NORM AND TRACE ESTIMATION WITH RANDOM RANK-ONE VECTORS[*]

ZVONIMIR BUJANOVIC[†] AND DANIEL KRESSNER[‡]

Abstract. A few matrix-vector multiplications with random vectors are often sufficient to obtain reasonably good estimates for the norm of a general matrix or the trace of a symmetric positive semi-definite matrix. Several such probabilistic estimators have been proposed and analyzed for standard Gaussian and Rademacher random vectors. In this work, we consider the use of rank-one random vectors, that is, Kronecker products of (smaller) Gaussian or Rademacher vectors. It is not only cheaper to sample such vectors but it can sometimes also be much cheaper to multiply a matrix with a rank-one vector instead of a general vector. In this work, theoretical and numerical evidence is given that the use of rank-one instead of unstructured random vectors still leads to good estimates. In particular, it is shown that our rank-one estimators multiplied with a modest constant constitute, with high probability, upper bounds of the quantity of interest. Partial results are provided for the case of lower bounds. The application of our techniques to condition number estimation for matrix functions is illustrated.

Key words. norm of a matrix, trace of a matrix, stochastic estimator, rank-one vectors

AMS subject classifications. 65F35, 60E15

DOI. 10.1137/20M1331718

1. Introduction. This work is concerned with estimating the norm of a matrix $A \in \mathbb{R}^{m \times n}$ or the trace of a symmetric positive semidefinite matrix $B \in \mathbb{R}^{n \times n}$, which are given implicitly via matrix-vector products. Given an integer factorization $n = \hat{n} \cdot \tilde{n}$, we say that $x \in \mathbb{R}^n$ has rank one if it takes the form $x = \tilde{x} \otimes \hat{x}$, where $\tilde{x} \in \mathbb{R}^{\tilde{n}}$, $\hat{x} \in \mathbb{R}^{\hat{n}}$ are nonzero and $\otimes$ denotes the Kronecker product. Equivalently, the $\hat{n} \times \tilde{n}$ matrix obtained from reshaping x has rank one. In this work, we specifically target a setting where it is (much) cheaper to multiply with a rank-one vector than with a general vector. Short sums of Kronecker products have this property, and such matrices arise in a variety of applications; see, e.g., [13, 29]. More intricately, condition number estimation for matrix equations and matrix functions involves linear operators that can often be cheaply applied to rank-one vectors; see section 4 for details. Another situation involving such operators arises in the context of error estimates for low-rank tensor approximation [25].

Existing work. Norm estimation is a classical topic in matrix analysis; see Chapter 15 in the book by Higham [17] for a comprehensive overview. A single matrix-vector product Ax with a random vector x chosen from a suitable distribution can already give a good first estimate of the norm of A. A classical result by Dixon [11] shows that the spectral norm $\|A\|_2$ is bounded by $\theta \|Ax\|_2$ with probability at least $1 - 0.8 \sqrt{n} \, \theta^{-1/2}$ if x is distributed uniformly on the unit sphere in $\mathbb{R}^n$. The normalization of x implies

[†]Faculty of Science, Department of Mathematics, University of Zagreb, Zagreb, 10000, Croatia (zbujanov@math.hr).

[‡]Institute of Mathematics, SB-MATHICSE-ANCHP, EPF Lausanne, Lausanne, CH-1015, Switzerland (daniel.kressner@epfl.ch).

that $\|Ax\|_2$ is always a lower bound, but it also leads to the unfavorable appearance of $\sqrt{n}$ in the tail probability. The latter can be avoided when choosing x from $\mathcal{N}(0, I_n)$. This choice of x is called standard Gaussian (or normal) random vector. In this case, Lemma 4.1 in [16] states that

$$(1.1) \qquad \|A\|_2 \leq \theta \|Ax\|_2 \quad \text{holds with probability at least } 1 - \sqrt{2\pi^{-1}}\,\theta^{-1}.$$

Note that the expected value of $\|Ax\|_2^2$ is not $\|A\|_2^2$ but $\|A\|_F^2$, where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. In turn, one expects that $\|Ax\|_2$ tends to overestimate $\|A\|_2$ by a factor $\|A\|_F/\|A\|_2$, which can be between 1 and $\sqrt{n}$ depending on the singular value distribution of A. For the setting in [16], which considers matrices A that can be well approximated by low-rank matrices, this factor is usually modest.

There are numerous approaches to go beyond the simple estimate $\|Ax\|_2$ and construct estimators that improve upon (1.1). A straightforward modification suggested, e.g., in [16] is to choose $k \geq 2$ independent Gaussian vectors $x_1, \ldots, x_k$ and return the maximum estimator $\max\{\|Ax_1\|_2, \ldots, \|Ax_k\|_2\}$. On the one hand, this increases the success probability in (1.1) to $1 - (\sqrt{2\pi^{-1}}\,\theta^{-1})^k$, but, on the other hand, this also increases the risk of overestimation.

A second approach starts with the observation that $\|A\|_F^2 = \text{trace}(B)$, where $B = A^T A$, which suggests the use of the stochastic trace estimator

$$\mathsf{Est}_k = \frac{1}{k} \sum_{i=1}^{k} x_i^T B x_i,$$

going back to Hutchinson [21]. Because Est_k is a subexponential random variable with mean $\text{trace}(B)$, one can apply Chernoff bounds to obtain concentration inequalities that yield relative upper and lower bounds [3, 14, 34]. For example, the proof of Corollary 3.3 in [14] establishes

(1.2)

$\text{trace}(B) \leq \dfrac{\mathsf{Est}_k}{1 - \varepsilon} \quad$ holds with prob. at least $1 - \exp(-k\rho_B \varepsilon^2/4)$ for all $0 < \varepsilon < 1$,

(1.3)

$\text{trace}(B) \geq \dfrac{\mathsf{Est}_k}{1 + \varepsilon} \quad$ holds with prob. at least $1 - \exp(-k\rho_B \varepsilon^2/2)$ for all $0 < \varepsilon$

with the stable rank $\rho_B = \|B\|_F^2/\|B\|_2^2$. One disadvantage of (1.2) compared to (1.1) is that the bound on the failure probability for fixed k does not converge to 0 when loosening the upper bound (that is, when letting $\varepsilon \to 1$). For example, for $\rho_B = 1$, the success probability cannot be larger than $1 - \exp(-k/4) \approx 1 - 0.78^k$. This issue is addressed in [14] by a combination with the techniques from [15].

The entries of a Rademacher vector are independent random variables that are either 1 or -1, both with probability $1/2$. As shown in [34, Theorem 1], each of the bounds (1.2) and (1.3) holds for independent Rademacher vectors $x_1, \ldots, x_k$ with probability at least $1 - \exp(-k(\varepsilon^2/4 - \varepsilon^3/6))$. A more recent result [9, Corollary 13] establishes that both bounds hold jointly with probability at least $1 - 2\exp(-k\rho_B\varepsilon^2(1 + \varepsilon)^{-1}/8)$.

A third approach to improve (1.1) is to apply a few steps of the power or Lanczos method to B with a standard Gaussian starting vector; see, e.g., [11, 19, 27]. We

will not consider such an approach in this work because repeated application of B to the same vector makes it difficult to benefit from rank-one structure of the starting vector.

Finally, the Hutch++ trace estimator, recently proposed and analyzed in [32], is a promising combination of randomized low-rank approximation with stochastic trace estimation. The extension of this idea to the rank-one estimators considered in this work is likely possible, but the analysis is nontrivial and deferred to future work.

New results. In this work, we develop and analyze estimators that operate with vectors $\tilde{x} \otimes \hat{x}$, where $\tilde{x}$ and $\hat{x}$ are independent random vectors. A simple calculation shows that $\|A(\tilde{x} \otimes \hat{x})\|_2^2$ retains the property of being an unbiased estimator of $\|A\|_F^2$ for common choices of distributions for the entries of $\tilde{x}, \hat{x}$; see Lemma 2.1. For the specific case that both $\tilde{x}, \hat{x}$ are standard Gaussian vectors, Theorem 2.2, one of our main results, shows that

$$(1.4) \quad \|A\|_2 \le \theta \|A(\tilde{x} \otimes \hat{x})\|_2 \quad \text{holds with prob. at least } 1 - 2\pi^{-1}(2 + \ln(1 + 2\theta))\theta^{-1}.$$

Compared to (1.1), the success probability only becomes slightly worse. Using the maximum estimator mentioned above, this probability can be easily improved, while the risk of overestimation can be quantified with the lower bounds from Theorems 2.3 and 2.4. We also explain why it is not possible to have a result of the form (1.4) when $\tilde{x}, \hat{x}$ are Rademacher vectors. In contrast, an upper bound of the form (1.2) holds for rank-one standard Gaussian *and* Rademacher vectors; see Theorem 3.1. We also establish lower bounds of the form (1.3) (see Theorems 3.2 and 3.4), but they come with additional unfavorable factors in the exponent.

We would like to stress that the rank-one structure significantly complicates the theory because the techniques used in existing work on norm and trace estimation do not carry over. For example, when $x, \tilde{x}, \hat{x}$ are standard Gaussian vectors, the distribution of x is invariant under orthogonal transformations, but the distribution of $\tilde{x} \otimes \hat{x}$ is not. The random variable $\|Ax\|_2^2$ is subexponential, but this property, which is frequently assumed in concentration inequalities, is not enjoyed by $\|A(\tilde{x} \otimes \hat{x})\|_2^2$.

Random rank-one vectors/measurements have been used and analyzed in compressed sensing, particularly in the context of matrix recovery [7, 8, 23, 28] and phase retrieval [38], as well as in dimensionality reduction [22, 37]. While much of the existing analyses cannot directly be used to address the questions discussed in this paper, a notable exception is recent work by Versyhnin [39]; see section 3.

2. Small-sample estimation. Hutchinson [21] has shown that $\|Ax\|_2^2$ is an unbiased estimator of $\|A\|_F^2$ if x contains independent random variables with mean zero and variance one, which include standard normal and Rademacher random variables. The following lemma extends this result to rank-one vectors.

LEMMA 2.1. *Let $A \in \mathbb{R}^{m \times n}$, $n = \hat{n} \cdot \tilde{n}$, and let $\hat{x}, \tilde{x}$ be random vectors that contain $\hat{n} + \tilde{n}$ independent random variables with mean zero and variance one. Then $\mathbb{E}\left[\|A(\tilde{x} \otimes \hat{x})\|_2^2\right] = \|A\|_F^2$.*

Proof. We let $a_i \in \mathbb{R}^{\hat{n}\tilde{n}}$ denote the ith column of A^T and reshape a_i into the $\hat{n} \times \tilde{n}$ matrix A_i, that is, $\text{vec}(A_i) = a_i$. By standard properties of the Kronecker product [12],

$$(2.1) \qquad \|A(\tilde{x} \otimes \hat{x})\|_2^2 = \sum_{i=1}^m ((\tilde{x} \otimes \hat{x})^T a_i)^2 = \sum_{i=1}^m (\hat{x}^T A_i \tilde{x})^2.$$

Because of the assumptions on $\hat{x}$ we have $\mathbb{E}[\hat{x}\hat{x}^T] = I_{\hat{n}}$ and, therefore,

$$\mathbb{E}\big[(\hat{x}^T A_i \tilde{x})^2\big] = \mathbb{E}\big[\tilde{x}^T A_i^T \hat{x}\hat{x}^T A_i \tilde{x}\big] = \mathbb{E}_{\tilde{x}}\big[\mathbb{E}_{\hat{x}}\big[\tilde{x}^T A_i^T \hat{x}\hat{x}^T A_i \tilde{x} \mid \tilde{x}\big]\big]$$
$$= \mathbb{E}_{\tilde{x}}\big[\tilde{x}^T A_i^T A_i \tilde{x}\big] = \mathbb{E}_{\tilde{x}}\big[\|A_i \tilde{x}\|_2^2\big] = \|A_i\|_F^2.$$

Inserted into (2.1), we obtain $\mathbb{E}[\|A(\tilde{x} \otimes \hat{x})\|_2^2] = \sum \|A_i\|_F^2 = \sum \|a_i\|_2^2 = \|A\|_F^2.$ □

2.1. One-sample estimation: Upper bound for rank-one Gaussian vectors. We proceed with our first main result, which bounds the risk of underestimating the spectral norm when using a single Kronecker product of standard Gaussian random vectors. Note that the derivation of our result relies on an anticoncentration inequality that exploits specific properties of the distribution.

