



Inapproximability of Positive Semidefinite Permanents and Quantum State Tomography

Alexander Meiburg¹

Received: 4 March 2023 / Accepted: 17 August 2023 / Published online: 3 September 2023
© The Author(s) 2023

Abstract

Matrix permanents are hard to compute or even estimate in general. It had been previously suggested that the permanents of Positive Semidefinite (PSD) matrices may have efficient approximations. By relating PSD permanents to a task in quantum state tomography, we show that PSD permanents are NP-hard to approximate within a constant factor, and so admit no polynomial-time approximation scheme (unless $P = NP$). We also establish that several natural tasks in quantum state tomography, even approximately, are NP-hard in the dimension of the Hilbert space. These state tomography tasks therefore remain hard even with only logarithmically few qubits.

Keywords Matrix permanent · Hermitian matrix · NP-hard · Quantum state tomography · Positive semidefinite

1 Introduction

1.1 Background

The permanent is a classical problem of intense interest in the study of counting problems. For a matrix $A \in \mathbb{C}^{n \times n}$, the permanent is defined as

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

summing over all permutations σ of rows and columns, products of matrix elements $A_{i, j}$. While directly evaluating the expression in Eq. (1) takes $O(n!)$ time, Ryser's

This work was originally presented at FOCS 2022, the 63rd IEEE Symposium on Foundations of Computer Science.

✉ Alexander Meiburg
alex@ohai.the.re

¹ Department of Physics, University of California, Santa Barbara, Santa Barbara, CA, USA

formula [1] gives an $O(2^n n)$ time algorithm. Valiant showed in 1979 that computing the permanent exactly is #P-hard, even for 0–1 matrices [2, 3]. However, it is amenable to efficient approximation in particular settings. In 2001, Jerrum, Sinclair and Vigoda [4] gave a fully-polynomial randomized approximation scheme (FPRAS) for permanents of nonnegative matrices. In 2002, Gurvits and Samorodnitsky [5] gave a polynomial time e^n multiplicative approximation to PSD mixed discriminants, which included permanents of nonnegative matrices as a special case.

When the matrix is Hermitian positive semidefinite (HPSD, or if purely real, PSD), the permanent is necessarily nonnegative, and this offers hope of efficient multiplicative approximation. HPSD permanents are of particular interest to the quantum information community - for reasons unrelated to quantum state tomography, but rather related to thermal BosonSampling experiments [6–8]. Computing PSD permanents exactly remains #P-hard [9]. It is known that by Stockmeyer counting [7, 9, 10] computing multiplicative approximations to PSD permanents is contained in FBPP^{NP}. In 1963, Marcus [11] observed that the product of the diagonal of a PSD matrix immediately gives an $n!$ approximation ratio to the permanent. In 2017, Anari et al. gave a polytime approximation to PSD permanents within a ratio of c^n with $c = e^{1+\gamma} \approx 4.85$ [12]. Yuan and Parillo [13] described a similar approach with the same approximation ratio. Chakhmakhchyan et al. [14] and Barvinok [15] gave algorithms for approximation when the spectrum of the matrix is small in radius, that is, when $\lambda_{\min}/\lambda_{\max}$ is not too small.

1.2 Main Results

Our main result is to show that there is no efficient approximation of PSD permanents. Precisely, we show that it is NP-hard to approximate within a particular subexponential factor.

Theorem 1 (Theorem 5, restated) *For any constant $\epsilon > 0$, it is NP-hard to approximate the permanent of $n \times n$ HPSD matrices within a factor of $2^{n^{1-\epsilon}}$.*

This implies the absence of a polynomial time approximation scheme (PTAS) or polynomial randomized approximation scheme (PRAS).

Corollary 1 (of Theorem 5) *There is no PTAS for HPSD permanents unless P=NP, and there is no PRAS for HPSD permanents unless RP=NP.*

In Sect. 3.5, we show that these theorems also hold for (purely real) PSD matrices.

Our work provides a lower bound on the difficulty of approximating PSD permanents, that almost matches known upper bounds. The algorithm of Anari et al. [12] shows that the singly exponential approximation ratio 4.85^n is possible within polynomial time, while we show that a subexponential approximation ratio $2^{n^{1-\epsilon}}$ is intractable. This primarily leaves the question whether $(1 + \epsilon)^n$ is polynomial-time computable for any $\epsilon > 0$. The algorithms of Chakhmakhchyan et al. [14] and Barvinok [15] fail on the hard instances that we construct: the matrices we construct are highly rank deficient, and therefore have $\lambda_{\min} = 0$.

Our most key connection is that between a permanent and an particular integral over the unit sphere. If a matrix M is Hermitian positive semidefinite, then it has a

matrix square root $V = V^\dagger$ satisfying $VV^\dagger = V^2 = M$. If M is $n \times n$, and rank d , then V is $n \times d$. We will show (Theorem 3) that

$$\text{Perm}(M) = \frac{(d + n - 1)!}{2\pi^n} \int_{\vec{x} \in \mathbb{C}^n, |\vec{x}|=1} d\vec{x} \prod_{k=1}^d |\vec{x} \cdot V_k|^2. \tag{2}$$

Here V_i are the rows of V , and the integral is the Haar measure over the unit complex sphere. As we will see, this integral occurs naturally in the context of Bayesian inference for quantum state tomography. In that context, the rows V_i correspond to an observation history, and the variable of integration \vec{x} represents an unknown quantum state. This intuition gained from viewing it as a quantum state tomography problem guided us towards finding our hard instances M . We analyze the problem by first establishing a concentrating construction (Lemmas 2 and 3). Informally, when V_i contains many copies of basis vectors \vec{e}_j and vectors of the form $\frac{\vec{e}_j \pm i\vec{e}_k}{\sqrt{2}}$, the integral concentrates at the points (up to a phase) of an appropriately scaled hypercube:

$$\int_{\vec{x} \in \mathbb{C}^n, |\vec{x}|=1} d\vec{x} \prod_{k=1}^d |\vec{x} \cdot V_k|^2 \propto \sum_{\vec{x} \in \{-1, +1\}^d} \prod_{k=1}^d |\vec{x} \cdot V_k|^2 \tag{3}$$

with some exponentially small error, and a simple constant of proportionality that depends only on d and n . This concentration will let us relate permanents to combinatorial problems (Lemma 4), specifically counting solutions to Not-All-Equal-3SAT, and ultimately let us prove hardness.

The connection to quantum state tomography means we also get results about the hardness of estimating quantum states given measurements.

Definition 1 (Maximum Pure State Likelihood) For a quantum system with Hilbert space dimension n and $\text{poly}(n)$ observations, the *maximum pure state likelihood* is the highest likelihood of those observations attainable over any pure state $|\psi\rangle$.

Theorem 2 (Theorem 9, informal) *For any constant $\epsilon > 0$, the following task is NP-hard: given a series of quantum observations, find a pure state with likelihood at least $2^{-n^{1-\epsilon}}$ times the maximum pure state likelihood.*

Unless $\text{RP}=\text{NP}$, this implies that there is no PRAS for maximum likelihood estimation (MLE) quantum state tomography; in fact, it is not even in APX. We have similar statements about the NP-hardness of computing the Bayesian average state and Bayesian average observables (Theorem 8). These results are unusual in that they imply exponential difficulty in dimension n in the Hilbert space \mathbb{C}^n . Most quantum problems are only considered tractable if they have efficient algorithms in the number of particles $q = \log(n)$, and have trivially polynomial solutions in n ; whereas we show that (assuming ETH [16]) quantum state tomography takes time exponential in n .

We stress that although our work has connections to quantum information through BosonSampling and tomography, our discussion of complexity is focused on classical computers. The NP-hardness are statements about classical hardness, and the algorithm

described in Sect. 4.3 for tomography in fixed dimension is a polynomial time *classical* algorithm. Unless $NP \subseteq BQP$ however, our results rule efficient permanent computations on quantum computers as well.

