

Linear Trace-Deficit Decay for Permanents of Positive Semidefinite Contractions*

Matěj Kripner Václav Rozhoň Robert Šámal
Martin Šindelář Adrián Zámečník

Abstract

Every Hermitian positive semidefinite contraction A satisfies $\text{per}(A) \leq \exp(-\text{tr}(I - A)/68)$. This proves the linear trace-deficit bound conjectured by Ebrahimnejad, Nagda, and Oveis Gharan, whose estimate had quadratic dependence on the normalized deficit. The proof combines a size-biased complex Gaussian representation, a block version of Hanson–Wright decoupling, and a rational Loewner majorant. The constant is uniform in the dimension and is not optimized.

1 Introduction

1.1 Background and the source problem

For $A = (A_{ij}) \in \mathbb{C}^{n \times n}$, its permanent is

$$\text{per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^n A_{i, \sigma(i)}.$$

When A is Hermitian positive semidefinite, the permanent is nonnegative and admits a useful complex Gaussian representation. The permanent of a positive semidefinite matrix is also monotone in the Loewner order; both facts will be proved below.

There are two related questions: how accurately a permanent can be approximated, and how small it must be when the matrix falls below the identity. For the first, Anari, Gurvits, Oveis Gharan, and Saberi [3, Theorem 1 and Definition 1] obtained a deterministic $e^{(\gamma+1)n}$ -approximation through a diagonal-majorant relaxation; here γ is Euler’s constant. Yuan and Parrilo [4, Corollary 1.2] sharpened the guarantee for that relaxation using its connection with products of linear forms and low-rank solutions. Ebrahimnejad, Nagda, and Oveis Gharan [1, Theorems 1.1–1.2] subsequently improved the approximation exponent below $\gamma + 1$ and established hardness below γ . Anari and Ebrahimnejad [2, Theorem 1] have now obtained the optimal $e^{(\gamma+o(1))n}$ exponent. The present paper concerns the second question, a structural inequality for each individual matrix.

In their work on approximating positive semidefinite permanents, Ebrahimnejad, Nagda, and Oveis Gharan established the following trace-sensitive estimate.

Theorem 1.1 (Ebrahimnejad–Nagda–Oveis Gharan [1, Lemma 1.3]). *Let $0 \leq \varepsilon \leq 1$, and let $A \in \mathbb{C}^{n \times n}$ be Hermitian with*

$$0 \preceq A \preceq I_n \quad \text{and} \quad \text{tr}(A) \leq (1 - \varepsilon)n.$$

Then

$$\text{per}(A) \leq \left(1 - \frac{\varepsilon^2}{20}\right)^n.$$

*This work was produced within the Bolzano project (<https://bolzano.app>), using AI systems for mathematical research and manuscript preparation. The listed authors contributed to the project and the preparation of this release. The manuscript received only limited human checking.

In Section 1.4 of [1], the authors identify the $O(\varepsilon^2)$ dependence in Theorem 1.1 as the bottleneck to a better approximation ratio and conjecture that it can be improved to $O(\varepsilon)$. Since the estimate in Theorem 1.1 is expressed through its base, the corresponding quantitative target is the following.

Conjectural target 1.2 (Linear form of the source-paper target [1, Section 1.4]). There is a universal constant $c > 0$ such that, for every $0 \leq \varepsilon \leq 1$ and every Hermitian matrix A satisfying

$$0 \preceq A \preceq I_n, \quad \text{tr}(A) \leq (1 - \varepsilon)n,$$

one has

$$\text{per}(A) \leq (1 - c\varepsilon)^n.$$

1.2 Contribution

The main result establishes Conjecture 1.2 through an exponential estimate in the actual trace deficit.

