58th Annual IEEE Symposium on Foundations of Computer Science

Simply Exponential Approximation of the Permanent of Positive Semidefinite Matrices

Nima Anari
Simons Institute
UC Berkeley
anari@berkeley.edu

Leonid Gurvits
Computer Science
City College of New York
gurvits@cs.ccny.cuny.edu

Shayan Oveis Gharan
Computer Science
University of Washington
shayan@cs.washington.edu

Amin Saberi
MS&E
Stanford University
saberi@stanford.edu

***Abstract*—We design a deterministic polynomial time c^n approximation algorithm for the permanent of positive semidefinite matrices where $c = e^{\gamma+1} \simeq 4.84$. We write a natural convex relaxation and show that its optimum solution gives a c^n approximation of the permanent. We further show that this factor is asymptotically tight by constructing a family of positive semidefinite matrices. We also show that our result implies an approximate version of the permanent-on-top conjecture, which was recently refuted in its original form; we show that the permanent is within a c^n factor of the top eigenvalue of the Schur power matrix.**

***Keywords*-permanent; positive semidefinite; permanent-on-top; approximation algorithm; semidefinite program; immanant;**

I. Introduction

Given a matrix $A \in \mathbb{C}^{n\times n}$, its permanent is defined as

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

where S_n is the set of permutations on $\{1,\dots,n\}$. There is a very rich body of work on permanent of matrices and its algebraic properties, see [1] for a recent survey on several theorems and open problems in this area.

The problem has been also studied from the point of view of computational complexity. Valiant [2] showed that it is #P-complete to compute the permanent of $\{0,1\}$-matrices. Aaronson [3] gave a new proof of the #P-hardness, using the model of linear optical quantum computing. In addition, he showed that it is #P-hard to compute the sign of $\mathrm{per}(A)$, essentially ruling out a multiplicative approximation. Grier and Schaeffer [4] extended Aaronson's proof and proved #P-hardness of computing the permanent of real orthogonal matrices. They also showed by a simple polynomial interpolation argument that it is #P-hard to compute the permanent of PSD matrices.

Given a general matrix $A \in \mathbb{R}^{n\times n}$, Gurvits [5] designed a randomized algorithm that in time $O(n^2/\epsilon^2)$ approximates $\mathrm{per}(A)$ within $\pm|A|^n$ additive error, where $|A|$ is the largest singular value of A. Chakhmakhchyan, Cerf, and Garcia-Patron [6] improve on Gurvits's algorithm if the matrix A is PSD and its eigenvalues satisfy a certain smoothness property.

If all entries of A are nonnegative then $\mathrm{per}(A) \geq 0$ by definition. In particular, if $A \in \{0,1\}^{n\times n}$, then $\mathrm{per}(A)$ is equal to the number of perfect matchings of the bipartite graph associated with A. Jerrum, Sinclair, and Vigoda [7], in a breakthrough, obtained a fully polynomial time randomized approximation scheme (FPRAS) for the permanent of matrices with nonnegative entries. In other words, they designed a randomized algorithm that for any given $\epsilon > 0$, outputs a $1+\epsilon$ multiplicative approximation of the permanent, in time polynomial in n and $1/\epsilon$. On the other hand, among deterministic polynomial time algorithms, the best known result is due to Gurvits and Samorodnitsky [8], who showed that the permanent of nonnegative matrices can be approximated within a factor of 2^n.

The focus of this paper is on the permanent of PSD matrices, which has received significant attention in the last decade because of its close connection to quantum optics. In particular, permanent of PSD matrices describe output probabilities of a boson sampling experiment in which the input is a tensor product of thermal states. They form the generating function of the quantum linear optical distribution [4].

It turns out that the permanent is a monotone function with respect to the Loewner order on the cone of PSD matrices and therefore the permanent of every PSD matrix is nonnegative (see corollaries 2 and 3). This fact is a priori not obvious considering that a PSD matrix can have negative entries. Since the permanent is nonnegative, unlike general matrices, there is no difficulty in computing the sign. So, it may be possible to design a polynomial time approximation scheme for the permanent of PSD matrices. This question has been posted as an open problem in several sources [9], [4]. Our main result can be seen as a first step along this line.

To this date, not much is known about multiplicative approximation of the permanent of PSD matrices. To the best of our knowledge, the only previous result is the work of Marcus [10] which shows that the product of the diagonal entries of a PSD matrix gives an $n!$ approximation of the

DOI 10.1109/FOCS.2017.89

permanent. For any PSD matrix $A \in \mathbb{R}^{n\times n}$,

$$\prod_{i=1}^{n} A_{i,i} \leq \operatorname{per}(A) \leq n! \prod_{i=1}^{n} A_{i,i}.$$

This approximation can be slightly improved using Lieb's permanent inequality [11]. Using this inequality one can show that $\operatorname{per}(A)$ can be approximated to within a factor of $n!/m!^{n/m}$ in time $2^{O(m+\log n)}$ for any desired m.

In this paper we design a c^n deterministic approximation algorithm for computing the permanent of PSD matrices, where $c > 0$ is a universal constant. Prior to our paper, no efficient algorithm (deterministic, randomized, or quantum) was known for simply exponential approximation of the permanent of general positive semidefinite matrices.

Theorem 1. *There is a deterministic polynomial time algorithm that for any given PSD matrix A returns a number $\operatorname{rel}(A)$ such that*

$$\operatorname{rel}(A) \geq \operatorname{per}(A) \geq c^{-n} \operatorname{rel}(A)$$

where $c = e^{\gamma+1}$ and γ is Euler's constant.

Our result uses a semidefinite relaxation. Because of the aforementioned monotonicity of the permanent with respect to the positive semidefinite order, a natural way to upper bound the permanent of a hermitian PSD matrix $A \in \mathbb{C}^{n\times n}$ is to find another matrix $D \succeq A$ whose permanent is easy to compute, and to use $\operatorname{per}(D)$ as the upper bound. For example if $D \succeq A$ is a diagonal matrix, then

$$\operatorname{per}(D) = D_{11}D_{22}\ldots D_{nn}$$

gives an easy-to-compute upper bound on $\operatorname{per}(A)$. This motivates the following natural relaxation for the permanent of PSD matrices.

Definition 1. *For an $n \times n$ hermitian PSD matrix A define*

$$\operatorname{rel}(A) := \inf\{\operatorname{per}(D) : D \text{ is diagonal and } D \succeq A\}. \quad (1)$$

Our main result is to prove that $\operatorname{rel}(A)$ also lower bounds $\operatorname{per}(A)$ up to a multiplicative factor. Additionally, we show that $\operatorname{rel}(A)$ can be efficiently computed using convex programming, thus giving a polynomial-time approximation algorithm for $\operatorname{per}(A)$.

Connection to the Permanent-On-Top Conjecture: As a byproduct of our main result we prove an approximate version of the permanent-on-top conjecture, originally formulated by Soules [12]. For an overview of the history of this conjecture, see, e.g., [13].

The permanent-on-top conjecture was motivated by an inequality between the permanent and the determinant of PSD matrices, first proved by Schur [14]. This inequality simply states that $\operatorname{per}(A) \geq \det(A)$ for any $A \succeq 0$. In fact, Schur proved a more general statement involving immanants: For any $A \succeq 0$,

$$\operatorname{imm}_\chi(A) \geq \det(A).$$

Here imm_χ is the immanant with respect to the character χ, which generalizes the notions of permanent and determinant. For any character χ of the symmetric group S_n, the function imm_χ is defined as

$$\operatorname{imm}_\chi(A) = \sum_{\sigma\in S_n} \chi(\sigma) \prod_{i=1}^{n} A_{i,\sigma(i)}.$$

For more detailed definitions, see, e.g. [13]. It is easy to see that we get the permanent when χ is the constant 1 character and we get the determinant when χ is the alternating character, i.e., the sign of permutations.

Schur's inequality shows that the determinant is the minimum amongst all immanants of a PSD matrix. This inspired Lieb's permanent dominance conjecture [11], which states that the permanent is the maximum amongst all immanants:

$$\operatorname{per}(A) \geq \operatorname{imm}_\chi(A).$$

To this date, the permanent dominance conjecture remains open.

The immanants of a matrix and their corresponding characters form eigenvalue/eigenvector pairs for the so-called Schur power matrix, which we define next. For a matrix $A \in \mathbb{C}^{n\times n}$, define the Schur power of A, which we denote by $A^{\otimes n}_{S_n,S_n}$, as the following $n! \times n!$ matrix: Let the rows and columns be indexed by S_n, and let the entry at row σ and column τ be simply

$$\prod_{i=1}^{n} A_{\sigma(i),\tau(i)}.$$

Note that the Schur power of A is a submatrix of $A^{\otimes n}$ which justifies the notation $A^{\otimes n}_{S_n,S_n}$.

It is not hard to verify that any character χ of S_n, viewed as an $n!$-dimensional vector, is an eigenvector of $A^{\otimes n}_{S_n,S_n}$ with eigenvalue $\operatorname{imm}_\chi(A)$. In fact Schur proved the stronger statement that $\det(A)$ is the smallest eigenvalue of this matrix. Motivated by this fact, Soules strengthened the permanent dominance conjecture to the following:

Conjecture 1 (Permanent-On-Top [12])**.** *For any $A \succeq 0$, the maximum eigenvalue of $A^{\otimes n}_{S_n,S_n}$ is equal to $\operatorname{per}(A)$:*

$$\|A^{\otimes n}_{S_n,S_n}\| = \operatorname{per}(A).$$

This conjecture would have implied the permanent dominance conjecture. Unfortunately the permanent-on-top conjecture was recently refuted [15]. We, on the other hand, in section V, prove the following theorem, which can be seen as an approximate form of the permanent-on-top conjecture.