2 Key Ideas of the Proof

We start with a lemma relating symmetric, multilinear functions to permanents. Similar lemmas have appeared in [15, 17], and they can broadly be viewed as alternate forms of Wick’s Theorem [18].

Lemma 1 *Suppose $f : (\mathbb{C}^d)^{2n} \rightarrow \mathbb{R}$ is a function of $2n$ vectors of dimension d , that is:*

- *Multilinear in its first n arguments*
- *Conjugate multilinear in its latter n arguments*
- *Symmetric in its first n arguments, and its latter n arguments*
- *Invariant under unitary change of basis: for any unitary $U \in \mathbb{C}^{d \times d}$,*

$$f(v_1, \dots; v_n, \dots) = f(Uv_1, \dots; Uv_n, \dots)$$

Then f is determined up to an overall constant C by the formula,

$$f(v_1, \dots; v_n, \dots) = C \text{Perm}(A_{ij}), \quad \text{where } A_{ij} = v_i \cdot v_j^* \tag{4}$$

and the constant C can be determined by

$$C = \frac{f(\vec{e}_1, \vec{e}_1, \vec{e}_1, \dots)}{n!} \tag{5}$$

where \vec{e}_1 is the unit basis vector in the first coordinate.

Proof Because f is invariant under a unitary change of basis, f can only depend on its inputs through inner products of vectors, $\langle v_i, v_j \rangle$. Since f is multilinear, it can be written as a sum of terms t_k , where each t_k is a product of terms from the vectors. The separate linearity and conjugate linearity means that the only permitted inner products are of covariant (first n) and contravariant (latter n) vectors. This means every term in the sum must be some product of the form $\prod_{i \in [n]} v_i \cdot v_{n+\sigma(i)}^*$ for some permutation σ of $[n]$. Then by symmetry of the arguments, all pairs must occur in the same relation to either, so all pairings must occur equally. This leaves only a single form, the result above.

Computing C can be found by substituting in \vec{e}_1 in Eq. (4) so that all dot products become 1. The permanent of the all-1’s matrix is just $n!$, so this becomes the normalizing factor. □

This lets us relate the permanent to a particular integral over unit-norm complex vectors:

Theorem 3 For any $L, R \in \mathbb{C}^{d \times n}$ be complex matrices, denoting the k th row of L as L_k and the k th row of R as R_k ,

$$\int_{\vec{x} \in \mathbb{C}^n, |\vec{x}|=1} d\vec{x} \left(\prod_{k=1}^d \vec{x}^\dagger L_k \right) \left(\prod_{k=1}^d R_k^\dagger \vec{x} \right) = \frac{2\pi^n \text{Perm}(LR^\dagger)}{(d+n-1)!} \tag{6}$$

Note that when $L = R$, the product in the integral becomes $\prod_k |\langle L_k, \vec{x} \rangle|^2$, and the product $M = LL^\dagger$ is PSD.

Proof Viewing the left side as a function f of the n rows of each L and R , we can see that it satisfies all the hypotheses of Lemma 1. It is linear in each row of L , conjugate linear in each row of R , and symmetric under permuting the rows of L or the rows of R . It is also invariant under a unitary change of basis:

$$f(UL, UR) = \int_{\vec{x} \in \mathbb{C}^n, |\vec{x}|=1} d\vec{x} \prod_{k=1}^d \vec{x}^\dagger (UL_k) \prod_{k=1}^d (UR_k)^\dagger \vec{x} \tag{7}$$

$$= \int_{\vec{x} \in \mathbb{C}^n, |\vec{x}|=1} d\vec{x} \prod_{k=1}^d (U^\dagger \vec{x})^\dagger L_k \prod_{k=1}^d R_k^\dagger (U^\dagger \vec{x}) \tag{8}$$

$$= \int_{\vec{u} \in \mathbb{C}^n, |\vec{u}|=1} d\vec{u} \prod_{k=1}^d \vec{u}^\dagger L_k \prod_{k=1}^d R_k^\dagger \vec{u} = f(L, R) \tag{9}$$

so that we’ve used the symmetry of the unit sphere in \mathbb{C}^n to remove the unitary via $\vec{u} = U\vec{x}$. Setting each $L_k = R_k = \vec{e}_1$, the spherical integral can be computed with standard formulae (e.g. [19]) to find the normalizing constant

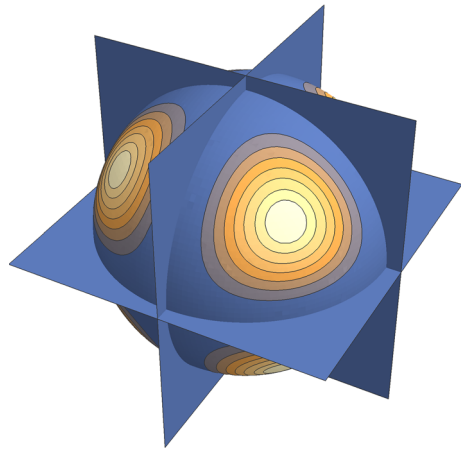
$$C = \frac{1}{n!} \int_{\vec{x} \in \mathbb{C}^n, |\vec{x}|=1} d\vec{x} (\vec{x}^\dagger \vec{e}_1)^n (\vec{e}_1^\dagger \vec{x})^n = \frac{2\pi^n}{(d+n-1)!}. \tag{10}$$

□

2.1 Outline of the Proof

Before diving into the proof of hardness itself, we aim to provide some intuition of the construction. We focus on the integral $F = \int_{\vec{x}} \prod_k |\langle V_k, \vec{x} \rangle|^2$ over the sphere of unit (complex) vectors, and build up a set of vectors V with desirable properties. The proof will involve gradually adding vectors to a list V_k , in turn modifying the integrand $G_V(\vec{x}) = \prod_k |\langle V_k, \vec{x} \rangle|^2$. This integrand $G_V(\vec{x})$ is nonnegative, so there cannot be any cancellation in the integral. Our goal will be only showing that certain regions have exponentially small magnitude, so that only particular regions with appreciable contribution remain, and they are primarily responsible for the overall value of F . Then, the magnitude of F will be used to understand the value of G_V on those particular regions, where large values of F indicate solutions to an NP-hard problem. And since

Fig. 1 Schematic of how we can create “corners” on the sphere by repeatedly cutting with planes. Blue represents lower magnitude. This shows only purely real \vec{x}



F can be computed by a HPSD permanent, computing that permanent must be hard as well.

How are we to choose the V in order to make an interesting function G_V ? Each vector V_k introduces zeroes on the sphere at all vectors orthogonal to V_k . All points approximately orthogonal to V_k will have a very small magnitude, and so contribute very little to the integral. We will start our collection of vectors by taking several copies of each standard basis vector \vec{e}_k . This creates high-degree zeros along each of d distinct perpendicular directions, slicing the sphere so that the only regions with appreciable magnitude form the corners of a cube.

After adding one copy of each basis vector \vec{e}_k , the magnitude at a given point $\vec{x} = (\alpha_1, \dots, \alpha_d)$ is the product of the absolute values of its entries in that basis: $G_V(\vec{x}) = \prod_j |\alpha_j|^2$. This is maximized when $|\alpha_j| = |\alpha_k| = \frac{1}{\sqrt{d}}$ for all j, k . If we then subsequently add several vectors of the form $\frac{\vec{e}_j + i\vec{e}_k}{\sqrt{2}}$ and $\frac{\vec{e}_j - i\vec{e}_k}{\sqrt{2}}$, together these rule out a purely imaginary phase between the j and k components, so that the maxima are at $\frac{\vec{e}_j \pm \vec{e}_k}{\sqrt{2}}$. After adding these two for each $j \neq k$, $G(\vec{x})$ will peak near $\vec{x} = \frac{e^{i\theta}}{\sqrt{d}}(1, \pm 1, \pm 1 \dots)$. Up to an overall phase of \vec{x} , we’ve focused G to a set of 2^{d-1} distinct points. These 2^{d-1} circles of “binarized” vectors will be our focus, and we call this set $B_0 = \{\frac{e^{i\theta}}{\sqrt{d}}(1, \pm 1, \pm 1 \dots)\}$. To force G_V to concentrate on B_0 , we had to add $d + 2_d C_2 = d^2$ vectors into our running list V_k . By analogy with quantum information, we will refer to these as the Z vectors and Y vectors respectively. Together, this set of d^2 vectors will form one “basic set” – “basic” in the set of “enforcing the basis”.

