On randomized trace estimates for indefinite matrices
with an application to determinants

Alice Cortinovis∗ Daniel Kressner†

Abstract

Randomized trace estimation is a popular and well studied technique that approximates the trace of a large-scale matrix $B$ by computing the average of $x^T B x$ for many samples of a random vector $X$. Often, $B$ is symmetric positive definite (SPD) but a number of applications give rise to indefinite $B$. Most notably, this is the case for log-determinant estimation, a task that features prominently in statistical learning, for instance in maximum likelihood estimation for Gaussian process regression. The analysis of randomized trace estimates, including tail bounds, has mostly focused on the SPD case. A notable exception is recent work by Ubaru, Chen, and Saad [34] on trace estimates for a matrix function $f(A)$ with Rademacher random vectors. In this work, we derive new tail bounds for randomized trace estimates applied to indefinite $B$ with Rademacher or Gaussian random vectors. These bounds significantly improve existing results for indefinite $B$, reducing the number of required samples by a factor $n$ or even more, where $n$ is the size of $A$. Even for an SPD matrix, our work improves an existing result by Roosta-Khorasani and Ascher [30] for Rademacher vectors.

This work also analyzes the combination of randomized trace estimates with the Lanczos method for approximating the trace of $f(A)$. Particular attention is paid to the matrix logarithm, which is needed for log-determinant estimation. We improve and extend an existing result, to not only cover Rademacher but also Gaussian random vectors.

Keywords: trace estimation, determinant, tail bounds, entropy method, Lanczos method.

AMS 2010 Subject Classification: 65C05, 65F40, 65F60, 68W20, 60E15.

1 Introduction

This paper is concerned with approximating the trace of a symmetric matrix $B \in \mathbb{R}^{n \times n}$ that is accessible only implicitly via matrix-vector products or, more precisely, (approximate) quadratic forms. If $X$ is a random vector of length $n$ such that $E[X] = 0$ and $E[XX^T] = I$, 

∗Institute of Mathematics, EPF Lausanne, 1015 Lausanne, Switzerland. E-mail: alice.cortinovis@epfl.ch.
The work of Alice Cortinovis has been supported by the SNSF research project Fast algorithms from low-rank updates, grant number: 200020_178806.
†Institute of Mathematics, EPF Lausanne, 1015 Lausanne, Switzerland. E-mail: daniel.kressner@epfl.ch.
then $\mathbb{E}[X^T B X] = \text{tr}(B)$. Based on this result, a stochastic trace estimator \cite{23} is obtained from sampling an average of $N$ quadratic forms:

$$\text{tr}_N(B) := \frac{1}{N} \sum_{i=1}^{N} (X^{(i)})^T B X^{(i)}, \quad (1)$$

where $X^{(i)}$, $i = 1, \ldots, N$, are independent copies of $X$. The most common choices for $X$ are standard Gaussian and Rademacher random vectors. The latter are defined by having i.i.d. entries that take values $\pm 1$ with equal probability. We will consider both choices in this paper and denote the resulting trace estimates by $\text{tr}_N^G(B)$ and $\text{tr}_N^R(B)$, respectively.

Hutchinson \cite{23} used $\text{tr}_N^R(B)$ to approximate the trace of the influence matrix of Laplacian smoothing splines. In this setting, $B = A^{-1}$ for a symmetric positive definite (SPD) matrix $A$ and, in turn, $A$ is SPD as well. Other applications, such as spectral density estimation \cite{27} and determinant computation \cite{7}, may feature a symmetric but indefinite matrix $B$. For approximating the determinant, one exploits the relation

$$\log(\det(A)) = \text{tr} (\log(A)),$$

where $\log(A)$ denotes the matrix logarithm of $A$. The need for estimating determinants arises, for instance, in statistical learning \cite{3, 13, 18}, lattice quantum chromodynamics \cite{33}, and Markov random fields models \cite{36}. Certain quantities associated with graphs can be formulated as determinants, such as the number of spanning trees and triangles, and various negative approximation results exist in this context; see, e.g., \cite{5, 14, 15, 29}. Relying on the Cholesky factorization, the exact computation of the determinant is often infeasible for a large matrix $A$. In contrast, the Hutchinson estimator combined with (2) bypasses the need for factorizing $A$ and instead requires to (approximately) evaluate the quadratic form $x^T \log(A)x$ for several vectors $x \in \mathbb{R}^n$. Compared to the task of estimating the trace of $A^{-1}$, the determinant computation via (2) is complicated by two issues: (a) Even when $A$ is SPD, the matrix $B = \log(A)$ may be indefinite; and (b) the quadratic forms $x^T \log(A)x$ themselves are expensive to compute exactly, so they need to be approximated. We mention in passing that there are other methods to approximate determinants, including subspace iteration \cite{31} and block Krylov methods \cite{26}, but they only work well in specific cases, e.g., when $A = \sigma I + C$ for a matrix $C$ of low numerical rank.

By the central limit theorem, the estimate (1) can be expected to become more reliable as $N$ increases; see, e.g., \cite{13, Corollaries 3.3 and 4.3} for such an asymptotic result as $N \to \infty$. Most existing non-asymptotic results for trace estimation are specific to an SPD matrix $B$; see \cite{6, 20, 30} for examples. They provide a bound on the estimated number $N$ of probe vectors to ensure a small relative error with high probability:

$$\mathbb{P} \left( \frac{|\text{tr}(B) - \text{tr}_N^G(B)|}{\text{tr}(B)} \geq \varepsilon \right) \leq \delta; \quad (3)$$

see Remark \cite{4} below for a specific example. As already mentioned, the assumption that $B$ is SPD is usually not met when computing the determinant of an SPD matrix $A$ via
\text{tr}(\log(A)) \text{ because this would require all eigenvalues of } A \text{ to be larger than one. For general indefinite } B, \text{ it is unrealistic to aim at a bound of the form } \text{(3)} \text{ for the relative error, because } \text{tr}(B) = 0 \text{ does not imply zero error. Ubaru, Chen, and Saad} \text{ instead derive a bound for the absolute error via rescaling, that is, the results from } \text{(3)} \text{ are applied to the matrix } C := -\log(\lambda A) \text{ for a value of } \lambda > 0 \text{ that ensures } C \text{ to be SPD. Specifically, for Rademacher vectors it is shown in } \text{(34, Corollary 4.5)} \text{ that}

\begin{equation}
P\left( | \text{tr}_N^R(\log(A)) - \log(\det(A)) | \geq \varepsilon \right) \leq \delta \tag{4}
\end{equation}

\text{is satisfied with fixed failure probability } \delta \text{ if the number of samples } N \text{ grow at most quadrat-
ically with } n. \text{ Unfortunately, this estimated number of samples compares unfavorably with a much simpler approach; computing the trace from the diagonal elements of } \log(A) \text{ only requires the evaluation of } n \text{ quadratic forms, using all } n \text{ unit vectors of length } n. \text{ To approximate the quadratic forms } x^T \log(A)x, \text{ a polynomial approximation of the logarithm can be used, see } \text{(21, 28)} \text{ for approximation by Chebyshev expansion/interpolation and } \text{(8, 12, 38)} \text{ for approximation by Taylor series expansion. Often, a better approximation can be obtained by the Lanczos method, which is equivalent to applying Gaussian quadrature to the integral } \int \log(x) d\mu(x) \text{ on the spectral interval of } A, \text{ for a suitably defined measure } \mu; \text{ see } \text{(19)} \text{. In this case, upper and lower bounds for the quantity } x^T \log(A)x \text{ can be determined without much additional effort } \text{(7)}. \text{ Moreover, the convergence of Gaussian quadrature for the quadratic form can be related to the best polynomial approximation of the logarithm on the spectral interval of } A; \text{ see } \text{(34, Theorem 4.2)}. \text{ By combining the polynomial approximation error with } \text{(4)}, \text{ one obtains a total error bound that takes into account both sources of errors. Such a result is presented in } \text{(34, Corollary 4.5)} \text{ for Rademacher vectors; the fact that all such vectors have bounded norm is essential in the analysis.}

\text{In this paper, we improve the results from } \text{(34)} \text{ by first showing that the number of samples required to achieve } \text{(4)} \text{ is much lower. In particular, we show for a general symmetric matrix } B \text{ that}

\begin{equation}
P\left( | \text{tr}_N^{G,R}(B) - \text{tr}(B) | \geq \varepsilon \right) \leq \delta \tag{5}
\end{equation}

\text{is satisfied with fixed failure probability } \delta \text{ if the number of samples } N \text{ grows proportionally
with the stable rank } \rho(B) := ||B||^2_F/||B||^2_2; \text{ as } \rho(B) \in [1, n], \text{ the growth is at most linear in } n \text{ (instead of quadratic). We derive such a result for both, Gaussian and Rademacher vectors, and demonstrate that the dependence on } n \text{ is asymptotically tight with an explicit example. For SPD matrices, our bound also improves the state-of-the-art result } \text{(30, Theorem 1)} \text{ for Rademacher vectors by establishing that the number of probe vectors is inversely proportional to the stable rank.}

\text{Specialized to determinant computation, we combine our results with an improved analysis of the Lanczos method, to get a sharper total error bound for Rademacher vectors. Finally, we extend this combined error bound to Gaussian vectors, which requires some additional consideration because of the unboundedness of such vectors. We remark that some of our results are potentially of wider interest, beyond stochastic trace and determinant estimation, such as a tail bound for Rademacher chaos (Theorem 8) and an error bound (Corollary 15 combined with Corollary 23) on the polynomial approximation of the logarithm.}
2 Tail bounds for trace estimates

In this section we derive tail bounds of the form (5) for the stochastic trace estimator applied to a symmetric, possibly indefinite matrix $B \in \mathbb{R}^{n \times n}$. We will analyze Gaussian and Rademacher vectors separately. In the following, we will frequently use a spectral decomposition $B = Q\Lambda Q^T$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ contains the eigenvalues of $B$ and $Q$ is an orthogonal matrix.