THEOREM 2.2. *Let $A \in \mathbb{R}^{m \times n}$ and $n = \hat{n} \cdot \tilde{n}$. Suppose that $\hat{x} \sim \mathcal{N}(0, I_{\hat{n}})$ and $\tilde{x} \sim \mathcal{N}(0, I_{\tilde{n}})$, and let $\theta > 1$. Then the inequality*

$$\|A\|_2 \leq \theta \|A(\tilde{x} \otimes \hat{x})\|_2 \tag{2.2}$$

holds with probability at least $1 - \frac{2}{\pi}\left(2 + \ln(1 + 2\theta)\right)\theta^{-1}$.

Proof. Let $x = \tilde{x} \otimes \hat{x}$. We will bound the probability that (2.2) fails:

$$\mathbb{P}\left\{\|A\|_2 > \theta\|Ax\|_2\right\} = \mathbb{P}\left\{\|A\|_2^2 > \theta^2\|Ax\|_2^2\right\} = \mathbb{P}\left\{\|A^T A\|_2 > \theta^2 x^T A^T A x\right\}.$$

Consider the spectral decomposition $A^T A = U \Lambda U^T = \lambda_1 u_1 u_1^T + \cdots + \lambda_n u_n u_n^T$, $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$, and note that

$$\theta^2 x^T A^T A x = \theta^2 \sum_{j=1}^n \lambda_j (x^T u_j)^2 \geq \theta^2 \lambda_1 (x^T u_1)^2.$$

Therefore, if $\|A^T A\|_2 > \theta^2 x^T A^T A x$, then $\|A^T A\|_2 > \theta^2 \lambda_1 (x^T u_1)^2$, and, since $\|A^T A\|_2 = \lambda_1$, we have

$$\mathbb{P}\left\{\|A^T A\|_2 > \theta^2 x^T A^T A x\right\} \leq \mathbb{P}\left\{\lambda_1 > \theta^2 \lambda_1 (x^T u_1)^2\right\} = \mathbb{P}\left\{(x^T u_1)^2 < \frac{1}{\theta^2}\right\}. \tag{2.3}$$

To bound the last probability, we need to control the random variable $X := x^T u_1$. For this purpose, reshape the vector $u_1 \in \mathbb{R}^{\hat{n}\tilde{n}}$ as a matrix: $u_1 = \text{vec}(U_1)$ with $U_1 \in \mathbb{R}^{\hat{n} \times \tilde{n}}$, and consider its singular value decomposition

$$U_1 = \hat{V}\Sigma\tilde{V}^T, \ \hat{V} = \big[\hat{v}_1, \ldots, \hat{v}_r\big], \ \tilde{V} = \big[\tilde{v}_1, \ldots, \tilde{v}_r\big], \ \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$$

with $r = \min\{\hat{n}, \tilde{n}\}$ and $\sum_{i=1}^r \sigma_i^2 = \|U_1\|_F^2 = \|u_1\|_2^2 = 1$. Using $x = \tilde{x} \otimes \hat{x} = \text{vec}(\hat{x}\tilde{x}^T)$, we obtain

$$x^T u_1 = \text{vec}(\hat{x}\tilde{x}^T)^T \text{vec}(U_1) = \text{trace}(\tilde{x}\hat{x}^T U_1) = \hat{x}^T U_1 \tilde{x} = \sum_{i=1}^r \sigma_i \hat{x}_i \tilde{x}_i,$$

where $\hat{x}_i := \hat{v}_i^T \hat{x}$ and $\tilde{x}_i := \tilde{v}_i^T \tilde{x}$ are independent standard normal random variables, because $\hat{V}^T \hat{x} \sim \mathcal{N}(0, I_r)$ and $\tilde{V}^T \tilde{x} \sim \mathcal{N}(0, I_r)$ follow from the orthogonality of $\hat{V}$ and $\tilde{V}$. For a random variable Z, let F_Z denote its cumulative distribution function, and let $\varphi_Z(t)$ denote its characteristic function. The characteristic function for the

product of two independent standard normal random variables is given by $1/\sqrt{1+t^2}$, and hence

$$\varphi_{\sigma_i \hat{x}_i \tilde{x}_i}(t) = \frac{1}{\sqrt{1+\sigma_i^2 t^2}}.$$

Since $X = \sum_{i=1}^r \sigma_i \hat{x}_i \tilde{x}_i$ is a sum of independent random variables, its characteristic function is given by

$$\varphi_X(t) = \prod_{i=1}^r \frac{1}{\sqrt{1+\sigma_i^2 t^2}}.$$

We now apply Lévy's theorem (see, e.g., [35, Corollary 2]) to reformulate (2.3) in terms of characteristic functions:

$$\mathbb{P}\left\{(x^T u_1)^2 < \frac{1}{\theta^2}\right\} = \mathbb{P}\left\{\frac{-1}{\theta} < x^T u_1 < \frac{1}{\theta}\right\} = F_X\left(\frac{1}{\theta}\right) - F_X\left(\frac{-1}{\theta}\right)$$

$$(2.4) \qquad = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin(t/\theta)}{t}\varphi_X(t)\,\mathrm{d}t = \frac{2}{\pi}\int_0^{\infty} \frac{\sin(t/\theta)}{t}\prod_{i=1}^r \frac{1}{\sqrt{1+\sigma_i^2 t^2}}\,\mathrm{d}t.$$

To bound this oscillatory integral, we use $\sin(t/\theta) \le t/\theta$ for $t \in [0,\theta]$ and $\sin(t/\theta) \le 1$ elsewhere. Also, using $\sum_{i=1}^r \sigma_i^2 = 1$, note that

$$\prod_{i=1}^r \frac{1}{\sqrt{1+\sigma_i^2 t^2}} = \frac{1}{\sqrt{1+\sum_{i=1}^r \sigma_i^2 t^2 + \text{positive terms}}} \le \frac{1}{\sqrt{1+t^2}}.$$

This gives

$$\int_0^{\infty} \frac{\sin(t/\theta)}{t}\prod_{i=1}^r \frac{1}{\sqrt{1+\sigma_i^2 t^2}}\,\mathrm{d}t$$

$$= \int_0^{\theta} \frac{\sin(t/\theta)}{t}\prod_{i=1}^r \frac{1}{\sqrt{1+\sigma_i^2 t^2}}\,\mathrm{d}t + \int_{\theta}^{\infty} \frac{\sin(t/\theta)}{t}\prod_{i=1}^r \frac{1}{\sqrt{1+\sigma_i^2 t^2}}\,\mathrm{d}t$$

$$\le \int_0^{\theta} \frac{t/\theta}{t}\prod_{i=1}^r \frac{1}{\sqrt{1+\sigma_i^2 t^2}}\,\mathrm{d}t + \int_{\theta}^{\infty} \frac{1}{t}\prod_{i=1}^r \frac{1}{\sqrt{1+\sigma_i^2 t^2}}\,\mathrm{d}t$$

$$\le \int_0^{\theta} \frac{1}{\theta}\frac{1}{\sqrt{1+t^2}}\,\mathrm{d}t + \int_{\theta}^{\infty} \frac{1}{t}\frac{1}{\sqrt{1+t^2}}\,\mathrm{d}t$$

$$= \theta^{-1}\ln(\theta + \sqrt{1+\theta^2}) + \ln\left(\theta^{-1} + \sqrt{1+\theta^{-2}}\right)$$

$$\le \theta^{-1}\ln(1+2\theta) + \ln\left(1+2\theta^{-1}\right) \le \theta^{-1}\ln(1+2\theta) + 2\theta^{-1}$$

$$= \theta^{-1}\left(2 + \ln(1+2\theta)\right),$$

which, when inserted into (2.4), implies the claim of the theorem. $\qquad\square$

Comparing the bounds (2.2) and (1.1), the use of rank-one Gaussian vectors instead of Gaussian vectors comes with a slight penalty. The following table shows the bounds for the failure probability $\mathcal{P} := \mathbb{P}\{\|A\|_2 > \theta\|Ax\|_2\}$ for different values of θ.

θ	Gaussian vectors: $\mathcal{P} \leq \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\theta}$	Rank-one Gaussian vectors: $\mathcal{P} \leq \frac{2}{\pi} \cdot \frac{1}{\theta} \cdot (2 + \ln(1 + 2\theta))$
5	0.159577	0.559957
10	0.079788	0.321144
20	0.039894	0.181869
30	0.026596	0.129677
50	0.015957	0.084226

Note that taking larger values of θ, such as 30 or 50, still makes sense for applications where one is interested in obtaining only a rough estimate of the matrix norm, such as the one described in section 4.2.

Quite trivially, Theorem 2.2 can also be applied to the Frobenius norm $\|A\|_F$. Multiplying both sides of (2.2) by $\sqrt{\rho_A} = \|A\|_F / \|A\|_2$, it follows that

$$(2.5) \qquad \mathbb{P}\left\{ \|A\|_F \leq \theta \sqrt{\rho_A} \cdot \|A(\tilde{x} \otimes \hat{x})\|_2 \right\} \geq 1 - \frac{2}{\pi} \left(2 + \ln(1 + 2\theta) \right) \theta^{-1}.$$

Given that $1 \leq \rho_A \leq n$, this bound suggests that, as the stable rank of A increases, $\|A(\tilde{x} \otimes \hat{x})\|_2$ needs to be multiplied with a larger constant in order to remain a reliable upper bound for $\|A\|_F$. However, this seems to be entirely an artifact of our derivation of the bound. The numerical experiments in section 2.5 below demonstrate that (2.5) tends to be tight when the stable rank is small but is overly pessimistic when the stable rank is large.

2.2. One-sample estimation: Lower bounds for rank-one Gaussian vectors. As discussed in the introduction, the result of Lemma 2.1 indicates that $\|A(\tilde{x} \otimes \hat{x})\|_2$ tends to overestimate $\|A\|_2$. The following result guarantees that an overestimation by a factor much larger than $\sqrt{n}$ is unlikely.

THEOREM 2.3. *Let* $A \in \mathbb{R}^{m \times n}$ *and* $n = \hat{n} \cdot \tilde{n}$ *with* $\tilde{n} \leq \hat{n}$. *Suppose that* $\hat{x} \sim \mathcal{N}(0, I_{\hat{n}})$ *and* $\tilde{x} \sim \mathcal{N}(0, I_{\tilde{n}})$, *and let* $\theta > 1$. *The inequality*

$$(2.6) \qquad \|A\|_2 \geq n^{-1/2}\theta^{-1}\|A(\tilde{x} \otimes \hat{x})\|_2$$

holds with probability at least $1 - 2e^{-\tilde{n}(\theta - \ln\theta - 1)/2}$.