This is visualized in Fig. 1. This plots $G(\vec{x})$, where $\vec{x} = (x, y, z)$. The unit vectors form a sphere. After adding one basic set, we add a zero plane at $x = 0, y = 0, z = 0$; these planes are drawn as squares intersecting the sphere. The orange/white circles show the points that maximize $G(\vec{x})$: $\frac{1}{\sqrt{3}}(\pm 1, \pm 1, \pm 1)$. Taking many copies of the basic set is equivalent to raising $G(\vec{x})$ to a high power, and the function will rapid fall off anywhere off the eight white hotspots, the binarized points that we concentrate to.

Once we have our basic vectors to concentrate G at these binarized points B_0 , we want to add vectors that will penalize some of these 2^{d-1} points, so that finding the optimum becomes a search problem over exponentially many points. Our functional G is only sensitive to the relative phase between components of a vector, and not to the signs of the components themselves. This is the same symmetry that appears in quantum mechanics, where multiplying a quantum state $|\psi\rangle$ by a phase yields as a physically identical state $e^{i\theta}|\psi\rangle$. Such a factor θ is a global phase, that we can and will neglect later. When we restrict to the binarized points B_0 , these phases are just sign differences, and the phase symmetry says that we don't care about any individual sign: only the pattern of relative signs in the vector.

This sign-flipping symmetry leads us most naturally to the problem of Not-All-Equal 3-Satisfiability, or NAE3SAT [20]:

Definition 2 (NAE3SAT) Given n boolean variables and a set of *clauses*, each of which are triple of variables (v_1, v_2, v_3) , is there an assignment such that each clause contains at least one true variable and at least one false variable?

NAE3SAT is known to be NP-complete. It also has the same global symmetry where all variables in an assignment can be negated, and the satisfaction of the clauses remains unchanged; this reflects our global phase symmetry. So now consider the impact of adding a triple of “clause vectors”,

$$\vec{v}_1 = (\sqrt{6})^{-1}(-2\vec{e}_1 + \vec{e}_2 + \vec{e}_3) \quad (11)$$

$$\vec{v}_2 = (\sqrt{6})^{-1}(\vec{e}_1 - 2\vec{e}_2 + \vec{e}_3) \quad (12)$$

$$\vec{v}_3 = (\sqrt{6})^{-1}(\vec{e}_1 + \vec{e}_2 - 2\vec{e}_3). \quad (13)$$

Each is orthogonal to $\frac{1}{\sqrt{3}}(\vec{e}_1 + \vec{e}_2 + \vec{e}_3)$, in which all the relative signs are positive (or equivalently, all negative). We call this collection of three vectors a “clause set”. This effectively rules out the possibility of all signs being the same. There are three not-all-equal points (up to phase):

$$\vec{p}_1 = (\sqrt{3})^{-1}(-\vec{e}_1 + \vec{e}_2 + \vec{e}_3) \quad (14)$$

$$\vec{p}_2 = (\sqrt{3})^{-1}(\vec{e}_1 - \vec{e}_2 + \vec{e}_3) \quad (15)$$

$$\vec{p}_3 = (\sqrt{3})^{-1}(\vec{e}_1 + \vec{e}_2 - \vec{e}_3) \quad (16)$$

while we forbid the the fourth point

$$\vec{q} = (\sqrt{3})^{-1}(\vec{e}_1 + \vec{e}_2 + \vec{e}_3). \quad (17)$$

We see that $|\vec{p}_i \cdot \vec{v}_j|^2 = \frac{2+6\delta_{i,j}}{9}$. So, when V consists of just the three \vec{v}_i , then $G_V(\vec{p}_i) = \frac{32}{729}$, while $G_V(\vec{q}) = 0$.

This is visualized in Fig. 2. This shows $G(\vec{x})$ in the same coordinates as Fig. 1, from three different viewpoints. The basic set from Fig. 1 are still present, and form the axis-aligned cuts visible in the first plot. The clause set adds three more planes,

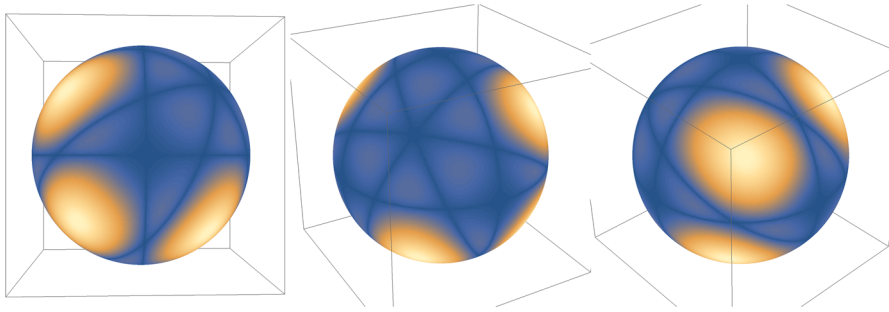


Fig. 2 Illustration of the effect of a clause. The eight corners $\{-1/\sqrt{3}, 1/\sqrt{3}\}^3$ represent four essentially distinct assignments, where opposite points are equivalent. The clause has three zero lines passing through one of the four points, while leaving the other three points untouched

and eliminate two of the eight orange spots. They form the six-way intersection in the second plot. That six-way intersection was one of the eight orange spots, but now $G(x)$ has been driven down to be zero there by the three incoming zero planes. The other six points are essentially unaffected.

By adding appropriate clause sets, the only remaining points with large values will be those satisfying an NAE3SAT problem, which is NP-hard. The other points will be too small to contribute to the integral, so that evaluating the integral tells us about the satisfiability of the NAE3SAT problem. With the outline complete, we now begin the steps of the proof, starting with the concentration.

3 Proof of Hardness

3.1 Concentration

After one basic set, each point in B_0 has a value $G_V(\vec{x})$ of $1/d^{d^2}$ (by direct calculation). We would like to show that any point far away from B_0 has a significantly lower value. For this reason, (and with the intuition that the integrand G_V represents likelihood values) we talk about *relative* values of G_V . By the value of $G_V(a)$ relative to $G_V(b)$, we simply mean $G_V(a)/G_V(b)$.

Any unit vector $\vec{x} \in \mathbb{C}^d$ can be written as

$$\vec{x} = \frac{e^{i\Theta}}{\sqrt{d}} \sum_{k=1}^d \sqrt{\alpha_k} e^{i\pi(\theta_k+n_k)} \vec{e}_k \tag{18}$$

where Θ , $\vec{\alpha}_k$ and $\vec{\theta}_k$ are all real, $\alpha_k \geq 0$, $\sum_k \alpha_k = d$, $\theta_1 = 0$, and all Θ , $\theta_k \in [-1/2, 1/2)$, and $n_k \in \{0, 1\}$. The $\vec{\alpha}$, $\vec{\theta}$, and \vec{n} respectively indicate the amplitudes, phases relative to the first component, and signs of the real part. This polar representation is unique except for when one of the $\alpha_0 = 0$, which is a measure-zero set. (We can neglect measure zero sets, as our concern is only with the integral $F(V) = \int G_V(x)$).

Lemma 2 Let \vec{x} be a unit vector with polar representation $\Theta, \vec{\alpha}, \vec{\theta},$ and \vec{n} . Let ϵ_α be the 2-norm distance of $\vec{\alpha} = (\alpha_1, \dots, \alpha_d)$ from $\vec{1}$. Then when V is one basic set, the value of $G_V(\vec{x})$ relative to any point in B_0 , is at most $1 - \frac{\epsilon_\alpha^2}{4d}$. If $\epsilon_\alpha \leq 1/2$, then the value is also at most $1 - 3\theta_i^2$ for all components θ_i of $\vec{\theta}$.