2.1 Standard Gaussian random vectors

The case of Gaussian vectors will be addressed by using a tailbound for sub-Gamma random variables, which follows from Chernoff bounds; see, e. g., [11].

**Definition 1.** A random variable $X$ is called sub-Gamma with variance parameter $\nu > 0$ and scale parameter $c > 0$ if

$$
\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{\nu \lambda^2}{2(1 - c\lambda)}\right) \quad \text{for all } 0 < \lambda < \frac{1}{c}.
$$

**Lemma 2 ([11, Section 2.4]).** Let $X$ be a sub-Gamma random variable with parameters $(\nu, c)$. Then, for all $\varepsilon \geq 0$, we have

$$
\mathbb{P}(X \geq \sqrt{2\varepsilon \nu + c\varepsilon}) \leq \exp(-\varepsilon).
$$

**Lemma 3 ([35, Proposition 2.10]).** Let $X$ be a random variable such that $\mathbb{E}[X] = 0$, and such that both $X$ and $-X$ are sub-Gamma with parameters $(\nu, c)$. Then, for all $\varepsilon \geq 0$, we have

$$
\mathbb{P}(|X| \geq \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{2(\nu + c\varepsilon)}\right).
$$

Lemma 3 implies the following result for the tail of a single-sample trace estimate. This result is similar, but not identical, to [11, Example 2.12] and [25, Lemma 1], which apply to symmetric matrices with zero diagonal and SPD matrices, respectively.

**Lemma 4.** For a Gaussian vector $X$ of length $n$ we have

$$
\mathbb{P}\left(|X^T BX - \text{tr}(B)| \geq \varepsilon\right) \leq 2 \exp\left(-\frac{\varepsilon^2}{4\|B\|_F^2 + 4\varepsilon\|B\|_2}\right) \quad (6)
$$

for all $\varepsilon > 0$.

**Proof.** We let

$$
Y := X^T BX - \text{tr}(B) = X^T Q\Lambda Q^T X - \text{tr}(B) = \sum_{i=1}^{n} \lambda_i (Z_i^2 - 1),
$$

where $Z_i$ are independent standard normal random variables.
where $Z_i \sim \mathcal{N}(0, 1)$ is the $i$th component of the Gaussian vector $Q^TX$. To show that $Y$ is sub-Gamma, we define for $\lambda \in \mathbb{R}$ the function

$$\psi(\lambda) := \log \mathbb{E}[\exp(\lambda(Z^2 - 1))], \quad Z \sim \mathcal{N}(0, 1).$$

By direct computation, it follows that $\psi(\lambda) = -\lambda - \frac{1}{2} \log(1 - 2\lambda)$ for $\lambda < \frac{1}{2}$. In particular, this implies $\psi(\lambda) \leq \frac{\lambda^2}{1 - 2\lambda}$ for $0 \leq \lambda < \frac{1}{2}$, and $\psi(\lambda) \leq \lambda^2 \leq \frac{\lambda^2}{1 + c\lambda}$ for all $c > 0$ when $\lambda < 0$. Using the independence of $Z_i$ for different $i$ we obtain

$$\log \mathbb{E}[\exp(\lambda Y)] = \sum_{i=1}^{n} \log \mathbb{E}[\exp(\lambda \lambda_i(Z_i^2 - 1))] = \sum_{i=1}^{n} \psi(\lambda \lambda_i) \leq \sum_{i=1}^{n} \frac{\lambda_i^2 \lambda^2}{1 - 2\lambda_i} \leq \frac{\|B\|^2 \lambda^2}{1 - 2\|B\|2\lambda}$$

for $0 < \lambda < \frac{1}{2\|B\|2}$. This shows that $Y$ is sub-Gamma with parameters $(\nu, c) = (2\|B\|^2_F, 2\|B\|_2)$. Moreover, $-Y = X^T(-B)X - \text{tr}(-B)$ is also sub-Gamma with the same parameters. Because $\mathbb{E}[Y] = 0$, Lemma 4 implies the desired result.

A diagonal embedding trick turns Lemma 4 into a tail bound for the stochastic trace estimator (1).

**Theorem 5.** Let $B \in \mathbb{R}^{n \times n}$ be symmetric. Then

$$\mathbb{P}\left(\left|\text{tr}^N(B) - \text{tr}(B)\right| \geq \varepsilon\right) \leq 2 \exp\left(-\frac{N\varepsilon^2}{4\|B\|^2_F + 4\varepsilon\|B\|_2}\right)$$

for all $\varepsilon > 0$. In particular, for $N \geq \frac{4\varepsilon^2(\|B\|^2_F + \varepsilon\|B\|_2)}{\log \frac{2}{\delta}}$ it holds that $\mathbb{P}(\left|\text{tr}^N(B) - \text{tr}(B)\right| \geq \varepsilon) \leq \delta$.

**Proof.** We apply Lemma 4 to the matrix

$$B := \text{diag}(N^{-1}B, \ldots, N^{-1}B) \in \mathbb{R}^{Nn \times Nn},$$

that is, the block diagonal matrix with the $N$ diagonal blocks containing rescaled copies of $B$. In turn, the trace estimate (1) equals $X^T B X$ for a Gaussian vector $X$ of length $Nn$. Noting that $\|B\|_F^2 = N^{-1/2}\|B\|_F^2$ and $\|B\|_2 = N^{-1}\|B\|_2$, the first part of the corollary follows from Lemma 4. Setting

$$\delta := 2 \exp\left(-\frac{\varepsilon^2}{4\|B\|^2_F + 4\varepsilon\|B\|_2}\right) = 2 \exp\left(-\frac{N\varepsilon^2}{4\|B\|^2_F + 4\varepsilon\|B\|_2}\right)$$

we obtain $N = \frac{4\varepsilon^2}{\delta}(\|B\|^2_F + \varepsilon\|B\|_2)^{1/2}$.

We recall that the stable rank of $B$ is defined as $\rho = \|B\|^2_F / \|B\|^2$ and satisfies $\rho \in [1, n]$. In particular, $\rho(B) = 1$ when $B$ has rank one and $\rho(B) = n$ when all singular values are equal. Intuitively, $\rho(B)$ tends to be large when $B$ has many singular values not significantly
smaller than the largest one. The minimum number of probe vectors required by Theorem 5 depends on the stable rank of $B$ in the following way:

$$\frac{4}{\varepsilon^2} (\rho \|B\|_2^2 + \varepsilon \|B\|_2) \log \frac{2}{\delta} \leq \frac{4}{\varepsilon^2} (n \|B\|_2^2 + \varepsilon \|B\|_2) \log \frac{2}{\delta}.$$ 

The upper bound indicates that $N$ may need to be chosen proportionally with $n$ to reach a fixed (absolute) accuracy $\varepsilon$ with constant success probability, provided that $\|B\|_2$ remains constant as well. The following lemma shows for a simple matrix $B$ that such a linear growth of $N$ can actually not be avoided.

**Lemma 6.** Let $n$ be even and consider the traceless matrix $B = \begin{bmatrix} I_n & 0 \\ 0 & -I_n \end{bmatrix}$. Then, for every $\varepsilon > 0$, it holds that

$$\mathbb{P}(|\text{tr}_N(B)| \leq \varepsilon) \leq \varepsilon \sqrt{\frac{N}{\pi n}}.$$

**Proof.** By the definition of $B$, the trace estimate takes the form

$$\text{tr}_N(B) = \frac{1}{N} \left( \sum_{i=1}^{\frac{nN}{2}} X_i^2 - \sum_{j=1}^{\frac{nN}{2}} Y_j^2 \right)$$

for independent $X_i, Y_j \sim N(0, 1)$. In other words,

$$N \cdot \text{tr}_N(B) = X - Y,$$

where $X, Y \sim \chi^2 \left( \frac{nN}{2} \right)$ are independent Chi-squared random variables with $\frac{nN}{2}$ degrees of freedom. The probability density function $f$ of $Z = X - Y$ can be expressed as

$$f(z) = \frac{1}{2^{nN/4} \sqrt{\pi nN/4}} |z|^{nN/4 - 1} K_{nN/4 - 1}(|z|),$$

where $K_{nN/4 - 1/2}$ is a modified Bessel function of the second kind. In particular,

$$f(0) = \frac{1}{4\sqrt{\pi}} \frac{\Gamma \left( \frac{nN}{4} - \frac{1}{2} \right)}{\Gamma \left( \frac{nN}{4} \right)} = \frac{1}{4\sqrt{\pi}} \frac{\sqrt{\pi}}{2^{nN/4} - 2} \left( \frac{nN}{2} - 2 \right) \left( \frac{nN}{4} - 1 \right) \leq \frac{1}{2\sqrt{\pi nN}},$$

where we used the duplication formula for Gamma functions and the inequality $\frac{1}{2\sqrt{\pi}} \left( \frac{2k}{k} \right) \leq \frac{1}{\sqrt{\pi k}}$, see [2].

