On the upper bounds for the matrix spectral norm

D. Ryapolov, A. Naumov, M. Rakhuba, S. Samsonov

June 16, 2025

Abstract

In this paper we consider the problem of estimating spectral norm of a matrix using only matrix-vector products. We propose a new *Counterbalance* estimator, which provides an upper bounds on the norm, and derive bounds on its underestimation probability. We show that it allows for constructing tighter upper bounds for the operator norm of the underlying matrix, compared to baseline methods.

1 Introduction

Estimating the spectral norm (largest singular value) of an implicitly defined matrix using only matrix-vector products is a fundamental problem in numerical linear algebra and machine learning. In many applications, one cannot access the entire matrix $A \in \mathbb{R}^{m\times n}$ explicitly, but one can compute products Ax and sometimes $A^\top x$ for chosen vectors x. The challenge is to estimate $\|A\|_2$ (or related norms) from above with few matrix-vector products (matvecs). This scenario arises, for example, in sensitivity analysis of matrix functions and condition number estimation [1, 2, 16], and in regularizing deep neural networks by enforcing Lipschitz continuity (bounding layer Jacobian norms) [7]. Traditional deterministic approaches like the power method estimate the norm from below and require multiple sequential iterations until convergence, which may be infeasible if each matvec is expensive. Randomized algorithms provide attractive alternatives: by probing A with random test vectors, one can obtain unbiased or statistically reliable estimates of various matrix norms and traces much faster in practice. The main advantage of randomized methods in this context is the fact that the required matvecs can be computed in parallel.

Perhaps the simplest approach to estimate $\|A\|_2$ is to compute matvec AX for a random vector X and use the norm $\|AX\|_2$ for an estimator. Specifically, let X be a random vector of appropriate dimension with $\|X\|_2 = 1$, and consider the scalar $\|AX\|_2$. This quantity is always at most $\|A\|_2$, and in fact, early work by Dixon [6] established probabilistic guarantees for this estimator: for a random vector X drawn uniformly from the sphere $\mathbb{S}_{n-1}$, one has

$$\mathbb{P}(\|A\|_2 \le \theta \|AX\|_2) \ge 1 - 0.8\sqrt{d}\theta^{-1/2}, \tag{1}$$

for any $\theta > 1$, where $d = \max(m, n)$. Equivalently, with high probability, $\|AX\|_2$ underestimates the true norm $\|A\|_2$ by at most a factor θ that grows only moderately with the ambient dimension d. In practice, one can boost the reliability of norm estimates (1) by averaging or taking maxima over multiple independent statistics, see [6, 8]. If $X_1, \dots, X_k$ are k independent random vectors, a natural improvement of (1) is the maximum observed norm:

$$T(X_1, \dots, X_k) = \max_{1\le i\le k} \|AX_i\|_2 . \tag{2}$$

This multi-sample approach reduces the risk of a "bad" draw that misses the top singular vector and can be straighforwadly parallelized.

It is important to note that these basic randomized estimators tend to underestimate the true norm. Indeed, $\max_i \|AX_i\|_2 \leq \|A\|_2$ for any unit vector X. In some applications, a probabilistic upper bound on $\|A\|_2$ is desired. One can obtain an upper bound by scaling the test statistic (1) or (2) by an appropriate factor. The formula above from Dixon's result provides one such factor: e.g., setting $\theta(\delta) = (4/5)^2 d/\delta^2$ in (1) yields

$$\mathbb{P}(\|A\|_2 \leq \theta(\delta)\,\|AX\|_2) \geq 1 - \delta\,.$$

This bound has unfavorable scaling with d, and can be formally improved if one considers the statistic $T_\theta(X) = \theta\,\|AX\|_2$, as suggested in [8]. With this estimator, one can prove the following bound on the underestimation probability of $\|A\|_2$, see [8, Lemma 4.1]: with $X \sim \mathcal{N}(0, \mathrm{I}_n)$, it holds that

$$\mathbb{P}(\theta\,\|AX\|_2 \leq \|A\|_2) \leq \sqrt{\frac{2}{\pi}}\frac{1}{\theta}\,. \tag{3}$$