Proof The set B_0 consists of the points with $\alpha_k = 1$ and θ_k is an integer. If \vec{x} has significant distance from all elements of B_0 , then either the amplitudes α_k or phases θ_k must differ significantly from these conditions. The value after the basic set is

$$\begin{aligned} G_V(\vec{x}) &= \left(\prod_k \left| \sqrt{\frac{\alpha_k}{d}} \right|^2 \right) \left(\prod_{j \leq k} \left| \frac{\sqrt{\alpha_j} e^{i\pi(\theta_j+n_j)} + i\sqrt{\alpha_k} e^{i\pi(\theta_k+n_k)}}{\sqrt{2d}} \right|^2 \left| \frac{\sqrt{\alpha_j} e^{i\pi(\theta_j+n_j)} - i\sqrt{\alpha_k} e^{i\pi(\theta_k+n_k)}}{\sqrt{2d}} \right|^2 \right) \\ &= \left(\prod_k \frac{\alpha_k}{d} \right) \left(\prod_{j \leq k} \frac{\alpha_j^2 + \alpha_k^2 + 2\alpha_j\alpha_k \cos(2\pi(\theta_j - \theta_k + n_j - n_k))}{4d^2} \right) \\ &= \frac{1}{d^d (2d)^{d^2-d}} \left(\prod_k \alpha_k \right) \left(\prod_{j \leq k} \alpha_j^2 + \alpha_k^2 + 2\alpha_j\alpha_k \cos(2\pi(\theta_j - \theta_k)) \right) \end{aligned}$$

The first factor coming from the Z vectors \vec{e}_k in the basic set, and the last two factors coming from the Y vectors $\frac{\vec{e}_j \pm i\vec{e}_k}{\sqrt{2}}$, for each $j < k$, in the basic set.

The first step is to bound the value in terms of the magnitudes α_k . Looking at the effect of the Z vectors, $\prod_k \alpha_k$, we have a convex function on the standard $(d - 1)$ -simplex $\sum \alpha_k = d$. It is clearly maximized at $\vec{\alpha}_{opt} = (1, 1, 1, \dots, 1)$, where it evaluates to 1. Suppose that our \vec{x} 's associated α -vector, $\vec{\alpha} = (\alpha_1, \dots, \alpha_d)$, is a distance at least ϵ_α away from $\vec{\alpha}_{opt}$, and that $\epsilon_\alpha \leq 1$. Then one of the coordinates must be at least ϵ_α/\sqrt{d} away from 1. With generality, let this coordinate be α_1 . If $\alpha_1 \leq 1 - \epsilon_\alpha/\sqrt{d}$, then the greatest the value could still be is when the other α_k are all equal at $1 + \epsilon_\alpha/\sqrt{d}(d - 1)$. Multiplying these together, the resulting value is upper-bounded by $1 - \frac{\epsilon_\alpha^2}{2(d-1)}$. If α_1 has instead been increased so that $\alpha_1 \geq 1 + \epsilon_\alpha/\sqrt{d}$, then the value is maximized when the other α_k are all equal at $1 - \epsilon_\alpha/\sqrt{d}(d - 1)$. Multiplying these together, the resulting value is upper-bounded by $1 - \frac{\epsilon_\alpha^2}{4d}$. Since the latter of these bounds is looser, we see that any state whose $\vec{\alpha}$ is at least ϵ_α away from the all-ones vector has a value at most $1 - \frac{\epsilon_\alpha^2}{4d}$ in these measurements.

This gives bounds on the Z vectors' contribution to the value. To keep this bound when the Y vectors are added, we need to check that they are also maximized at $\vec{\alpha} = \vec{1}$. Each factor

$$\prod_{j \leq k} \alpha_j^2 + \alpha_k^2 + 2\alpha_j\alpha_k \cos(2\pi(\theta_j - \theta_k))$$

is maximized when $\theta_j - \theta_k$ is an integer, at which point it becomes $\prod_{j \leq k} (\alpha_j + \alpha_k)^2 = (\prod \alpha_j + \alpha_k)^2$. This is in turn globally maximized by $\alpha_j = \alpha_k = 1$, so the error bound on $\vec{\alpha}$ holds.

The next step is to bound the value in terms of the $\vec{\theta}$. We only care about the degree to which $\theta_i - \theta_j$ is not an integer, let $r_{ij} = \theta_i - \theta_j$ to the nearest integer, so $r_{ij} \in [-1/2, 1/2]$. Given that $\cos(2\pi r) \leq 1 - 8r^2$ for all $r \in [-1/2, 1/2]$, we have a relative value of

$$\begin{aligned} \frac{G_V(\vec{x})}{G_V(B_0)} &= \frac{\alpha_j^2 + \alpha_k^2 + 2\alpha_j\alpha_k \cos(2\pi r_{jk})}{\alpha_j^2 + \alpha_k^2 + 2\alpha_j\alpha_k} \leq \frac{\alpha_j^2 + \alpha_k^2 + 2\alpha_j\alpha_k(1 - 8r_{jk}^2)}{\alpha_j^2 + \alpha_k^2 + 2\alpha_j\alpha_k} \\ &= 1 - \frac{16\alpha_j\alpha_k}{(\alpha_j + \alpha_k)^2} r_{jk}^2 \end{aligned}$$

Let's assume that each α_j is in the interval $[1/2, 3/2]$ – which is implied by them being sufficiently close to the all-ones vector, that is, $\epsilon_\alpha \leq 1/2$. Then the expression $\frac{16\alpha_j\alpha_k}{(\alpha_j + \alpha_k)^2}$ is at least 3, so

$$\frac{G_V(\vec{x})}{G_V(B_0)} \leq 1 - 3r_{jk}^2$$

which tells us that every phase θ_i should be close to 0 for G_V to be large, or else suffer a $1 - 3r^2$ penalty in the value. □

This sets upper bounds on the integrand where are not close to B_0 . We will also need lower bounds on the integrand, if we are close to B_0 :

Lemma 3 *If a vector \vec{x} is within distance $\epsilon \leq 0.1$ of some point b in B_0 , and V is one basic set, then \vec{x} has value at least*

$$G_V(\vec{x}) \geq \frac{1 - 2\epsilon d^{5/2}}{d^{d^2}}$$

or in terms of the relative value,

$$G_V(\vec{x})/G_V(B_0) \geq 1 - 2\epsilon d^{5/2}.$$

Proof We will again use polar representation for \vec{x} :

$$G_V(\vec{x}) = \frac{1}{d^d (2d)^{d^2-d}} \left(\prod_k \alpha_k \right) \left(\prod_{j \leq k} \alpha_j^2 + \alpha_k^2 + 2\alpha_j\alpha_k \cos(2\pi(\theta_j - \theta_k)) \right)$$

If our point \vec{x} is within distance $\epsilon < 1$ of B_0 , then each of the α_i must individually be within $\epsilon\sqrt{d}$ of 1, and each θ_i satisfies

$$\cos(\pi\theta_i) > \sqrt{1 - \epsilon^2} \implies |\theta_i| < \epsilon/2$$

and so

$$\cos(2\pi(\theta_j - \theta_k)) \geq 1 - \frac{(2\pi(\theta_j - \theta_k))^2}{2} \geq 1 - \frac{(2\pi\epsilon)^2}{2}.$$

Then the value is bounded by,

$$\begin{aligned} G_V(\vec{x}) &\geq \frac{1}{d^d(2d)^{d^2-d}} \left(\prod_k 1 - \epsilon\sqrt{d} \right) \\ &\quad \left(\prod_{j \leq k} (1 - \epsilon\sqrt{d})^2 + (1 - \epsilon\sqrt{d})^2 + 2(1 - \epsilon\sqrt{d})^2 \left(1 - \frac{(2\pi\epsilon)^2}{2} \right) \right) \\ &\geq \frac{1}{d^d(2d)^{d^2-d}} (1 - \epsilon\sqrt{d})^{(d^2+d)/2} (4 - 4\pi^2\epsilon^2)^{(d^2-d)/2} \\ &\geq \frac{1}{d^{d^2}} \left(1 - \epsilon\sqrt{d}(d^2 + d)/2 - \pi^2\epsilon^2(d^2 - d)/2 \right) \end{aligned}$$

If $\epsilon < \sqrt{d}/\pi^2$, which is implied by $\epsilon < 0.1$, then the term $\pi^2\epsilon^2(d^2 - d)$ is smaller than $\epsilon\sqrt{d}(d^2 + d)/2$, so we can combine the two. We can also bound $d^2 + d < 2d^2$.

$$G_V(\vec{x}) \geq \frac{1}{d^{d^2}} \left(1 - 2\epsilon\sqrt{d}(d^2 + d)/2 \right) = \frac{1 - 2\epsilon d^{5/2}}{d^{d^2}}$$

□

Together, these two lemmas establish a form of concentration: points close to B_0 have large (lower-bounded) values of G_V , and points far from B_0 have small (upper-bounded) values of G_V .