As $f$ is an autocorrelation function (of the density function of a Chi-squared variable with $nN/2$ degrees of freedom), its maximum is at 0. We can therefore estimate the probability of $X - Y$ being in the interval $[-N\varepsilon, N\varepsilon]$ in the following way:

$$\mathbb{P}(|\text{tr}_N(B)| \leq \varepsilon) = \mathbb{P}(|X - Y| \leq N\varepsilon) \leq 2N\varepsilon f(0) \leq \varepsilon \sqrt{\frac{N}{\pi n}}.$$ 

\[ \square \]
Figure 1: Asterisks: Errors $|\text{tr}^G_{10}(B) - \text{tr}(B)|$ of 100 samples for each $n = 2^k$ with $k = 2, \ldots, 23$, where $B$ is the matrix from Lemma 6. Blue line: Error bound $\varepsilon(B, 0.01, 10)$ from (8).

We can reformulate Theorem 5 in such a way that, given a number $N$ of probe vectors and a failure probability $\delta \in (0, 1)$, we have $\varepsilon = \varepsilon(B, N, \delta)$ such that with probability at least $1 - \delta$ one has $\text{tr}^G_N(B) \in [\text{tr}(B) - \varepsilon, \text{tr}(B) + \varepsilon]$. The random variable $X^T B X - \text{tr}(B)$, where $B$ is defined as in (7) and $X$ is a Gaussian vector of length $nN$, is sub-Gamma with parameters $(2 \frac{\|B\|_F^2}{N}, 2 \frac{\|B\|_F^2}{N})$, and the same holds for $-X^T B X$. By Lemma 2 we have

$$\varepsilon \equiv \varepsilon(B, \delta, N) = \frac{2}{\sqrt{N}} \|B\|_F \sqrt{\log \frac{2}{\delta} + \frac{2}{N} \|B\|_2 \log \frac{2}{\delta}} \leq 2 \sqrt{\log \frac{2}{\delta} + \frac{2}{N} \log \frac{2}{\delta}} \|B\|_2. \hspace{1cm} (8)$$

As the example in Lemma 6 shows, the potential growth of $\varepsilon$ with $\sqrt{n}$ cannot be avoided in general. Figure 1 illustrates this growth.

Also the dependence of $N$ on $\log \frac{2}{\delta}$ and $\frac{1}{\varepsilon^2}$ in Theorem 5 cannot be improved, as shown in [37, Theorem 3].

Remark 7. For a nonzero symmetric positive semidefinite (SPSD) matrix $B$, Theorem 5 yields the following relative error estimate:

$$\mathbb{P}\left( \frac{|\text{tr}^G_N(B) - \text{tr}(B)|}{\text{tr}(B)} \geq \varepsilon \right) \leq \delta \quad \text{for} \quad N \geq \frac{4}{\varepsilon^2} \left(1 + \varepsilon\right) \rho^{-1} \log \frac{2}{\delta}. \hspace{1cm} (9)$$

State-of-the-art results of a similar form are Theorem 3 in [30], which requires $N \geq \frac{8}{\varepsilon^2} \mu \log \frac{2}{\delta}$, with $\mu := \frac{\|B\|_2}{\text{tr}(B)} \leq 1$, and Corollary 3.3 in [20], which requires $N \geq \frac{2}{\varepsilon^2} \rho^{-1} \log \frac{2}{\delta}$ and $\varepsilon \in (0, \frac{1}{2})$. Compared to [30], our result imposes no restriction on $\varepsilon$ at the expense of a somewhat larger constant. On the other hand, because $\mu$ can be much smaller than $\rho^{-1}$, the result from [30] is more favorable than ours for SPSD matrices.
2.2 Rademacher random vectors

The quadratic form $X^T BX$ for a Rademacher vector $X$ is called Rademacher chaos of order 2. We will first consider the homogeneous case, corresponding to a matrix $B$ with zero diagonal, which has been studied extensively in the literature, see, e.g., [11, 17, 22, 24, 32]. The non-homogeneous case is easily obtained from the homogeneous case; see Corollary 10 below. We make use of the the entropy method [11] to establish the following tail bound for a single-sample trace estimate.

**Theorem 8.** Let $X$ be a Rademacher vector of length $n$ and let $B$ be a nonzero symmetric matrix such that $B_{ii} = 0$ for $i = 1, \ldots, n$. Then, for all $\varepsilon > 0$,

$$P(\lvert X^T BX \rvert \geq \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{8\|B\|_F^2 + 8\varepsilon\|B\|_2}\right).$$  \hspace{1cm} (10)

**Proof.** The proof follows closely [3, Theorem 6] and [11, Theorem 17]; see Remark 9 for a comparison with these results. The main idea of the proof is as follows. Using the logarithmic Sobolev inequalities discussed in the appendix, a bound on the entropy of the random variable $X^T BX$ is obtained. Using a (modified) Herbst argument, we derive a bound on the moment generating function (MGF) of $X^T BX$, establishing that it is sub-Gamma with certain constants, which then allows us to apply Lemma [3].

Without loss of generality, we may assume $\|B\|_2 = 1$; the general case follows from applying the result to $\tilde{B} := B/\|B\|_2$. Let us consider the function $f : \{-1, 1\}^n \to \mathbb{R}$ defined as

$$f(x) = x^T Bx = \sum_{i \neq j} x_i x_j B_{ij}.$$  

We want to apply the logarithmic Sobolev inequality (22) from Theorem 21 to $f(X)$. For this purpose, we let

$$\tilde{X}^{(i)} = [X_1, \ldots, X_{i-1}, -X_i, X_{i+1}, \ldots, X_n]^T = X - 2X_i e_i, \quad i = 1, \ldots, n,$$

where $e_i$ denotes the $i$th unit vector. Using that $B$ has zero diagonal entries, we obtain

$$f(X) - f(\tilde{X}^{(i)}) = \langle BX, X \rangle - \langle BX - 2X_i Be_i, X - 2X_i e_i \rangle = 4X_i \langle Be_i, X \rangle.$$  

Therefore, denoting

$$Y := \|BX\|_2^2 = \frac{1}{16} \sum_{i=1}^n \left( \sum_{j=1}^n B_{ij} X_j \right)^2,$$

Theorem 21 establishes, for all $\lambda > 0$,

$$\mathbb{H}(\exp(\lambda f(X))) \leq 2\lambda^2 \mathbb{E}[Y \exp(\lambda f(X))].$$  \hspace{1cm} (11)

The decoupling inequality in [17, Lemma 8.50], which follows from Jensen’s inequality, gives

$$\lambda \mathbb{E}[Y \exp(\lambda f(X))] \leq \mathbb{H}(\exp(\lambda f(X))) + \mathbb{E}[\exp(\lambda f(X))] \log \mathbb{E}[\exp(\lambda Y)].$$

8
Combined with (11), this implies
\[ \mathbb{H}(\exp(\lambda f(X))) \leq \frac{2\lambda}{1-2\lambda} \mathbb{E}[\exp(\lambda f(X))] \cdot \log \mathbb{E}[\exp(\lambda Y)] \text{ for } 0 < \lambda < \frac{1}{2}. \] (12)

To find an upper bound on the MGF of \( Y \), we use again a logarithmic Sobolev inequality, then transform the obtained bound on the entropy into a bound on the MGF by Herbst argument. We do so by applying the inequality (21) from Theorem 21 to the function \( h : \mathbb{R}^n \to \mathbb{R} \) defined by \( h(x) := \|Bx\|_2^2 \). For this purpose, note that
\[
h(X) - h(\bar{X}^{(i)}) = \langle BX, BX \rangle - \langle B\bar{X}^{(i)}, B\bar{X}^{(i)} \rangle = \langle B(X - \bar{X}^{(i)}), B(X + \bar{X}^{(i)}) \rangle
\[
= 4\langle X_i B e_i, B X - X_i B e_i \rangle \leq 4\langle B e_i, B X \rangle^2 \leq 16\|B\|_F^2 \|B X\|_2^2.
\]
and, hence,
\[
\sum_{i=1}^n \left( h(X) - h(\bar{X}^{(i)}) \right)^2 \leq 16 \sum_{i=1}^n \langle B e_i, B X \rangle^2 = 16\|B^T B X\|_2^2 \leq 16\|B X\|_2^2.
\]
Therefore Theorem 21 gives
\[ \mathbb{H}(\exp(\lambda Y)) \leq 4\lambda^2 \mathbb{E}[Y \exp(\lambda Y)]. \]