Proof. Proceeding with the spectral decomposition of $A^T A$ as in the proof of Theorem 2.2, it follows that the probability of (2.6) failing for $x = \tilde{x} \otimes \hat{x}$ is

$$(2.7) \quad \mathbb{P}\left\{ \|Ax\|_2 > \sqrt{n}\theta \, \|A\|_2 \right\} = \mathbb{P}\left\{ \sum_{j=1}^{n} \lambda_j (x^T u_j)^2 > \lambda_1 n\theta^2 \right\} \leq \mathbb{P}\left\{ \|x\|_2^2 > n\theta^2 \right\},$$

where the last inequality uses $\sum_{j=1}^{n} \lambda_j (x^T u_j)^2 \leq \lambda_1 \sum_{j=1}^{n} (x^T u_j)^2 = \lambda_1 \|x\|_2^2$. Because $\|x\|_2 = \|\tilde{x}\|_2 \|\hat{x}\|_2$ and $n = \hat{n} \cdot \tilde{n}$, it follows that $\|x\|_2^2 > n\theta^2$ is only possible if $\|\tilde{x}\|_2^2 > \tilde{n}\theta$ or $\|\hat{x}\|_2^2 > \hat{n}\theta$. Thus,

$$\mathbb{P}\left\{ \|x\|_2^2 > n\theta^2 \right\} \leq \mathbb{P}\left\{ \|\tilde{x}\|_2^2 > \tilde{n}\theta \right\} + \mathbb{P}\left\{ \|\hat{x}\|_2^2 > \hat{n}\theta \right\} \leq 2 \cdot \mathbb{P}\left\{ \|\tilde{x}\|_2^2 > \tilde{n}\theta \right\}.$$

Now, $\|\tilde{x}\|_2^2$ is a chi-square random variable with $\tilde{n}$, and its properties (see Lemma A.1 in the appendix) imply

$$\mathbb{P}\left\{ \|\tilde{x}\|_2^2 > \tilde{n}\theta \right\} \leq (\theta e^{1-\theta})^{\tilde{n}/2},$$

which concludes the proof. $\qquad\square$

As already discussed for the upper bound, one directly obtains a corresponding result for the Frobenius norm by multiplying (2.6) with $\sqrt{\rho_A}$:

$$(2.8) \qquad \mathbb{P}\left\{\|A\|_F \geq \sqrt{\rho_A} n^{-1/2} \theta^{-1} \|A(\tilde{x} \otimes \hat{x})\|_2\right\} \geq 1 - 2e^{-\tilde{n}(\theta - \ln\theta - 1)/2}.$$

Another bound for the Frobenius norm estimates is obtained by using a different approach.

THEOREM 2.4. *Let $A \in \mathbb{R}^{m \times n}$ and $n = \hat{n} \cdot \tilde{n}$. Suppose that $\hat{x} \sim \mathcal{N}(0, I_{\hat{n}})$ and $\tilde{x} \sim \mathcal{N}(0, I_{\tilde{n}})$, and let $\theta > 2$. The inequality*

$$\|A\|_F \geq \theta^{-1} \|A(\tilde{x} \otimes \hat{x})\|_2$$

holds with probability at least $1 - \sqrt{2\theta}\, e^{-\theta + 2}$.

Proof. Using, once again, the spectral decomposition of $A^T A$ as in the proof of Theorem 2.2 and denoting $x = \tilde{x} \otimes \hat{x}$, the failure probability equals

$$\mathbb{P}\left\{\|Ax\|_2 > \theta\|A\|_F\right\} = \mathbb{P}\left\{x^T A^T A x > \theta^2 \cdot \mathrm{trace}(A^T A)\right\}$$
$$= \mathbb{P}\left\{\sum_{i=1}^{n} \lambda_i (x^T u_i)^2 > \theta^2 \sum_{i=1}^{n} \lambda_i\right\} = \mathbb{P}\left\{\sum_{i=1}^{n} \mu_i (x^T u_i)^2 > \theta^2\right\},$$

where $\mu_i := \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} \in [0, 1]$, which satisfies $\sum_{i=1}^{n} \mu_i = 1$. Exploiting that the function $f(\xi) = e^{\sqrt{1+\xi}}$ is convex on $[0, +\infty\rangle$, Jensen's inequality gives

$$f\left(\sum_{i=1}^{n} \mu_i \cdot t^2 (x^T u_i)^2\right) \leq \sum_{i=1}^{n} \mu_i f\left(t^2 (x^T u_i)^2\right) = \sum_{i=1}^{n} \mu_i e^{\sqrt{1 + t^2 (x^T u_i)^2}}$$
$$\leq \sum_{i=1}^{n} \mu_i e^{1 + t|x^T u_i|} = e \sum_{i=1}^{n} \mu_i e^{t|x^T u_i|}$$

for all $t > 0$. Combined with the monotonicity of f as well as Markov's inequality, we obtain

$$\mathbb{P}\left\{\sum_{i=1}^{n} \mu_i (x^T u_i)^2 > \theta^2\right\} = \mathbb{P}\left\{f\left(\sum_{i=1}^{n} \mu_i \cdot t^2 (x^T u_i)^2\right) > f(t^2 \theta^2)\right\}$$
$$\leq \mathbb{P}\left\{e \sum_{i=1}^{n} \mu_i e^{t|x^T u_i|} > e^{\sqrt{1 + t^2 \theta^2}}\right\}$$
$$\leq \mathbb{E}\left[\sum_{i=1}^{n} \mu_i e^{t|x^T u_i|}\right] e^{1 - \sqrt{1 + t^2 \theta^2}}$$
$$(2.9) \qquad = \left(\sum_{i=1}^{n} \mu_i \mathbb{E}\left[e^{t|x^T u_i|}\right]\right) e^{1 - \sqrt{1 + t^2 \theta^2}}.$$

Since $e^{|a|} \leq e^a + e^{-a}$, we obtain for $0 < t < 1$ that

$$\mathbb{E}\left[e^{t|x^T u_i|}\right] \leq \mathbb{E}\left[e^{t(x^T u_i)}\right] + \mathbb{E}\left[e^{t(x^T \cdot (-u_i))}\right] \leq 2\frac{1}{\sqrt{1 - t^2}},$$

where we used Corollary A.3 in the last step. Plugged into (2.9), it follows that

$$\mathbb{P}\left\{\sum_{i=1}^{n}\mu_i(x^Tu_i)^2 > \theta^2\right\} \leq 2e\frac{e^{-\sqrt{1+t^2\theta^2}}}{\sqrt{1-t^2}} \leq \sqrt{2\theta}\,e^{-\theta+2},$$

where the last inequality follows from setting $t = \sqrt{1-2/\theta}$.   □

Example 2.5. To illustrate the theoretical results presented above, we generate matrices with different singular value decompositions and compare their norms with randomized estimates. More precisely, we consider the following seven $n \times n$ matrices:
  • rank-one matrices $A_1 = U_1(e_1e_1^T)$, $A_2 = U_2(e_1e_1^T)V_2^T$, where $e_1 = [1, 0, \ldots, 0]^T$;
  • matrices $A_3 = U_3D$, $A_4 = U_4DV_4^T$, where D is diagonal with $D_{ii} = e^{-i/2}$;
  • the matrix $A_5 = U_5D$, where D is diagonal with $D_{ii} = i^2$;
  • the matrix A_6 is a random orthogonal matrix;
  • the matrix A_7 is a random Gaussian matrix.
Here, U_i, V_i, and A_6 are chosen randomly from the uniform distribution on orthogonal matrices.

For each of these matrices, we sample $100\,000$ vectors $x = \tilde{x} \otimes \hat{x}$ for $\tilde{x}, \hat{x} \sim \mathcal{N}(0, I)$ in order to estimate the probability that one of the inequalities $\|A_i\| \leq \tau\|A_ix\|_2$ or $\|A_i\| \geq \tau^{-1}\|A_ix\|_2$ fails, with $\tau \in [1, 100]$. The obtained results are shown in Figure 1. We have chosen $n = 16$ and $n = 196$ for $\|\cdot\| \equiv \|\cdot\|_2$ and $\|\cdot\| \equiv \|\cdot\|_F$, respectively.

Figure 1(a) displays failure probabilities for the upper bound $\|A_i\|_2 \leq \tau\|A_ix\|_2$ as well as the corresponding bound by Theorem 2.2, with $\tau = \theta$. The bound happens to be quite tight for the rank-one matrices A_1 and A_2. Although the bound is significantly less tight for the other matrices, for which the spectral and Frobenius norms are different, this demonstrates that the result of Theorem 2.3 cannot be improved significantly without taking additional properties of the matrix into account. This panel also highlights the impact of structure in the singular vectors. For example, the right singular vectors of A_3 all have rank one (when reshaped as matrices). While the matrix A_4 has the same singular values, its singular vectors are not of low (numerical) rank. The estimator is observed to perform significantly better for A_4 than for A_3. An analogous observation holds for A_1 and A_2.

Figure 1(b) displays failure probabilities for the lower bound $\|A_i\|_2 \geq \tau^{-1}\|A_ix\|_2$ as well as the corresponding bound by Theorem 2.3, with $\tau = \sqrt{n}\,\theta$. Clearly, the bound is not sharp, but it correctly captures the exponential decay of the probabilities with respect to θ.

Figure 1(c) displays failure probabilities for the upper bound $\|A_i\|_F \leq \tau\|A_ix\|_2$ as well as the corresponding bound from (2.5), with $\tau = \sqrt{\rho_A}\,\theta$. As the bound depends on the stable rank, the dashed lines shown in the panel differ for matrices with different stable rank. Solid and dashed lines of the same color and mark belong to the same matrix A_i. Again, the bounds are quite tight for low (stable) rank but become increasingly loose as the stable rank increases. In fact, the bound increases with larger stable rank while estimated failure probabilities actually *decrease*.

Figure 1(d) displays failure probabilities for the upper bound $\|A_i\|_F \leq \tau^{-1}\|A_ix\|_2$ as well as the corresponding bound (2.8), with $\tau = \sqrt{n/\rho_A}\,\theta$. Again, this bound depends on the stable rank. Although far off for matrices of low (stable) rank, it correctly captures the observation that the exponential decay becomes faster as the stable rank ρ_A increases. In contrast, the bound from Theorem 2.4 does not depend on ρ_A and is clearly preferable when ρ_A is small or no estimate of ρ_A is available.
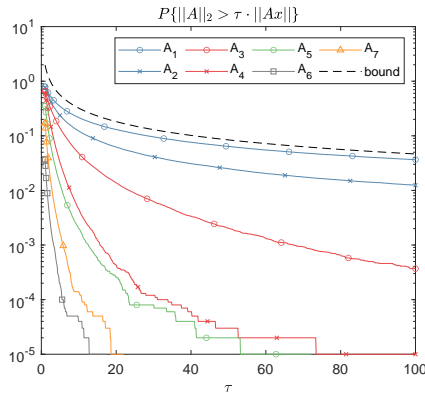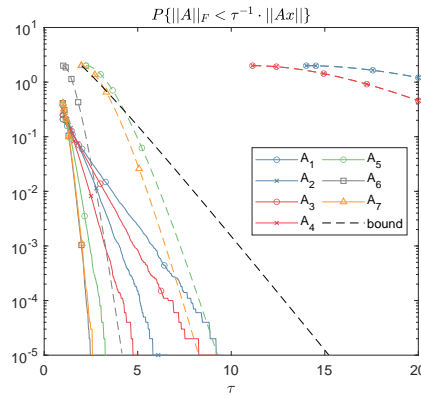
(a) Theorem 2.2 (dashed line) vs. estimated failure probabilities for $\|A_i\|_2 > \tau\|A_i x\|_2$ (solid line).
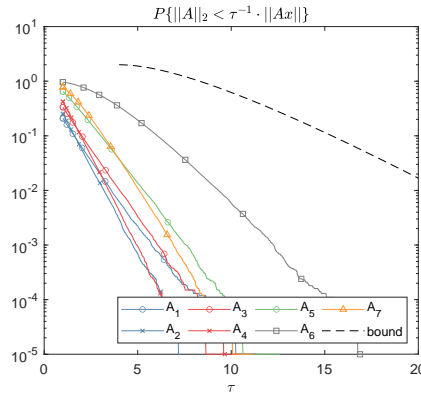
(b) Theorem 2.3 (dashed line) vs. estimated failure probabilities for $\|A_i\|_2 < \tau^{-1}\|A_i x\|_2$ (solid line).
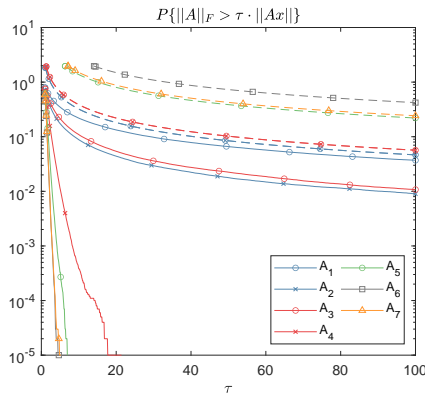
(c) Bound (2.5) (dashed lines) vs. estimated failure probabilities for $\|A_i\|_F > \tau\|A_i x\|_2$ (solid lines).