3.2 Restricting to Neighborhoods of B_0

Now we consider the effect of clause sets. A clause C is defined by a triple of integers (C_1, C_2, C_3) . A point $b \in B_0$ with coordinates (b_1, b_2, \dots, b_d) , each $b_k = \pm e^{i\theta}$, is “good” for the clause C if $\{b_{C_1}, b_{C_2}, b_{C_3}\}$ are not all equal. A point in B_0 is “good” for a set of clauses if it is good for each of them, and a point is “bad” if it is not good. Each clause C has an associated set of three clause vectors

$$\begin{aligned} \vec{v}_1 &= (\sqrt{6})^{-1}(-2\vec{e}_{C_1} + \vec{e}_{C_2} + \vec{e}_{C_3}) \\ \vec{v}_2 &= (\sqrt{6})^{-1}(\vec{e}_{C_1} - 2\vec{e}_{C_2} + \vec{e}_{C_3}) \\ \vec{v}_3 &= (\sqrt{6})^{-1}(\vec{e}_{C_1} + \vec{e}_{C_2} - 2\vec{e}_{C_3}). \end{aligned}$$

Lemma 4 *Take a clause $C = (C_1, C_2, C_3)$ and let V be its three clause vectors. Nowhere does G_V exceed 1. At any point \vec{x} within a distance ϵ of a good point,*

$G_V(\vec{x}) \geq \frac{32}{27d^3} (1 - 12\epsilon\sqrt{d})$. At any point \vec{x} within a distance ϵ of a bad point, $G_V(\vec{x}) \leq \frac{4096}{27}\epsilon^6$.

Proof To see that 1 is an upper bound on G_V , note that G_V is a product of dot products of unit vectors, each of which is at most 1, so that $G_V \leq 1$.

For the second claim, we have a point \vec{x} close to a good point \vec{g} . Since we only care about the value of G_V and the distance between $|\vec{x} - \vec{g}|$, we may adjust the phase of \vec{x} and \vec{g} jointly so that \vec{g} is entirely real, and all of its entries are ± 1 . We decompose \vec{x} in the form

$$\vec{x} = \alpha\vec{e}_{C_1} + \beta\vec{e}_{C_2} + \gamma\vec{e}_{C_3} + \vec{x}_\perp$$

where \vec{x}_\perp is the support of \vec{x} on all component besides the first three. Then the likelihood factor due to the three clause vectors is,

$$G_V(\vec{x}) = \frac{1}{6^3} |\alpha + \beta - 2\gamma|^2 \cdot |\alpha - 2\beta + \gamma|^2 \cdot |-2\alpha + \beta + \gamma|^2$$

We seek to bound this value in the vicinity of good points. A good B_0 point has not all signs equal. Since we can permute the elements of C without affecting the value of G_V , a general good point \vec{g} can be written as

$$\vec{g} = \frac{1}{\sqrt{d}} \left(-\vec{e}_{C_1} + \vec{e}_{C_2} + \vec{e}_{C_3} + \sqrt{d-3}\vec{g}_\perp \right)$$

where \vec{g}_\perp contains the support on all the other basis vectors. It has $G_V(\vec{g}) = \frac{32}{27d^3}$, by direct computation. Then for our other point \vec{x} within a distance ϵ of \vec{g} , each coordinate must also be within ϵ of the corresponding coordinate in \vec{g} . So

$$\begin{aligned} \Re[\alpha + \beta - 2\gamma] &\leq \frac{1}{\sqrt{d}} \left((-1 + \epsilon\sqrt{d}) + (1 + \epsilon\sqrt{d}) - 2(1 - \epsilon\sqrt{d}) \right) \\ &= -2(1 - 2\epsilon\sqrt{d})/\sqrt{d} \end{aligned}$$

and similarly

$$\begin{aligned} \Re[-2\alpha + \beta + \gamma] &\geq \frac{1}{\sqrt{d}} \left(-2(-1 + \epsilon\sqrt{d}) + (1 - \epsilon\sqrt{d}) + (1 - \epsilon\sqrt{d}) \right) \\ &= 4(1 - \epsilon\sqrt{d})/\sqrt{d} \geq 4(1 - 2\epsilon\sqrt{d})/\sqrt{d}. \end{aligned}$$

Putting together the six factors,

$$G_V(\vec{x}) = \frac{1}{6^3} |\alpha + \beta - 2\gamma|^2 \cdot |\alpha - 2\beta + \gamma|^2 \cdot |-2\alpha + \beta + \gamma|^2 \tag{19}$$

$$\geq \frac{1}{6^3} \Re[\alpha + \beta - 2\gamma]^2 \Re[\alpha - 2\beta + \gamma]^2 \Re[-2\alpha + \beta + \gamma]^2 \tag{20}$$

$$\geq \frac{32}{27d^3} \times (1 - 2\epsilon\sqrt{d})^6 \tag{21}$$

$$\geq \frac{32}{27d^3} \times (1 - 12\epsilon\sqrt{d}) \tag{22}$$

which is the second claim. For the third claim, take a bad point \vec{h} in B_0 , for which we can correct the phase to put it in the form

$$\vec{h} = \frac{1}{\sqrt{d}} \left(+\vec{e}_{C_1} + \vec{e}_{C_2} + \vec{e}_{C_3} + \sqrt{d-3} \vec{h}_\perp \right)$$

Then for a nearby point only ϵ away, each coordinate is at most ϵ away. This means

$$\begin{aligned} \Re[\alpha + \beta - 2\gamma] &\leq \left(\frac{1}{\sqrt{d}} + \epsilon \right) + \left(\frac{1}{\sqrt{d}} + \epsilon \right) + \left(\frac{-2}{\sqrt{d}} + 2\epsilon \right) = 4\epsilon \\ \Im[\alpha + \beta - 2\gamma] &\leq 4\epsilon \\ \implies |\alpha + \beta - 2\gamma|^2 &\leq 32\epsilon^2 \end{aligned}$$

and similarly for the other two permutations, so that

$$G_V(\vec{x}) \leq \frac{1}{6^3} (32\epsilon^2)^3 = \frac{4096}{27} \epsilon^6.$$

□

3.3 $F = \int_x G_V(\vec{x})$ Detects NAE3SAT

With these bounds, we will be able to relate the number of solutions to a NAE3SAT instance to the integral $F = \int_x G_V(\vec{x})$.