Letting \( g(\lambda) := 4\mathbb{E}[Y \exp(\lambda Y)] / \mathbb{E}[\exp(\lambda Y)] \), we have obtained a bound of the form (20), as required by Lemma 20. Note that \( g(\lambda) = 4\psi'(\lambda) \), where \( \psi(\lambda) := \log \mathbb{E}[\exp(\lambda Y)] \). The result of Lemma 20 gives
\[ \log \mathbb{E}[\exp(\lambda Y)] \leq \frac{\lambda}{1-4\lambda}\|B\|_F^2 \text{ for } \lambda \in \left( 0, \frac{1}{4} \right). \]

Inserting this inequality into (12) gives
\[ \mathbb{H}(\exp(\lambda f(X))) \leq \frac{2\lambda^2\|B\|_F^2}{(1-4\lambda)(1-2\lambda)} \mathbb{E}[\exp(\lambda f(X))] \text{ for } \lambda \in \left( 0, \frac{1}{4} \right). \]

The random variable \( f(X) \) satisfies (20) for the function \( g(\lambda) := \frac{2\|B\|_F^2}{(1-4\lambda)(1-2\lambda)} \) in the interval \([0, 1/4]\). Recalling that \( \mathbb{E}[f(X)] = 0 \), the result of Lemma 20 gives
\[ \log \mathbb{E}[\exp(\lambda f(X))] \leq \lambda\|B\|_F^2 \log \frac{1-2\lambda}{1-4\lambda} \leq \frac{2\|B\|_F^2 \lambda^2}{1-4\lambda}, \quad \lambda \in [0, 1/4), \]
where we used \( \log(1 + x) \leq x \) in the last inequality.

Replacing \( f \) by \( -f \) and \( B \) by \( -B \), we also obtain
\[ \log \mathbb{E}[\exp(-\lambda f(X))] \leq \frac{2\|B\|_F^2 \lambda^2}{1-4\lambda}, \quad \lambda \in [0, 1/4). \]

Therefore the random variables \( f(X) \) and \( -f(X) \) are sub-Gamma with parameters \((4\|B\|_F^2, 4)\).

Applying Lemma 3 concludes the proof. \( \square \)
Remark 9. The proof of Theorem 8 follows the proof of 3, Theorem 6], which in turn refines a result from [10, Theorem 17] (see also [11]) by substituting the more general logarithmic Sobolev inequality from [10]. Proposition 10] with the ones from Theorem 21 specific for Rademacher random variables. However, let us stress that the results in [3, 10] feature larger constants partly because they deal with the more general Rademacher chaos.

\[ f(X) = \sup_{B \in \mathcal{B}} \sum_{i \neq j} X_i X_j B_{ij}, \]  

(13)

where \( \mathcal{B} \) is a set of symmetric matrices with zero diagonal. Restricted to the case \( \mathcal{B} = \{ B \} \), the results stated in 3, Theorem 6] and [11, Exercise 6.9] give \( \mathbb{P}(\|X^T BX\| \geq \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{16\|B\|_F^2 + 16\|B\|_2^2}\right) \) and \( \mathbb{P}(\|X^T BX\| \geq \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{32\|B\|_F^2 + 128\|B\|_2^2}\right) \), respectively. Proposition 8.13 in [17], which also aims at the more general \( \mathbb{P}(\|X^T BX\| \geq \varepsilon) \leq 2 \exp\left(-\min\left\{\frac{\varepsilon^2}{128\|X\|_F^2}, \frac{\varepsilon^2}{32\|X\|_2^2}\right\}\right) \).

As for Gaussian vectors, the result of Theorem 8 can be turned into a tail bound for \( \text{tr}_N^R(B) \) by block diagonal embedding. In the following, let \( D_B \) denote the diagonal matrix containing the diagonal entries of \( B \).

Corollary 10. Let \( B \) be a nonzero symmetric matrix. Then

\[ \mathbb{P}(\|\text{tr}_N^R(B) - \text{tr}(B)\| \geq \varepsilon) \leq 2 \exp\left(-\frac{N\varepsilon^2}{8\|B - D_B\|_F^2 + 8\|B - D_B\|_2}\right) \]

for every \( \varepsilon > 0 \). In particular, for

\[ N \geq \frac{8}{\varepsilon^2} \left(\|B - D_B\|^2_F + \varepsilon\|B - D_B\|_2\right) \log \frac{2}{\delta} \]

it holds that \( \mathbb{P}(\|\text{tr}_N^R(B) - \text{tr}(B)\| \geq \varepsilon) \leq \delta \).

Proof. Let \( C := B - D_B \) and \( C := \text{diag}(N^{-1}C, \ldots, N^{-1}C) \in \mathbb{R}^{Nn \times Nn} \). Then, \( \text{tr}_N^R(B) - \text{tr}(B) = X^T CX \) for a Rademacher vector \( X \) of length \( Nn \). The matrix \( C \) has zero diagonal, \( \|C\|_F = N^{-1/2} \), and \( \|C\|_2 = N^{-1}\|C\|_2 \). Now, the first part of the corollary directly follows from Theorem 8. Imposing a failure probability of \( \delta \) in (10) gives

\[ \delta := 2 \exp\left(-\frac{\varepsilon^2}{8\|C\|^2_F + 8\varepsilon\|C\|_2}\right) = 2 \exp\left(-\frac{N\varepsilon^2}{8\|C\|^2_F + 8\varepsilon\|C\|_2}\right), \]

and hence \( N = \frac{\delta}{\varepsilon^2} \left(\|C\|^2_F + \varepsilon\|C\|_2\right) \log \frac{2}{\delta} \). \( \square \)

Remark 11. It is instructive to compare the result of Corollary 10 to the straightforward application of Bernstein’s inequality, which gives

\[ \mathbb{P}(\|\text{tr}_N^R(B) - \text{tr}(B)\| \geq \varepsilon) \leq 2 \exp\left(-\frac{N\varepsilon^2}{4\|B - D_B\|^2_F + 4\varepsilon\|B - D_B\|_2}\right). \]

Clearly, a disadvantage of this bound is the explicit dependence of the denominator on \( n \), which does not appear in Corollary 10.
An alternative expression for the lower bound on $N$ is obtained by noting that $\|B - DB\|_F \leq \|B\|_F$ and $\|B - DB\|_2 \leq 2\|B\|_2$ (the factor 2 in the latter inequality is asymptotically tight, see, e.g., [9]). The result of Corollary [10] thus states that $N$ needs to be at least in the following way:

$$\frac{8}{\varepsilon^2}(\rho\|B\|^2 + 2\varepsilon\|B\|_2) \log \frac{2}{\delta} \leq \frac{8}{\varepsilon^2}(n\|B\|^2 + 2\varepsilon\|B\|_2) \log \frac{2}{\delta},$$

where $\rho$ is the stable rank of $B$. In analogy to the Gaussian case, the following lemma shows that a potential linear dependence of $N$ on $n$ cannot be avoided in general.

**Lemma 12.** Let $n$ be even and consider the traceless matrix $B = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$. Then

$$\Pr\left( |\text{tr}_N R(B)\| \leq \varepsilon \right) \leq \varepsilon \sqrt{\frac{N}{\pi n}},$$

for every $\varepsilon > 0$.

**Proof.** We first note that $\text{tr}_N R(B) = \frac{2}{N} \sum_{i=1}^{nN/2} Z_i$ with independent Rademacher random variables $Z_i$. In turn, $\Pr\left( |\text{tr}_N R(B)\| \leq \varepsilon \right) = \Pr\left( \left| \sum_{i=1}^{nN/2} Z_i \right| \leq \frac{N\varepsilon}{2} \right)$ equals the probability that the number of variables satisfying $Z_i = 1$ is at least $\frac{n-\varepsilon}{4} N$ and at most $\frac{n+\varepsilon}{4} N$. Therefore,

$$\Pr\left( |\text{tr}_N R(B)\| \leq \varepsilon \right) \leq \frac{1}{2^{nN/2}} \sum_{i=\frac{N}{2}(1-\varepsilon)}^{\frac{N}{2}(1+\varepsilon)} \binom{nN/2}{i} \leq \frac{N\varepsilon}{2} \cdot \frac{1}{2^{nN/2}} \cdot \binom{nN/2}{nN/4} \leq \frac{N\varepsilon}{2} \cdot \frac{2}{\sqrt{\pi nN}} = \varepsilon \sqrt{\frac{N}{\pi n}},$$

where we used the inequality $\frac{1}{2\pi} \left( \frac{2k}{\pi} \right) \leq \frac{1}{\sqrt{\pi k}}$.

We do not report a figure analogous to Figure [1] because the observed errors are very similar to the Gaussian case.

For SPSD matrices, a relative error estimate follows from Corollary [10] similarly to what has been discussed in Remark [7] for Gaussian vectors. We recall that $\rho = \|B\|^2_F/\|B\|^2_2$ denotes the stable rank of $B$.