(d) Theorem 2.4 (dashed black line) versus estimated failure probabilities for $\|A_i\|_F < \tau^{-1}\|A_i x\|_2$. Upper bounds (2.8) are also shown (dashed colored lines).

Fig. 1. *Theoretical bounds on failure probabilities versus estimates; see Example* 2.5 *for details.*

2.3. Small-sample estimation: Rank-one Gaussian vectors. As discussed in the introduction, a simple way to reduce the failure probability of upper bounds is to use a maximum estimator. In our setting, this translates into

$$\mathsf{Max}_k := \max_{j=1,\dots,k} \|A(\tilde{x}_j \otimes \hat{x}_j)\|_2$$

with independent standard Gaussian vectors $\tilde{x}_j$, $\hat{x}_j$ for $j = 1, \dots, k$.

Using Theorem 2.2 one obtains

$$(2.10) \qquad \mathbb{P}\left\{\|A\|_2 \leq \theta \cdot \mathsf{Max}_k\right\} \geq 1 - \left(\frac{2}{\pi}\left(2 + \ln(1 + 2\theta)\right)\theta^{-1}\right)^k.$$

For example, for $\theta = 10$, choosing $k = 7$ is sufficient to guarantee a success probability of more than 99.9%.

Of course, taking the maximum for several samples increases the risk of overestimation. But this increase can be easily mitigated by accepting a slight increase of

the overestimation factor. By Theorem 2.4, the probability that $\|A(\tilde{x} \otimes \hat{x})\|_2$ overestimates $\|A\|_F$ by more than a factor 10 is less than 0.15%. For the maximum estimator one has (again by Theorem 2.4) that

$$\mathbb{P}\left\{\|A\|_F \geq \theta^{-1} \cdot \mathsf{Max}_k\right\} \geq \left(1 - \sqrt{2\theta}\, e^{-\theta+2}\right)^k.$$

For $k = 7$, the probability that Max_k overestimates $\|A\|_F$ by more than a factor 12.1 is again less than 0.15%.

2.4. One-sample estimation: Rank-one Rademacher vectors. In this section, we discuss whether the results from sections 2.1 and 2.2 can be extended when we choose $\tilde{x}$ and $\hat{x}$ to be Rademacher instead of standard Gaussian vectors in the rank-one vector $x = \tilde{x} \otimes \hat{x}$.

It turns out that it is not possible to have an upper bound of the form presented in Theorem 2.2 for Rademacher vectors without further assumptions on A. To see this, consider the matrix $A = uu^T \in \mathbb{R}^{n \times n}$ with $u = e/\sqrt{n}$, where e denotes the vector of all ones. Then $A^T A = A$ and

$$\begin{aligned}
\mathbb{P}\left\{\|A\|_2 > \theta\|Ax\|_2\right\} &= \mathbb{P}\left\{\|A^T A\|_2 > \theta^2 x^T A^T Ax\right\} \\
&= \mathbb{P}\left\{(x^T u)^2 < \theta^{-2}\right\} = \mathbb{P}\left\{-\theta^{-1} < x^T u < \theta^{-1}\right\}.
\end{aligned}$$

Now $x^T u = (\tilde{x} \otimes \hat{x})^T u = n^{-1/2}(\hat{x}^T e)(\tilde{x}^T e)$, which implies

$$\mathbb{P}\left\{x^T u = 0\right\} = \mathbb{P}\left\{(\hat{x}^T e)(\tilde{x}^T e) = 0\right\} \geq \mathbb{P}\left\{\hat{x}^T e = 0\right\}.$$

For even $\hat{n}$, the sum $\hat{x}^T e = \hat{x}_1 + \cdots + \hat{x}_{\hat{n}}$ equals zero when exactly $\hat{n}/2$ of the entries are equal to 1. As this happens with probability $\binom{\hat{n}}{\hat{n}/2}/2^{\hat{n}}$, we obtain the lower bound

$$\mathbb{P}\left\{\|A\|_2 > \theta\|Ax\|_2\right\} \geq \mathbb{P}\left\{x^T u = 0\right\} \geq 2^{-\hat{n}}\binom{\hat{n}}{\hat{n}/2} \approx \frac{1}{\sqrt{\hat{n}\pi/2}}$$

for any $\theta > 1$. In particular, and in contrast to the result of Theorem 2.2, the failure probability does not converge to zero as θ increases. Let us stress that this is not an artifact of using rank-one vectors; an analogous negative result can be obtained when using an (unstructured) Rademacher vector x.

On the other hand, because Rademacher vectors are bounded, the risk of overestimation becomes zero beyond a certain threshold. More specifically, it always holds that $\|\hat{x}\|_2 = \sqrt{\hat{n}}$, $\|\tilde{x}\|_2 = \sqrt{\tilde{n}}$, and hence, for all matrices A, by (2.7) we have

$$\mathbb{P}\left\{\|Ax\|_2 > \sqrt{n}\|A\|_2\right\} \leq \mathbb{P}\left\{\|x\|_2 > \sqrt{n}\right\} = \mathbb{P}\left\{\|\hat{x}\|_2 \cdot \|\tilde{x}\|_2 > \sqrt{n}\right\} = 0.$$

This is in contrast to the result of Theorem 2.3, which yields a small but nonzero risk for Gaussian vectors.

3. Large-sample estimation of the trace. In this section, we analyze the use of random rank-one vectors in stochastic trace estimators for estimating $\text{trace}(B) = \|A\|_F^2$ with $B = A^T A$:

$$\text{(3.1)} \qquad \mathsf{Est}_k = \frac{1}{k}\sum_{i=1}^{k}(\tilde{x}_i \otimes \hat{x}_i)^T B(\tilde{x}_i \otimes \hat{x}_i).$$

By Lemma 2.1, we have $\mathbb{E}[\mathsf{Est}_k] = \text{trace}(B)$. The following theorem shows that Est_k times a modest factor is an upper bound of $\text{trace}(B)$ with high probability for larger k.

THEOREM 3.1. *Let $0 < \varepsilon < 1$, and consider the trace estimator Est_k defined in (3.1), where $\tilde{x}_i$, $\hat{x}_i$ are either independent standard Gaussian or independent Rademacher random vectors. Then the bound*

$$\mathrm{trace}(B) \leq \frac{1}{1-\varepsilon}\mathsf{Est}_k$$

holds with probability at least $1 - \exp(-k\varepsilon^2/18)$.

Proof. The proof is divided into two parts. First, we follow the arguments from [34] and use a Chernoff bound. We then arrive at the problem of bounding the moment generating function of decoupled Gaussian/Rademacher chaos, which will be discussed in the second part.

Consider the spectral decomposition $B = U\Lambda U^T$ with U orthogonal and Λ diagonal containing the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ on the diagonal. Letting u_j denote the jth column of U and $x_i = \tilde{x}_i \otimes \hat{x}_i$, we obtain

$$\mathsf{Est}_k = \frac{1}{k}\sum_{i=1}^{k} x_i^T U\Lambda U^T x_i = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n} \lambda_j (x_i^T u_j)^2 = \frac{1}{k}\sum_{j=1}^{n} \lambda_j \sum_{i=1}^{k} z_{ij}^2,$$

where we set $z_{ij} := x_i^T u_j$. The statement of the theorem is equivalent to showing that $\exp(-k\varepsilon^2/18)$ is an upper bound for

$$P := \mathbb{P}\left\{\mathsf{Est}_k < (1-\varepsilon)\,\mathrm{trace}(B)\right\} = \mathbb{P}\left\{\sum_{j=1}^{n} \frac{\lambda_j}{\mathrm{trace}(B)}\sum_{i=1}^{k} z_{ij}^2 < k(1-\varepsilon)\right\}.$$

By the Chernoff bound, it holds for arbitrary $t > 0$ that

$$P \leq \exp(tk(1-\varepsilon))\mathbb{E}\left[\exp\left(-\sum_{j=1}^{n}\frac{\lambda_j}{\mathrm{trace}(B)}\sum_{i=1}^{k} tz_{ij}^2\right)\right]$$

$$\leq \exp(tk(1-\varepsilon))\sum_{j=1}^{n}\frac{\lambda_j}{\mathrm{trace}(B)}\mathbb{E}\left[\exp\left(-\sum_{i=1}^{k} tz_{ij}^2\right)\right]$$

$$= \exp(tk(1-\varepsilon))\sum_{j=1}^{n}\frac{\lambda_j}{\mathrm{trace}(B)}\prod_{i=1}^{k}\mathbb{E}\left[\exp\left(-tz_{ij}^2\right)\right],$$

where we used Jensen's inequality and the convexity of the exponential in the second inequality, and the independence of z_{ij} for different i in the equality.

It remains to bound the moment generating function $\mathbb{E}\left[\exp\left(-tz_{ij}^2\right)\right]$. To simplify the notation we let $Q \in \mathbb{R}^{\hat{n}\times\tilde{n}}$ denote the matrix such that $\mathrm{vec}(Q) = u_j$ and drop indices. Then $z = \hat{x}^T Q\tilde{x}$, a random variable that is sometimes called decoupled (order-2) chaos. Using that $\exp(-t\alpha^2) \leq 1 - t\alpha^2 + \frac{1}{2}t^2\alpha^4$ holds for any fixed $\alpha \in \mathbb{R}$, we obtain

$$\mathbb{E}\left[\exp(-tz^2)\right] \leq 1 - t\mathbb{E}[z^2] + \frac{t^2}{2}\mathbb{E}[z^4].$$

For both Rademacher and Gaussian random vectors $\hat{x}$ and $\tilde{x}$, we have that

$$\mathbb{E}[z^2] = \|Q\|_F^2, \quad \mathbb{E}[z^4] \leq 9\|Q\|_F^4.$$

For the Rademacher case, this follows from Khintchine inequalities; see, e.g., [33, section 6.8]. For the Gaussian case, see Lemma A.2 in the appendix. Because of $\|u_j\|_2 = 1$, the matrix Q has Frobenius norm 1, and we thus arrive at

$$P \leq \exp(tk(1-\varepsilon))\Big(1 - t + \frac{9}{2}t^2\Big)^k.$$

Taking the logarithm and applying Taylor expansion, we obtain

$$\frac{1}{k}\log P \leq t(1-\varepsilon) + \log\Big(1 - t + \frac{9}{2}t^2\Big) \leq t(1-\varepsilon) + \Big(-t + \frac{9}{2}t^2\Big).$$

For $t = \varepsilon/9$, the right-hand side equals $-\varepsilon^2/18$, which concludes the proof. □

When $\tilde{n} = 1$ or $\hat{n} = 1$ and $\tilde{x}_i$, $\hat{x}_i$ are Rademacher vectors, the Kronecker product $\tilde{x}_i \otimes \hat{x}_i$ is again a Rademacher vector, and Theorem 3.1 becomes a result on the standard stochastic trace estimator. Compared to the bound $\exp(-k(\varepsilon^2/4 - \varepsilon^3/6))$ on the failure probability established in [34] for this case, the bound of Theorem 3.1 is only modestly worse. Compared to the bound from [9], it can become significantly worse when the stable rank of A is known a priori to be large.