Theorem 4 *Given an instance of NAE3SAT with d variables and k clauses, let the set of vectors V be given by $K_1 = 1600d^7 \ln^2(d)$ copies of basic vectors (Z and Y vectors), together with $K_2 = d^2 \ln(d)$ copies of the clause vectors for each clause. For sufficiently large d , there is a function $p(n, k)$ such that, if there is at least one solution to the NAE3SAT, $F = \int_x G_V(\vec{x}) \geq pd^{-22d}$, and if there are no solutions, $F \leq pd^{-d^2}$.*

Proof The theorem will hold if we take p as the value of G_V at a good point, or

$$p = d^{-K_1 d^2} \left(\frac{32}{27d^3} \right)^{K_2}.$$

If the original NAE3SAT instance has a satisfying assignment $(1, 0, 0, 1, \dots)$, there is a corresponding good point

$$\vec{g} = \frac{1}{\sqrt{d}} (+\vec{e}_1 - \vec{e}_2 - \vec{e}_3 + \vec{e}_4 \dots)$$

with a large value of $G_V(\vec{g})$. Each set of basic vectors introduces a factor of $1/d^{d^2}$ in G , and each set of clause vectors introduces a factor of $32/27d^3$. Thus

$$G_V(\vec{g}) = d^{-K_1 d^2} \left(\frac{32}{27d^3} \right)^{K_2} = p$$

Further, we want to show that around this good point \vec{g} , there is an appreciable volume with large G_V , that will contribute substantially to F . Around each good point, take the ball of radius

$$\epsilon_g = \frac{1}{3200d^9(1+d)}.$$

Then by Lemma 3, each set of basic observations gives a factor in G of at least

$$G_1 \geq \frac{1 - 2\epsilon_g d^{5/2}}{d^{d^2}}$$

and by Lemma 4, each set of clause observations gives a factor at least

$$G_2 \geq \frac{32}{27d^3} (1 - 12\epsilon_g \sqrt{d})$$

so that the final G_V value of each point in the ball is at least

$$G_0 = G_1^{K_1} G_2^{K_2} \geq p(1 - 2\epsilon_g d^{5/2})^{K_1} (1 - 12\epsilon_g \sqrt{d})^{K_2} \tag{23}$$

$$\geq p(1 - 2\epsilon_g K_1 d^{5/2})(1 - 12K_2 \epsilon_g \sqrt{d}) \tag{24}$$

$$= p \left(1 - 2 \frac{1}{3200d^9(1+d)} (1600d^7 \ln^2(d)) d^{5/2} \right) \left(1 - 12(d^2 \ln(d)) \frac{1}{3200d^9(1+d)} \sqrt{d} \right) \tag{25}$$

$$\geq p \left(1 - \frac{\ln^2 d}{\sqrt{d}} \right) \tag{26}$$

This means the total contributed to F by the ball around this good point is then at least $p(1 - \ln^2 d/\sqrt{d})$ times the volume of this ball around \vec{g} . The ball is not actually a sphere in \mathbb{R}^{2d} , as it lies on the manifold of normalized states, which is curved; it’s the intersection of a ball centered at \vec{g} and the unit sphere. But since $\epsilon_g < 1/2$, this deformation reduces the volume by less than a factor of $1/2$, and then we can use the standard volume of the ball. So the volume obeys

$$\text{Vol} \geq \frac{1}{2} \cdot \frac{2(d-1)!(4\pi)^{(d-1)}}{(2d-1)!} \epsilon_g^{2d-1}$$

and a single good point contributes a total to the integral $F(V)$ at least

$$\text{Vol} \cdot G_0 \geq pc_1 c_2^{-d} d^9 d^{-21d}$$

for some particular constants $c_1, c_2 > 1$; the d^{-21d} term clearly dominates the scale for large d . For sufficiently large d then we can write

$$F \geq \text{Vol} \cdot G_0 \geq pd^{-22d}$$

which establishes the first claim. The second claim concerns when there are no good points. Suppose for contradiction that there is some point \vec{x} (not necessarily in B_0) so that $G_V(\vec{x}) > p/d^{d^2}$. Applying Lemma 2, we know that it must have $\epsilon_\alpha = |\vec{\alpha} - \vec{1}| < 0.1/d^2$, otherwise it would have at most

$$G_V(\vec{x}) \leq \left(d^{-d^2} (1 - 0.1^2/4d^5)\right)^{K_1} < d^{-K_1d^2} \exp(-K_1/400d^5) \tag{27}$$

$$= d^{-K_1d^2} \exp(-4d^2 \ln^2 d) \tag{28}$$

$$< d^{-K_1d^2} \exp(-4d^2 \ln^2 d + d^2 \ln d \ln(32/27)) \tag{29}$$

$$= d^{-K_1d^2} \exp(-d^2 \ln^2 d + d^2 \ln d \ln(32/27d^3)) \tag{30}$$

$$= d^{-K_1d^2} \exp\left(-d^2 \ln^2 d + \ln\left(\left(\frac{32}{27d^3}\right)^{K_2}\right)\right) \tag{31}$$

$$= d^{-K_1d^2} \left(\frac{32}{27d^3}\right)^{K_2} / d^{d^2 \ln d} = p/d^{d^2 \ln d} \tag{32}$$

$$< p/d^{d^2} \tag{33}$$

Since $\epsilon_\alpha \leq 1/2$, we can also apply the second part of Lemma 2 and check that the all phases $|\theta_i| < 0.1/d$, otherwise our point would have G_V at most

$$\left(d^{-d^2} (1 - 3\theta_i^2)\right)^{K_1} < \left(d^{-d^2} (1 - 0.03/d^2)\right)^{K_1} \tag{34}$$

$$< \left(d^{-d^2} (1 - 0.1^2/4d^5)\right)^{K_1} \tag{35}$$

$$< p/d^{d^2} \tag{36}$$

Since the amplitudes are all within ϵ_a of $1/\sqrt{d}$, and the phases are all within $0.1/d$ of 0, the point's distance to the nearest point b in B_0 is at most

$$\begin{aligned} \text{dist}_{B_0} &\leq \sqrt{d} \left(\epsilon_a + \left(\frac{1}{\sqrt{d}} + \epsilon_a \right) \left((1 - \cos(\theta_i))^2 + \sin^2(\theta_i) \right) \right) \\ &\leq \sqrt{d} \left(\frac{0.1}{d^2} + \left(\frac{1}{\sqrt{d}} + \frac{0.1}{d^2} \right) (2 - 2 \cos(0.1/d)) \right) \leq \sqrt{d} \left(\frac{0.1}{d^2} + \frac{2}{\sqrt{d}} (0.1/d)^2 \right) \\ &\leq \frac{0.11}{d^{3/2}} \end{aligned}$$

If that point b is bad, then by Lemma 4 our point would have G_V at most

$$\begin{aligned} d^{-K_1 d^2} \left(\frac{4096}{27} \left(\frac{0.11}{d^{3/2}} \right)^6 \right)^{K_2} &= d^{-K_1 d^2} \left(\frac{0.00027}{d^9} \right)^{K_2} \\ &= d^{-K_1 d^2} \left(\frac{32}{27 d^3} \right)^{K_2} \left(\frac{0.00027}{(32/27) d^6} \right)^{K_2} \\ &< p \times 0.00025^{K_2} = p/4000^{d^2 \ln d} < p/(4000d)^{d^2} < p/d^{d^2}. \end{aligned}$$

We’ve shown that all points have $G_V \leq p/d^{d^2}$. The volume of integration is $S_{2n-1} < 1$, so the total integral F is less than p/d^{d^2} . □

As an aside, we might hope that F actually detects MAX3SAT as well: that, if there are no solutions to the original constraint problem, that perhaps the value of F would be determined by the *maximum* number of clauses satisfied at once. This is considerably harder to analyze, however, since G_V itself is actually zero at any binarized point, since at least one constraint is not satisfied. Accordingly, the value of F will be based on the relatively small values of G_V away from binarized points. The fact that we cannot reasonably analyze this case does not hurt our results of course, since MAX3SAT can still be reduced to NAE3SAT by virtue of its NP-completeness.

3.4 NP Hardness

We can now prove our main result.