**Corollary 13.** For a nonzero SPSD matrix $B$, we have

$$\Pr\left( \frac{|\text{tr}_N R(B) - \text{tr}(B)|}{\text{tr}(B)} \geq \varepsilon \right) \leq \delta \quad \text{for} \quad N \geq \frac{8}{\varepsilon^2} (1 + \varepsilon)\rho^{-1} \log \frac{2}{\delta}. $$

11
Proof. First of all, it is immediate that \( \|B - D_B\|_F \leq \|B\|_F \). As shown, e.g., in [3], Theorem 4.1, the same holds for the spectral norm when \( B \) is SPSD. For convenience, we provide a short proof: For every \( y \in \mathbb{R}^n \) it holds that

\[
|y^T (B - D_B) y| \leq \max \{ y^T B y, y^T D_B y \} \leq \max \{ \|B\|_2, \|D_B\|_2 \} \leq \|B\|_2,
\]

where the first inequality uses that both \( y^T B y \) and \( y^T D_B y \) are nonnegative. By taking the maximum with respect to all vectors of norm 1 one obtains \( \|B - D_B\|_2 \) on the left-hand side, which shows that it is bounded by \( \|B\|_2 \).

Now, Corollary 10 implies that \( P \left( |\text{tr}(R_N(B)) - \text{tr}(B)| \geq \varepsilon \text{tr}(B) \right) \leq \delta \) for

\[
N \geq \frac{8}{\varepsilon^2 \text{tr}(B)^2} \left( \|B\|_F^2 + \varepsilon \text{tr}(B) \|B\|_2 \right) \log \frac{2}{\delta}.
\]

The proof is concluded by noting that \( \|B\|_F^2 / \text{tr}(B)^2 \leq \rho^{-1} \) and \( \|B\|_2 / \text{tr}(B) \leq \rho^{-1} \).

Corollary 13 improves the result from [30, Theorem 1], which requires \( N \geq \frac{\rho}{\varepsilon^2} \log \frac{2}{\delta} \); a lower bound that does not improve as the stable rank of \( B \) increases.

3 Lanczos method to approximate quadratic forms

Let us now consider the problem of estimating the log determinant through \( \log(\det(A)) = \text{tr}(\log(A)) \), or more generally the problem of computing the trace of \( f(A) \) for an analytic function \( f \).

Applying a trace estimator to \( \text{tr}(f(A)) \) requires the (approximate) computation of the quadratic forms \( x^T f(A) x \) for fixed vectors \( x \in \mathbb{R}^n \). Following [34], we use the Lanczos method, Algorithm 1, for this purpose.

**Algorithm 1** Lanczos method to approximate quadratic form \( x^T f(A) x \)

**Input:** Matrix \( A \in \mathbb{R}^{n \times n} \), nonzero vector \( x \in \mathbb{R}^n \), number of iterations \( m \)

**Output:** Approximation of \( x^T f(A) x \)

1. Initialize \( u_1 \leftarrow x / \|x\|_2 \) and \( \beta_0 \leftarrow 0 \)
2. for \( i = 1, \ldots, m \) do
   3. \( \alpha_i \leftarrow u_i^T A u_i \)
   4. \( r_i \leftarrow A u_i - \alpha_i u_i - \beta_{i-1} u_{i-1} \)
   5. \( \beta_i \leftarrow \|r_i\|_2 \)
   6. \( u_{i+1} \leftarrow r_i / \beta_i \)
3. end for
4. \( T_m \leftarrow \begin{bmatrix} \alpha_1 & \beta_1 \\ \beta_1 & \alpha_2 \\ \vdots & \vdots \\ \beta_{m-1} & \alpha_m \end{bmatrix} \)
5. Return \( \|x\|_2^2 \cdot e_1^T f(T_m) e_1 \)
For theoretical considerations, it is helpful to view the quadratic form as an integral:

\[ x^T f(A)x = I := \int_{\lambda_{\min}}^{\lambda_{\max}} f(x) \, d\mu(x), \]

where \( \lambda_{\min}, \lambda_{\max} \) are the smallest/largest eigenvalues of \( A \) and \( \mu \) is the measure

\[ \mu(x) := \sum_{i=1}^{n} z_i^2 \delta_{\lambda_i}(x), \quad z := \frac{Q^T x}{\| x \|_2} \]

defined by a spectral decomposition \( A = Q \Lambda Q^T, \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n) \). It is well known [19, Theorem 6.2] that the approximation \( I_m \) returned by the \( m \)-points Gaussian quadrature rule applied to \( I \) is identical to the approximation returned by \( m \) steps of the Lanczos method:

\[ I_m := \| x \|_2^2 \cdot f(T_m)_{11}. \]

To bound the error \( |I - I_m| \), the analysis in [34] proceeds by using existing results on the polynomial approximation error of analytic functions. Although our analysis is along the same lines, it differs in a key technical aspect; we derive and use an improved error bound for the approximation of the logarithm; see Corollary 16. We have also noted two minor erratas in [34]; see the proof of Theorem 14 and the remark after Corollary 15 for details.

**Theorem 14.** Let \( f : [-1, 1] \to \mathbb{R} \) admit an analytic continuation to a Bernstein ellipse \( E_{\rho_0} \) with foci \( \pm 1 \) and elliptical radius \( \rho_0 \). For \( 1 < \rho < \rho_0 \), let \( M_\rho \) be the maximum of \( |f(z)| \) on \( E_\rho \). Then

\[ |I - I_m| \leq \| x \|_2^2 \cdot \frac{4M_\rho}{1 - \rho^{-1}} \rho^{-2m}. \]

**Proof.** As in [34], this result follows directly from bounds on the polynomial approximation error of analytic functions via Chebyshev expansion, combined with the fact that \( m \)-points Gaussian quadrature is exact for polynomials up to degree \( 2m - 1 \). However, the proof of [34, Theorem 4.2] uses an extra ingredient, which seems to be wrong. It claims that the integration error for odd-degree Chebyshev polynomials is zero thanks to symmetry. While this fact is indeed true for the standard Lebesgue measure, it does not hold for the measure (14). In turn, one obtains the slightly worse factor \( 1 - \rho^{-1} \) in the denominator, compared to the factor \( 1 - \rho^{-2} \) that would have been obtained from [34, Theorem 4.2] translated into our setting. \( \square \)

The affine linear transformation

\[ \varphi : [\lambda_{\min}, \lambda_{\max}] \to [-1, 1], \quad x \mapsto \frac{2}{\lambda_{\max} - \lambda_{\min}} t - \frac{\lambda_{\max} + \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}, \]

is used to map an interval \([\lambda_{\min}, \lambda_{\max}]\) containing the eigenvalues of \( A \) to the interval \([-1, 1]\) of Theorem 14. Defining \( g := f \circ \varphi^{-1} \), one has

\[ x^T g(\varphi(A))x = x^T f(A)x, \quad e_1^T g(\varphi(T_m))e_1 = e_1^T f(T_m)e_1. \]

By its shift invariance, the Lanczos method with \( g, \varphi(A) \), and \( x \) returns the approximation \( e_1^T g(\varphi(T_m))e_1 \). This allows us to apply Theorem 14. Combined with the relations (15), the following result is obtained.
Corollary 15. With the notation introduced above, it holds that

\[ |x^T f(A)x - \|x\|^2 \cdot e_1^T f(T_m)e_1| \leq \|x\|^2 \cdot \frac{4M_\rho}{1-\rho} \rho^{-2m}, \]

Note that \( M_\rho \) is the maximum of \( g \) on \( E_\rho \), which is equal to the maximum of \( f \) on the transformed ellipse with foci \( \lambda_{\min}, \lambda_{\max} \), and elliptical radius \((\lambda_{\max} - \lambda_{\min})\rho/2\). The result of Corollary 15 differs from the corresponding result in [34, page 1087], which features an additional, erroneous factor \((\lambda_{\max} - \lambda_{\min})(\lambda_{\max} - \lambda_{\min})/2\).

For the special case of the logarithm, the following result is obtained.

Corollary 16. Let \( A \in \mathbb{R}^{n \times n} \) be SPD with condition number \( \kappa(A) \), \( f \equiv \log \) and \( x \in \mathbb{R}^n \setminus \{0\} \). Then the error of the Lanczos method after \( m \) steps satisfies

\[ |x^T \log(A)x - \|x\|^2 \cdot e_1^T \log(T_m)e_1| \leq \|x\|^2 \cdot 2 \sqrt{\kappa(A)} \left(\frac{\sqrt{\kappa(A)} + 1 - 1}{\sqrt{\kappa(A)} + 1 + 1}\right)^{2m}. \]

Proof. The proof consists of applying Corollary 15 to a rescaled matrix. More specifically, we choose \( B := \lambda A \) with \( \lambda := 1/(2\lambda_{\min}) > 0 \). The tridiagonal matrix returned by the Lanczos method with \( A \) replaced by \( B \) satisfies \( T_m^B = \lambda T_m \). Together with the identity \( \log(A) = \log(\lambda I + A) \), this implies

\[ x^T \log(A)x - \|x\|^2 \cdot e_1^T \log(T_m)e_1 = x^T \log(B)x - \|x\|^2 \cdot e_1^T \log(T_m^B)e_1. \]

Note that the smallest/largest eigenvalues of \( B \) are given by \( 1/2 \) and \( \kappa(A)/2 \), respectively.