Theorem 2. *For any PSD matrix $A \in \mathbb{C}^{n\times n}$, the multiplicative gap between the maximum eigenvalue of $A^{\otimes n}_{S_n,S_n}$ and $\operatorname{per}(A)$ can be at most c^n where $c = e^{\gamma+1}$:*

$$\|A^{\otimes n}_{S_n,S_n}\| \leq c^n \operatorname{per}(A).$$

This implies an approximate version of the permanent dominance conjecture as well.

Corollary 1. *For any PSD matrix $A \in \mathbb{C}^{n\times n}$ and any character χ of S_n we have*

$$\operatorname{imm}_\chi(A) \le c^n \operatorname{per}(A).$$

We remark that in light of the fact that the permanent-on-top conjecture is false, one cannot hope to show much stronger results than theorem 2.

Claim 1. *There exists a universal constant $\hat{c} > 1$ such that for infinitely many n we have a PSD matrix $A \in \mathbb{C}^{n\times n}$ with*

$$\|A^{\otimes n}_{S_n,S_n}\| \ge \hat{c}^n \operatorname{per}(A).$$

We prove theorem 2 and claim 1 in section V.

II. Preliminaries

We denote the set $\{1,\ldots,n\}$ by $[n]$. We use S_n to denote the set of permutations on $[n]$.

A. Linear Algebra

We identify vectors $v \in \mathbb{C}^n$ with $n\times 1$ matrices. For a matrix $A \in \mathbb{C}^{n\times m}$ we let $A^\dagger \in \mathbb{C}^{m\times n}$ denote its conjugate transpose; in other words $(A^\dagger)_{ij} = \overline{A_{ji}}$. A matrix $A \in \mathbb{C}^{n\times n}$ is called hermitian iff $A = A^\dagger$. A hermitian matrix A is called positive semidefinite (PSD) iff $v^\dagger A v \ge 0$ for all $v \in \mathbb{C}^n$. We let $\succeq$ denote the usual Loewner order on hermitian matrices, i.e., $A \succeq B$ iff $A-B$ is PSD. For a vector $v \in \mathbb{C}^n$, we let $\operatorname{diag}(v) \in \mathbb{C}^{n\times n}$ denote the diagonal matrix with coordinates of v as its main diagonal, i.e.,

$$\operatorname{diag}(v) := \begin{bmatrix} v_1 & 0 & \ldots & 0 \\ 0 & v_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & v_n \end{bmatrix}.$$

For matrices $A \in \mathbb{C}^{n\times m}$ and $B \in \mathbb{C}^{p\times q}$ we let $A\otimes B$ denote the Kronecker product, i.e., the following block matrix:

$$A\otimes B := \begin{bmatrix} A_{11}B & \ldots & A_{1m}B \\ \vdots & \ddots & \vdots \\ A_{n1}B & \ldots & A_{nm}B \end{bmatrix}.$$

For a matrix A and $n \ge 0$, we define $A^{\otimes n}$ as $\overbrace{A\otimes A\otimes\cdots\otimes A}^{n}$. The Kronecker product respects the Loewner order on hermitian PSD matrices:

Fact 1. *If $A \succeq B \succeq 0$ and $C \succeq D \succeq 0$, then $A\otimes C \succeq B\otimes D \succeq 0$.*

B. Standard Complex Normal Distribution

We say that a complex-valued random variable $g = \operatorname{Re}(g) + i\operatorname{Im}(g)$ is distributed according to a standard complex normal, which we denote by $g \sim \mathbb{C}\mathcal{N}(0,1)$, iff $(\operatorname{Re}(g),\operatorname{Im}(g)) \sim \mathcal{N}(0,\frac{1}{2}I)$. The probability density function (over $\mathbb{C}\simeq\mathbb{R}^2$) for this distribution is given by

$$\frac{1}{\pi}e^{-(\operatorname{Re}(g)^2+\operatorname{Im}(g)^2)} = \frac{1}{\pi}e^{-|g|^2}.$$

Fact 2. *If $g \sim \mathbb{C}\mathcal{N}(0,1)$, then for integers $n,m \ge 0$ we have*

$$\mathbb{E}[g^n\overline{g}^m] = \begin{cases} 0 & \text{if } n \ne m, \\ n! & \text{if } n = m. \end{cases}$$

Proof: The distribution of g is circularly symmetric, i.e. for $u \in \mathbb{C}$ with $|u| = 1$, we have $ug \sim \mathbb{C}\mathcal{N}(0,1)$. This means that

$$\mathbb{E}[g^n\overline{g}^m] = \mathbb{E}[(ug)^n(\overline{ug})^m] = u^{n-m}\mathbb{E}[g^n\overline{g}^m].$$

Therefore, unless $n-m=0$, we have $\mathbb{E}[g^n\overline{g}^m]=0$. When $m=n$, we have $g^n\overline{g}^m = |g|^{2n}$. If we let $r = |g| \in \mathbb{R}_{\ge 0}$, then the probability density function of r is given by $2\pi r\frac{1}{\pi}e^{-r^2} = 2re^{-r^2}$. Therefore we have

$$\begin{aligned}\mathbb{E}[|g|^{2n}] &= \int_0^\infty r^{2n}\cdot 2re^{-r^2}dr \\ &= -r^{2n}\cdot e^{-r^2}\Big|_0^\infty + \int_0^\infty 2nr^{2n-1}\cdot e^{-r^2}dr \\ &= n\int_0^\infty r^{2n-2}\cdot 2re^{-r^2}dr = n\cdot\mathbb{E}[|g|^{2n-2}],\end{aligned}$$

where we used integration by parts. We can finally derive

$$\begin{aligned}\mathbb{E}[|g|^{2n}] &= n\cdot\mathbb{E}[|g|^{2n-2}] = n(n-1)\cdot\mathbb{E}[|g|^{2n-4}] = \ldots \\ &= n!\cdot\mathbb{E}[|g|^0] = n!.\end{aligned}$$

■

Fact 3. *If $g \sim \mathbb{C}\mathcal{N}(0,1)$, then*

$$\mathbb{E}[\ln(|g|^2)] = -\gamma,$$

where γ is Euler's constant.

Proof: Note that $|g|^2 = \operatorname{Re}(g)^2 + \operatorname{Im}(g)^2 = \frac{1}{2}(2\operatorname{Re}(g)^2 + 2\operatorname{Im}(g)^2)$. Since $(\operatorname{Re}(g),\operatorname{Im}(g)) \sim \mathcal{N}(0,\frac{1}{2}I)$, the random variable $2\operatorname{Re}(g)^2 + 2\operatorname{Im}(g)^2$ is distributed according to a χ^2-distribution with 2 degrees of freedom, which is identical to a $\Gamma(1,2)$ distribution [16]. Therefore we have

$$\mathbb{E}[\ln(2|g|^2)] = \psi(1) + \ln(2),$$

where ψ is the digamma function [16]. This implies that $\mathbb{E}[\ln(|g|^2)] = \psi(1)$, and the latter is equal to $-\gamma$ [17]. ■

We say that a random vector $v \in \mathbb{C}^n$ is distributed according to a standard complex normal, which we denote by $v \sim \mathbb{C}\mathcal{N}(0,I)$, iff $v_1,\ldots,v_n$ are independent standard complex normals.

Fact 4. *If $v \sim \mathbb{C}\mathcal{N}(0, I)$, and $u \in \mathbb{C}^n$ is a unit vector, i.e., $|u|^2 = u^\dagger u = 1$, then $u^\dagger v \sim \mathbb{C}\mathcal{N}(0, 1)$.*

Proof: Note that $(\mathrm{Re}(u^\dagger v), \mathrm{Im}(u^\dagger v))$ are linear combinations of the real and imaginary parts of v; as such, this 2-dimensional vector is distributed according to $\mathcal{N}(\mu, \Sigma)$ for some $\mu \in \mathbb{R}^2$ and $\Sigma \in \mathbb{R}^{2\times 2}$.

The distribution of $u^\dagger v$ is circularly symmetric; i.e., if $\phi \in \mathbb{C}$ is such that $|\phi| = 1$, then $\phi u^\dagger v$ is distributed the same way as $u^\dagger v$. This is true because $\phi u^\dagger v = u^\dagger(\phi v)$, and ϕv has the same distribution as v. Being circularly symmetric implies that $\mu = 0$ and $\Sigma = cI$ for some constant c. On the other hand, we have

$$\begin{aligned} 2c &= \mathbb{E}[|u^\dagger v|^2] = \mathbb{E}[u^\dagger v v^\dagger u] = u^\dagger \mathbb{E}[vv^\dagger]u \\ &= u^\dagger I u = |u|^2 = 1. \end{aligned}$$

Therefore $(\mathrm{Re}(u^\dagger v), \mathrm{Im}(u^\dagger v)) \sim \mathcal{N}(0, \frac{1}{2}I)$ or in other words, $u^\dagger v \sim \mathbb{C}\mathcal{N}(0, 1)$. ■

C. Permanent and Loewner Order

For a matrix $A \in \mathbb{C}^{n\times n}$, its permanent is defined as

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

Permanent is a monotone function on the space of PSD matrices w.r.t. the Loewner order. For completeness we sketch the proof given in [1] here.