The bound (3) can not be improved in general, as it is tight when considering matrices A with $\operatorname{rk} A = 1$. Note that the estimator $\theta\,\|AX\|_2$ typically behaves much better, compared to the upper bound (3), since

$$\mathsf{E}[\|AX\|_2^2] = \|A\|_{Fr}^2\,.$$

Using the apparatus of concentration of measure inequalities [4, 15], one can study the concentration properties of $\|AX\|_2$ around its expectation, which can be shown to be close to $\|A\|_{Fr}$. Thus, when we consider the matrix A with $\|A\|_{Fr} \gg \|A\|_2$, one can expect that $\|AX\|_2$ will significantly overestimate $\|A\|_2$. This property can be formalized using the concept of *effective rank* $\rho = \|A\|_{Fr}^2/\|A\|_2^2$. At the same time, the upper bound of the underestimation probability (3) does not depend upon ρ. This raises the following dilemma, which can be formulated as follows. Suppose that we use a statistic $T_\theta(X)$ (for example, $T_\theta(X) = \theta\,\|AX\|_2$) to provide an upper bound on $\|A\|_2$. Suppose also that we are able to provide an upper bound

$$\mathbb{P}(T_\theta(X) \leq \|A\|_2) \leq g(\theta) \tag{4}$$

for some function $g(\theta)$ that is independent of A. Then tighter expressions for $g(\theta)$ requires additional information about A, such as knowledge about the effective rank ρ. Typically, this information is not available for a statistician. Hence, a natural question arises:

Can we construct a simple estimator for $\|A\|_2$, which is agnostic to its structure and allows for providing tighter upper bound on $\|A\|_2$, compared to (3)*?*

It is important to provide a precise quantification of "tighter upper bound" here. Suppose, that for a fixed computational budget (that is, for a fixed number of matvecs), we construct two (randomized) estimators $T_1(X)$ and $T_2(X)$, such that for a fixed $\delta \in (0, 1)$, it holds that

$$\mathbb{P}(T_1(X) \leq \|A\|_2) \leq \delta\,,\ \text{and}\ \mathbb{P}(T_2(X) \leq \|A\|_2) \leq \delta\,. \tag{5}$$

Then it is natural to say that $T_1(X)$ is tighter then $T_2(X)$, if at least $\mathsf{E}[T_1(X)] \leq \mathsf{E}[T_2(X)]$. We postpone detailed discussion on the subject to Section 3. In our paper, we aim to present an estimator of $\|A\|_2$, which is tighter than the one, which is based on (3), but does not require any prior knowledge on the decay of singular values of A. Our main contributions are as follows:

- We propose a new *Counterbalance* (CB for short) estimator $T_\theta^c(X)$, which provides an upper bounds on $\|A\|_2$, and derive bounds on its underestimation probability.

$$\mathbb{P}(T_\theta^c(X) \leq \|A\|_2)\,.$$

  The term "counterbalance" comes from the additional summand, that we add to the classical estimator of a form (3), which noticeably sharpens our bound for low-rank matrices A. We also determine the way of selecting the parameter θ for our estimator, which does not depend upon the structure of singular values of A. We study the concentration properties of the novel estimator and show that it allows for constructing tighter upper bounds for the operator norm of the underlying matrix and fixed underestimation probability, compared to baseline methods.
- We perform a number of numerical simulations, illustrating the benefits of our method on various matrices with various ranks, effective ranks, and structure of singular values. We also consider the example of upper bounding the spectral norm of Jacobian of layers of ResNet neural networks. We show that our estimator allows for constructing tighter upper bounds for the operator norm of the underlying matrices for fixed underestimation probability, compared to baseline methods.

Notations. For a matrix $A \in \mathbb{R}^{m\times n}$, we write $\|A\|_2$ for its spectral norm and $\|A\|_F$ for its Frobenius norm. We denote by ρ its *effective rank* (also known as stable rank) $\rho = \|A\|_F^2 / \|A\|_2^2$. Given that A admits a singular value decomposition (SVD) $A = U\Sigma V^\star$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$, where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ are the corresponding singular values, the effective rank can be rewritten as $\rho = (\sum_{i=1}^k \sigma_i^2)/\sigma_1^2$. We also denote by $\chi^2(t) = \mathbb{P}\left(\xi^2 \leq t\right)$, where $\xi \sim N(0,1)$.