Deriving a lower bound estimate for Est_k turns out to be more difficult. The following theorem only considers the case of Rademacher vectors; it is shown that $n^{-1} \cdot \mathsf{Est}_k$ times a modest factor is a lower bound of $\mathrm{trace}(B)$ with high probability.

THEOREM 3.2. *Let $\varepsilon > 0$, and consider the trace estimator Est_k defined in* (3.1) *for independent Rademacher random vectors $\tilde{x}_i$, $\hat{x}_i$. Then the bound*

$$\mathrm{trace}(B) \geq \frac{1}{1+\varepsilon}\mathsf{Est}_k$$

holds with probability at least $1 - \exp(-k\varepsilon/(n-1))$, provided that $n - 1 \geq 48 \cdot \varepsilon^{-1}$.

Proof. Along the lines of the first part of the proof of Theorem 3.1, it can be shown that the statement of the theorem is equivalent to showing that $\exp(-k\varepsilon)$ is an upper bound on

$$\mathbb{P}\bigg\{\sum_{j=1}^n \frac{\lambda_j}{\mathrm{trace}(B)}\sum_{i=1}^k z_{ij}^2 > k(\varepsilon+1)\bigg\} = \mathbb{P}\bigg\{\sum_{i=1}^k X_i > k\varepsilon\bigg\}.$$

Here,

$$X_i = \sum_{j=1}^n \frac{\lambda_j}{\mathrm{trace}(B)}z_{ij}^2 - 1$$

are zero-mean independent random variables that are bounded; using

$$|z_{ij}| \leq |\hat{x}_i|^T|U_j||\tilde{x}_i| \leq \sqrt{n}$$

with $\mathrm{vec}(U_j) = u_j$, one obtains $|X_i| \leq n - 1$. Using that $\mathbb{E}[z_{ij}^2] = 1$ and $\mathbb{E}[z_{ij}^4] = 9$, it can be shown that $\mathbb{E}[X_i^2] \leq 8$. Plugging these bounds into Bernstein's inequality completes the proof:

$$\mathbb{P}\bigg\{\sum_{i=1}^k X_i > k\varepsilon\bigg\} \leq \exp\bigg[-\frac{k^2\varepsilon^2/2}{8k + (n-1)k\varepsilon/3}\bigg] \leq \exp(-k\varepsilon/(n-1)),$$

where we used the imposed condition on n in the second inequality. □

The proof technique of Theorem 3.2 does not extend to the Gaussian case. For $k = 1$, Vershynin [39] has derived two-sided bounds for the more general setting in which the vectors $\tilde{x}_1, \hat{x}_1$ have independent and identically distributed entries from a sub-Gaussian distribution. In the following, we will use the techniques from [39] to establish lower bounds that, on the one hand, apply to general $k \geq 1$ and, on the other hand, provide specific constants for Gaussian and Rademacher vectors. For this purpose, we will make use of the following result on coupled second-order Gaussian and Rademacher chaos, which can be extracted from the proofs of Lemma 1 in [31] and Theorem 8 in [9], respectively.

LEMMA 3.3. *Let $B \in \mathbb{R}^{n \times n}$ be symmetric positive semidefinite. Then*

$$\log \mathbb{E}\big[\exp\big(t(x^T B x - \operatorname{trace} B)\big)\big] \leq \frac{ct^2\|B\|_F^2}{1 - 2ct\|B\|_2} \quad for \quad 0 \leq t < \frac{1}{2c\|B\|_2}$$

holds with $c = 1$ when x is a standard Gaussian random vector and with $c = 2$ when x is a Rademacher vector.

In the following, we will use the weaker bound $2ct^2\|B\|_2 \cdot \operatorname{trace} B$, which holds for $t \leq 1/(4c\|B\|_2)$.

THEOREM 3.4. *Consider the trace estimator Est_k defined in (3.1), and suppose that $\tilde{n} \leq \hat{n}$. Then the bound*

$$\operatorname{trace}(B) \geq \frac{1}{1 + \varepsilon}\mathsf{Est}_k$$

holds for independent standard Gaussian random vectors $\tilde{x}_i, \hat{x}_i$ with probability at least

$$(3.2) \qquad 1 - \exp\Big(-\frac{k\rho_A\varepsilon^2}{50\hat{n}}\Big) - k\frac{1}{5^{\hat{n}/2}} \quad if \quad 0 < \varepsilon < \frac{25}{16}$$

and for independent Rademacher vectors $\tilde{x}_i, \hat{x}_i$ with probability at least

$$(3.3) \qquad 1 - \exp\Big(-\frac{k\rho_A\varepsilon^2}{52\hat{n}}\Big) \quad if \quad 0 < \varepsilon < \frac{13}{4}.$$

Here, $\rho_A = \operatorname{trace} B/\|B\|_2$ is the stable rank of A.

Proof. We first consider the Gaussian case. As the proof follows closely the arguments in [39], we will keep it relatively brief. The main idea of [39] is to use marginalization to separate $\tilde{x}_i$ from $\hat{x}_i$. In order to express this conveniently, we let $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_k)$, $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_k)$ and set $f(\tilde{\mathbf{x}}, \hat{\mathbf{x}}) := \mathsf{Est}_k$.

Considering the event $\mathcal{E} = \{(\tilde{\mathbf{x}}, \hat{\mathbf{x}}) : \|\hat{x}_j\| \leq 2\hat{n}^{1/2} \text{ for all } j = 1, \ldots, k\}$, we have that

$$\mathbb{P}\{\mathsf{Est}_k - \operatorname{trace} B \geq \theta\} \leq \mathbb{P}\{\mathsf{Est}_k - \operatorname{trace} B \geq \theta \text{ and } \mathcal{E}\} + \mathbb{P}\{\mathcal{E}^c\}$$

$$= \mathbb{P}\Big\{e^{t(\mathsf{Est}_k - \operatorname{trace} B)} \geq e^{t\theta} \text{ and } \mathcal{E}\Big\} + \mathbb{P}\{\mathcal{E}^c\}$$

$$= \mathbb{P}\Big\{e^{t(\mathsf{Est}_k - \operatorname{trace} B)} \cdot \mathbf{1}_{\mathcal{E}}(\tilde{\mathbf{x}}, \hat{\mathbf{x}}) \geq e^{t\theta}\Big\} + \mathbb{P}\{\mathcal{E}^c\}$$

$$(3.4) \qquad \leq e^{-t\theta}\mathbb{E}[e^{t(\mathsf{Est}_k - \operatorname{trace} B)} \cdot \mathbf{1}_{\mathcal{E}}(\tilde{\mathbf{x}}, \hat{\mathbf{x}})] + \mathbb{P}\{\mathcal{E}^c\},$$

where $\mathbf{1}$ denotes the indicator function. By Lemma A.1, $\mathbb{P}\{\mathcal{E}^c\} \leq k(0.2)^{\hat{n}/2}$.

The exponent in (3.4) is separated via conditional expectation:

$$
\begin{aligned}
\mathsf{Est}_k - \operatorname{trace} B &= f(\tilde{\mathbf{x}}, \hat{\mathbf{x}}) - \mathbb{E}[f(\tilde{\mathbf{x}}, \hat{\mathbf{x}})] \\
&= f(\tilde{\mathbf{x}}, \hat{\mathbf{x}}) - \mathbb{E}_{\tilde{\mathbf{x}}}[f(\tilde{\mathbf{x}}, \hat{\mathbf{x}})] + \mathbb{E}_{\tilde{\mathbf{x}}}[f(\tilde{\mathbf{x}}, \hat{\mathbf{x}})] - \mathbb{E}[f(\tilde{\mathbf{x}}, \hat{\mathbf{x}})] \\
&= f(\tilde{\mathbf{x}}, \hat{\mathbf{x}}) - \mathbb{E}_{\tilde{\mathbf{x}}}[f(\tilde{\mathbf{x}}, \hat{\mathbf{x}})] + \hat{f}(\hat{\mathbf{x}}) - \mathbb{E}_{\hat{\mathbf{x}}}[\hat{f}(\hat{\mathbf{x}})], \quad \hat{f}(\hat{\mathbf{x}}) := \mathbb{E}_{\tilde{\mathbf{x}}}[f(\tilde{\mathbf{x}}, \hat{\mathbf{x}})].
\end{aligned}
$$

We now consider $\hat{\mathbf{x}}$ fixed such that $\|\hat{x}_j\|_2 \leq 2\hat{n}^{1/2}$ for $j = 1, \ldots, k$. Viewing $\tilde{\mathbf{x}}$ as a Gaussian vector of length $k\tilde{n}$, we can write

$$
f(\tilde{\mathbf{x}}, \hat{\mathbf{x}}) = \tilde{\mathbf{x}}^T \tilde{B} \tilde{\mathbf{x}}, \quad \tilde{B} := \frac{1}{k} \operatorname{diag}\big((I \otimes \hat{x}_1)^T B(I \otimes \hat{x}_1), \ldots, (I \otimes \hat{x}_k)^T B(I \otimes \hat{x}_k)\big).
$$

Noting that $\operatorname{trace} \tilde{B} = \mathbb{E}_{\tilde{\mathbf{x}}}[f(\tilde{\mathbf{x}}, \hat{\mathbf{x}})]$ and $\|\tilde{B}\|_2 \leq \frac{4}{k}\hat{n}\|B\|_2$, Lemma 3.3 applied to $\tilde{\mathbf{x}}^T \tilde{B} \tilde{\mathbf{x}}$ implies for all $0 \leq t \leq \frac{k}{16\hat{n}\|B\|_2}$ that

$$
\log \mathbb{E}_{\tilde{\mathbf{x}}}\big[\exp\big(t(f(\tilde{\mathbf{x}}, \hat{\mathbf{x}}) - \mathbb{E}_{\tilde{\mathbf{x}}}[f(\tilde{\mathbf{x}}, \hat{\mathbf{x}})])\big)\big] \leq \frac{8t^2\hat{n}}{k}\|B\|_2 \cdot \mathbb{E}_{\tilde{\mathbf{x}}}[f(\tilde{\mathbf{x}}, \hat{\mathbf{x}})]
$$

or, equivalently,

$$
\log \mathbb{E}_{\tilde{\mathbf{x}}}\big[\exp\big(tf(\tilde{\mathbf{x}}, \hat{\mathbf{x}}) - t_1 \mathbb{E}_{\tilde{\mathbf{x}}}[f(\tilde{\mathbf{x}}, \hat{\mathbf{x}})]\big)\big] \leq 0
$$

with $t_1 = t + \frac{8t^2\hat{n}}{k}\|B\|_2$. Next, we consider $\hat{\mathbf{x}}$ variable again and note that $\hat{f}(\hat{\mathbf{x}}) = \operatorname{trace}(\tilde{B}) = \frac{1}{k}\sum_{i=1}^{k} \hat{x}_i^T \hat{B} \hat{x}_i$, where $\hat{B}$ is obtained from B by partitioning it into $\hat{n} \times \hat{n}$ blocks and summing up the diagonal blocks. We have $\|\hat{B}\|_2 \leq \tilde{n}\|B\|_2$ and $\operatorname{trace} \hat{B} = \operatorname{trace} B$. Using independence of $\hat{x}_i$ and applying Lemma 3.3 to each $\hat{x}_i^T \hat{B} \hat{x}_i$, we obtain

$$
\log \mathbb{E}_{\hat{\mathbf{x}}}\big[\exp\big(t_1(\hat{f}(\hat{\mathbf{x}}) - \mathbb{E}_{\hat{\mathbf{x}}}[\hat{f}(\hat{\mathbf{x}})])\big)\big] \leq \frac{2t_1^2\tilde{n}}{k}\|B\|_2 \cdot \operatorname{trace} B.
$$