Applying Corollary 15 to \( B \) with \[ \rho := \frac{\sqrt{\kappa(A)} + 1 + 1}{\sqrt{\kappa(A)} + 1 - 1} \] thus gives

\[ |x^T \log(A)x - \|x\|^2 \cdot e_1^T \log(T_m)e_1| \leq \|x\|^2 \cdot 4M_\rho \rho^{-2m}. \]

The constant \( M_\rho \) is the maximum absolute value of the logarithm on the ellipse with foci \( 1/2 \) and \( \kappa(A)/2 \) that intersects the real axis at \( \alpha := \frac{1}{2\kappa(A)} \) and \( \beta := \frac{\kappa(A)^2 + \kappa(A) - 1}{2\kappa(A)} \). By Corollary 23, \( M_\rho = \|\log(\alpha)\| = \log(2\kappa(A)) \), where we used \( \alpha \leq 1/\beta \leq 1 \). Noting that

\[ \frac{4M_\rho}{1-\rho} = 2(\sqrt{\kappa(A)} + 1 + 1) \log(2\kappa(A)) \leq 4\sqrt{\kappa(A)} \]

concludes the proof.

4 Combined bounds for determinant estimation

Combining randomized trace estimation with the Lanczos method, we obtain the following (stochastic) estimate for \( \log(\det(A)) \):

\[ \text{est}_{N,m}^{G,R} := \sum_{i=1}^{N} \|X^{(i)}\|^2 \cdot e_1^T \log(T_m^{(i)})e_1, \]

\[ ^{1}\text{In fact, it is possible to choose } \rho = \frac{\sqrt{\kappa(A)} + \varepsilon + 1}{\sqrt{\kappa(A)} + \varepsilon - 1} \text{ for arbitrary } \varepsilon > 0. \]
where $X^{(1)}, \ldots, X^{(N)}$ are independent Gaussian or Rademacher random vectors and $T_n^{(i)}$ is the tridiagonal matrix obtained from the Lanczos method with starting vector $X^{(i)}/\|X^{(i)}\|_2$. By combining the results obtained so far, we now derive new bounds on the number of samples and number of Lanczos steps needed to ensure an approximation error of at most $\varepsilon$ (with high probability).

4.1 Standard Gaussian random vectors

**Theorem 17.** Suppose that the following holds for $N$ (number of Gaussian probe vectors) and $m$ (number of Lanczos steps per probe vector):

(i) $N \geq 16\varepsilon^{-2}(\rho_\log \|\log(A)\|^2_2 + \varepsilon \|\log(A)\|_2) \log \frac{4}{\delta}$, where $\rho_\log$ denotes the stable rank of $\log(A)$;

(ii) $m \geq \sqrt{\kappa(A) + 1 + 4 \log \left(\frac{8\varepsilon - 1}{n^2} \sqrt{\kappa(A)}\right)}$.

If, additionally, $N < \frac{\delta}{2} \exp \left(\frac{1}{16n}\right)$ then $\Pr(\|\text{est}_{N,m} G - \log \det(A)\| \geq \varepsilon) \leq \delta$.

**Proof.** For a Gaussian vector $X$, the squared norm $\|X\|^2_2$ is a Chi-squared random variable with $n$ degrees of freedom. Therefore, by [25, Lemma 1] we have

$$\Pr(\|X\|^2_2 \geq n + 2\sqrt{n}t + 2t) \leq \exp(-t)$$

for every $t > 0$. For $t = \log \frac{2N}{\delta}$, the assumption of the theorem implies

$$n + 2\sqrt{n}t + 2t \leq n + 2\sqrt{n} \cdot \frac{n}{4} + 2 \cdot \frac{n^2}{16} < n^2,$$

and therefore $\Pr(\|X\|^2_2 \geq n^2) \leq \frac{\delta}{2N}$. By the union bound, it holds that

$$\Pr\left(\exists i \in \{1, \ldots, N\} \text{ s.t. } \|X^{(i)}\|^2_2 \geq \frac{n^2}{2}\right) \leq \frac{\delta}{2}.$$  \hfill (16)

Corollary [16] together with condition (ii) and (16) imply that $\|\text{est}_{N,m} G - \log \det(A)\| \leq \frac{\delta}{2}$ holds with probability at least $1 - \delta/2$, where we also used that $\log \left(\frac{\sqrt{\kappa(A) + 1 + 1}}{\sqrt{\kappa(A) + 1 - 1}}\right) \geq \frac{2}{\sqrt{\kappa(A) + 1}}$.

Applying Theorem 5 to the matrix $\log(A)$, for which $\|\log(A)\|^2_2 = \rho(\log(A))\alpha^2 \leq n\alpha^2$, we find that $|\text{tr}_G(\log(A)) - \log \det(A)| \leq \frac{\varepsilon}{2}$ holds with probability at least $1 - \delta/2$. The proof is concluded by applying the triangle inequality.

4.2 Rademacher random vectors

**Theorem 18.** Suppose that the following holds for $N$ (number of Rademacher probe vectors) and $m$ (number of Lanczos steps per probe vector):

(i) $N \geq 32\varepsilon^{-2} (\rho_\log \|\log(A) - D_{\log(A)}\|^2_2 + \frac{\varepsilon}{2} \|\log(A) - D_{\log(A)}\|_2) \log \frac{2}{\delta}$, where $\rho_\log$ denotes the stable rank of $\log(A) - D_{\log(A)}$ and $D_{\log(A)}$ is the diagonal matrix containing the diagonal entries of $\log(A)$;
(ii) $m \geq \frac{\sqrt{\kappa(A)+1}}{\sqrt{\kappa(A)+1}} \log \left( 8\varepsilon^{-1}n\sqrt{\kappa(A)} \right)$.

Then $\mathbb{P}(|\text{est}_{N,m}^R - \log \det(A)| \geq \varepsilon) \leq \delta$.

**Proof.** Using Corollary 15 and the fact that Rademacher random vectors have norm $\sqrt{n}$, the bound $|\text{est}_{N,m}^R - \text{tr}^R_{N}(\log(A))| \leq \frac{\varepsilon}{2}$ holds if

$$m \geq \frac{1}{2} \log \left( 8\varepsilon^{-1}n\sqrt{\kappa(A)} \right) / \log \left( \frac{\sqrt{\kappa(A)+1}}{\sqrt{\kappa(A)+1}} \right).$$

Because of $\log \left( \frac{\sqrt{\kappa(A)+1}}{\sqrt{\kappa(A)+1}} \right) \geq \frac{2}{\sqrt{\kappa(A)+1}}$, condition (ii) ensures that this inequality is satisfied.

Applying Corollary 10 to $\log(A)$ and with $\varepsilon$ replaced by $\varepsilon/2$, immediately shows

$$|\text{tr}^R_{N}(\log(A)) - \log \det(A)| \leq \frac{\varepsilon}{2}$$

with probability at least $1 - \delta$ if condition (i) is satisfied. The proof is concluded by applying the triangle inequality. \qed

**Comparison with existing result.** To compare Theorem 18 with an existing result from [34], it is helpful to first derive a simpler (but usually stronger) condition on $N$.

**Lemma 19.** The statement of Theorem 18 holds with condition (i) replaced by $N \geq 8\varepsilon^{-2} \left( n \log^2 \kappa(A) + 2\varepsilon \log \kappa(A) \right) \log \frac{2}{\delta}$.

**Proof.** We set $B := \lambda A$ with $\lambda := 1/\sqrt{\lambda_{\min}(A)\lambda_{\max}(A)}$ and note that

$$\text{tr}^R_{N}(\log(A)) - \log \det(A) = \text{tr}^R_{N}(\log(\lambda A)) - \log \det(\lambda A).$$

Using $\lambda_{\max}(B) = \sqrt{\kappa(A)}$, $\lambda_{\min}(B) = 1/\sqrt{\kappa(A)}$, and $\kappa(B) = \kappa(A)$, we obtain

$$\|\log(B) - D_{\log(A)}\|_2 \leq 2\|\log(B)\|_2 = \log \kappa(A);$$

$$\|\log(B) - D_{\log(A)}\|_F^2 \leq \|\log(B)\|_F^2 = \rho(\log(B)) \frac{\log^2 \kappa(A)}{4} \leq \frac{n}{4} \log^2 \kappa(A).$$

An application of Corollary 10 to $\log(B)$ therefore yields (17) with probability at least $1 - \delta$ for $N \geq 8\varepsilon^{-2} \left( n \log^2 \kappa(A) + 2\varepsilon \log \kappa(A) \right) \log \frac{2}{\delta}$. \qed

Correcting for the two minor errata explained above, the result from [34, Corollary 4.5] states that $\mathbb{P}(|\text{est}_{N,m}^R - \text{tr}(\log(A))| \geq \varepsilon) \leq \delta$ holds if

$$N \geq 24\varepsilon^{-2} n^2 \left( \log(1 + \kappa(A)) \right)^2 \log \frac{2}{\delta}$$

and

$$m \geq \frac{\sqrt{3\kappa(A)}}{4} \log \left( 20\varepsilon^{-1}n \left( \sqrt{2\kappa(A)+1} + 1 \right) \log(2\kappa(A)+2) \right).$$