2 Literature review

Randomized estimates of type (1) were suggested in [6] with test vectors X following the uniform distribution on a unit sphere. Then this estimate was generalized for the setting of normal and Rademacher vectors in [8]. Special setting of rank-1 test vectors and structured matrices A (admitting a Kroenecker product structure) has been considered in [5].

Another idea for improving (1) comes for symmetric matrices A from the power method. Kuczyński and Woźniakowski [12] studied the power method with random start, including probability bounds on how quickly the largest eigenvalue of a symmetric A is approximated. In particular, they showed that the expected error decays with each iteration proportional to the ratio $(\sigma_2/\sigma_1)^2$ and derived distributions for the approximation error after t steps [12]. Their work also compared power iteration with Lanczos (Krylov subspace) methods initiated by a random vector. Both of these methods are require sequential matvecs. Building on this idea, Hochstenbach [10] developed Krylov-based methods for constructing an upper bound on $\|A\|_2$ with high probability. At the same time, resulting estimates are relatively complicated, which justifies their utility mostly for the setting of small underestimation probabilities δ in (5). [9, 8] suggested using a block of multiple random vectors in parallel, that is, to compute $Y = AX$ for randomly generated $X \sim \mathbb{R}^{n\times k}$ consisting of k random vectors. Then $\|A\|_2$ is approximated by $\|Y\|_2$. Bujanović and Kressner [5] considered the problem os estimating $\|A\|_2$ for structured matrices A based on rank-one random vectors X. We briefly mention a number of papers, which focus on estimating Frobenius norm or trace of a matrix A, see Hutchinson's method [11] and its later developments [3, 14, 13].

3 Algorithm

We start from the upper bound on $\|A\|_2$ given by (3), that is, with $X \sim \mathcal{N}(0, \mathrm{I}_n)$, we consider the statistics $\theta \|AX\|_2$ and use an upper bound from [8]:

$$\mathbb{P}(\theta \|AX\|_2 \le \|A\|_2) \le \sqrt{\frac{2}{\pi}} \frac{1}{\theta}. \tag{6}$$

This estimator has the main advantage: right-hand side of (6) does not depend on the structure of the matrix A. At the same time, as already mentioned, for matrices A with large effective rank ρ, $\|AX\|_2$ significantly overestimates A. Tighter upper bounds on the underestimation probability in (6) typically depend on ρ and can not be applied without additional assumptions on A. This generally yields to too large values of θ and loose upper bound $\theta \|AX\|_2$. As a result, the estimator $\|AX\|$ is ineffective for matrices with low effective rank and overly conservative for those with high stable rank.

Our method. We propose to use the estimator, which is defined, for independent random vectors $X, Y \sim \mathcal{N}(0, \mathrm{I}_n)$ as

$$T_\theta^c(X, Y) = \theta \sqrt{\left(\frac{\|A^\top AY\|}{\|AY\|}\right)^2 + \|AX\|^2}. \tag{7}$$

The first additive term $\frac{\|A^\top AY\|}{\|AY\|}$ provides a tight (lower) bound on $\|A\|_2$ if its effective rank ρ is close to 1. In particular, if $\operatorname{rk} A = 1$, and A is symmetric, $\frac{\|A^\top AY\|}{\|AY\|} = \|A\|_2$. At the same time, $\|AX\|_2$ tends to overestimate $\|A\|_2$ as ρ increases.

Although $T_\theta^c(X, Y)$ is always greater than $\|AX\|$, we illustrate the benefits of using this estimator on Figure 1. It is important to notice that (7) requires to compute 3 matrix-vector products.

The pseudocode of the algorithm is summarized in Algorithm 1.

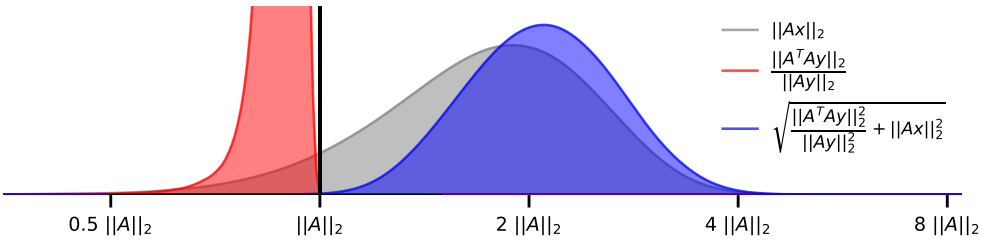

Figure 1: Density functions for $T_\theta^c(X)$ and its summands in (7). The first term is shifted to the right from $0.5 \|A\|_2$ and has negligible tails. This causes $T_\theta^c(X)$ to shift right as well, and its left tail becomes lighter than that of $\|Ax\|_2$.