Theorem 1.3 (Trace-deficit decay). *For every $n \geq 1$ and every Hermitian matrix $A \in \mathbb{C}^{n \times n}$ with $0 \preceq A \preceq I_n$,*

$$\text{per}(A) \leq \exp\left(-\frac{1}{68} \text{tr}(I_n - A)\right). \quad (1.1)$$

Corollary 1.4 (Linear dependence on ε). *Under the hypotheses of Conjecture 1.2,*

$$\text{per}(A) \leq e^{-\varepsilon n/68} \leq (1 - c\varepsilon)^n, \quad c = 1 - e^{-1/68} > 0.$$

The proof has two principal components.

1. The local estimate in Theorem 4.1 shows that, for $K \succeq 0$ of sufficiently small operator norm,

$$\log \text{per}((I_n + K)^{-1}) \leq -\text{tr}(K) + 17 \text{tr}(K^2).$$

Its linear term comes from an exact Gaussian density calculation, while Lemma 3.1 keeps the quadratic error proportional to $\text{tr}(K^2)$, rather than to $(\text{tr } K)^2$.

2. Lemma 5.2 compares a general contraction $A = I_n - D$ to the rational majorant $(I_n + \theta D)^{-1}$ in Loewner order. Taking $\theta = 1/34$, applying Theorem 4.1, and using $D^2 \preceq D$ yields Theorem 1.3.

1.3 Intuition and related methods

For a diagonal contraction, the scalar inequality $a \leq e^{-(1-a)}$ immediately gives $\text{per}(A) \leq e^{-\text{tr}(I-A)}$. Off-diagonal entries introduce additional permutation terms, and Example 6.2 shows that this constant cannot hold for all contractions. Nevertheless, the total spectral deficit remains the right quantity. The spectral cutoff in the proof of [1, Lemma 1.3] first finds sufficiently many eigenvalues a sufficient distance below one. Both quantities cost a factor of the normalized deficit. The rational majorant above instead shrinks every deficit by the same fixed factor and retains their entire sum.

The local calculation explains why this is possible. Gaussian Wick identities for PSD permanents are standard; see [1, Lemma 2.9]. Barvinok [5, Theorem 1.1 and Section 2.1] uses such an integral representation for an approximation scheme when the eigenvalues lie in $[1, 2]$, and [1, Lemma 3.2] uses it to bound permanents by determinants. Here a change of density absorbs the product of squared coordinates into independent size-biased complex Gaussian variables. Their expected quadratic form is $2 \text{tr } K$, whereas the determinant factor contributes at most $\text{tr } K$. This leaves a negative linear term. The essential estimate is that its fluctuation costs only $O(\text{tr } K^2)$.

The random partition and Gaussian linearization used to control that fluctuation follow the proof strategy of Rudelson and Vershynin [6, proof of Theorem 1.1, Steps 2–4]. The auxiliary-Gaussian

comparison also appears in Hsu, Kakade, and Zhang [7, Theorem 1 and Remark 2]. These are precedents for the method, rather than statements of the permanent inequality. Each complex coordinate here gives an independent two-dimensional real block; its real and imaginary parts are dependent after size biasing. Lemma 3.1 therefore gives the block argument explicitly, with constants sufficient for the application. The exact Gamma transform handles the diagonal part separately.

Corollary 1.4 provides the replacement requested in [1, Section 1.4]. It is not needed to claim the current best approximation ratio, which is already supplied by [2]. The entropic upper certificate in [2, Eq. (4) and Theorem 7] does not by itself give trace-deficit decay: on $A = aI_n$, $a > 0$, it equals $(4a/e)^n$. Indeed, its logarithm is the maximum of

$$n \log a + \sum_i \log X_{ii} + \log \det X - \operatorname{tr} X + n \quad (X \succ 0).$$

Hadamard's determinant inequality bounds this by $n \log a + \sum_i (2 \log X_{ii} - X_{ii} + 1)$. Each summand is maximized at $X_{ii} = 2$, and $X = 2I_n$ attains equality. Thus this certificate exceeds one for $e/4 < a < 1$, even though the trace deficit is positive. Theorem 1.3 supplies a direct pointwise estimate; its constant is not asserted to be optimal.