Theorem 5 *For any constant $C < 1$, it is NP-Hard to approximate the permanent of an $n \times n$ Hermitian positive semidefinite matrix within a factor of 2^{n^C} .*

Proof We can reduce from NAE3SAT. Given an NAE3SAT instance on d variables, we can use the set of vectors V described in Theorem 4 and examine the resulting value F . As we have $O(d^9)$ vectors in V , the quantity F can be represented as a permanent of a matrix of size $O(d^9)$. The NAE3SAT instance is satisfiable if $F \geq pd^{-22d}$ and unsatisfiable if $F \leq pd^{-d^2}$, which can be distinguished if approximating within a factor of $d^{d^2-22d} = O(d^{d^2})$, and so $O(2^{d^2})$ will suffice. If we had an oracle that could approximate permanents of size n PSD matrices within a factor of 2^{n^C} for some $C < 1$, then we could do the replica trick: take the matrix corresponding to F , and repeat it $M = d^{(9C-2)/(1-C)}$ many times along the diagonal. We assume that $9C - 2 \geq 0$, so that the exponent on M is positive; otherwise we can freely increase C up past $2/9$ and this would only make the problem easier. The result is a matrix of size Md^9 , which is then approximated within a factor of $2^{(Md^9)^C}$. The resulting matrix size Md^9 is still $poly(d)$ for any fixed C . Then we raise this approximate answer to the power $1/M$ to recover an approximation to the original permanent, and it has multiplicative error

$$\left(2^{(Md^9)^C} \right)^{1/M} = 2^{d^{9C} M^{C-1}} = 2^{d^{9C} d^{2-9C}} = 2^{d^2}$$

which is sufficient to distinguish between satisfiable and unsatisfiable instances. As #NAE3SAT is NP hard, so is approximating HPSD permanents with this accuracy. \square

This result is complementary to one of Anari et al. [12], where they show that one *can* approximate within a factor of $\exp((1 + \gamma + o(1))n)$ where γ is the Euler-Mascheroni constant, while we showed that permanents cannot be approximated with subexponential error. Our hard instances circumvent the fast approximation schemes of Barvinok [15] and Chakhmakhchyan et al. [14], which both have requirements on the spectrum of the matrix, and perform more favorably when $\lambda_{max}/\lambda_{min}$ is smaller. Our instances are of low rank (only rank d , which is much smaller than the matrix size n) so that $\lambda_{min} = 0$. Finally, we conjecture that the reduction above is approximation preserving: that each good point contributes an equal amount to the integral that can easily be estimated beforehand. Showing this would require tighter error bounds.

Conjecture 1 *With an appropriate choice of polynomial-scaling K_1 and K_2 , the construction used in Theorem 4 is an approximation-preserving reduction from #NAE3SAT to HPSD permanents, such that approximating HPSD Permanents within a factor C is as hard as approximating #NAE3SAT (or #3SAT) within a factor C .*

It is known that by Stockmeyer counting [7, 9, 10] computing multiplicative approximations to PSD permanents is contained in FBPP^{NP}. If approximating PSD permanents is indeed as hard as approximating #3SAT, it seems unlikely to be significantly easier than FBPP^{NP}.

3.5 Real Matrices

The arguments above all involve complex vectors, complex matrices, and integrals over the complex unit sphere. The arguments however can easily be adapted to show that PSD permanents remain hard even for purely real matrices. We could have proved the results only for the real case and this would of course imply hardness for the more general complex case, but the proof for the real case was less symmetric, aesthetic, or intuitive than the complex case, which is why we delayed to this section.

Theorem 6 *For any constant $C < 1$, it is NP-Hard to approximate the permanent of an $n \times n$ real positive semidefinite matrix within a factor of 2^{n^C} .*

Proof The construction proceeds very similarly to above, by reducing from NAE3SAT. However, we now use one dimension more in the space: a d -variable NAE3SAT problem is mapped to a $(d + 1)$ -dimensional spherical integral $\int G_V(\vec{x})$. The clauses are mapped, as before, with K_2 many sets of clause vectors, connecting the variables 1 through d in the original problem with dimensions 1 through d in the spherical integral $G(x)$. The “basic sets” still include K_1 many instances of the unit vectors \vec{e}_k in each basis direction $k \in [d + 1]$, what we previously referred to as the Z vectors.

The Y vectors were, in previous proofs, of the form $\frac{\vec{e}_j \pm i\vec{e}_k}{\sqrt{2}}$, for $j \neq k$. This was the sole source of complex terms in our vectors, and the reasons the resulting matrices were complex. Instead now we use four copies of each of $\frac{\vec{e}_j \pm \vec{e}_{d+1}}{\sqrt{2}}$. These each softly enforce the constraint that the component of \vec{x} in the j direction and the $d + 1$ direction

have relative phase $\pm i$ (that is, $\pm\sqrt{-1}$). Since each j has $\pm i$ relative to $d + 1$, this implies that each $j \neq k$ have relative phase ± 1 .

To make this quantitative and precise, we refer to the proof of Lemma 2. The bound of $1 - \frac{\epsilon_\alpha^2}{4d}$ applies as before, since the \vec{e}_k vectors occur just as before. As proved in Lemma 2, if θ_j and θ_{d+1} differ by a phase (up to ± 1) of $\Delta\theta_j = \theta_j - \theta_{d+1}$, then the likelihood $G_V(x)$ is reduced by a factor of $1 - 3\Delta\theta_j^2$; since we use each vector eight times, this becomes $(1 - 3\Delta\theta_j^2)^8$. Then for two $j \neq k, j, k \leq d$, the likelihood is at most

$$\begin{aligned} (1 - 3\Delta\theta_j^2)^4(1 - 3\Delta\theta_k^2)^4 &\leq \left(1 - 3\left(\frac{|\Delta\theta_j| + |\Delta\theta_k|}{2}\right)^2\right)^8 \\ &\leq \left(1 - 3\left(\frac{|\theta_j - \theta_k|}{2}\right)^2\right)^8 \leq 1 - 3(\theta_j - \theta_k)^2 \end{aligned}$$

which gives us the same bound on the relative phases as before, so that an analogous statement to Lemma 2 also holds for our new basis set. The proof of Lemma 3 holds with few modifications: in the proof above, the Y terms

$$\prod_{j \leq k} \alpha_j^2 + \alpha_k^2 + 2\alpha_j\alpha_k \cos(2\pi(\theta_j - \theta_k))$$

as bounded by a factor

$$\begin{aligned} \prod_{j \leq k} (1 - \epsilon\sqrt{d})^2 + (1 - \epsilon\sqrt{d})^2 + 2(1 - \epsilon\sqrt{d})^2 \left(1 - \frac{(2\pi\epsilon)^2}{2}\right) \\ = \left((1 - \epsilon\sqrt{d})(4 - 4\pi^2\epsilon^2)\right)^{(d^2-d)/2} \geq 1 - (\epsilon\sqrt{d} + \pi^2\epsilon^2)\frac{d^2 - d}{2}. \end{aligned}$$

Here instead we have four copies of each phase constraint, but only between $j \leq d$ and $d + 1$. So the penalty from

$$\prod_{j \leq d} \left(\alpha_j^2 + \alpha_{d+1}^2 + 2\alpha_j\alpha_{d+1} \cos(2\pi(\theta_j - \theta_{d+1}))\right)^4$$

becomes

$$\begin{aligned} \prod_{j \leq d} \left((1 - \epsilon\sqrt{d})^2 + (1 - \epsilon\sqrt{d})^2 + 2(1 - \epsilon\sqrt{d})^2 \left(1 - \frac{(2\pi\epsilon)^2}{2}\right)\right)^4 \\ = \left((1 - \epsilon\sqrt{d})(4 - 4\pi^2\epsilon^2)\right)^{4d} \\ \geq 1 - (\epsilon\sqrt{d} + \pi^2\epsilon^2)(4d) \geq 1 - (\epsilon\sqrt{d} + \pi^2\epsilon^2)\frac{d^2 - d}{2} \end{aligned}$$

as before, as long as $d \geq 9$. The resulting conclusion of the lemma that the relative value $G_V(\vec{x})/G_V(B_0) \geq 1 - 2\epsilon d^{5/2}$ thus still holds.