Lemma 1. *For any matrix $M \in \mathbb{C}^{n\times n}$, there is a vector $1_{S_n} \in \mathbb{C}^{n^n}$ such that*

$$\mathrm{per}(M) := \frac{1}{n!} 1_{S_n}^\dagger M^{\otimes n} 1_{S_n}.$$

Proof: The vector $1_{S_n} \in \mathbb{C}^{n^n}$ is constructed in the following way: Index each of the n^n coordinates by $\sigma \in [n]^n$ in the usual way (so that the indices respect the Kronecker product); we can think of σ as a function from $[n]$ to $[n]$. Then let the σ-th coordinate of 1_{S_n} be 1 iff σ is a permutation on $[n]$, and let it be 0 otherwise. Then, for a matrix M we have

$$\begin{aligned} 1_{S_n}^\dagger M^{\otimes n} 1_{S_n} &= \sum_{\sigma \in S_n} \sum_{\sigma' \in S_n} \prod_{i=1}^n M_{\sigma(i),\sigma'(i)} \\ &= \sum_{\sigma \in S_n} \mathrm{per}(M) = n! \cdot \mathrm{per}(M). \end{aligned}$$

■

Corollary 2. *If $A, B \in \mathbb{C}^{n\times n}$ are hermitian and $A \succeq B \succeq 0$, then*

$$\mathrm{per}(A) \geq \mathrm{per}(B).$$

Proof: The statement of the lemma follows, because $A \succeq B \succeq 0$ implies that $A^{\otimes n} \succeq B^{\otimes n} \succeq 0$ by fact 1. So, by lemma 1,

$$\mathrm{per}(A) = \frac{1}{n!} 1_{S_n}^\dagger A^{\otimes n} 1_{S_n} \geq \frac{1}{n!} 1_{S_n}^\dagger B^{\otimes n} 1_{S_n} = \mathrm{per}(B)$$

as desired. ■

Corollary 3. *For any hermitian PSD matrix $A \in \mathbb{C}^{n\times n}$, $\mathrm{per}(A) \geq 0$.*

Proof: This follows from corollary 2 by setting $B = 0$. ■

There is another way to show nonnegativity of the permanent over the PSD cone with the help of the complex normal distribution. For a vector $v \in \mathbb{C}^n$ define

$$|v|_\Pi := \sqrt{\prod_{i=1}^n |v_i|^2} \geq 0.$$

Then with the help of $|\cdot|_\Pi$ we can express the permanent of a PSD matrix as an expectation of a nonnegative value.

Lemma 2. *Let $U \in \mathbb{C}^{d\times n}$ be arbitrary and let $x \in \mathbb{C}^d$ be a random vector distributed according to the standard complex normal $\mathbb{C}\mathcal{N}(0, I)$. Then*

$$\mathrm{per}(U^\dagger U) = \mathbb{E}_{x\sim\mathbb{C}\mathcal{N}(0,I)}[|U^\dagger x|_\Pi^2].$$

Lemma 2 is a sepcial case of the relationship between the so-called G-norm and the quantum permanent shown in [18]. In particular if the rows of U are $u_1^\dagger, \ldots, u_d^\dagger$, then

$$|U^\dagger x|_\Pi^2 = |\det(\sum_{i=1}^d x_i \,\mathrm{diag}(u_i))|^2,$$

and therefore $\mathbb{E}_x[|U^\dagger x|_\Pi^2]$ is the same as the G-norm of the polynomial $\det(\sum_{i=1}^d x_i \,\mathrm{diag}(u_i))$. In [18] this is shown to be equal to the quantum permanent of the linear operator with Choi form given by the matrices $\mathrm{diag}(u_1), \ldots, \mathrm{diag}(u_d)$. It can be further shown that in this special case, the quantum permanent reduces to $\mathrm{per}(U^\dagger U)$. For exact definitions and further details see [18].

For the sake of completeness, we give a self-contained proof of lemma 2 below.

Proof of lemma 2: We will use the fact that the expression $|U^\dagger x|_\Pi^2$ is a polynomial in $x_1, \ldots, x_d$ and $\overline{x_1}, \ldots, \overline{x_d}$; therefore we can evaluate its expectation with the help of fact 2. We have

$$|U^\dagger x|_\Pi^2 = |\prod_{i=1}^n \sum_{j=1}^d \overline{U_{ji}} x_j|^2$$

If we define

$$p(x) := \prod_{i=1}^n \sum_{j=1}^d \overline{U_{ji}} x_j,$$

then $|U^\dagger x|_\Pi^2 = p(x)\overline{p(x)}$. Note that $p(x)$ is a polynomial in terms of $x_1, \ldots, x_d$. We can expand $p(x)$ as follows:

$$p(x) = \sum_{\sigma:[n]\to[d]} \prod_{i=1}^n \overline{U_{\sigma(i),i}} x_{\sigma(i)},$$

where the sum is taken over all n^d functions $\sigma : [n] \to [d]$. For a function $\sigma : [n] \to [d]$, let $\mathrm{sig}(\sigma)$ be $(k_1, \dots, k_d) \in \mathbb{Z}^d$ where k_j is the number of $i \in [n]$ such that $\sigma(i) = j$. Then we can alternatively write

$$p(x) = \sum_{\substack{k_1+\dots+k_d=n\\k_1,\dots,k_d\geq 0}} \left(x_1^{k_1} \dots x_d^{k_d} \sum_{\substack{\sigma:[n]\to[d]\\\mathrm{sig}(\sigma)=(k_1,\dots,k_d)}} \prod_{i=1}^{n} \overline{U_{\sigma(i),i}} \right).$$

For $(k_1, \dots, k_d) \neq (k'_1, \dots, k'_d)$, by fact 2 we have $\mathbb{E}_x[x_1^{k_1} \dots x_d^{k_d} \overline{x_1^{k'_1} \dots x_d^{k'_d}}] = 0$. Therefore we can write

$$\mathbb{E}_x[p(x)\overline{p(x)}] = \sum_{\substack{k_1+\dots+k_d=n\\k_1,\dots,k_d\geq 0}} \left(k_1! \dots k_d! \sum_{\substack{\sigma:[n]\to[d]\\\mathrm{sig}(\sigma)=(k_1,\dots,k_d)}} \sum_{\substack{\sigma':[n]\to[d]\\\mathrm{sig}(\sigma')=(k_1,\dots,k_d)}} \prod_{i=1}^{n} \overline{U_{\sigma(i),i}} U_{\sigma'(i),i} \right),$$

where we used that $\mathbb{E}[x_1^{k_1} \dots x_d^{k_d} \overline{x_1^{k_1} \dots x_d^{k_d}}] = k_1! \dots k_d!$ by fact 2. Note that when $\mathrm{sig}(\sigma) = \mathrm{sig}(\sigma')$, there is a permutation $\pi \in S_n$ such that $\sigma' = \sigma \circ \pi$. In fact if $\mathrm{sig}(\sigma) = \mathrm{sig}(\sigma') = (k_1, \dots, k_d)$, then the number of $\pi \in S_n$ for which $\sigma' = \sigma \circ \pi$ is exactly equal to $k_1! \dots k_d!$. Therefore we can rewrite the above sum as

$$\begin{aligned}
\mathbb{E}_x[p(x)\overline{p(x)}] &= \sum_{\sigma:[n]\to[d]} \sum_{\pi\in S_n} \prod_{i=1}^{n} \overline{U_{\sigma(i),i}} U_{\sigma(\pi(i)),i} \\
&= \sum_{\pi\in S_n} \sum_{\sigma[n]\to[d]} \prod_{i=1}^{n} (U^\dagger)_{i,\sigma(i)} U_{\sigma(i),\pi^{-1}(i)} \\
&= \sum_{\pi\in S_n} \prod_{i=1}^{n} \sum_{j=1}^{d} (U^\dagger)_{i,j} U_{j,\pi^{-1}(i)} \\
&= \sum_{\pi\in S_n} \prod_{i=1}^{n} (U^\dagger U)_{i,\pi^{-1}(i)} = \mathrm{per}(U^\dagger U).
\end{aligned}$$

■

III. Approximation of Permanent on the PSD Cone

In this section we prove theorem 1. Recall the definition of $\mathrm{rel}(A)$ from definition 1. Our first step is to prove that for every $n \times n$ hermitian PSD matrix $A \succeq 0$:

$$c^n \,\mathrm{per}(A) \geq \mathrm{rel}(A), \tag{2}$$

where $c = e^{\gamma+1}$.

We will prove a stronger statement. We find a vector $v \in \mathbb{C}^n$ such that $A \succeq vv^\dagger$. By corollary 2, $\mathrm{per}(A) \geq \mathrm{per}(vv^\dagger)$. So eq. (2) is implied by the following:

Theorem 3. *For a hermitian PSD matrix $A \in \mathbb{C}^{n\times n}$, there exists $v \in \mathbb{C}^n$ such that $A \succeq vv^\dagger$ and*

$$c^n \,\mathrm{per}(vv^\dagger) \geq \mathrm{rel}(A),$$

where $c = e^{\gamma+1}$.