Another algorithmic upper bound is the permanent process of Laddha and Pittu [8, Algorithm 1 and Theorem 1.3], valid for real PSD matrices as well as nonnegative matrices. Its elimination-type certificate need not decay on contractions: on the rank-one projection J_n/n , where J_n is the all-ones matrix, each step doubles all entries of the remaining submatrix. The product of its pivots is therefore $2^{n(n-1)/2}/n^n$, which already exceeds one at $n = 7$. Yi's recent deterministic approximation scheme [9, Theorem 1.1(b)] instead treats nonnegative matrices with dense support and nonzero entries in a fixed positive interval. That input regime differs from arbitrary complex PSD contractions.

2 Gaussian preliminaries

We write X^* for the conjugate transpose of a complex matrix X . The symbols $\|\cdot\|_{\text{op}}$ and $\|\cdot\|_{\text{F}}$ denote the operator and Frobenius norms. A standard circular complex Gaussian random variable has density

$$d\gamma(w) = \frac{1}{\pi} e^{-|w|^2} d^2w \quad (w \in \mathbb{C}).$$

Equivalently, $W = (G_1 + iG_2)/\sqrt{2}$, where G_1, G_2 are independent standard real Gaussian variables.

2.1 Wick's formula

The following representation is the complex version of Wick's formula. It also appears as Lemma 2.9 in [1].

Lemma 2.1 (Complex Gaussian representation). *Let $B \in \mathbb{C}^{n \times n}$ be Hermitian positive semidefinite, and let Z be a centered circular complex Gaussian vector with covariance $\mathbb{E}[ZZ^*] = B$. Then*

$$\operatorname{per}(B) = \mathbb{E} \prod_{i=1}^n |Z_i|^2. \quad (2.1)$$

Proof. For a circular complex Gaussian vector, the only nonzero pairings in the mixed moment

$$\mathbb{E}[Z_1 \cdots Z_n \overline{Z_1} \cdots \overline{Z_n}]$$

pair each unbarred coordinate with a barred coordinate. Such a pairing is specified by a permutation $\sigma \in S_n$, and its contribution is

$$\prod_{i=1}^n \mathbb{E}[Z_i \overline{Z_{\sigma(i)}}] = \prod_{i=1}^n B_{i, \sigma(i)}.$$

Summing over all $\sigma \in S_n$ gives (2.1). A singular covariance follows either from the same moment identity or by approximation with $B + \delta I_n$. $\square$

2.2 A size-biased complex Gaussian law

Define a probability measure ν_1 on $\mathbb{C}$ by

$$d\nu_1(w) = \frac{1}{\pi} |w|^2 e^{-|w|^2} d^2w. \quad (2.2)$$

It is normalized because $\mathbb{E}_\gamma |W|^2 = 1$. The next lemma records both a subgaussian estimate in real coordinates and the exact radial Laplace transform.

Lemma 2.2 (Size-biased Gaussian identities). *Let W have law ν_1 , and set*

$$\xi = (\sqrt{2} \Re W, \sqrt{2} \Im W) \in \mathbb{R}^2.$$

Then, for every $u \in \mathbb{R}^2$,

$$\mathbb{E} e^{u^\top \xi} = \left(1 + \frac{\|u\|_2^2}{2}\right) e^{\|u\|_2^2/2} \leq e^{\|u\|_2^2}. \quad (2.3)$$

Moreover, $R = |W|^2$ has the Gamma(2, 1) density xe^{-x} on $[0, \infty)$, and therefore

$$\mathbb{E} e^{-aR} = (1+a)^{-2}, \quad \mathbb{E} R = 2 \quad (a \geq 0). \quad (2.4)$$

Proof. Under the untilted law γ , the vector $(\sqrt{2} \Re W, \sqrt{2} \Im W)$ is a standard real Gaussian vector $G = (G_1, G_2)$, while $|W|^2 = (G_1^2 + G_2^2)/2$. Hence, for every nonnegative measurable function f ,

$$\mathbb{E}_{\nu_1} f(\xi) = \frac{1}{2} \mathbb{E} [(G_1^2 + G_2^2) f(G)].$$

Differentiating the standard Gaussian moment-generating function gives

$$\mathbb{E} [G_j^2 e^{u^\top G}] = (1 + u_j^2) e^{\|u\|_2^2/2}.$$

Summing over $j = 1, 2$ and dividing by 2 yields the equality in (2.3). Its inequality follows from $1 + t/2 \leq e^{t/2}$ for $t \geq 0$.