This requires $0 \leq t_1 \leq \frac{k}{4\tilde{n}\|B\|_2}$, which is implied by $\tilde{n} \leq \hat{n}$ and the condition on t. Lemma 4.1 in [39] now allows us to merge both bounds and obtain

$$
\log \mathbb{E}\big[e^{t\mathsf{Est}_k - t_1 \operatorname{trace} B} \cdot \mathbf{1}_{\mathcal{E}}(\tilde{\mathbf{x}}, \hat{\mathbf{x}})\big] \leq \frac{2t_1^2\tilde{n}}{k}\|B\|_2 \cdot \operatorname{trace} B.
$$

Using $t_1 \leq \frac{3}{2}t$, this implies $\log \mathbb{E}\big[e^{t(\mathsf{Est}_k - \operatorname{trace} B)} \cdot \mathbf{1}_{\mathcal{E}}(\tilde{\mathbf{x}}, \hat{\mathbf{x}})\big] \leq \frac{25t^2\hat{n}}{2k}\|B\|_2 \cdot \operatorname{trace} B$, which, when plugged into (3.4), gives

$$
\mathbb{P}\{\mathsf{Est}_k - \operatorname{trace} B \geq \theta\} \leq e^{-t\theta + t^2 \cdot \frac{25}{2k}\hat{n}\|B\|_2 \cdot \operatorname{trace} B} + k(0.2)^{\hat{n}/2}.
$$

Choosing the minimizing $t = \frac{\theta k}{25\hat{n}\|B\|_2 \cdot \operatorname{trace} B}$, which is valid for $\theta \leq 25/16 \cdot \operatorname{trace} B$, and setting $\varepsilon = \theta / \operatorname{trace} B$ yields the desired result (3.2).

The result (3.3) for Rademacher vectors is proven in the same fashion, with the simplification that $\|\hat{x}_j\|_2 = \hat{n}^{-1/2}$ holds with probability 1. □

A few remarks on the results of Theorem 3.4 are in order. Although the third term in (3.2) increases with k, it can be expected to remain negligible for all practical purposes; norm and trace estimates are primarily of interest for larger values of $\hat{n}$. In the Rademacher case, the bound (3.3) of Theorem 3.4 features the factor $\rho_A/(52\hat{n})$ in the exponent, which appears to be more favorable than the corresponding factor $1/n$ in Theorem 3.2. On the other hand, even for small values of k a rough lower bound with high probability can be obtained via Theorem 3.2 by choosing ε sufficiently

large. This is not possible with the results of Theorem 3.4 because of the imposed restrictions on ε.

4. Numerical experiments.

4.1. Performance of standard and rank-one estimators. The purpose of this section is to numerically assess the impact of using rank-one random vectors instead of unstructured random vectors on the performance of the trace estimator Est_k; see (3.1). For this purpose, the following 8 examples have been chosen to illustrate different aspects of the estimators.

Ones, matrix of all ones; $n = 2\,500$, $\tilde{n} = \hat{n} = 50$, estimation of $\mathrm{trace}(A) = 2\,500$.

Rank-one, matrix vv^T where $v = \mathrm{vec}(I_{\tilde{n}})$ with the identity matrix $I_{\tilde{n}}$; $n = 2\,500$, $\tilde{n} = \hat{n} = 50$, estimation of $\mathrm{trace}(A) = 50$.

ACTIVSg2000, from the SuiteSparse Matrix Collection [10] originating from a synthetic electric grid model [5]; $n = 4\,000$, $\tilde{n} = 80$, $\hat{n} = 50$, estimation of $\|A^{-1}\|_F^2 = \mathrm{trace}(A^{-T}A^{-1}) \approx 1.5 \times 10^4$.

ACTIVSg10K, same source as **ACTIVSg2000** but now $n = 20\,000$, $\tilde{n} = 200$, $\hat{n} = 100$, estimation of $\|A^{-1}\|_F^2 \approx 1.3 \times 10^5$.

CFD, from the SuiteSparse Matrix Collection [10] (matrix Rothberg/cfd1) and originating from the discretization of a fluid dynamics problem; $n = 70\,656$, $\tilde{n} = 276$, $\hat{n} = 256$, estimation of $\mathrm{trace}(A) = 70\,656$.

CFDinv, identical with **CFD** but estimation of $\mathrm{trace}(A^{-1}) \approx 6.0 \times 10^5$ instead.

Laplace, matrix from second-order finite difference discretization of Poisson equation on unit square; $n = 2\,500$, $\tilde{n} = \hat{n} = 50$, estimation of $\mathrm{trace}(A^{-1}) \approx 0.61$.

Convdiff, matrix from finite difference discretization of convection-diffusion equation on unit square (matrix from [26, section 7.2] with $c_s = 1$); $n = 2\,500$, $\tilde{n} = \hat{n} = 50$, estimation of $\|A^{-1}\|_F^2 \approx 0.0042$.

In all examples for which the trace is estimated the involved matrix is symmetric positive semidefinite. The matrices **Ones** and **Rank-one** both have rank one, but the stable rank of their (only) singular vector reshaped as a matrix is very different. The discussion in section 2.4 has singled out **Ones** as a bad example when estimating with Rademacher vectors. **CFD** is an example used in [3]. For both **Laplace** and **Convdiff**, the matrix A can be represented in the form $C_1 \otimes I_{\hat{n}} + I_{\tilde{n}} \otimes C_2$ for smaller, sparse matrices C_1, C_2. Such matrices are of particular interest for our new rank-one estimators because the application of A^{-1} (or A^{-T}) corresponds to the solution of a matrix Sylvester equation and such an equation can be solved much more efficiently when the right-hand side has low rank [36].

When the trace of A itself was estimated, we used its exact trace as reference value **Exact**. When the trace or Frobenius norm of A^{-1} was estimated, we used $\mathsf{Est}_{1\,000}$ with standard Gaussian random vectors as reference value **Exact**. For each matrix and each type of random vector, we repeated 10 000 times the computation of Est_k with k ranging from 1 to up to 50 and computed the minimum of $-\mathsf{Exact}/\mathsf{Est}_k$ and the maximum of $\mathsf{Est}_k/\mathsf{Exact}$ across all 10 000 runs for each k. The lower and upper curves in each plot of Figure 2 display the minima/maxima vs. k for four different types of random vectors: standard Gaussian, rank-one Gaussian, Rademacher, and rank-one Rademacher vectors. For example, for the matrix **Ones** and $k = 20$, it can be seen that all curves stay between -20 and 10, which implies that for all 10 000 samples of Est_k, the bounds $\mathsf{Est}_k/10 \leq \mathsf{Exact} \leq 20 \cdot \mathsf{Est}_k$ were satisfied. For nearly all configurations, using rank-one random vectors instead of unstructured random vectors only has a modest impact on these worst-case under- and overestimation factors. Additional
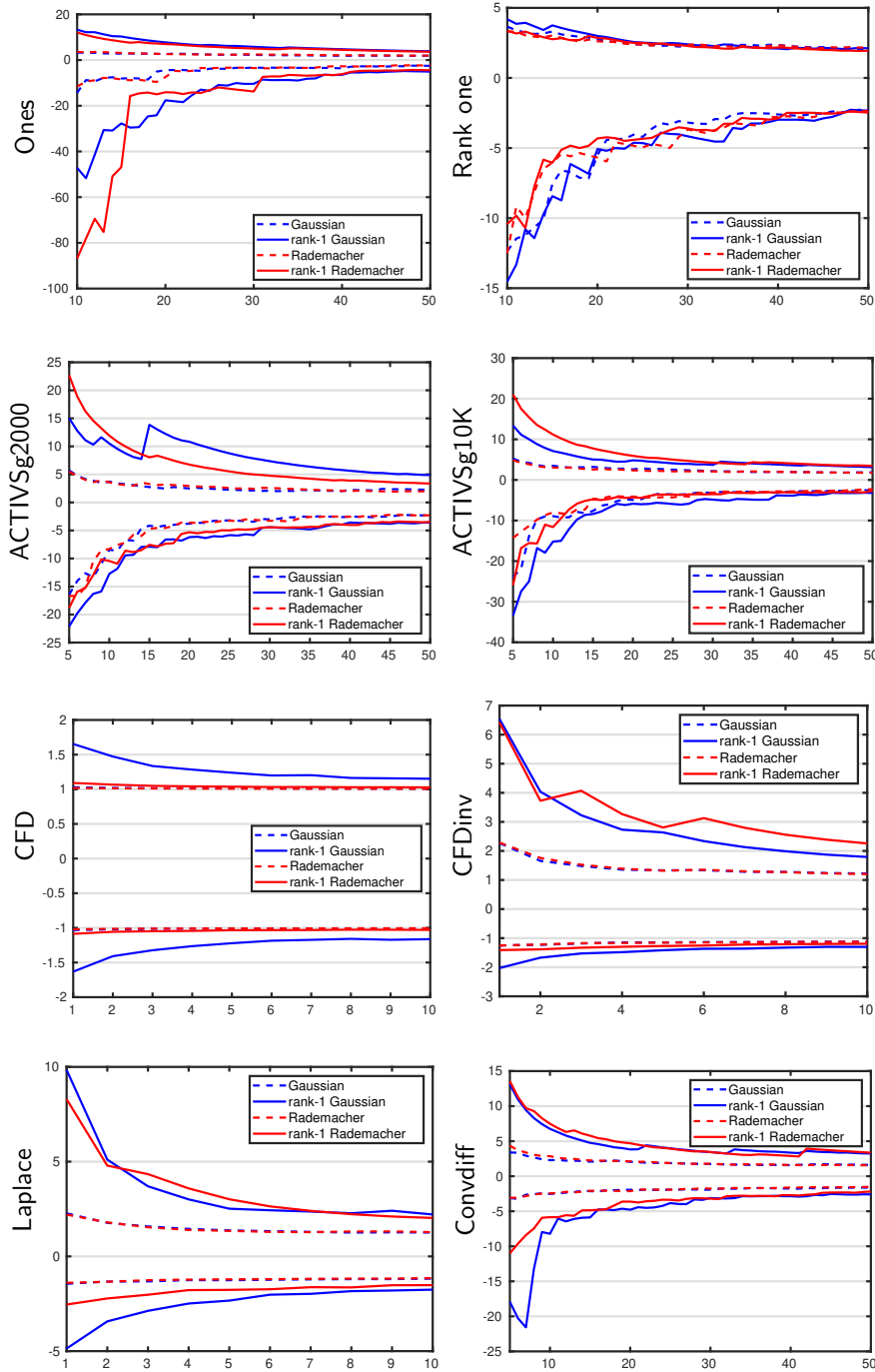
FIG. 2. *Performance of trace estimators for 8 different matrices. See section* 4.1 *for details.*

data is given in the supplementary material supplement.pdf [local/web 201KB], which shows that the inequalities

$$\mathsf{Est}_{10}/30 \leq \mathsf{Exact} \leq 30 \cdot \mathsf{Est}_{10}$$

are nearly always satisfied across all matrices and all types of random vectors.