Note that the above shows that for every hermitian PSD matrix $A \in \mathbb{C}^{n\times n}$, there exists a diagonal matrix D and a rank 1 matrix $vv^\dagger$ such that

$$D \succeq A \succeq vv^\dagger,$$

and $\mathrm{per}(D) \leq c^n \,\mathrm{per}(vv^\dagger)$ for $c = e^{\gamma+1}$. Thus $\mathrm{per}(A)$ is sandwiched between $\mathrm{per}(D)$ and $\mathrm{per}(vv^\dagger)$, two quantities that differ by at most a simply exponential factor.

It is also worth noting that there is no additional loss in approximating $\mathrm{per}(A)$ by the permanent of a rank one matrix. In section IV, we will show that the constant $e^{\gamma+1}$ is not only asymptotically tight in theorem 3, but also in eq. (2).

Another interesting corollary of theorem 3 is that that instead of $\mathrm{rel}(A)$ we can use $\mathrm{per}(vv^\dagger)$ as an approximation of $\mathrm{per}(A)$, with the same $e^{n(\gamma+1)}$ approximation factor:

$$\sup\{\mathrm{per}(vv^\dagger) : v \in \mathbb{C}^n \text{ and } A \succeq vv^\dagger\}. \tag{3}$$

Moreover, $\mathrm{per}(vv^\dagger)$ is easily computable.

Fact 5. *For a vector $v \in \mathbb{C}^n$, we have $\mathrm{per}(vv^\dagger) = n! \cdot \prod_{i=1}^{n} |v_i|^2$.*

Proof: For any permutation $\sigma \in S_n$ we have

$$\prod_{i=1}^{n} (vv^\dagger)_{i,\sigma(i)} = \prod_{i=1}^{n} v_i \overline{v_{\sigma(i)}} = \prod_{i=1}^{n} v_i \cdot \prod_{i=1}^{n} \overline{v_i} = \prod_{i=1}^{n} |v_i|^2.$$

Since $\mathrm{per}(vv^\dagger)$ is the sum of the above quantity for all $\sigma \in S_n$, we get that $\mathrm{per}(vv^\dagger) = n! \cdot \prod_{i=1}^{n} |v_i|^2$. ■

Even though $\mathrm{per}(vv^\dagger)$ has a closed form, we do not have an efficient way of computing the sup in eq. (3), whereas, as we show in section III-B, $\mathrm{rel}(A)$ can be computed efficiently.

The next section is dedicated to proving theorem 3. To finish up the proof of theorem 1 we need to design an algorithm to compute $\mathrm{rel}(A)$ for a given PSD matrix A.

Theorem 4. *There is an algorithm that outputs an $e^{n(\gamma+1)}$-approximation of $\mathrm{per}(A)$ for any hermitian PSD $A \in \mathbb{C}^{n\times n}$ in time $\mathrm{poly}(n + \langle A\rangle)$, where $\langle A\rangle$ represents the bit complexity of A.*

We will prove the above theorem in section III-B. Theorems 3 and 4 together complete the proof of theorem 1. In section IV we show that the constant $c = e^{\gamma+1}$ in eq. (2) is asymptotically tight.

A. Proof of the Main Result

In order to prove theorem 3, we use a seemingly unrelated quantity about distributions on unit vectors $\{u \in \mathbb{C}^d : |u|^2 = u^\dagger u = 1\}$. Let us define this quantity below.

Definition 2. *For a discrete distribution $\mathcal{U}$ supported on the sphere $\{u \in \mathbb{C}^d : |u|^2 = u^\dagger u = 1\}$, define*

$$f(\mathcal{U}) := \sup_{x \in \mathrm{span}(\mathcal{U})} \left\{ \frac{e^{\mathbb{E}_{u\sim\mathcal{U}}[\ln(|u^\dagger x|^2)]}}{\mathbb{E}_{u\sim\mathcal{U}}[|u^\dagger x|^2]} \right\},$$

where $\mathrm{span}(\mathcal{U})$ is the span of the support of $\mathcal{U}$, i.e., the set of vectors for which the denominator is nonzero.

We will prove theorem 3 by showing that there exists $v \in \mathbb{C}^n$ such that $A \succeq vv^\dagger$ and

$$\mathrm{per}(vv^\dagger) \geq \frac{n!}{n^n} f(\mathcal{U})^n \cdot \mathrm{rel}(A),$$

where $\mathcal{U}$ is an appropriately constructed distribution on unit vectors. The expression $n!/n^n$ is lower bounded by e^{-n}. Thus if we show that $f(\mathcal{U}) \geq e^{-\gamma}$, the above inequality would imply the multiplicative factor of $e^{n(\gamma+1)}$ desired in theorem 3.

To gain some intuition about $f(\mathcal{U})$, note that by Jensen's inequality, applied to the concave function $\ln$, it is easy to see that $f(\mathcal{U}) \leq 1$:

$$\frac{e^{\mathbb{E}_{u\sim\mathcal{U}}[\ln(|u^\dagger x|^2)]}}{\mathbb{E}_{u\sim\mathcal{U}}[|u^\dagger x|^2]} \leq \frac{e^{\ln(\mathbb{E}_{u\sim\mathcal{U}}[|u^\dagger x|^2])}}{\mathbb{E}_{u\sim\mathcal{U}}[|u^\dagger x|^2]} = 1.$$

On the other hand, we will show that for all $\mathcal{U}$, $f(\mathcal{U}) \geq e^{-\gamma}$.

Proposition 1. *For all discrete distributions $\mathcal{U}$ supported on the sphere $\{u \in \mathbb{C}^d : |u|^2 = u^\dagger u = 1\}$,*

$$f(\mathcal{U}) \geq e^{-\gamma}.$$

This universal lower bound is independent of the dimension d or the size of the support of $\mathcal{U}$. We defer the proof of proposition 1 to the end of this section. It is worth mentioning that the $\sup$ in the definition of $f(\mathcal{U})$ can be replaced by $\max$, since an appropriate power of $f(\mathcal{U})$ can be written as the $\sup$ of a rational function with no poles, over the unit sphere.

Let us now prove theorem 3, assuming correctness of proposition 1.

Proof of theorem 3: Let us break down the proof into a series of claims, and then prove them one by one.

Claim 2. *The infimum in eq.* (1) *is achieved by some diagonal matrix $\hat{D} = \hat{D}(A)$. In other words there exists a diagonal matrix $\hat{D} \succeq A$ such that $\mathrm{per}(\hat{D}) = \mathrm{rel}(A)$.*

Claim 3. *We may assume without loss of generality that $\hat{D} = I$.*

Claim 4. *The first-order optimality condition of $\hat{D}$ implies that there exists a correlation matrix $B \in \mathbb{C}^{n\times n}$, i.e., a hermitian PSD matrix with 1s on its main diagonal, such that $AB = B$.*

We may use the Cholesky decomposition to write $B = U^\dagger U$ where $U \in \mathbb{C}^{d\times n}$ for $d = \mathrm{rank}(B)$.

Claim 5. *For any $x \in \mathbb{C}^d$ the vector $v = U^\dagger x / |U^\dagger x|$ satisfies*

$$A \succeq vv^\dagger.$$

Naturally we may want to choose x so as to maximize $\mathrm{per}(vv^\dagger)$.

Claim 6. *We have*

$$\sup_{x\in\mathbb{C}^d} \{\mathrm{per}(vv^\dagger)\} = \frac{n!}{n^n} f(\mathcal{U})^n,$$

where $\mathcal{U}$ is the uniform distribution on the columns of U.

And now the statement of theorem 3 follows, because $\mathrm{rel}(A) = \mathrm{per}(\hat{D}) = 1$ when $\hat{D} = I$; we have found $v \in \mathbb{C}^n$ such that $A \succeq vv^\dagger$ and

$$e^{n(\gamma+1)}\,\mathrm{per}(vv^\dagger) \geq \frac{n^n}{n!} f(\mathcal{U})^{-n}\,\mathrm{per}(vv^\dagger) \geq 1 = \mathrm{rel}(A).$$

We remark that these claims can be understood as analogues to the classical Van der Waerden conjecture. Claim 4 provides us with the correlation matrix B. Claim 5 and claim 6 can be thought of as providing a lower bound for the permanent of the orthogonal projector onto the image of B. This is analogous to the Van der Waerden conjecture which provides a lower bound of $n!/n^n$ for the permanent of doubly stochastic matrices. Here, claim 5, claim 6, and proposition 1 prove that the permanent of the orthogonal projector onto the image of any correlation matrix is at least

$$\frac{n!}{n^n} e^{-n\gamma}.$$

Let us now prove the claims one by one.

Proof of claim 2: We divide the proof into two cases. First assume that $A_{ii} > 0$ for all $i \in [n]$. Let $\lambda \geq 0$ be larger than the maximum eigenvalue of A. Then $\lambda I \succeq A$. This proves that $\mathrm{rel}(A) \leq \lambda^n$. Note that $D \succeq A$ implies $D_{ii} \geq A_{ii}$ for all $i \in [n]$. If any entry D_{ii} of D satisfies

$$D_{ii} > \frac{\lambda^n A_{ii}}{\prod_{j=1}^n A_{jj}},$$

then

$$\mathrm{per}(D) > \frac{\lambda^n A_{ii}}{\prod_{j=1}^n A_{jj}} \prod_{j\neq i} A_{jj} = \lambda^n.$$

This effectively eliminates such a D as a candidate for the inf in eq. (1). Therefore we may take inf of $\mathrm{per}(D)$ over the set of all diagonal matrices D which in addition to $D \succeq A$ satisfy

$$D_{ii} \leq \frac{\lambda^n A_{ii}}{\prod_{j=1}^n A_{jj}}$$

for all $i \in [n]$. This is a compact set, and $\mathrm{per}(D)$ is a continuous function. Therefore the inf is achieved by some matrix $\hat{D}$.