Under γ , the variable $|W|^2$ has density e^{-x} on $[0, \infty)$. The size bias in (2.2) changes this to xe^{-x} . Direct integration gives

$$\int_0^\infty x e^{-(1+a)x} dx = (1+a)^{-2},$$

and differentiating at $a = 0$ gives $\mathbb{E} R = 2$. $\square$

Corollary 2.3 (Product-block linear MGF). *Let $\xi_1, \dots, \xi_n$ be independent copies of the vector in Lemma 2.2, and put $\Xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^{2n}$. Then*

$$\mathbb{E} e^{v^\top \Xi} \leq e^{\|v\|_2^2} \quad \text{for every } v \in \mathbb{R}^{2n}.$$

Proof. Apply (2.3) to each two-dimensional block of v and multiply the resulting estimates. $\square$

3 An off-diagonal quadratic MGF estimate

The local permanent estimate will separate the diagonal and off-diagonal parts of a quadratic form. The diagonal transform is explicit by (2.4). The following dimension-free estimate controls the off-diagonal part.

Lemma 3.1 (Block-off-diagonal quadratic estimate). *Let $\Xi = (\xi_1, \dots, \xi_n)$ be as in Corollary 2.3. Let R be a real symmetric $2n \times 2n$ matrix whose diagonal 2×2 blocks vanish. If*

$$64s^2 \|R\|_{\text{op}}^2 \leq \frac{1}{2}, \quad (3.1)$$

then

$$\log \mathbb{E} e^{s\Xi^\top R \Xi} \leq 32s^2 \|R\|_{\text{F}}^2. \quad (3.2)$$

Proof. Choose a random set $\Lambda \subseteq \{1, \dots, n\}$ by including each block index independently with probability $1/2$, and define

$$X_\Lambda = \sum_{\substack{i \in \Lambda \\ j \notin \Lambda}} \xi_i^\top R_{ij} \xi_j.$$

Because the diagonal blocks of R vanish, every ordered pair (i, j) with $i \neq j$ occurs in this sum with probability $1/4$. Consequently,

$$\Xi^\top R \Xi = 4\mathbb{E}_\Lambda X_\Lambda. \quad (3.3)$$

For every fixed value of Ξ , convexity of the exponential gives

$$e^{s\Xi^\top R \Xi} \leq \mathbb{E}_\Lambda e^{4sX_\Lambda}. \quad (3.4)$$

Fix Λ . Write $a = \Xi_\Lambda$ and $b = \Xi_{\Lambda^c}$, and let T_Λ be the rectangular cross-cut block of R for which

$$X_\Lambda = b^\top T_\Lambda a.$$

For $u \in \mathbb{R}$, conditioning on a and applying Corollary 2.3 to the independent blocks in b yields

$$\mathbb{E}_b[e^{uX_\Lambda} \mid a] \leq \exp\left(u^2 \|T_\Lambda a\|_2^2\right). \quad (3.5)$$

Let g be a standard real Gaussian vector of the row dimension of T_Λ , independent of all other variables. Gaussian linearization and another use of Corollary 2.3 give

$$e^{u^2 \|T_\Lambda a\|_2^2} = \mathbb{E}_g e^{\sqrt{2}|u|g^\top T_\Lambda a}, \quad (3.6)$$

$$\mathbb{E}_a e^{\sqrt{2}|u|g^\top T_\Lambda a} \leq \exp\left(2u^2 \|T_\Lambda^\top g\|_2^2\right). \quad (3.7)$$

Combining (3.5)–(3.7), and denoting the singular values of T_Λ by $\sigma_1, \dots, \sigma_m$, gives

$$\mathbb{E} e^{uX_\Lambda} \leq \prod_{r=1}^m (1 - 4u^2 \sigma_r^2)^{-1/2}, \quad (3.8)$$

provided $4u^2 \|T_\Lambda\|_{\text{op}}^2 < 1$.