Lemma 19 improves (18) roughly by a factor $n$. However, let us stress that often the lower bound from Theorem 18 can be expected to be even better; below we describe a situation in which the lower only depends logarithmically on $n$. Condition (ii) of Theorem 18 improves (19) clearly but less drastically, roughly by a factor $\sqrt{3}$. \hfill 16
A situation leading to low stable ranks. We consider a family of matrices \( \{A_n\} \) of increasing dimension, such that the condition number and the spectral norm remain bounded. For a fixed failure probability \( \delta \) and a fixed accuracy \( \varepsilon \), the number of matrix-vector multiplications required to get 
\[
P(|\text{est}^G - \text{tr}(\log(A_n))| \geq \varepsilon) \leq \delta
\]
is proportional to \( O(\rho_{\log} \log(n)) \), where \( \rho_{\log} \) is the stable rank of \( \log(A_n) \). In certain applications, including regularized kernel matrices (see, e.g., [18]), also the stable rank remains constant when the matrix size increases. For example, this is the case for \( A_n := I + B_n \) where \( B_n \) has exponentially decaying eigenvalues, that is, \( \lambda_i(B_n) \leq C\alpha^i \) for some constant \( C > 0 \) and \( 0 < \alpha < 1 \), for all \( i < n \). For all \( n \) we have 
\[
\| \log(A_n) \|_2 = \log(1 + C\alpha), \quad \kappa(A_n) \approx \log(1 + C\alpha),
\]
and 
\[
\| \log(A_n) \|_F^2 \leq \sum_{i=1}^{n} \log^2(1 + C\alpha^i) \leq \sum_{i=1}^{n} C^2\alpha^{2i} \leq \frac{C^2\alpha^2}{1 - \alpha^2}.
\]
Therefore, the number of required matrix-vector multiplications required to attain accuracy \( \varepsilon \) with probability \( 1 - \delta \) is \( O(\log(n)) \).

Acknowledgments. We thank Radosław Adamczak, Rasmus Kyng, and Shashanka Ubaru for helpful discussions on topics related to this work.

References

[1] Distribution of difference of chi-squared variables. 
https://math.stackexchange.com/questions/85249/distribution-of-difference-of-chi-squared-variables
Accessed: 06/03/2020.

[2] Upper limit on the central binomial coefficient. 
https://mathoverflow.net/questions/133732/upper-limit-on-the-central-binomial-coefficient
Accessed: 23/03/2020.

[3] R. Adamczak. The entropy method and concentration of measure in product spaces. Master’s thesis, University of Warsaw and Vrije Universiteit van Amsterdam, 2003. Available at http://duch.mimuw.edu.pl/~radamcz/Old/Papers/master.pdf.

[4] R. H. Affandi, E. Fox, R. Adams, and B. Taskar. Learning the parameters of determinantal point process kernels. In International Conference on Machine Learning, pages 1224–1232, 2014.

[5] H. Avron. Counting triangles in large graphs using randomized matrix trace estimation. In Workshop on Large-scale Data Mining: Theory and Applications, volume 10, pages 10–9, 2010.

[6] H. Avron and S. Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. J. ACM, 58(2):Art. 8, 17, 2011.

[7] Z. Bai, M. Fahey, and G. Golub. Some large-scale matrix computation problems. volume 74, pages 71–89. 1996. TICAM Symposium (Austin, TX, 1995).
[8] R. P. Barry and R. K. Pace. Monte Carlo estimates of the log determinant of large sparse matrices. volume 289, pages 41–54. 1999. Linear algebra and statistics (Istanbul, 1997).

[9] R. Bhatia, M. D. Choi, and C. Davis. Comparing a matrix to its off-diagonal part. In The Gohberg anniversary collection, Vol. I (Calgary, AB, 1988 ), volume 40 of Oper. Theory Adv. Appl., pages 151–164. Birkhäuser, Basel, 1989.

[10] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. Ann. Probab., 31(3):1583–1614, 2003.

[11] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[12] C. Boutsidis, P. Drineas, P. Kambadur, E.-M. Kontopoulou, and A. Zouzias. A randomized algorithm for approximating the log determinant of a symmetric positive definite matrix. Linear Algebra Appl., 533:95–117, 2017.

[13] J. Chen. How accurately should I compute implicit matrix-vector products when applying the Hutchinson trace estimator? SIAM J. Sci. Comput., 38(6):A3515–A3539, 2016.

[14] D. Durfee, J. Peebles, R. Peng, and A. B. Rao. Determinant-preserving sparsification of SDDM matrices. SIAM Journal on Computing, 0(0):FOCS17–350–FOCS17–408, 0.

[15] T. Eden, A. Levi, D. Ron, and C. Seshadhri. Approximately counting triangles in sublinear time. SIAM J. Comput., 46(5):1603–1646, 2017.

[16] J. Fitzsimons, D. Granziol, K. Cutajar, M. Osborne, M. Filippone, and S. Roberts. Entropic trace estimates for log determinants. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 323–338. Springer, 2017.

[17] S. Foucart and H. Rauhut. A mathematical introduction to compressive sensing. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.

[18] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson. GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. In Advances in Neural Information Processing Systems, volume 2018-December, pages 7576–7586, 2018.

[19] G. H. Golub and G. Meurant. Matrices, moments and quadrature with applications. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2010.

[20] S. Gratton and D. Titley-Peloquin. Improved bounds for small-sample estimation. SIAM J. Matrix Anal. Appl., 39(2):922–931, 2018.
[21] I. Han, D. Malioutov, and J. Shin. Large-scale log-determinant computation through stochastic Chebyshev expansions. In International Conference on Machine Learning, pages 908–917, 2015.

[22] D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.*, 42:1079–1083, 1971.

[23] M. F. Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Comm. Statist. Simulation Comput.*, 18(3):1059–1076, 1989.

[24] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.

[25] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5):1302–1338, 2000.

[26] H. Li and Y. Zhu. Randomized block Krylov space methods for trace and log-determinant estimators. *arXiv preprint arXiv:2003.00212*, 2020.

[27] L. Lin, Y. Saad, and C. Yang. Approximating spectral densities of large matrices. *SIAM Rev.*, 58(1):34–65, 2016.

[28] R. K. Pace and J. P. LeSage. Chebyshev approximation of log-determinants of spatial weight matrices. *Comput. Statist. Data Anal.*, 45(2):179–196, 2004.

[29] W. Peng and H. Wang. A general scheme for log-determinant computation of matrices via stochastic polynomial approximation. *Comput. Math. Appl.*, 75(4):1259–1271, 2018.

[30] F. Roosta-Khorasani and U. Ascher. Improved bounds on sample size for implicit matrix trace estimators. *Found. Comput. Math.*, 15(5):1187–1212, 2015.

[31] A. K. Saibaba, A. Alexanderian, and I. C. F. Ipsen. Randomized matrix-free trace and log-determinant estimators. *Numer. Math.*, 137(2):353–395, 2017.

[32] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.

[33] C. Thron, S. J. Dong, K. F. Liu, and H. P. Ying. Padé-Z2 estimator of determinants. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 57(3):1642–1653, 1998.

[34] S. Ubaru, J. Chen, and Y. Saad. Fast estimation of $tr(f(A))$ via stochastic Lanczos quadrature. *SIAM J. Matrix Anal. Appl.*, 38(4):1075–1099, 2017.

[35] M. J. Wainwright. *High-dimensional statistics*, volume 48 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019. A non-asymptotic viewpoint.
A Auxiliary results

A.1 Herbst argument and logarithmic Sobolev inequalities

This section contains auxiliary results used in the proof of Theorem 8. We recall that the entropy of a random variable $Z$ is defined as

$$H(Z) := \mathbb{E}[Z \log Z] - \mathbb{E}[Z] \log \mathbb{E}[Z],$$

provided that all the involved expected values exist.

The Herbst argument (see, e.g., [11, page 11], [17, pages 239–240], and [35, Section 3.1.2]) turns a bound on the entropy of a random variable into a bound on the moment generating function. By Chernoff’s bound, the latter implies a bound on the tail of the random variable. Specifically, we use the following (modified) Herbst argument.

**Lemma 20.** Let $Z$ be a random variable and $g : [0, a) \rightarrow \mathbb{R}$ such that

$$H(\exp(\lambda Z)) \leq \lambda^2 g(\lambda) \mathbb{E}\exp(\lambda Z).$$

(20)

Then for all $\lambda \in [0, a)$ it holds

$$\log \mathbb{E}\exp(\lambda Z) \leq \lambda \mathbb{E}[Z] + \lambda \int_0^\lambda g(\xi)d\xi.$$

**Proof.** For $\psi(\lambda) := \log \mathbb{E}[\exp(\lambda Z)]$, it holds that $\psi'(\lambda) = \mathbb{E}[\lambda Z \exp(\lambda Z)]/\mathbb{E}[\exp(\lambda Z)]$. Recalling the definition of entropy, this allows us to rewrite (20) as

$$\lambda \psi'(\lambda) \exp(\psi(\lambda)) - \psi(\lambda) \exp(\psi(\lambda)) \leq \lambda^2 g(\lambda) \exp(\psi(\lambda)),$$

which is equivalent to

$$\frac{d}{d\lambda} \left( \frac{\psi(\lambda)}{\lambda} \right) \leq f(\lambda).$$

Integration on the interval $[0, \lambda]$ gives

$$\frac{\psi(\lambda)}{\lambda} - \lim_{\lambda \to 0^+} \frac{\psi(\lambda)}{\lambda} \leq \int_0^\lambda f(\xi)d\xi.$$

We conclude by noting that $\lim_{\lambda \to 0^+} \frac{\psi}{\lambda} = \mathbb{E}[Z]$. \qed
For deriving bounds on the entropy, we need the following two variations of Gross’ logarithmic Sobolev inequality.