For the second case, assume that $A_{ii} = 0$ for some i. Then since A is PSD, the i-th row and the i-th column of A are both zero. Let λ be larger than the largest eigenvalue of A. Define $\hat{D}$ by $\hat{D}_{ii} = 0$ and $\hat{D}_{jj} = \lambda$ for $j \neq i$. It is easy to see that $\hat{D} \succeq A$ and $\mathrm{per}(\hat{D}) = 0$. Therefore $\mathrm{rel}(A) = 0$ and it is achieved at $\hat{D}$. ■

Proof of claim 3: First note that without loss of generality we may assume $\hat{D}(A) \succ 0$, since otherwise $\mathrm{rel}(A) = 0$ and the conclusion of theorem 3 is trivial.

Now let $\lambda \in \mathbb{R}^n_{>0}$ be an arbitrary positive vector and define $T_\lambda : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ by

$$T_\lambda(M) = \mathrm{diag}(\lambda) M \,\mathrm{diag}(\lambda).$$

Note that T_λ respects the Loewner order and maps diagonal matrices to diagonal matrices. It is one-to-one and surjective on the space of diagonal matrices. The matrix $T_\lambda(M)$ is obtained from M by multiplying column i by λ_i for $i \in [n]$ and then row i by λ_i for $i \in [n]$. Therefore

$$\mathrm{per}(T_\lambda(M)) = \lambda_1^2 \dots \lambda_n^2 \,\mathrm{per}(M).$$

This implies that

$$\begin{aligned}
&\lambda_1^2 \dots \lambda_n^2 \,\mathrm{rel}(A)\\
&= \inf\{\lambda_1^2 \dots \lambda_n^2 \,\mathrm{per}(D) : D \text{ diagonal and } D \succeq A\}\\
&= \inf\{\mathrm{per}(T_\lambda(D)) : T_\lambda(D) \text{ diagonal and } T_\lambda(D) \succeq T_\lambda(A)\}\\
&= \mathrm{rel}(T_\lambda(A)).
\end{aligned}$$

It is also easy to see that the above also implies $\hat{D}(T_\lambda(A)) = T_\lambda(\hat{D}(A))$. In particular if λ is set so that $\lambda_i = 1/\sqrt{\hat{D}_{ii}}$, then $\hat{D}(T_\lambda(A)) = I$. So we can replace A by $T_\lambda(A)$ and continue the proof of theorem 3 to find $v \in \mathbb{C}^n$ satisfying

$$T_\lambda(A) \succeq vv^\dagger,$$

and $c^n \,\mathrm{per}(vv^\dagger) \geq \mathrm{rel}(T_\lambda(A)) = 1$ with $c = e^{\gamma+1}$. Let $w = \mathrm{diag}(\lambda)^{-1} v$. Then $T_\lambda(ww^\dagger) = vv^\dagger$. This implies that

$$A \succeq ww^\dagger,$$

and

$$\begin{aligned}
c^n \,\mathrm{per}(ww^\dagger) &= \frac{1}{\lambda_1^2 \dots \lambda_n^2} c^n \,\mathrm{per}(vv^\dagger)\\
&\geq \frac{1}{\lambda_1^2 \dots \lambda_n^2} \,\mathrm{rel}(T_\lambda(A)) = \mathrm{rel}(A).
\end{aligned}$$

■

Proof of claim 4: We use the first-order optimality condition of $\mathrm{per}(D)$ at $D = I$. Let us see how $\mathrm{per}(I + X)$ compares to $\mathrm{per}(I)$ where X is a diagonal matrix. If X is small enough $\mathrm{per}(I + X) \simeq 1 + \mathrm{tr}(X)$. More precisely, we have

$$\begin{aligned}
\frac{d}{dt} \mathrm{per}(I + tX)\Big|_{t=0} &= \frac{d}{dt} \prod_{i=1}^n (1 + tX_{ii})\Big|_{t=0}\\
&= \sum_{i=1}^n X_{ii} = \mathrm{tr}(X).
\end{aligned}$$

If $I + X \succeq A$ then $I + tX \succeq A$ for all $t \in [0, 1]$. If $\mathrm{tr}(X) < 0$, then for small enough t, $\mathrm{per}(I + tX) < \mathrm{per}(I)$ which contradicts the fact that $\hat{D}(A) = I$. This implies that the optimal solution of the following SDP is 0:

$$\begin{array}{lll}
\min_X & \mathrm{tr}(X) & \\
\text{subject to} & I + X \succeq A & \\
& X_{ij} = 0 & \forall i \neq j
\end{array}$$

The dual of this SDP has the variable $B \succeq 0$, corresponding to the constraint $I + X \succeq A$:

$$\begin{array}{lll}
\max_{B,\mu_{ij}} & \mathrm{tr}((A - I)B) & \\
\text{subject to} & B_{ii} = 1 & \forall i\\
& B \succeq 0 &
\end{array}$$

Because of strong duality, the optimum of this SDP is 0. The optimal B satisfies $B \succeq 0$ and $B_{ii} = 1$ for $i \in [n]$, i.e., B is a correlation matrix. We also have $\mathrm{tr}((I - A)B) = 0$. But since $I - A \succeq 0$ and $B \succeq 0$, this implies that $(I - A)B = 0$ or in other words $AB = B$. ■

Proof of claim 5: We have $B = U^\dagger U$ with $U \in \mathbb{C}^{d\times n}$ and $\mathrm{rank}(B) = d$. This implies that $UU^\dagger \in \mathbb{C}^{d\times d}$ is invertible. Now we have

$$BU^\dagger (UU^\dagger)^{-1} x = U^\dagger UU^\dagger (UU^\dagger)^{-1} x = U^\dagger x.$$

This together with $AB = B$ implies that

$$AU^\dagger x = ABU^\dagger (UU^\dagger)^{-1} x = BU^\dagger (UU^\dagger)^{-1} x = U^\dagger x.$$

In other words, $U^\dagger x$ is an eigenvector of A with eigenvalue 1. This means that $v = U^\dagger x / |U^\dagger x|$ is also such an eigenvector. So $Av = v$ and $|v| = 1$. We conclude that $A \succeq vv^\dagger$. ■

Proof of claim 6: Let us compute $\mathrm{per}(vv^\dagger)$. By fact 5 we have

$$\mathrm{per}(vv^\dagger) = n! \cdot \prod_{i=1}^n |v_i|^2.$$

Let the columns of U be $u_1, \dots, u_n \in \mathbb{C}^d$. Then $v_i = u_i^\dagger x / |U^\dagger x|$, and note that $|U^\dagger x|^2 = \sum_{i=1}^n |u_i^\dagger x|^2$. We can rewrite $\mathrm{per}(vv^\dagger)$ as

$$\begin{aligned}
\mathrm{per}(vv^\dagger) &= n! \cdot \frac{\prod_{i=1}^n |u_i^\dagger x|^2}{(\sum_{i=1}^n |u_i^\dagger x|^2)^n}\\
&= \frac{n!}{n^n} \cdot \left(\frac{\sqrt[n]{\prod_{i=1}^n |u_i^\dagger x|^2}}{\frac{1}{n} \sum_{i=1}^n |u_i^\dagger x|^2} \right)^n .
\end{aligned}$$

Now if we let $\mathcal{U}$ be the uniform distribution on $u_1, \ldots, u_n$, we can rewrite the above as

$$\operatorname{per}(vv^\dagger) = \frac{n!}{n^n} \cdot \left(\frac{\exp(\mathbb{E}_{u \sim \mathcal{U}}[\ln(|u^\dagger x|^2)])}{\mathbb{E}_{u \sim \mathcal{U}}[|u^\dagger x|^2]} \right)^n$$

Therefore

$$\sup_{x \in \mathbb{C}^d} \{\operatorname{per}(vv^\dagger)\} = \frac{n!}{n^n} f(\mathcal{U})^n.$$

■

This concludes the proof of theorem 3. ■

It only remains to prove proposition 1.

Proof of proposition 1: Without loss of generality we may assume that $\operatorname{span}(\mathcal{U}) = \mathbb{C}^d$; if that is not the case, we can identify $\operatorname{span}(\mathcal{U})$ with $\mathbb{C}^{d'}$ for some $d' < d$ using a unitary transformation and nothing changes.

Let $x \sim \mathbb{C}\mathcal{N}(0, I)$ be a d-dimensional standard complex normal. Let

$$g(x) = \exp(\mathbb{E}_{u \sim \mathcal{U}}[\ln(|u^\dagger x|^2)]),$$
$$h(x) = \mathbb{E}_{u \sim \mathcal{U}}[|u^\dagger x|^2].$$

Then our goal is to prove that $\mathbb{P}_x[g(x)/h(x) \geq e^{-\gamma}] > 0$ or equivalently $\mathbb{P}_x[g(x) - e^{-\gamma} h(x) \geq 0] > 0$. To this end, we will prove that $\mathbb{E}_x[g(x) - e^{-\gamma} h(x)] \geq 0$, and the conclusion follows.