For $0 \leq x \leq 1/2$, the elementary inequality $-\frac{1}{2} \log(1 - x) \leq x$ holds. It follows from (3.8) that, whenever $4u^2 \|T_\Lambda\|_{\text{op}}^2 \leq 1/2$,

$$\log \mathbb{E} e^{uX_\Lambda} \leq 4u^2 \|T_\Lambda\|_{\text{F}}^2. \quad (3.9)$$

The cross-cut block is a rectangular compression of R , so

$$\|T_\Lambda\|_{\text{op}} \leq \|R\|_{\text{op}}.$$

In the decomposition into the coordinates indexed by Λ and Λ^c , the two cross-cut blocks of the symmetric matrix R are T_Λ and $T_\Lambda^\top$. Thus

$$2 \|T_\Lambda\|_F^2 \leq \|R\|_F^2. \quad (3.10)$$

Set $u = 4s$. The hypothesis (3.1) implies the condition for (3.9), and hence

$$\log \mathbb{E} e^{4sX_\Lambda} \leq 64s^2 \|T_\Lambda\|_F^2 \leq 32s^2 \|R\|_F^2.$$

This bound is uniform in Λ . Taking expectations in (3.4) and then applying the uniform bound proves (3.2). $\square$

4 Local stability near the identity

We now combine the exact radial transform and Lemma 3.1. The resulting estimate is local only in operator norm and has no dependence on the dimension.

Theorem 4.1 (Local permanent stability). *Let $K \in \mathbb{C}^{n \times n}$ be Hermitian positive semidefinite and satisfy*

$$\|K\|_{\text{op}} \leq \frac{\sqrt{2}}{17}.$$

Then

$$\log \text{per}((I_n + K)^{-1}) \leq -\text{tr}(K) + 17 \text{tr}(K^2). \quad (4.1)$$

Proof. Put $B = (I_n + K)^{-1}$. The matrix B is positive definite. If Z is circular complex Gaussian with covariance B , then Lemma 2.1 and the Gaussian density give

$$\begin{aligned} \text{per}(B) &= \frac{1}{\pi^n \det B} \int_{\mathbb{C}^n} \left(\prod_{i=1}^n |z_i|^2 \right) e^{-z^* B^{-1} z} d^{2n} z \\ &= \det(I_n + K) \mathbb{E}_\nu e^{-W^* K W}, \end{aligned} \quad (4.2)$$

where $\nu = \nu_1^{\otimes n}$, and the coordinates $W_1, \dots, W_n$ are independent with the density (2.2).

Write $k_i = K_{ii} \geq 0$ and decompose the real random variable

$$Q = W^* K W = Q_d + Q_o,$$

where

$$Q_d = \sum_{i=1}^n k_i |W_i|^2, \quad Q_o = \sum_{i \neq j} \overline{W_i} K_{ij} W_j.$$

By (2.4),

$$\mathbb{E}_\nu Q_d = 2 \text{tr}(K), \quad \mathbb{E}_\nu Q_o = 0.$$

Set $U_d = Q_d - 2 \text{tr}(K)$, and choose the conjugate Hölder exponents

$$p = 17, \quad q = \frac{17}{16}.$$

Hölder's inequality yields

$$\mathbb{E}_\nu e^{-(Q - 2 \text{tr}(K))} \leq (\mathbb{E}_\nu e^{-p U_d})^{1/p} (\mathbb{E}_\nu e^{-q Q_o})^{1/q}. \quad (4.3)$$