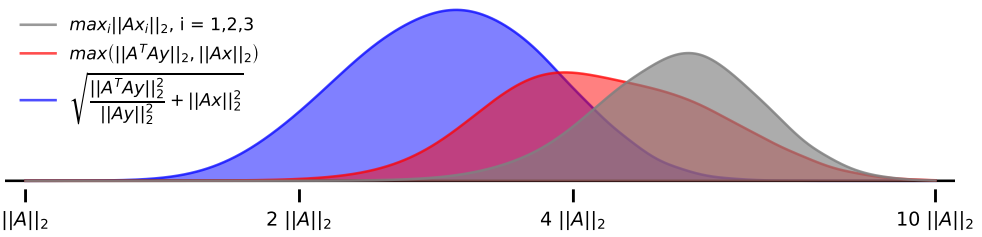

Figure 2: Density statistics using 3 matvecs and θ values for $p = 0.05$. Our method achieves the smallest $|\mathbb{E}(T_\theta(X)) - \|A\|_2|$ and a left-shifted distribution with a negligible right tail.

Algorithm 1 Operator Norm Estimation via Matvecs

Require: Implicit matrix $A \in \mathbb{R}^{n\times n}$; access to matvecs Ax and $A^\top x$
Require: Confidence level $p \in (0, 1)$
Require: Number of matvecs m; sample matrix $X \in \mathbb{R}^{n\times m}$ with columns $X_1, \ldots, X_m$
1: Generate sample matrix $X \in \mathbb{R}^{n\times m}$ (e.g., with i.i.d. $\mathcal{N}(0, 1)$ entries)
2: Compute the statistic $T^c_\theta(X)$ based on the samples
3: Determine the multiplier θ such that $\mathbb{P}(T^c_\theta(X) \le \|A\|_2) \le p$
4: **return** $T^c_\theta(X, Y)$

4 Theoretical analysis

In this section we provide a theoretical bounds, which allows to set the parameter θ in (7) in order to have a controllable underestimation probability. In the theorem below we use the notation $p_\zeta(t)$ for a density of random variable ζ at point t.

Theorem 1. *Let X, Y be independent $N(0, I_n)$ random vectors and $\xi_1, \eta \sim N(0, 1)$ be independent random variables, which are independent of X, Y. Then it holds that*

$$\mathbb{P}\left(T^c_\theta(X, Y) \le \|A\|_2\right) \le \begin{cases} \displaystyle\int_0^{\theta^{-2}} \chi^2\left(\frac{(\rho-1)t}{1-t}\right) p_{\xi_1^2+(\rho-1)\eta^2}\left(\theta^{-2} - t\right) dt, & (\rho-1) \ge \theta^{-2}, \\ \displaystyle\int_0^{\theta^{-2}} \chi^2\left(\frac{(\rho-1)t}{1-t}\right) p_{\xi_1^2}\left(\frac{\theta^{-2}-t}{\rho}\right) dt, & (\rho-1) < \theta^{-2}. \end{cases}$$

It is noticable that the right-hand side of the above inequality at fixed point θ depends only on the effective rank ρ, so this function is easy to maximize on a finite segment numerically. The table with the values of theta with respect to probabilities is given in the next section with experiments. Empirically, we also reduced the variance of the classic estimator expectancy on the matrix space, which is especially valuable for low-rank matrices. We highlight, that we aim to select θ, which guarantees that the right-hand side of the bound of Theorem 1 is smaller then a given $p \in (0, 1)$, *uniformly* in the effective rank ρ.