By fact 4, for each fixed u in the support of $\mathcal{U}$, $u^\dagger x \sim \mathbb{C}\mathcal{N}(0, 1)$. Therefore we have

$$\mathbb{E}_x[h(x)] = \mathbb{E}_x \mathbb{E}_u[|u^\dagger x|^2] = \mathbb{E}_u \mathbb{E}_x[|u^\dagger x|^2] = \mathbb{E}_u[1] = 1.$$

On the other hand by fact 3 we have

$$\begin{aligned}\mathbb{E}_x[g(x)] &= \mathbb{E}_x[\exp(\mathbb{E}_u[\ln(|u^\dagger x|^2)])] \\ &\geq \exp(\mathbb{E}_x \mathbb{E}_u[\ln(|u^\dagger x|^2)]) \\ &= \exp(\mathbb{E}_u \mathbb{E}_x[\ln(|u^\dagger x|^2)]) = \exp(\mathbb{E}_u[-\gamma]) = e^{-\gamma},\end{aligned}$$

where the inequality is an application of Jensen's to the convex function exp. Putting these together we get that $\mathbb{E}_x[g(x) - e^{-\gamma} h(x)] \geq e^{-\gamma} - e^{-\gamma} = 0$ as desired. ■

B. Computing the Approximation

In this section we show how to approximately compute $\operatorname{rel}(A)$. The main result of this section will be theorem 4.

The main ingredient of the proof is transforming $\operatorname{rel}(A)$ to the objective of a convex program. The original optimization problem that computes $\operatorname{rel}(A)$ is the following:

$$\begin{array}{ll} \min_D & D_{11} \ldots D_{nn} \\ \text{subject to} & D \succeq A \\ & D \text{ is diagonal} \end{array}$$

The objective is not concave, even if we apply ln to it. The trick is to change from the variables $D_{11}, \ldots, D_{nn}$ to $D_{11}^{-1}, \ldots, D_{nn}^{-1}$. If we have the Cholesky decomposition $A = V^\dagger V$ for some $V \in \mathbb{C}^{d \times n}$, then $D \succeq A$ if and only if

$$I \succeq V D^{-1} V^\dagger.$$

So we can turn the optimization problem into the following by identifying D^{-1} with $\operatorname{diag}(x)$.

$$\begin{array}{lll} \min_{x \in \mathbb{R}^n} & -\ln(x_1 \ldots x_n) & \\ \text{subject to} & I \succeq V \operatorname{diag}(x) V^\dagger & \\ & x_i \geq 0 & \forall i \end{array} \tag{4}$$

If the objective of the above program is OPT, then $\operatorname{rel}(A) = e^{\mathrm{OPT}}$. Note that $-\ln(x_1 \ldots x_n)$ is convex over $\mathbb{R}^n_{\geq 0}$, so the above is a valid convex program.

Proof of theorem 4: We can detect whether $\operatorname{rel}(A) = 0$ by checking whether any of A's main diagonal entries are 0. See the proof of claim 2.

When all of the main diagonal entries of A are strictly positive, similar to the proof of claim 2, we can determine upper and lower bounds on the optimum x_i. In particular if λ is a number larger than the largest eigenvalue of A, for the optimum x_i we have

$$A_{ii}^{-1} \geq x_i \geq \frac{\prod_{j=1}^n A_{jj}}{\lambda^n A_{ii}}.$$

Thus, we can restrict the domain of the convex program in eq. (4) to a compact bounded domain. We can compute the Cholesky decomposition of A and then use our favorite convex programming technique, such as the ellipsoid method, to find the optimum value of eq. (4) to within accuracy ϵ in time $\operatorname{poly}(n + \langle A \rangle + \log(1/\epsilon))$. This gives us a $1 + \epsilon$ approximation of $\operatorname{rel}(A)$ which by eq. (2) is a $(1 + \epsilon) c^n$ approximation of $\operatorname{per}(A)$ for $c = e^{\gamma+1}$.

As a final remark, we note that the approximation factor $e^{n(\gamma+1)}$ in eq. (2) can in fact be slightly strengthened to

$$\frac{n^n}{n!} e^{n\gamma},$$

if one carefully reviews the proof. The term $n^n/n!$ is at most e^n, but the difference allows us to absorb $1 + \epsilon$ into the approximation factor for an appropriately chosen ϵ. This allows us to state an ϵ-free result: We can find an $e^{n(\gamma+1)}$ approximation to $\operatorname{per}(A)$ in time $\operatorname{poly}(n + \langle A \rangle)$. ■

IV. Asymptotically Tight Examples

In this section we show that the constant $c = e^{\gamma+1}$ cannot be replaced by anything smaller in eq. (2). In other words we will construct $n \times n$ hermitian PSD matrices A such that

$$\sqrt[n]{\frac{\operatorname{rel}(A)}{\operatorname{per}(A)}} \to e^{\gamma+1}.$$

The construction will begin with a distribution $\mathcal{U}$ that is uniform over n unit vectors $u_1, \ldots, u_n \in \mathbb{C}^d$. We will later show how we can construct $\mathcal{U}$ so that $f(\mathcal{U})$ is arbitrarily close to $e^{-\gamma}$.

Lemma 3. *For any $\epsilon > 0$ there exists a distribution $\mathcal{U}$ that is uniform over n unit vectors $u_1, \ldots, u_n \in \mathbb{C}^d$ for some n and d that satisfies*

$$f(\mathcal{U}) \leq e^{-\gamma} + \epsilon.$$

We postpone the proof of lemma 3 to the end of this section. For now we use it to show the following. The following proposition together with lemma 3 show that $e^{\gamma+1}$ cannot be improved in eq. (2).

Proposition 2. *Given a distribution $\mathcal{U}$ that is uniform over a finite number of unit vectors $u_1, \dots, u_n$, we can construct a sequence of matrices $A_1, A_2, \dots$ of sizes $n_1 \times n_1, n_2 \times n_2, \dots$ such that*

$$\sqrt[n_k]{\frac{\mathrm{rel}(A_k)}{\mathrm{per}(A_k)}} \to e f(\mathcal{U})^{-1}.$$

Proof: Our goal is to construct a PSD matrix A and relate $\mathrm{rel}(A)/\mathrm{per}(A)$ to $f(\mathcal{U})$. We will assume without loss of generality that $\mathrm{span}\{u_1, \dots, u_n\} = \mathbb{C}^d$; otherwise, we use a unitary transformation to map $u_1, \dots, u_n$ onto a lower dimensional space and $f(\mathcal{U})$ would not change.

Consider the matrix $U \in \mathbb{C}^{d\times n}$ whose columns are $u_1, \dots, u_n$. Note that $\mathrm{rank}(U) = d$ and $U^\dagger U \succeq 0$ has 1s on the main diaognal. In other words $U^\dagger U$ is a correlation matrix of rank d. Since $\mathrm{rank}(U) = d$, the matrix $UU^\dagger$ is invertible and we can define

$$V := (UU^\dagger)^{-1/2} U,$$

and

$$A := V^\dagger V = U^\dagger (UU^\dagger)^{-1} U.$$

We will study $\mathrm{rel}(A)$ and $\mathrm{per}(A)$ and relate them to $f(\mathcal{U})$.

As observed in the proof of claim 4, correlation matrices can be used as optimality certificates for rel, albeit in that context first order optimality was just a necessary condition. We now make a formal claim by certifying that $\mathrm{rel}(A) = 1$ using $U^\dagger U$ as the certificate.

Claim 7. *If A is constructed as above, then*

$$\mathrm{rel}(A) = \mathrm{rel}(V^\dagger V) = 1.$$

Proof: We clearly have $I \succeq U^\dagger (UU^\dagger)^{-1} U = V^\dagger V$. This implies that $\mathrm{rel}(A) \le 1$. Now consider a diagonal matrix $D \succeq A = V^\dagger V$. We need to show that $\mathrm{per}(D) \ge 1$. Without loss of generality, by adding a small multiple of I if necessary, we may assume that $D \succ 0$. Now $D \succeq V^\dagger V$ implies that

$$I \succeq V D^{-1} V^\dagger,$$

which in turn implies

$$\begin{aligned} UU^\dagger &= (UU^\dagger)^{1/2} (UU^\dagger)^{1/2} \\ &\succeq (UU^\dagger)^{1/2} V D^{-1} V^\dagger (UU^\dagger)^{1/2} = U D^{-1} U^\dagger. \end{aligned}$$

By taking the trace we get

$$\mathrm{tr}(U^\dagger U) = \mathrm{tr}(UU^\dagger) \ge \mathrm{tr}(U D^{-1} U^\dagger) = \mathrm{tr}(D^{-1} U^\dagger U).$$

Since $U^\dagger U$ has 1s on the diagonal and D is diagonal the above becomes

$$n \ge \sum_{i=1}^n D_{ii}^{-1}.$$

By using the AM-GM inequality we get

$$(D_{11}^{-1} \dots D_{nn}^{-1})^{1/n} \le \frac{\sum_{i=1}^n D_{ii}^{-1}}{n} \le 1.$$

This means that $\mathrm{per}(D) = D_{11} \dots D_{nn} \ge 1$. ■

Next we study $\mathrm{per}(A)$. This is where the term $f(\mathcal{U})$ appears.

