_J\|_p^p \|A_J\|_{op}^{2-p}$. Using (18), we obtain that

$$\|A_J\|_p \ge \frac{\|A_J\|_F^{2/p}}{\|A_J\|_{op}^{2/p-1}} \ge \left(\frac{t}{2}\right)^{\frac{1}{p}-\frac{1}{2}} \|A_J\|_F.$$

On the other hand, it follows from Property (c) of G that

$$\|GA_J\|_q \le r^{\frac{1}{q}-\frac{1}{2}}\|GA_J\|_F \le (1+\eta) r^{\frac{1}{q}-\frac{1}{2}}\|A_J\|_F.$$

Therefore,

$$\|GA_J\|_q \le (1+\eta) r^{\frac{1}{q}-\frac{1}{2}} \left(\frac{2}{t}\right)^{\frac{1}{p}-\frac{1}{2}} \|A_J\|_p = (1+\eta)(Cbqt)^{\frac{1}{q}-\frac{1}{2}} \left(\frac{2}{t}\right)^{\frac{1}{p}-\frac{1}{2}} \|A_J\|_p \lesssim \frac{1}{t^{\frac{1}{p}-\frac{1}{q}}}\|A_J\|_p. \blacktriangleleft$$

Now we are ready to show Theorem 9.

Proof of Theorem 9. It follows from the subspace embedding property of G that

$$(1-\eta)\|A_{I_i}\|_q \le \|GA_{I_i}\|_q \le (1+\eta)\|A_{I_i}\|_q, \quad 1 \le i \le b$$

and thus

$$\frac{1-\eta}{t^{\frac{1}{p}-\frac{1}{q}}}\|A_{I_i}\|_p \le \|GA_{I_i}\|_q \le (1+\eta)\|A_{I_i}\|_p.$$

When (17) holds, there exists i^* ($1 \le i^* \le b$) such that

$$\|A_{I_{i^*}}\|_p \ge \frac{1}{2^{\frac{1}{p}} b}\|A\|_p$$

and thus

$$\frac{1}{bt^{\frac{1}{p}-\frac{1}{q}}}\|A\|_p \lesssim \|GA_{I_{i^*}}\|_q \lesssim \|A\|_p.$$

When (17) does not hold, let J be as defined in Lemma 13 and

$$\frac{1}{2^{\frac{1}{p}}}\|A\|_p \le \|A_J\|_p \le \|A\|_p.$$

The claimed upper and lower bounds follow from combining the bounds above, together with Lemma 13, Lemma 14, and

$$\max\left\{\|GA_{I_1}\|_q, \ldots, \|GA_{I_b}\|_q, \|GA_J\|_q\right\} \le \|GA\|_q \le \sum_{i=1}^{b} \left\|GA_{[I_i]}\right\|_q + \|GA_J\|_q,$$

noticing that $t^{\frac{1}{2}-\frac{1}{p}}/n^{\frac{1}{2}-\frac{1}{p}} \le 1/t^{\frac{1}{p}-\frac{1}{q}}$. ◀

References

1 Alexandr Andoni. Nearest neighbor search in high-dimensional spaces. In *the workshop: Barriers in Computational Complexity II*, 2010. `http://www.mit.edu/~andoni/nns-barriers.pdf`.

2 Alexandr Andoni, Robert Krauthgamer, and Ilya Razenshteyn. Sketching and embedding are equivalent for norms. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 479–488. ACM, 2015.

3 Jaroslaw Blasiok, Vladimir Braverman, Stephen R. Chestnut, Robert Krauthgamer, and Lin F. Yang. Streaming symmetric norms via measure concentration. arXiv:1511.01111 [cs.DS], 2016.

4 Vladimir Braverman, Stephen R. Chestnut, Robert Krauthgamer, and Lin F. Yang. Sketches for matrix norms: Faster, smaller and more general. arXiv:1609.05885 [cs.DS], 2016.

5 Emmanuel Candes and Benjamin Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012.

6 Kenneth L Clarkson and David P Woodruff. Low rank approximation and regression in input sparsity time. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 81–90. ACM, 2013.

7 Michael B. Cohen, Jelani Nelson, and David P. Woodruff. Optimal approximate matrix product in terms of stable rank. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 11:1–11:14, 2016.

8 Aram W Harrow, Ashley Montanaro, and Anthony J Short. Limitations on quantum dimensionality reduction. In *International Colloquium on Automata, Languages, and Programming*, pages 86–97. Springer, 2011.

9 Weihao Kong and Gregory Valiant. Spectrum estimation from samples. arXiv:1602.00061 [cs.LG], 2016.

10 Kasper Green Larsen and Jelani Nelson. The Johnson-Lindenstrauss Lemma Is Optimal for Linear Dimensionality Reduction. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 82:1–82:11, 2016.

11 Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Springer-Verlag, Berlin, 1991.

12 Yi Li, Huy L. Nguyen, and David P. Woodruff. On sketching matrix norms and the top singular vector. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1562–1581, 2014.

13 Yi Li and David P. Woodruff. On approximating functions of the singular values in a stream. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 726–739, 2016.

14 Yi Li and David P. Woodruff. Tight bounds for sketching the operator norm, schatten norms, and subspace embeddings. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2016, September 7-9, 2016, Paris, France*, pages 39:1–39:11, 2016.

15 Françoise Lust-Piquard. Inégalités de khintchine dans c_p $(1 < p < \infty)$. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 303:289–292, 1986.

16 Xiangrui Meng and Michael W. Mahoney. Low-distortion subspace embeddings in input-sparsity time and applications to robust linear regression. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 91–100, 2013.

17 Cameron Musco, Praneeth Netrapalli, Aaron Sidford, Shashanka Ubaru, and David P. Woodruff. Spectral sums beyond fast matrix multiplication: Algorithms and hardness. arXiv:1704.04163 [cs.DS], 2017.

18 Jelani Nelson and Huy L. Nguyen. OSNAP: faster numerical linear algebra algorithms via sparser subspace embeddings. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 117–126, 2013.

19 Shashanka Ubaru, Jie Chen, and Yousef Saad. Fast estimation of $\mathrm{tr}(F(A))$ via stochastic lanczos quadrature. 2016. URL: `http://www-users.cs.umn.edu/~saad/PDF/ys-2016-04.pdf`.

20 Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Yonina C. Eldar and Gitta Kutyniok, editors, *Compressed Sensing: Theory and Practice*, pages 210–268. Cambridge University Press, 2012.

21 Andreas J. Winter. Quantum and classical message identification via quantum channels. *Quantum Information & Computation*, 5(7):605–606, 2005.

22 David P. Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1-2):1–157, 2014.