4.2. Fréchet derivative norm estimation. In this section we describe an application in which random vectors of the form $\tilde{x} \otimes \hat{x}$ are exploited to significantly speed up computation. Given a matrix function $f(A)$ for a matrix $A \in \mathbb{R}^{n \times n}$, our goal is to estimate the operator norm of the Fréchet derivative, which is a linear map $Df\{A\} : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ uniquely defined by the property $f(A + X) = f(A) + Df\{A\}(X) + \mathcal{O}(\|X\|_2^2)$. This quantity measures the first-order sensitivity of the matrix function $f(A)$ under perturbation of A [1, 2, 18].

It is well known [18, section 3.2] that under certain conditions on the smoothness of f,

$$(4.1) \qquad f\left(\begin{bmatrix} A & X \\ 0 & A \end{bmatrix} \right) = \begin{bmatrix} f(A) & Df\{A\}(X) \\ 0 & f(A) \end{bmatrix}.$$

By vectorizing the matrices in its domain and range, the action of $Df\{A\}$ can be represented by an $n^2 \times n^2$ matrix K_A such that $\mathrm{vec}(Df\{A\}(X)) = K_A x$, where $x = \mathrm{vec}(X)$. This yields

$$\|Df\{A\}\| := \sup_{\|X\|_F = 1} \|Df\{A\}(X)\|_F = \sup_{\|x\|_2 = 1} \|K_A x\|_2 = \|K_A\|_2 = \sqrt{\lambda_{\max}(K_A^T K_A)}.$$

Therefore, to compute $\|Df\{A\}\|$, one may apply the power method to the matrix $K_A^T K_A$ to compute its largest eigenvalue. The power method requires evaluating $K_A x$ and $K_A^T x$ for several vectors x, i.e., evaluating $Df\{A\}X$ and $Df\{A^T\}X$ for several matrices X, where $\mathrm{vec}(X) = x$. This is executed via (4.1): the matrix function f is evaluated at a $2n \times 2n$ matrix, and the top right block is read from the resulting matrix. For larger n, such computation may be very demanding.

If we are only interested in an upper bound for $\|Df\{A\}\|$ instead of its exact value, we can apply techniques discussed in section 2.3: let $x_j = \tilde{x}_j \otimes \hat{x}_j$, where $\tilde{x}_j$ and $\hat{x}_j$ are standard Gaussian vectors of length n, for $j = 1, \dots, k$. Using the maximum estimator

$$(4.2) \qquad \mathsf{Max}_k = \max_{j=1,\dots,k} \|K_A(\tilde{x}_j \otimes \hat{x}_j)\|_2 = \max_{j=1,\dots,k} \|Df\{A\}(\hat{x}_j \tilde{x}_j^T)\|_F$$

and applying (2.10) then guaranties the following:

$$\mathbb{P}\{\|Df\{A\}\| \leq \theta \cdot \mathsf{Max}_k\} = \mathbb{P}\{\|K_A\|_2 \leq \theta \cdot \mathsf{Max}_k\} \geq 1 - \left(\frac{2}{\pi} \left(2 + \ln(1 + 2\theta)\right) \theta^{-1} \right)^k.$$

Computing the matrix-vector products in (4.2) reduces to evaluating $Df\{A\}(X_j)$ for rank-one matrices $X_j = \hat{x}_j \tilde{x}_j^T$. This can be done far more efficiently than evaluating $Df\{A\}(X)$ for general matrices X by using Algorithm 1, slightly adapted from [24]. Algorithm 1 also needs to evaluate the function f, but for matrices of sizes at most $2\ell \times 2\ell$, where the final dimension ℓ of the Krylov subspace is significantly smaller than n. This is the source of the significant speedup when compared to the power method. Also note that only the first iteration of the power method can benefit from Algorithm 1, as the later iterates are generally not matrices of rank one.

Algorithm 1: Arnoldi method for approximating $Df\{A\}(cd^T)$.

---

1 **for** $\ell = 1, 2, \ldots$ **do**

2   Perform one step of the Arnoldi method to obtain an orthonormal basis
     U_ℓ of the Krylov subspace $\mathcal{K}_\ell(A, c)$ and $G_\ell = U_\ell^T A U_\ell$, $\tilde{c} = U_\ell^T c$.

3   Perform one step of the Arnoldi method to obtain an orthonormal basis
     V_ℓ of the Krylov subspace $\mathcal{K}_\ell(A^T, d)$ and $H_\ell = V_\ell^T A^T V_\ell$, $\tilde{d} = V_\ell^T d$.

4   Compute $F_\ell = f\left(\begin{bmatrix} G_\ell & \tilde{c}\tilde{d}^T \\ 0 & H_\ell^T \end{bmatrix}\right)$ and set $X_\ell = F_\ell(1\!:\!\ell, \ell+1\!:\!2\ell)$.

5   **if** *converged* **then**

6   Stop the loop.

7 Return $U_\ell X_\ell V_\ell^T$.

To illustrate this difference, we ran both the power method and the 99.9% confidence maximum estimator ($\theta = 10$, $k = 7$) to estimate $\|Df\{A\}\|$, where f is the matrix exponential and $A = -0.01(I_n \otimes T_n + T_n \otimes I_n)$. Here T_n is the tridiagonal matrix with $2/(n-1)^2$ on the main diagonal and $-1/(n-1)^2$ on the first upper and lower suddiagonal, i.e., the 1D discrete Laplacian on $[0,1]$. Running in MATLAB R2019b on an Intel i5 4690K processor, we obtain the following results:

n	Time(power method)	$\|Df\{A\}\|$	Time(Max_k)	$\theta \cdot \mathsf{Max}_k$	Maximum ℓ
10	0.12	0.86	0.05	147.10	20
20	1.81	0.85	0.10	151.00	30
30	16.49	0.78	0.38	115.44	45
40	74.06	0.80	1.18	102.13	55
50	275.52	0.82	2.95	93.38	70

Note that A is an $n^2 \times n^2$ matrix. We ran 7 iterations of the power method; the third column shows the approximation of $\|Df\{A\}\|$ it reported. The fifth column shows the 99.9% confidence upper bound for $\|Df\{A\}\|$ as reported by the maximum estimator. In the last column is the maximum dimension ℓ of all Krylov subspaces needed for the computation of Max_k. We stop the Arnoldi iteration once $\|F_\ell - \begin{bmatrix} F_{\ell-1} & 0 \\ 0 & 0 \end{bmatrix}\| < 10^{-8}$, as suggested in [4, section 2.3]. All times are given in seconds. While the upper bounds provided by the maximum estimators are, in this case, about 100 times larger than the actual norm of the Fréchet derivative, this may be sufficient as a rough estimate. Such an estimate can be rapidly computed by using random vectors studied in this paper.

5. Conclusions. In this work we have provided theoretical and experimental evidence that rank-one random vectors are suited for norm and trace estimation. While their performance is consistently worse compared to unstructured vectors, this can be easily mitigated by, e.g., increasing the constants or increasing the number of samples in the stochastic trace estimator.

It is tempting to ask whether the results of this paper have a meaningful extension to higher-order tensors, that is, norm and trace estimation with Kronecker products of d vectors. Without further assumptions on A, the techniques used in this work will lead to estimates of the success probability that vanish exponentially

fast as d increases. This is explicit in the results from [39] but also the Khintchine inequalities [30] and, more generally, moments of order-d chaos exhibit exponential growth.

Appendix A.

The following lemma provides a Chernoff bound for chi-square distributions. This result is certainly well known; we include it for completeness.

LEMMA A.1. *Let X be a random variable having a chi-square distribution with k degrees of freedom. For $\theta > 1$, it holds that*

$$\mathbb{P}\{X > k\theta\} \leq (\theta e^{1-\theta})^{k/2}.$$

Proof. For $t > 0$, we let $M_X(t) = (1 - 2t)^{-k/2}$ denote the moment generating function of X. By the Markov inequality,

$$\mathbb{P}\{X > k\theta\} = \mathbb{P}\{tX/k > t\theta\} = \mathbb{P}\left\{e^{tX/k} > e^{t\theta}\right\}$$
$$\leq \mathbb{E}[e^{tX/k}] \cdot e^{-t\theta} = M_X(t/k)e^{-t\theta} = (1 - 2t/k)^{-k/2}e^{-t\theta}, \qquad \square$$

which holds for $t/k < 1/2$. Choosing $t = k/2 - k/2\theta$ implies

$$\mathbb{P}\{X > k\theta\} \leq \theta^{k/2}e^{k/2(1-\theta)} = (\theta e^{1-\theta})^{k/2}.$$

The following two results on the moments and the moment generating function of decoupled second-order Gaussian chaos are closely related to existing results by Latała [30]; see also the monograph [6]. We include these results for the convenience of the reader.

LEMMA A.2. *Let $Q \in \mathbb{R}^{\hat{n} \times \tilde{n}}$, and let $\hat{x} \sim \mathcal{N}(0, I_{\hat{n}})$, $\tilde{x} \sim \mathcal{N}(0, I_{\tilde{n}})$. For $Z = \hat{x}^T Q \tilde{x}$ it holds that*

$$\mathbb{E}[Z^2] = \|Q\|_F, \qquad \mathbb{E}[Z^4] = 3(2\|Q\|_{(4)}^4 + \|Q\|_F^4) \leq 9\|Q\|_F^4,$$

where $\|\cdot\|_{(4)}$ denotes the Schatten 4-norm [20, section 7.4] of a matrix. For any even k, we have

(A.1)
$$\mathbb{E}[Z^k] \leq \big((k-1)!!\big)^2 \|Q\|_F^k,$$

where $(k - 1)!! = (k - 1)(k - 3) \cdots 3 \cdot 1$ denotes the double factorial. For odd k, $\mathbb{E}[Z^k] = 0$.

Proof. For the second moment, we obtain

$$\mathbb{E}[Z^2] = \mathbb{E}_{\hat{x}}\big[\mathbb{E}_{\tilde{x}}\big[(\hat{x}^T Q \tilde{x})^2 | \hat{x}\big]\big] = \mathbb{E}_{\hat{x}}\big[\|Q^T \hat{x}\|_2^2\big] = \|Q\|_F^2,$$

where the second equality follows from the fact that for $y = Q^T \hat{x}$ with fixed $\hat{x}$, the random variable $y^T \tilde{x}$ is normal with zero mean and variance $\|y\|_2^2$. Noting that the fourth moment of such a normal random variable is $3\|y\|_2^4$, an analogous argument shows $\mathbb{E}[Z^4] = 3 \cdot \mathbb{E}\big[\|Q^T \hat{x}\|_2^4\big]$. To proceed from here, we may assume—without loss of generality—that $\hat{n} \leq \tilde{n}$ and that Q is a diagonal matrix with the singular values $\sigma_1 \geq \cdots \geq \sigma_{\hat{n}} \geq 0$ on the diagonal; see, e.g., the proof of Theorem 2.2. This gives

$$\mathbb{E}\big[\|Q^T \hat{x}\|_2^4\big] = \mathbb{E}\big[\big(\sigma_1^2 \hat{x}_1^2 + \cdots + \sigma_{\hat{n}}^2 \hat{x}_{\hat{n}}^2\big)^2\big] = \sum_{ij} \sigma_i^2 \sigma_j^2 \mathbb{E}[\hat{x}_i^2 \hat{x}_j^2]$$

$$= \sum_i \sigma_i^4 \mathbb{E}[\hat{x}_i^4] + \sum_{i \neq j} \sigma_i^2 \sigma_j^2 \mathbb{E}[\hat{x}_i^2] \cdot \mathbb{E}[\hat{x}_j^2] = 3 \sum_i \sigma_i^4 + \sum_{i \neq j} \sigma_i^2 \sigma_j^2$$

$$= 2 \sum_i \sigma_i^4 + \sum_{ij} \sigma_i^2 \sigma_j^2 = 2\|Q\|_{(4)}^4 + \|Q\|_F^4,$$

which establishes the claimed expression for $\mathbb{E}[Z^4]$. The upper bound $9\|Q\|_F^4$ follows from $\|Q\|_{(4)} \leq \|Q\|_F$.