**Theorem 21.** Let \( f : \{-1, 1\}^n \to \mathbb{R} \) and let \( X \) be a Rademacher vector with components \( X_1, \ldots, X_n \). Define \( f(\bar{X}^{(i)}) := f(X_1, \ldots, X_{i-1}, -X_i, X_{i+1}, \ldots, X_n) \) for \( i = 1, \ldots, n \). Then for all \( \lambda > 0 \) we have

\[
H(\exp(\lambda f(X))) \leq \frac{\lambda^2}{4} \mathbb{E} \left[ \sum_{i=1}^{n} \left( f(X) - f(\bar{X}^{(i)}) \right)^2 \right]
\]  

(21)

and

\[
H(\exp(\lambda f(X))) \leq \frac{\lambda^2}{8} \mathbb{E} \left[ \sum_{i=1}^{n} \left( f(X) - f(\bar{X}^{(i)}) \right)^2 \right].
\]

(22)

**Proof.** The inequality (21) is a standard result and can be found, e.g., in [11, page 122]. The inequality (22) is a variation of the same argument; see also [11, Exercise 5.5] for a related (but not identical) result. The inequality (22) can, in fact, be found in a Master’s thesis [3, Theorem 5]. For convenience of the reader, we provide a proof of (22) based on the textbook [11].

In [11, page 122] it is proven that

\[
H(\exp(\lambda f(X))) \leq \frac{1}{2} \mathbb{E} \left[ \sum_{i=1}^{n} \left( \exp(\lambda f(X)/2) - \exp(\lambda f(\bar{X}^{(i)})/2) \right)^2 \right].
\]

(23)

For \( a \geq b \) we have

\[
\exp \left( \frac{a}{2} \right) - \exp \left( \frac{b}{2} \right) = \int_{b/2}^{a/2} \exp(t)dt \leq \frac{a-b}{2} \left( \exp \left( \frac{a}{2} \right) + \exp \left( \frac{b}{2} \right) \right)
\]

\[
\leq \frac{a-b}{2} \sqrt{\exp \left( \frac{a}{2} \right) + \exp \left( \frac{b}{2} \right)},
\]

where the first inequality follows from the concavity of the exponential and the Hermite-Hadamard inequality. Therefore, for all \( a, b \in \mathbb{R} \) we have

\[
(\exp(a/2) - \exp(b/2))^2 \leq \frac{1}{8} (a-b)^2 (\exp(a) + \exp(b)).
\]

(24)

Applying (21) to each summand in Equation (23) one obtains

\[
H(\exp(\lambda f(X))) \leq \frac{\lambda^2}{16} \sum_{i=1}^{n} \mathbb{E} \left[ (f(X) - f(\bar{X}^{(i)}))^2 \left( \exp(\lambda f(X)) + \exp(\lambda f(\bar{X}^{(i)})) \right) \right]
\]

\[
= \frac{\lambda^2}{16} \sum_{i=1}^{n} \mathbb{E} \left[ (f(X) - f(\bar{X}^{(i)}))^2 \exp(\lambda f(X)) \right]
\]

\[
+ \frac{\lambda^2}{16} \sum_{i=1}^{n} \mathbb{E} \left[ (f(X) - f(\bar{X}^{(i)}))^2 \exp(\lambda f(\bar{X}^{(i)})) \right]
\]

\[
= \frac{\lambda^2}{8} \mathbb{E} \left[ \exp(\lambda f(X)) \sum_{i=1}^{n} \left( f(X) - f(\bar{X}^{(i)}) \right)^2 \right],
\]

21
where the last equality follows from the fact that \( f(X) \) and \( f(\tilde{X}^{(i)}) \) have the same distribution and changing the sign of the \( i \)th entry of \( \tilde{X}^{(i)} \) gives \( X \) again.

\[ \square \]

### A.2 Bounds on the complex logarithm

The following two elementary results on the complex logarithm are needed in the convergence proofs of the Lanczos method in Section 3.

**Lemma 22.** Consider a circle in the complex plane with center \( a \in \mathbb{R}^+ \), \( a > 1 \) and radius \( b \) such that \( b^2 = a^2 - 1 \). Then the maximum absolute value of the logarithm on this circle is attained on the real axis.

**Proof.** By symmetry, we can restrict ourselves to the upper half of the circle. For fixed \( \theta \in [0, \arctan(b)] \) the line \( re^{i\theta} \) for \( r > 0 \) intersects the circumference when \((r \cos \theta - a)^2 + r^2 \sin^2 \theta = b^2\). Clearly, this equality holds for \( r = a \cos \theta \pm \sqrt{a^2 \cos^2 \theta - 1} \). Note that these points parametrize the entire upper semi-circle and we have \( r = \frac{1}{\sqrt{r^2 + \sin^2 (\arctan(b))}} \), so

\[
|\log (r e^{i\theta})| = |\log (r) + i\theta| = | - \log (r) + i\theta| = |\log (r) + i\theta| = |\log (r + e^{i\theta})|.
\]

Therefore, to prove the lemma it is sufficient to show that the function \( g : [0, \arctan(b)] \to \mathbb{R} \) given by \( g(\theta) := |f(\theta)|^2 \), where \( f(\theta) = \log (r + e^{i\theta}) \), attains its maximum for \( \theta = 0 \). We will establish this fact by showing that \( g \) decreases monotonically. We have

\[
f(\theta) = \log \left( (a \cos \theta + \sqrt{a^2 \cos^2 \theta - 1}) e^{i\theta} \right) = \arcsinh \left( \sqrt{a^2 \cos^2 \theta - 1} \right) + i\theta;
\]

\[
f'(\theta) = \frac{-a^2 \cos \theta \sin \theta}{a \cos \theta \sqrt{a^2 \cos^2 \theta - 1}} + i = -\frac{a \sin \theta}{\sqrt{a^2 \cos^2 \theta - 1}} + i,
\]

and therefore

\[
g'(\theta) = 2 \text{Re} \left( f'(t) \cdot \overline{f(t)} \right) = -2 \frac{a \sin \theta \cdot \arcsinh \left( \sqrt{a^2 \cos^2 \theta - 1} \right)}{\sqrt{a^2 \cos^2 \theta - 1}} + 2 \theta.
\]

For \( \theta \in (0, \arctan(b)) \) we have that

\[
g'(\theta) \leq 0 \iff \frac{\theta}{a \sin \theta} \leq \frac{\arcsinh \left( \sqrt{a^2 \cos^2 \theta - 1} \right)}{\sqrt{a^2 \cos^2 \theta - 1}}.
\]

Using the facts that \( x \mapsto \frac{x}{\sin(x)} \) is increasing for \( x \in [0, \pi] \), \( \arctan(x) < \arcsinh(x) \) for \( x > 0 \), and \( x \mapsto \frac{\arcsinh(x)}{x} \) is decreasing for \( x > 0 \), one obtains

\[
\frac{\theta}{a \sin \theta} \leq \frac{\arctan(b)}{b} < \frac{\arcsinh(b)}{b} \leq \frac{\arcsinh \left( \sqrt{a^2 \cos^2 \theta - 1} \right)}{\sqrt{a^2 \cos^2 \theta - 1}},
\]

for every \( \theta \in (0, \arctan(b)) \). In particular, this shows \( g'(\theta) < 0 \) for \( \theta \in (0, \arctan(b)) \) and hence \( g \) is decreasing.

\[ \square \]
Corollary 23. Consider an ellipse $\mathcal{E}$ in the open right-half complex plane, with foci on the real axis. Then the maximum absolute value of the logarithm on this ellipse is attained on the real axis.

Proof. Let $0 < \alpha < \beta$ be the two intersections of the ellipse with the real axis. If $|\log \alpha| \geq |\log \beta|$ then $\mathcal{E}$ is contained in the circle $C_1$ of center $a := \frac{1}{2} \left( \frac{1}{\alpha} + \alpha \right)$ and radius $b := \frac{1}{2} \left( \frac{1}{\alpha} - \alpha \right) = \sqrt{a^2 - 1}$, and $\mathcal{E}$ is tangent to $C_1$ in $\alpha$; otherwise $\mathcal{E}$ is contained in the circle $C_2$ of center $a := \frac{1}{2} \left( \beta + \frac{1}{\beta} \right)$ and radius $b := \frac{1}{2} \left( \beta - \frac{1}{\beta} \right) = \sqrt{a^2 - 1}$, and $\mathcal{E}$ is tangent to $C_2$ in $\beta$. In both cases, the result follows from Lemma 22. □