For the diagonal factor, the exact transform (2.4) gives

$$\mathbb{E} e^{-p k_i (|W_i|^2 - 2)} = e^{2p k_i} (1 + p k_i)^{-2}.$$

Independence therefore implies

$$\frac{1}{p} \log \mathbb{E}_\nu e^{-pU_d} = \sum_{i=1}^n \left(2k_i - \frac{2}{p} \log(1 + pk_i) \right). \quad (4.4)$$

The elementary inequality $\log(1+x) \geq x - x^2/2$ for $x \geq 0$ turns (4.4) into

$$\frac{1}{p} \log \mathbb{E}_\nu e^{-pU_d} \leq 17 \sum_{i=1}^n k_i^2. \quad (4.5)$$

It remains to control the off-diagonal factor. Write $K = P + iS$, where P is real symmetric and S is real skew-symmetric. Its realification is

$$\mathcal{R}(K) = \begin{pmatrix} P & -S \\ S & P \end{pmatrix}.$$

For $\xi_i = (\sqrt{2} \Re W_i, \sqrt{2} \Im W_i)$ and $\Xi = (\xi_1, \dots, \xi_n)$, after the harmless orthogonal permutation that groups coordinates by i , one has

$$W^* K W = \Xi^\top M \Xi, \quad M = \frac{1}{2} \mathcal{R}(K).$$

Let $H = K - \text{diag}(K)$, and let M_o be obtained from M by deleting its diagonal 2×2 blocks. Then

$$Q_o = \Xi^\top M_o \Xi, \quad M_o = \frac{1}{2} \mathcal{R}(H)$$

up to the same coordinate permutation. The realification identities give

$$\|M_o\|_{\text{op}} = \frac{1}{2} \|H\|_{\text{op}}, \quad \|M_o\|_{\text{F}}^2 = \frac{1}{2} \|H\|_{\text{F}}^2. \quad (4.6)$$

Since $K \succeq 0$,

$$\|\text{diag}(K)\|_{\text{op}} = \max_i k_i \leq \|K\|_{\text{op}}.$$

It follows from (4.6) that

$$\|M_o\|_{\text{op}} \leq \frac{1}{2} (\|K\|_{\text{op}} + \|\text{diag}(K)\|_{\text{op}}) \leq \|K\|_{\text{op}}. \quad (4.7)$$

The norm hypothesis and $q = 17/16$ now give the exact threshold

$$64q^2 \|M_o\|_{\text{op}}^2 \leq 64 \left(\frac{17}{16} \right)^2 \left(\frac{\sqrt{2}}{17} \right)^2 = \frac{1}{2}.$$

Lemma 3.1, applied with $R = M_o$ and $s = -q$, therefore yields

$$\log \mathbb{E}_\nu e^{-qQ_o} \leq 32q^2 \|M_o\|_{\text{F}}^2.$$

Because H is Hermitian with zero diagonal,

$$\|M_o\|_{\text{F}}^2 = \frac{1}{2} \|H\|_{\text{F}}^2 = \sum_{i < j} |K_{ij}|^2.$$

Since $32q = 34$, we obtain

$$\frac{1}{q} \log \mathbb{E}_\nu e^{-qQ_o} \leq 34 \sum_{i < j} |K_{ij}|^2. \quad (4.8)$$

Combining (4.3), (4.5), and (4.8), and using

$$\operatorname{tr}(K^2) = \sum_i k_i^2 + 2 \sum_{i < j} |K_{ij}|^2,$$

gives

$$\log \mathbb{E}_\nu e^{-(Q-2\operatorname{tr} K)} \leq 17 \operatorname{tr}(K^2). \quad (4.9)$$

Equivalently,

$$\log \mathbb{E}_\nu e^{-Q} \leq -2 \operatorname{tr}(K) + 17 \operatorname{tr}(K^2).$$

Finally, applying $\log(1+x) \leq x$ to the eigenvalues of K gives $\log \det(I_n + K) \leq \operatorname{tr}(K)$. Together with (4.2), this implies

$$\log \operatorname{per}(B) \leq -\operatorname{tr}(K) + 17 \operatorname{tr}(K^2),$$

which is (4.1). $\square$

5 Loewner comparison and the global estimate

For completeness, we first prove the monotonicity used in the source paper as Lemma 2.8 [1].