For general even k, the statement and proof of (A.1) are contained in the proof of Lemma 7.1 in [7]. The proof that follows is slightly simpler. We first note that the kth moment of a centered normal random variable with variance σ^2 is given by $(k-1)!!\sigma^k$. In turn, $\mathbb{E}[Z^k] = (k-1)!! \cdot \mathbb{E}\big[\|Q^T\hat{x}\|_2^k\big]$. We proceed as above and obtain

$$\mathbb{E}\big[\|Q^T\hat{x}\|_2^k\big] = \sum_{i_1,\ldots,i_{k/2}} \sigma_{i_1}^2 \cdots \sigma_{i_{k/2}}^2 \mathbb{E}[\hat{x}_{i_1}^2 \cdots \hat{x}_{k/2}^2].$$

Using that $\mathbb{E}[\hat{x}_i^{2p}\hat{x}_j^{2q}] = (2p-1)!! \cdot (2q-1)!! \leq (2(p+q)-1)!! = \mathbb{E}[\hat{x}_i^{2(p+q)}]$ for any $p,q \in \mathbb{N}$ and $i \neq j$, we obtain

$$\mathbb{E}\big[\|Q^T\hat{x}\|_2^k\big] \leq (k-1)!! \sum_{i_1,\ldots,i_{k/2}} \sigma_{i_1}^2 \cdots \sigma_{i_{k/2}}^2 = (k-1)!! \cdot \|Q\|_F^k,$$

which concludes the proof of (A.1).

The statement on odd k follows from the symmetry of the distribution: Z and $-Z = \hat{x}^T Q(-\tilde{x})$ have the same distribution, and hence

$$\mathbb{E}\big[Z^k\big] = \mathbb{E}\big[(-Z)^k\big] = \mathbb{E}\big[-Z^k\big] = -\mathbb{E}\big[Z^k\big].$$

This shows $\mathbb{E}\big[Z^k\big] = 0$. $\qquad\square$

COROLLARY A.3. *For Z as in Lemma A.2 with $\|Q\|_F = 1$, the moment generating function exists and is bounded by*

$$\mathbb{E}\big[\exp(tZ)\big] \leq \frac{1}{\sqrt{1-t^2}}$$

for $|t| < 1$.

Proof. Using the result of Lemma A.2, it follows that

$$\mathbb{E}\big[\exp(tZ)\big] = \sum_{k=0}^{\infty} \frac{1}{k!}\mathbb{E}\big[Z^k\big]t^k = \sum_{k=0}^{\infty} \frac{1}{(2k)!}\mathbb{E}\big[Z^{2k}\big]t^{2k}$$

$$\leq \sum_{k=0}^{\infty} \frac{\big((k-1)!!\big)^2}{(2k)!}t^{2k} = \sum_{k=0}^{\infty} \frac{1}{4^k}\binom{2k}{k}t^{2k} = \frac{1}{\sqrt{1-t^2}},$$

where the last step follows from Taylor expansion. $\qquad\square$

REFERENCES

[1] A. H. AL-MOHY AND N. J. HIGHAM, *Computing the Fréchet derivative of the matrix exponential, with an application to condition number estimation*, SIAM J. Matrix Anal. Appl., 30 (2009), pp. 1639–1657, https://doi.org/10.1137/080716426.

[2] A. H. Al-Mohy, N. J. Higham, and S. D. Relton, *Computing the Fréchet derivative of the matrix logarithm and estimating the condition number*, SIAM J. Sci. Comput., 35 (2013), pp. C394–C410, https://doi.org/10.1137/120885991.

[3] H. Avron and S. Toledo, *Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix*, J. ACM, 58 (2011), 8, https://doi.org/10.1145/1944345.1944349.

[4] B. Beckermann, D. Kressner, and M. Schweitzer, *Low-rank updates of matrix functions*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 539–565, https://doi.org/10.1137/17M1140108.

[5] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, *Grid structural characteristics as validation criteria for synthetic networks*, IEEE Trans. Power Syst., 32 (2017), pp. 3258–3265.

[6] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities*, Oxford University Press, Oxford, 2013, https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

[7] T. T. Cai and A. Zhang, *ROP: Matrix recovery via rank-one projections*, Ann. Statist., 43 (2015), pp. 102–138, https://doi.org/10.1214/14-AOS1267.

[8] Y. Chen, Y. Chi, and A. J. Goldsmith, *Exact and stable covariance estimation from quadratic sampling via convex programming*, IEEE Trans. Inform. Theory, 61 (2015), pp. 4034–4059, https://doi.org/10.1109/TIT.2015.2429594.

[9] A. Cortinovis and D. Kressner, *On Randomized Trace Estimates for Indefinite Matrices with an Application to Determinants*, preprint, arXiv:2005.10009, 2020.

[10] T. A. Davis and Y. Hu, *The University of Florida Sparse Matrix Collection*, ACM Trans. Math. Software, 38 (2011), 1, https://doi.org/10.1145/2049662.2049663, see also https://sparse.tamu.edu/.

[11] J. D. Dixon, *Estimating extremal eigenvalues and condition numbers of matrices*, SIAM J. Numer. Anal., 20 (1983), pp. 812–814, https://doi.org/10.1137/0720053.

[12] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 2013.

[13] L. Grasedyck, D. Kressner, and C. Tobler, *A literature survey of low-rank tensor approximation techniques*, GAMM-Mitt., 36 (2013), pp. 53–78, https://doi.org/10.1002/gamm.201310004.

[14] S. Gratton and D. Titley-Peloquin, *Improved bounds for small-sample estimation*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 922–931, https://doi.org/10.1137/17M1137541.

[15] T. Gudmundsson, C. S. Kenney, and A. J. Laub, *Small-sample statistical estimates for matrix norms*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 776–792, https://doi.org/10.1137/S0895479893243876.

[16] N. Halko, P. G. Martinsson, and J. A. Tropp, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Rev., 53 (2011), pp. 217–288, https://doi.org/10.1137/090771806.

[17] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, PA, 2002.

[18] N. J. Higham, *Functions of Matrices*, SIAM, Philadelphia, PA, 2008, https://doi.org/10.1137/1.9780898717778.

[19] M. E. Hochstenbach, *Probabilistic upper bounds for the matrix two-norm*, J. Sci. Comput., 57 (2013), pp. 464–476, https://doi.org/10.1007/s10915-013-9716-x.

[20] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed., Cambridge University Press, Cambridge, 2013.

[21] M. F. Hutchinson, *A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines*, Comm. Statist. Simulation Comput., 19 (1990), pp. 433–450, https://doi.org/10.1080/03610919008812864.

[22] R. Jin, T. G. Kolda, and R. Ward, *Faster Johnson–Lindenstrauss transforms via Kronecker products*, Inf. Inference, (2020), iaaa028, https://doi.org/10.1093/imaiai/iaaa028.

[23] M. Kliesch, R. Kueng, J. Eisert, and D. Gross, *Improving compressed sensing with the diamond norm*, IEEE Trans. Inform. Theory, 62 (2016), pp. 7445–7463, https://doi.org/10.1109/TIT.2016.2606500.

[24] D. Kressner, *A Krylov subspace method for the approximation of bivariate matrix functions*, in Structured Matrices in Numerical Linear Algebra, Springer INdAM Ser. 30, Springer, Cham, 2019, pp. 197–214.

[25] D. Kressner and L. Periša, *Recompression of Hadamard products of tensors in Tucker format*, SIAM J. Sci. Comput., 39 (2017), pp. A1879–A1902, https://doi.org/10.1137/16M1093896.

[26] D. Kressner and C. Tobler, *Krylov subspace methods for linear systems with tensor product structure*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 1688–1714, https://doi.org/10.1137/090756843.

[27] J. Kuczyński and H. Woźniakowski, *Estimating the largest eigenvalue by the power and Lanczos algorithms with a random start*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 1094–1122, https://doi.org/10.1137/0613066.

[28] R. Kueng, H. Rauhut, and U. Terstiege, *Low rank matrix recovery from rank one measurements*, Appl. Comput. Harmon. Anal., 42 (2017), pp. 88–116, https://doi.org/10.1016/j.acha.2015.07.007.

[29] A. N. Langville and W. J. Stewart, *The Kronecker product and stochastic automata networks*, J. Comput. Appl. Math., 167 (2004), pp. 429–447, https://doi.org/10.1016/j.cam.2003.10.010.

[30] R. Latała, *Estimates of moments and tails of Gaussian chaoses*, Ann. Probab., 34 (2006), pp. 2315–2331, https://doi.org/10.1214/009117906000000421.

[31] B. Laurent and P. Massart, *Adaptive estimation of a quadratic functional by model selection*, Ann. Statist., 28 (2000), pp. 1302–1338, https://doi.org/10.1214/aos/1015957395.

[32] R. A. Meyer, C. Musco, C. Musco, and D. P. Woodruff, *Hutch++: Optimal Stochastic Trace Estimation*, preprint, arXiv:2010.09649, 2020.

[33] H. Rauhut, *Compressive sensing and structured random matrices*, in Theoretical Foundations and Numerical Methods for Sparse Recovery, Radon Ser. Comput. Appl. Math. 9, Walter de Gruyter, Berlin, 2010, pp. 1–92, https://doi.org/10.1515/9783110226157.1.

[34] F. Roosta-Khorasani and U. Ascher, *Improved bounds on sample size for implicit matrix trace estimators*, Found. Comput. Math., 15 (2015), pp. 1187–1212, https://doi.org/10.1007/s10208-014-9220-1.

[35] N. G. Shephard, *From characteristic function to distribution function: A simple framework for the theory*, Econometric Theory, 7 (1991), pp. 519–529, https://doi.org/10.1017/S0266466600004746.

[36] V. Simoncini, *Computational methods for linear matrix equations*, SIAM Rev., 58 (2016), pp. 377–441, https://doi.org/10.1137/130912839.

[37] Y. Sun, Y. Guo, J. A. Tropp, and M. Udell, *Tensor random projection for low memory dimension reduction*, in Proceedings of the NeurIPS Workshop on Relational Representation Learning, 2018.

[38] J. A. Tropp, *Convex recovery of a structured signal from independent random linear measurements*, in Sampling Theory, a Renaissance, Appl. Numer. Harmon. Anal., Birkhäuser/Springer, Cham, 2015, pp. 67–101.

[39] R. Vershynin, *Concentration inequalities for random tensors*, Bernoulli, 26 (2020), pp. 3139–3162, http://dx.doi.org/10.3150/20-BEJ1218.