5 Numerics

We first describe our pipeline for comparing statistics $T_1(X)$ and $T_2(X)$, depending on their own set of parameters θ_1 and θ_2, respectively. We fix the desired underestimation probability $p \in (0, 1)$

(typically, we use $p = 0.05$). Namely, we find $\hat{\theta}_1$ and $\hat{\theta}_2$, that guarantees underestimation probability p and compare the empirical densities of T_1 and T_2.

Since we use two sequential matvecs and a single one on independent X, Y, our comparison involves two baseline algorithms. Vanilla estimator $\max_i \|AX_i\|_2$, $i \in \{1, 2, 3\}$, with underestimation probability $\sqrt{8/\pi^3}\theta^{-3}$. Dixon estimator $\max\left(\sqrt{\|A^\top AY\|_2}, \|AX\|_2\right)$, with underestimation probability $\frac{2}{\pi}\theta^{-3}$.

We evaluate our method on the following representative matrices:

- **Truncated SVD ($\rho = 0.05$, Olivetti Faces dataset)**: Matrix with the lowest ρ in our set. Used to demonstrate the algorithm's performance under small values of ρ.
- **Hilbert matrix ($\rho = 0.15$)**: A classical example of a matrix with small approximate rank.
- **Dominant matrix 0.1 ($\rho = 0.1$)**: Similarly to the Hilbert matrix in terms of ρ, but with lower variance due to more uniformly distributed singular values.
- **Dominant matrix 0.5 ($\rho = 2.5$)**: A mid-rank variant of the previous matrix with similarly low norm variance despite increased ρ.
- **Fréchet derivative matrix ($\rho = 40$)**: A high-rank matrix common in applications. Its singular values are uniformly distributed, showing no strong concentration.
- **ResNet convolution layer matrix ($\rho = 30$)**: Another high-rank matrix, typical in deep learning. Like the Fréchet matrix, its singular values are spread uniformly on $[0, 1]$.

This selection spans a wide range of ρ values and matrix types, covering diverse application contexts and spectral characteristics. We demonstrate the consistent advantages of our method across all the cases.

All figures use $\log_2$ scale to show both left and right tails in detail.

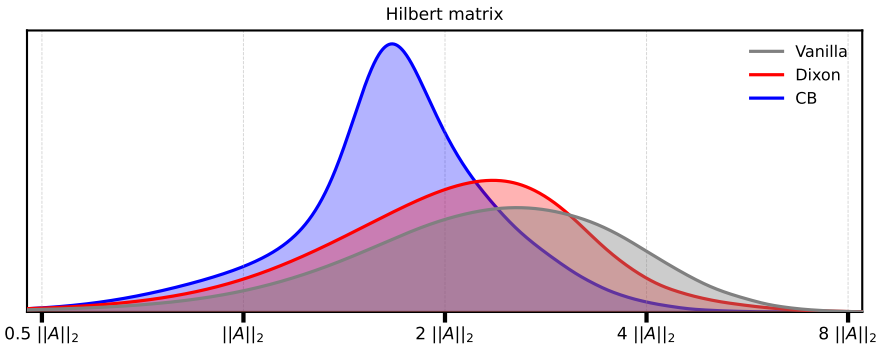

Figure 3: Empirical density for a Hilbert matrix of size 100. This matrix has $\rho \sim 1.15$, so the first term approximates $\|A\|_2$ well, and the second term does not overestimate it. This results in lighter tails and significantly reduced variance.

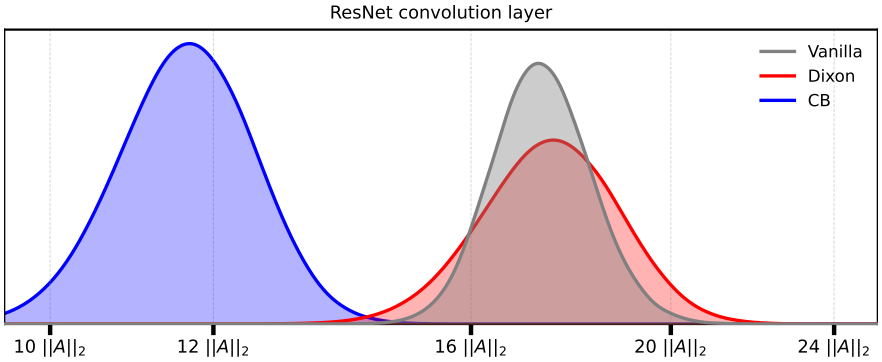

Figure 4: Density comparison for a convolution layer in ResNet. All algorithms overestimate $\|A\|_2$, but our method shows a left shift and lower variance.