Lemma 5.1 (Loewner monotonicity). *If $0 \preceq A \preceq B$ are Hermitian $n \times n$ matrices, then*

$$\operatorname{per}(A) \leq \operatorname{per}(B).$$

Proof. In $(\mathbb{C}^n)^{\otimes n}$, define

$$v = \frac{1}{\sqrt{n!}} \sum_{\sigma \in S_n} e_{\sigma(1)} \otimes \cdots \otimes e_{\sigma(n)}.$$

Expanding the matrix coefficient and grouping pairs of permutations by their relative permutation gives

$$\langle v, A^{\otimes n} v \rangle = \operatorname{per}(A). \quad (5.1)$$

Moreover, the telescoping identity

$$B^{\otimes n} - A^{\otimes n} = \sum_{j=1}^n A^{\otimes(j-1)} \otimes (B - A) \otimes B^{\otimes(n-j)}$$

expresses the difference as a sum of positive semidefinite matrices. Hence $A^{\otimes n} \preceq B^{\otimes n}$. Taking the matrix coefficient at v and using (5.1) proves the claim. $\square$

Lemma 5.2 (Rational Loewner majorant). *Let $0 \preceq A \preceq I_n$, put $D = I_n - A$, and let $0 \leq \theta \leq 1$. Then*

$$A \preceq (I_n + \theta D)^{-1}.$$

Proof. For every $d \in [0, 1]$,

$$\frac{1}{1+\theta d} - (1-d) = \frac{d(1-\theta+\theta d)}{1+\theta d} \geq 0. \quad (5.2)$$

Since the spectrum of D is contained in $[0, 1]$, applying (5.2) through functional calculus gives

$$I_n - D \preceq (I_n + \theta D)^{-1}.$$

$\square$

Proof of Theorem 1.3. Set $D = I_n - A$, so that $0 \preceq D \preceq I_n$, and choose

$$\theta = \frac{1}{34}, \quad K = \theta D, \quad B = (I_n + K)^{-1}.$$

Lemma 5.2 gives $A \preceq B$. Furthermore,

$$\|K\|_{\text{op}} \leq \frac{1}{34} \leq \frac{\sqrt{2}}{17},$$

so Theorem 4.1 applies. Since $D^2 \preceq D$,

$$\text{tr}(K) = \theta \text{tr}(D), \quad \text{tr}(K^2) = \theta^2 \text{tr}(D^2) \leq \theta^2 \text{tr}(D).$$

Therefore,

$$\begin{aligned} \log \text{per}(B) &\leq -\text{tr}(K) + 17 \text{tr}(K^2) \\ &\leq -(\theta - 17\theta^2) \text{tr}(D) = -\frac{1}{68} \text{tr}(D). \end{aligned}$$

By Lemma 5.1,

$$\text{per}(A) \leq \text{per}(B) \leq \exp\left(-\frac{1}{68} \text{tr}(I_n - A)\right).$$

□

Proof of Corollary 1.4. The trace hypothesis gives $\text{tr}(I_n - A) \geq \varepsilon n$, so Theorem 1.3 implies

$$\text{per}(A) \leq e^{-\varepsilon n/68}.$$

The function $x \mapsto e^{-x/68}$ is convex. On $[0, 1]$, it therefore lies below the chord joining its endpoint values:

$$e^{-\varepsilon/68} \leq (1 - \varepsilon) + \varepsilon e^{-1/68} = 1 - (1 - e^{-1/68})\varepsilon.$$

Raising this scalar inequality to the n th power proves the result. □

6 Examples and remarks

Example 6.1 (Diagonal contractions). For $A = \text{diag}(a_1, \dots, a_n)$ with $0 \leq a_i \leq 1$,

$$\text{per}(A) = \prod_{i=1}^n a_i \leq \exp\left(-\sum_{i=1}^n (1 - a_i)\right) = e^{-\text{tr}(I_n - A)}.$$

Here the first inequality is the scalar bound $\log a \leq a - 1$, with the case $a = 0$ understood by continuity. Thus the trace deficit is the natural first-order quantity already in the diagonal case, although the universal constant furnished by the noncommutative argument is smaller.