Claim 8. *If A is constructed as above, then*

$$\mathrm{per}(A) \le \frac{n!}{n^n} \binom{n+d-1}{d-1} \cdot f(\mathcal{U})^n.$$

Before proving claim 8, let us show why it suffices to finish the proof of proposition 2. By claim 7 and claim 8 we have

$$\sqrt[n]{\frac{\mathrm{rel}(A)}{\mathrm{per}(A)}} \ge \sqrt[n]{\frac{n^n}{n!}} \cdot \sqrt[n]{\binom{n+d-1}{d-1}^{-1}} \cdot f(\mathcal{U})^{-1}.$$

This is not quite the same as $e f(\mathcal{U})^{-1}$ yet. However we have one degree of freedom we have not used. Initially we assumed $\mathcal{U}$ was a uniform distribution over n unit vectors. But we might have as well assumed that it is a uniform distribution over nk unit vectors for any integer k, by simply repeating the vectors in the support of $\mathcal{U}$. Therefore we may make n as large as we would like without changing d or $f(\mathcal{U})$. As $n \to \infty$, by Stirling's formula we have

$$\sqrt[n]{\frac{n^n}{n!}} \to e,$$

and by a simple bound for large enough n

$$\sqrt[n]{\binom{n+d-1}{d-1}^{-1}} \ge \sqrt[n]{n^{-d}} \to 1.$$

Therefore as $n \to \infty$ we have

$$\sqrt[n]{\frac{\mathrm{rel}(A)}{\mathrm{per}(A)}} \to e f(\mathcal{U})^{-1}.$$

It only remains to prove claim 8.

Proof of claim 8: We will use lemma 2 to write down $\mathrm{per}(A) = \mathrm{per}(V^\dagger V)$. Let $x \in \mathbb{C}^d$ be distributed according to a d-dimensional standard complex normal $\mathbb{C}\mathcal{N}(0, I)$. Then according to lemma 2 we have

$$\mathrm{per}(A) = \mathbb{E}_{x\sim\mathbb{C}\mathcal{N}(0,I)}[|V^\dagger x|_\Pi^2].$$

Our goal is to use $f(\mathcal{U})$ to bound $|V^\dagger x|_\Pi$. According to the definition of $f(\mathcal{U})$, for the vector $y = (UU^\dagger)^{-1/2} x$ we have

$$\frac{\sqrt[n]{\prod_{i=1}^n |u_i^\dagger y|^2}}{\frac{1}{n}\sum_{i=1}^n |u_i^\dagger y|^2} = \frac{\exp(\mathbb{E}_{u\sim\mathcal{U}}[\ln(|u^\dagger y|^2)]}{\mathbb{E}_{u\sim\mathcal{U}}[|u^\dagger y|^2]} \le f(\mathcal{U}).$$

Note that

$$u_i^\dagger y = (U^\dagger y)_i = (U^\dagger (UU^\dagger)^{-1/2} x)_i = (V^\dagger x)_i.$$

This means that $\prod_{i=1}^n |u_i^\dagger y|^2 = |V^\dagger x|_\Pi^2$. We also have

$$\sum_{i=1}^n |u_i^\dagger y|^2 = x^\dagger V V^\dagger x = x^\dagger (UU^\dagger)^{-1/2} UU^\dagger (UU^\dagger)^{-1/2} x$$
$$= x^\dagger x = |x|^2.$$

Putting these together we get

$$|V^\dagger x|_\Pi^2 \le \left(\frac{f(\mathcal{U})|x|^2}{n}\right)^n = \left(\frac{f(\mathcal{U})}{n}\right)^n |x|^{2n}.$$

Let us now compute $\mathbb{E}_x[|x|^{2n}]$. We have

$$\mathbb{E}_x[|x|^{2n}] = \mathbb{E}_x[\prod_{j=1}^n (\sum_{i=1}^d |x_i|^2)]$$
$$= \sum_{\substack{k_1,\dots,k_d \ge 0 \\ k_1+\dots+k_d=n}} \binom{n}{k_1,\dots,k_d} \mathbb{E}_x[|x_1|^{2k_1} \dots |x_d|^{2k_d}].$$

According to fact 2, we have $\mathbb{E}_x[|x_1|^{2k_1} \dots |x_d|^{2k_d}] = k_1! \dots k_d!$. Therefore

$$\mathbb{E}_x[|x|^{2n}] = \sum_{\substack{k_1,\dots,k_d \ge 0 \\ k_1+\dots+k_d=n}} \binom{n}{k_1,\dots,k_d} k_1! \dots k_d!$$
$$= \sum_{\substack{k_1,\dots,k_d \ge 0 \\ k_1+\dots+k_d=n}} n! = n! \binom{n+d-1}{d-1},$$

where in the last equality we used the fact the number of ways to write n as a sum of d nonnegative integers is $\binom{n+d-1}{d-1}$. We conclude by getting

$$\mathrm{per}(A) \le \left(\frac{f(\mathcal{U})}{n}\right)^n \mathbb{E}_x[|x|^{2n}] = \frac{n!}{n^n} \binom{n+d-1}{d-1} f(\mathcal{U})^n.$$

■

This finishes the proof of proposition 2. ■

Now we switch gears and construct the distribution $\mathcal{U}$ promised by lemma 3.

Proof of lemma 3: The idea is to make $\mathcal{U}$ be close to the uniform distribution on the sphere $\{u \in \mathbb{C}^d : |u| = 1\}$ for some large d. If we were allowed to pick $\mathcal{U}$ to be uniform over the sphere, then intuitively all choices of x in the definition of $f(\mathcal{U})$ would yield the same value and we would be able to argue about this common value using the same tricks as in the proof of proposition 1. Instead we use the uniform distribution on a large number of samples from the sphere to serve as the proxy for the uniform distribution on the sphere itself. We further need the dimension d to grow, to make the uniform distribution on the sphere similar to a (scaled) normal distribution. We now make these formal.

Let us fix some d and let $\mathcal{S}$ denote the uniform distribution on the sphere $\{u \in \mathbb{C}^d : |u| = 1\}$. For any fixed distance ϵ we can cover the sphere by a finite number of balls $B(o_1, \epsilon), \dots, B(o_m, \epsilon)$ where $o_1, \dots, o_m$ are unit vectors and

$$B(o, \epsilon) = \{v \in \mathbb{C}^d : |o - v| \le \epsilon\}.$$

Let n be a large number and draw n random points $u_1, \dots, u_n$ from $\mathcal{S}$. We will let $\mathcal{U}$ be the uniform distribution over $u_1, \dots, u_n$. We would like to argue that $f(\mathcal{U})$ is with high probability close to $f(\mathcal{S})$. Because the sphere was covered by the balls around o_i's, for each unit vector x we have $|x - o_i| \le \epsilon$ for some i. This implies that

$$\mathbb{E}_{u\sim\mathcal{U}}[\ln(|u^\dagger x|^2)] \le \mathbb{E}_{u\sim\mathcal{U}}[\ln((|u^\dagger o_i| + \epsilon)^2)],$$
$$\mathbb{E}_{u\sim\mathcal{U}}[|u^\dagger x|^2] \ge \mathbb{E}_{u\sim\mathcal{U}}[\max(0, |u^\dagger o_i| - \epsilon)^2].$$

On the other hand by the law of large numbers for each o_i we have with high probability as $n \to \infty$

$$\mathbb{E}_{u\sim\mathcal{U}}[\ln((|u^\dagger o_i| + \epsilon)^2)] \to \mathbb{E}_{u\sim\mathcal{S}}[\ln((|u^\dagger o_i| + \epsilon)^2)],$$
$$\mathbb{E}_{u\sim\mathcal{U}}[\max(0, |u^\dagger o_i| - \epsilon)^2] \to \mathbb{E}_{u\sim\mathcal{S}}[\max(0, |u^\dagger o_i| - \epsilon)^2].$$

Let us condition on the event that the LHS of the above are sufficiently close to the RHS for all o_i. This event happens with high probability as $n \to \infty$. Note that because of symmetry, the RHS of the above are independent of the choice of o_i. Under this condition we have for all unit vectors x

$$\frac{\exp(\mathbb{E}_{u\sim\mathcal{U}}[\ln(|u^\dagger x|^2)])}{\mathbb{E}_{u\sim\mathcal{U}}[|u^\dagger x|^2]} \le \frac{\exp(\mathbb{E}_{u\sim\mathcal{S}}[\ln((|u^\dagger o| + \epsilon)^2)])}{\mathbb{E}_{u\sim\mathcal{S}}[\max(0, |u^\dagger o| - \epsilon)^2]} + \delta,$$

where o is any arbitrary vector and $\delta \to 0$ as $n \to \infty$. The above bounds the LHS for unit vectors x. However note that the LHS does not change if we scale x by any constant. Therefore $f(\mathcal{U})$ is bounded by the RHS. As we take the limit with $\epsilon \to 0$ and $\delta \to 0$ we get $\mathcal{U}$ with $f(\mathcal{U})$ asymptotically bounded by $f(\mathcal{S})$.