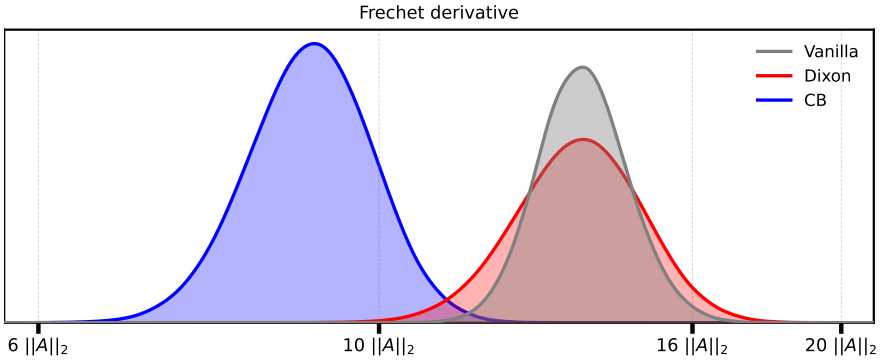

Figure 5: Density comparison for the Fréchet derivative matrix. Despite $\rho \gg 1$, our method maintains comparative advantages. Matrices with large stable ranks (e.g., Fréchet, convolution layers) show similar patterns.

In the third series of plots we show convergence of different algorithms as the number of matvecs increases. The line plot displays the mean estimate, and confidence bands reflect empirical variance. Since our method uses 3 matvecs per iteration, the number of total matvecs is a multiple of 3.

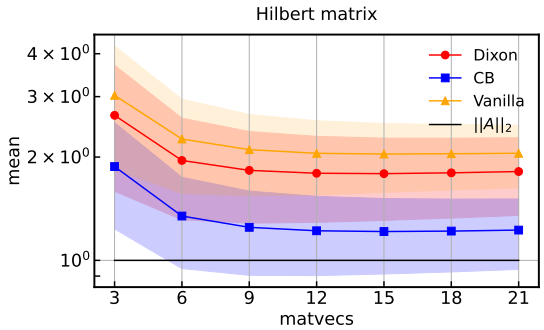

Figure 6: Convergence for the Hilbert matrix ($\rho = 1.15$).

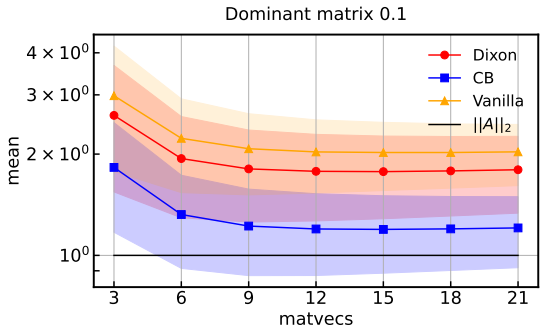

Figure 7: Convergence for a matrix with 10 singular values $\sigma_i^2/\sigma_1^2 = 0.1$ ($\rho = 1.05$).

p	$\theta_{\mathbf{cb}}$	$\theta_{\mathbf{v}}$
0.1	1.28	1.73
0.05	1.46	2.17
0.01	2.46	4.71
0.001	5.10	7.90

Table 1: Values of θ for Counterbalance and Vanilla estimators at different p.

As shown in Figures 6 and 7, all methods stabilize after 9 matvecs, but our method maintains a higher convergence rate for any number of matvecs. Despite structural differences, both matrices behave similarly.

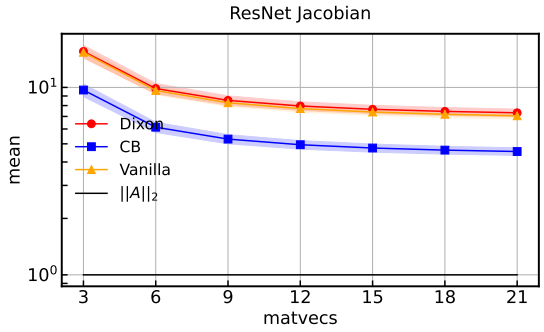

Figure 8: Convergence for the ResNet Jacobian matrix ($\rho = 7$).