Example 6.2 (A restriction on the universal constant). The rank-one projection

$$P = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

has eigenvalues zero and one, $\text{per}(P) = 1/4 + 1/4 = 1/2$, and $\text{tr}(I_2 - P) = 1$. For the block diagonal contraction consisting of m copies of P , a nonzero permutation term must preserve each block. The permanent is therefore 2^{-m} , while the trace deficit is m . Consequently any universal inequality $\text{per}(A) \leq e^{-\kappa \text{tr}(I - A)}$ must have $\kappa \leq \log 2$. In particular, the diagonal-case constant one does not extend to all PSD contractions.

Remark 6.3 (Where the dimension-free error enters). The key feature of Lemma 3.1 is its Frobenius-norm error. After realification, this becomes the off-diagonal part of $\text{tr}(K^2)$. Together with the exact Gamma transform of the diagonal part, the entire second-order error in Theorem 4.1 is $17 \text{tr}(K^2)$. For $K = \theta D$ and $0 \preceq D \preceq I_n$, this is at most $17\theta^2 \text{tr}(D)$, which can be absorbed into the linear term uniformly in n .

Remark 6.4 (Constants). The choices $p = 17$, $q = 17/16$, and $\theta = 1/34$ were made to keep the threshold arithmetic transparent. They are not expected to be optimal. The purpose of the estimate is the linear dependence on $\text{tr}(I_n - A)$, which is exactly what converts the quadratic ε -dependence in Theorem 1.1 into the linear dependence requested in [1, Section 1.4].

References

- [1] F. Ebrahimnejad, A. Nagda, and S. Oveis Gharan, *On approximability of the permanent of PSD matrices*, STOC 2025, 625–630. Numbered references use the full version, arXiv:2404.10959v1, 2024. <https://arxiv.org/abs/2404.10959>.
- [2] N. Anari and F. Ebrahimnejad, *Optimal $e^{(\gamma+o(1))n}$ -Approximation of the Permanent of Positive Semidefinite Matrices*, arXiv:2605.21946, 2026. <https://arxiv.org/abs/2605.21946>.
- [3] N. Anari, L. Gurvits, S. Oveis Gharan, and A. Saberi, *Simply exponential approximation of the permanent of positive semidefinite matrices*, FOCS 2017, 914–925. <https://arxiv.org/abs/1704.03486>.
- [4] C. Yuan and P. A. Parrilo, *Maximizing products of linear forms, and the permanent of positive semidefinite matrices*, Mathematical Programming **193** (2022), 499–510. <https://doi.org/10.1007/s10107-021-01616-3>.
- [5] A. Barvinok, *A remark on approximating permanents of positive definite matrices*, arXiv:2005.06344, 2020. <https://arxiv.org/abs/2005.06344>.
- [6] M. Rudelson and R. Vershynin, *Hanson–Wright inequality and sub-Gaussian concentration*, Electronic Communications in Probability **18** (2013), no. 82, 1–9. <https://arxiv.org/abs/1306.2872>.
- [7] D. Hsu, S. M. Kakade, and T. Zhang, *A tail inequality for quadratic forms of subgaussian random vectors*, Electronic Communications in Probability **17** (2012), no. 52, 1–6. <https://arxiv.org/abs/1110.2842>.
- [8] A. Laddha and M. R. Pittu, *An algorithmic upper bound for permanents via a permanental Schur inequality*, arXiv:2509.08121, 2025. <https://arxiv.org/abs/2509.08121>.
- [9] Z. Yi, *Diffuse Gaussian truncation for deterministic approximate counting*, arXiv:2609.04079, 2026. <https://arxiv.org/abs/2609.04079>.