Now it only remains to show that as the dimension d grows $f(\mathcal{S}) \to e^{-\gamma}$. Let o be an arbitrary point with $|o|^2 = d$ such as $\sqrt{d}e_1$ where e_1 is the first element of the standard basis. When $u \sim \mathcal{S}$ is a random point on the sphere, we would like to argue that $u^\dagger o$ is almost distributed like $\mathbb{C}\mathcal{N}(0,1)$. If this were the case we would have

$$f(S) = \frac{\exp(\mathbb{E}_u[\ln(|u^\dagger o|^2)])}{\mathbb{E}_u[|u^\dagger o|^2]}$$
$$\simeq \frac{\exp(\mathbb{E}_{g\sim\mathbb{C}\mathcal{N}(0,1)}[\ln(|g|^2)])}{\mathbb{E}_{g\sim\mathbb{C}\mathcal{N}(0,1)}[|g|^2]} = e^{-\gamma},$$

where in the last equality we used fact 3.

To make this approximation rigorous, let us generate the random point u on the sphere by the following process: We sample a standard d-dimensional complex normal $v \sim \mathbb{C}\mathcal{N}(0, I)$ and then we let $u = v/|v|$. We have $u^\dagger o = v_1 \frac{d}{|v|}$. Therefore

$$\mathbb{E}_u[\ln(|u^\dagger o|^2)] = \mathbb{E}_v[\ln(|v_1|^2)] + 2\ln(d) - 2\mathbb{E}_v[\ln(|v|^2)].$$

The random variable $|v|^2$ is distributed according to a $\frac{1}{2}$-scaled χ^2-distribution with $2d$ degrees of freedom which is the same as $\Gamma(d, 1)$. We can therefore write

$$\mathbb{E}_v[\ln(|v|^2)] = \psi(d) = \ln(d-1) + o(1),$$

where ψ is the digamma function [16], [17]. We therefore have $\mathbb{E}_u[\ln(|u^\dagger o|^2)] = -\gamma + o(1)$.

For $\mathbb{E}_u[|u^\dagger o|^2]$ we observe that

$$\mathbb{E}_u[|u^\dagger o|^2] = d \cdot \mathbb{E}_v \left[\frac{|v_1|^2}{|v|^2} \right].$$

The random variables $|v_i|^2/|v|^2$ are identically distributed for different i. As such we have

$$\begin{aligned}\mathbb{E}_u[|u^\dagger o|^2] &= d \cdot \mathbb{E}_v \left[\frac{|v_1|^2}{|v|^2} \right] \\ &= \mathbb{E}_v \left[\frac{|v_1|^2}{|v|^2} \right] + \cdots + \mathbb{E}_v \left[\frac{|v_d|^2}{|v|^2} \right] \\ &= \mathbb{E}_v \left[\frac{|v|^2}{|v|^2} \right] = 1.\end{aligned}$$

Therefore

$$\frac{\exp(\mathbb{E}_u[\ln(|u^\dagger o|^2)])}{\mathbb{E}_u[|u^\dagger o|^2]} = e^{-\gamma + o(1)}.$$

This shows that $f(\mathcal{S}) \to e^{-\gamma}$ as $d \to \infty$ and concludes the proof. ■

V. Approximate Permanent-On-Top

In this section we first prove theorem 2. The proof follows simply from the fact that the Schur power of a diagonal matrix is a multiple of the identity matrix.

Proof of theorem 2: Let $A \succeq 0$ be given and let $D \succeq A$ be a diagonal matrix such that $\mathrm{per}(D) = \mathrm{rel}(A)$. Because $D \succeq A \succeq 0$, we have

$$D^{\otimes n} \succeq A^{\otimes n}.$$

The Schur powers of A and D are submatrices of $D^{\otimes n}$ and $A^{\otimes n}$ respectively. Therefore we have

$$D^{\otimes n}_{S_n, S_n} \succeq A^{\otimes n}_{S_n, S_n}.$$

The matrix $D^{\otimes n}$ is diagonal, and so is $D^{\otimes n}_{S_n, S_n}$. The entry on the diagonal corresponding to row and column $\sigma \in S_n$ is by definition

$$\prod_{i=1}^{n} D_{\sigma(i), \sigma(i)} = \prod_{i=1}^{n} D_{ii} = \mathrm{per}(D),$$

which is independent of σ. This shows that $D^{\otimes n}_{S_n, S_n} = \mathrm{per}(D) I$. Therefore $A^{\otimes n}_{S_n, S_n} \preceq \mathrm{per}(D) I$ which means that

$$\|A^{\otimes n}_{S_n, S_n}\| \leq \mathrm{per}(D) = \mathrm{rel}(A) \leq c^n \, \mathrm{per}(A).$$

■

Now we prove claim 1. This shows that in the worst case there is at least an exponential gap between the largest eigenvalue of the Schur power matrix and the permanent.

Proof of claim 1: We start with the counterexample given in [15] of an $m \times m$ matrix B whose permanent is not equal to the largest eigenvalue of its Schur power matrix. We use standard tricks in order to construct larger counterexamples, by building block diagonal matrices. We let $A = B \otimes I_k$ where I_k is the $k \times k$ identity matrix for some k and let $n = mk$. It is clear from the block diagonal form of A that

$$\mathrm{per}(A) = \mathrm{per}(B)^k.$$

It is also not hard to see that the Schur power of A can be expressed in terms of the Schur power of B via the following identity, up to a reindexing of the rows and columns:

$$A^{\otimes n}_{S_n, S_n} = (B^{\otimes m}_{S_m, S_m})^{\otimes k} \otimes I_{\binom{mk}{m, m, \ldots, m}}.$$

It follows that

$$\|A^{\otimes n}_{S_n, S_n}\| = \|B^{\otimes m}_{S_m, S_m}\|^k \cdot 1^{\binom{n}{m, \ldots, m}} = \|B^{\otimes m}_{S_m, S_m}\|^k$$

Now we have

$$\frac{\|A^{\otimes n}_{S_n, S_n}\|}{\mathrm{per}(A)} = \left(\frac{\|B^{\otimes m}_{S_m, S_m}\|}{\mathrm{per}(B)} \right)^k = \hat{c}^n,$$

where we let

$$\hat{c} = \sqrt[m]{\frac{\|B^{\otimes m}_{S_m, S_m}\|}{\mathrm{per}(B)}} > 1$$

be a universal constant. ■

We conclude with the following open question.

Problem 1. *What is the smallest constant c_* such that for every $A \succeq 0$ we have*

$$\|A^{\otimes n}_{S_n, S_n}\| \leq c_*^n \, \mathrm{per}(A)?$$

References

[1] R. Bapat, "Recent developments and open problems in the theory of permanents," *The Mathematics student*, vol. 76, no. 1, p. 55, 2007.

[2] L. G. Valiant, "The complexity of computing the permanent," *Theoretical computer science*, vol. 8, no. 2, pp. 189–201, 1979.

[3] S. Aaronson, "A linear-optical proof that the permanent is# p-hard," in *Proc. R. Soc. A*, vol. 467, no. 2136. The Royal Society, 2011, pp. 3393–3405.

[4] D. Grier and L. Schaeffer, "New hardness results for the permanent using linear optics," electronic Colloquium on Computational Complexity (ECCC).

[5] L. Gurvits, "On the complexity of mixed discriminants and related problems," in *Mathematical Foundations of Computer Science*, 2005, pp. 447–458.

[6] L. Chakhmakhchyan, N. Cerf, and R. Garcia-Patron, "A quantum-inspired algorithm for estimating the permanent of positive semidefinite matrices," *arXiv preprint arXiv:1609.02416*, 2016.

[7] M. Jerrum, A. Sinclair, and E. Vigoda, "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries," *Journal of the ACM (JACM)*, vol. 51, no. 4, pp. 671–697, 2004.

[8] L. Gurvits and A. Samorodnitsky, “Bounds on the permanent and some applications,” in *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*. IEEE, 2014, pp. 90–99.

[9] S. Aaronson, see point (4) of Comment #84 at http://www.scottaaronson.com/blog/?p=2408.

[10] M. Marcus, “The permanent analogue of the hadamard determinant theorem,” *Bull. Amer. Math. Soc.*, vol. 69, no. 4, pp. 494–496, 07 1963. [Online]. Available: http://projecteuclid.org/euclid.bams/1183525360

[11] E. H. Lieb, “Proofs of some conjectures on permanents,” *Journal of Mathematics and Mechanics*, vol. 16, no. 2, pp. 127–134, 1966.

[12] G. W. Soules, “Matrix functions and the laplace expansion theorem,” Ph.D. dissertation, University of California, Santa Barbara, 1966.

[13] F. Zhang, “An update on a few permanent conjectures,” *Special Matrices*, vol. 4, no. 1, 2016.

[14] I. Schur, “über endliche gruppen und hermitesche formen,” *Mathematische Zeitschrift*, vol. 1, no. 2-3, pp. 184–207, 1918.

[15] V. S. Shchesnovich, “The permanent-on-top conjecture is false,” *Linear Algebra and its Applications*, vol. 490, pp. 196–201, 2016.

[16] S. P. Chan, “A statistical study of log-gamma distribution,” Ph.D. dissertation, 1993.

[17] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Corporation, 1964, vol. 55.

[18] L. Gurvits, “Classical deterministic complexity of edmonds’ problem and quantum entanglement,” in *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. ACM, 2003, pp. 10–19.