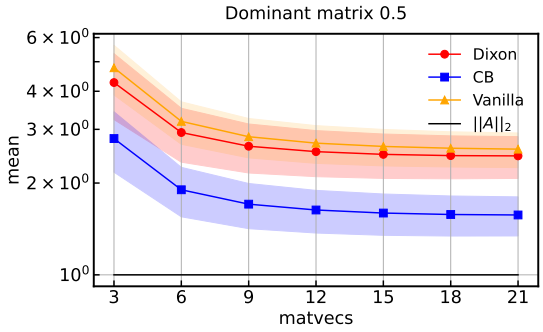

Figure 9: Convergence for matrix with 10 $\sigma_i^2/\sigma_1^2 = 0.5$ ($\rho = 1.8$).

Compared to 6 and 7, all estimators in Figures 8 and 9 show slower convergence on matrices with larger ρ. However, the comparative advantage of our method remains stable. Notably, for the ResNet Jacobian (largest ρ), our method achieves minimal variance—attributed to the suppression of the second term's tail.

6 Conclusion

We presented the Counterbalance estimator $T_\theta^c(X)$, a hybrid norm estimator that improves upon classical randomized methods by reducing underestimation probability across a wide range of matrix types. Our theoretical guarantees and empirical evaluations demonstrate that this approach achieves lower variance and tighter bounds without requiring structural knowledge of the matrix.

Beyond outperforming baseline estimators on both low- and high-rank matrices, our method remains computationally efficient, making it suitable for modern applications such as neural network analysis and large-scale matrix computations.

7 Acknowledgements

This work was supported by the Ministry of Economic Development of the Russian Federation (code 25-139-66879-1-0003).

References

[1] Awad H Al-Mohy and Nicholas J Higham. Computing the Fréchet derivative of the matrix exponential, with an application to condition number estimation. *SIAM J. Matrix Anal. Appl.*, 30(4):1639–1657, 2009.

[2] Awad H Al-Mohy, Nicholas J Higham, and Samuel D Relton. Computing the Fréchet derivative of the matrix logarithm and estimating the condition number. *SIAM J. Sci. Comput.*, 35(4):C394–C410, 2013.

[3] H. Avron and S. Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semidefinite matrix. *Journal of the ACM*, 58(2):8:1–8:17, 2011.

[4] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

[5] Z. Bujanović and D. Kressner. Norm and trace estimation with random rank-one vectors. *SIAM Journal on Matrix Analysis and Applications*, 42(1):202–223, 2021.

[6] John Dixon. Estimating extremal eigenvalues and condition numbers of matrices. *Siam*, pages 812–814, 1983.

[7] Henry Gouk, Eibe Frank, Bernhard Pfahringer, and Michael J Cree. Regularisation of neural networks by enforcing lipschitz continuity. *Machine Learning*, 110:393–416, 2021.

[8] N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.

[9] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.

[10] M. E. Hochstenbach. Probabilistic upper bounds for the matrix two-norm. *Journal of Scientific Computing*, 57(3):464–476, 2013.

[11] M. F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics - Simulation and Computation*, 19(2):433–450, 1990.

[12] J. Kuczyński and H. Woźniakowski. Estimating the largest eigenvalues by the power and lanczos algorithms with a random start. *SIAM Journal on Matrix Analysis and Applications*, 13(4):1094–1122, 1992.

[13] R. A. Meyer, C. Musco, C. Musco, and D. P. Woodruff. Hutch++: Optimal stochastic trace estimation. In *Proceedings of the SIAM Symposium on Simplicity in Algorithms (SOSA)*, pages 142–148, 2021. arXiv:2010.09649.

[14] F. Roosta-Khorasani and U. Ascher. Improved bounds on sample size for implicit matrix trace estimators. *Foundations of Computational Mathematics*, 15:1187–1212, 2015.

[15] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[16] Shaoxin Wang, Hu Yang, and Hanyu Li. Condition numbers for the nonlinear matrix equation and their statistical estimation. *Linear Algebra and its Applications*, 482:221–240, 2015.