Finally, Lemma 4 remains unmodified in this setting, as the form of the clause vectors is unchanged. As all the necessary lemmas hold as before, and the proofs of Theorems 4 and 5 only care about relative values, they will all hold in the real-valued PSD setting. \square

4 Quantum State Tomography

The author initially found the above construction while investigating the worst-case hardness of quantum state tomography, and the hardness implies that several problems in the context of tomography are NP-hard as well.

Quantum State Tomography (QST) is the procedure of estimating an unknown quantum state from a set of measurements on an identically prepared ensemble. The procedure can encompass both the choosing of measurement bases as well as estimating the resulting state from the measurements; in adaptive settings, the running estimate is also used to inform future measurement choices [21, 22]. We focus on the latter task, of building an estimate of the state. We look at four related forms of what “estimation” can qualify as:

1. Finding the Maximum Likelihood Estimator (MLE): the pure state ρ most likely to produce the observations.
2. Finding the Bayesian expected state ρ_{Avg} : assuming a prior over the possible pure states, finding the mixed state presenting the mixture of appropriately weighted possible states.
3. Computing the expectation value of some future observation(s).
4. Finding the probability that the unknown state is in fact some particular ρ_0 . (As there are infinitely many different pure states, we are actually asking for the probability *density* at ρ_0 .)

The first three estimations problems have all been extensively studied with various heuristics. MLE can be attempted by linear inversion [23, 24], iterative search [25, 26], or even neural networks [27]. Bayesian estimation can be accomplished by direct numerical integration [28] or particle based sampling [21], possibly with neural networks guiding the particles [22]. Directly estimating future samples has also been attempted with neural networks [29] or classical shadows [30–32]. The author is not aware of any prior work on computing estimation problem 4.

We can show that estimation problems 2, 3, and 4 are essentially as hard as approximating PSD permanents, and that task 1 is also NP-hard. The exponential difficulty (assuming ETH [16]) is in fact in the dimension d of the underlying Hilbert space. Many questions in quantum information appear to be “exponentially” hard, in the sense that it is hard to analyze a system of q qubits faster than $O(2^q)$. But here $d = 2^q$, so that even when the number of qubits is a logarithmically small $q = \log(d)$, the problem of state estimation remains exponentially hard.

4.1 Outline of Tomography Results

Of the four forms above, we focus first on estimation problem 4. Although it is likely the question least relevant to experiment, it is the easiest to manipulate algebraically. We call it QUANTUM- BAYESIAN- UPDATE, or simply QBU, and define it in Sect. 4.2. In Sect. 4.3, we give an exponential time algorithm for QBU, showing that it is at least possible. In Sect. 4.4 we show that estimation problems 2, 3, and 4 are equivalent. In Sect. 4.5 we explain QBU’s connection to HPSD permanents, and show it is NP-Hard to approximate within subexponential error. In Sect. 4.6 we show how the construction of difficult PSD permanents can also be modified shows that the MLE problem (estimation problem 1) is also NP-hard to approximate: it is NP-hard to check the existence of a state with likelihood within a subexponential factor.

4.2 Quantum Bayesian Update

We define the QBU problem as follows: given a series of observations \mathcal{O}_i each taken from a copy of ρ , and a guess ρ_0 , what is the probability density that $\rho = \rho_0$? The actual probability of equality is zero—unless we have some other powerful information about the state—which is why we ask for the probability density in the space of candidate density matrices.

Bayes’ theorem lets us compute the probability density of a true state ρ in terms of the likelihood of the observations $P(\mathcal{O}|\rho)$, a prior belief distribution $P(\rho)$, and the total probability of the sequence of observations $P(\mathcal{O})$. It reads,

$$P(\rho_0|\mathcal{O}) = \frac{P(\mathcal{O}|\rho_0)P(\rho_0)}{P(\mathcal{O})}$$

In order for the equation to be meaningful and not identically zero on both sides, we can read ρ as representing a small volume in the space of density matrices. While there are many natural priors on the space of density matrices, we focus on the case where we know the unknown state ρ is pure. This models, for instance, where we are trying to identify the output of a unitary quantum channel. The most natural prior is then the uniform distribution over all pure states, given by the Haar measure. Then all $P(E)$ are equal. The likelihood of a given observation \mathcal{O}_i is simply $\text{Tr}[\mathcal{O}_i \rho]$, so our goal is to compute

$$P(\rho|\mathcal{O}) = \frac{\prod_{i \in [n]} \text{Tr}[\mathcal{O}_i \rho]}{P(\mathcal{O})}$$

In general \mathcal{O}_i could be operators of any rank, and could belong to POVMs. For hardness, it will suffice it consider only observations with rank 1 and trace 1, but for now we allow them to be general. For any particular ρ and sequence \mathcal{O}_i , the likelihood $\prod \text{Tr}[\mathcal{O}_i \rho]$ can be evaluated directly in $O(nd^2)$ operations. The difficulty then lies in the normalizing factor,

$$p_{norm} = P(\mathcal{O})$$

so that

$$P(\rho|\mathcal{O}) = p_{norm}^{-1} \prod_{i \in [n]} \text{Tr}[\mathcal{O}_i \rho]$$

This indicates the probability of an entire sequence of observations. While a single observation has the simple form of $P(\mathcal{O}_i) = \text{Tr}[\mathcal{O}_i]$, the expression rapidly becomes more complicated as we consider sequences of observations.

A brief example is useful for understanding what p_{norm} represents. Suppose that we measure a qubit 1000 times along each of the X, Y, and Z axes: we expect to see a particular amount of bias. Observing 1000 results each of +X, +Y, and +Z would be very unlikely, as the qubit cannot be in the +1 eigenstate of all three axes at once. It would be similarly surprising to see exactly 500 counts each of +X, -X, +Y, -Y, +Z, and -Z: this state shows no tendency of a particular orientation, but a pure qubit state must show a bias towards some orientation. This would have a small value of p_{norm} , as there is no good state to explain the sequence observed. A sequence of 1000 +Z observations, and 500 each of +X, -X, +Y, and -Y is much more likely, as it can be well explained by the $|\uparrow\rangle$ state, and so has a larger value of p_{norm} .

As we just saw, computing the probability density $P_{density}(\rho = \rho_0|\mathcal{O})$ is easy if p_{norm} is known, and conversely p_{norm} can be easily computed from the probability density. p_{norm} is a more attractive goal for our problem, as it doesn't depend on ρ_0 . It can be computed by summing up all unnormalized probabilities:

$$p_{norm} = \int_{\vec{x} \in \mathbb{C}_1^d} \prod_{i \in [n]} \text{Tr}[\mathcal{O}_i x x^\dagger] dx$$

where the integral is over the Hilbert space \mathbb{C}^d restricted to length-1 vectors. This leads to the definition,

Definition 3 (Quantum-Bayesian-Update) Given a collection of observations $\mathcal{O} = (\mathcal{O}_1, \dots \mathcal{O}_n)$ in a Hilbert space of dimension d , compute

$$p_{norm} = \frac{\int_{\vec{x} \in \mathbb{C}_1^d} \prod_{i \in [n]} \text{Tr}[\mathcal{O}_i x x^\dagger] dx}{\text{Tr}[\mathcal{O}_i]} \tag{37}$$

4.3 Polynomial Time QBU for Fixed d

This space of state vectors \mathbb{C}_1^d has the geometry of a real $(2d - 1)$ -sphere, and the entries of ρ are quadratic in the Cartesian coordinates for this sphere. Thus, p_{norm} becomes an integral over a $(2d - 1)$ -sphere of a homogeneous $2n$ degree polynomial in the $2d$ variables. The expansion of the polynomial into monomials takes $O((2n)^{2d})$ time, and each monomial can then be immediately integrated over the sphere using the formula [19]

$$\int_{S^k} x_1^{\alpha_1} x_1^{\alpha_2} \dots x_k^{\alpha_k} = \begin{cases} 0 & \text{if any } \alpha_i \text{ are odd} \\ \frac{2 \prod_i \Gamma(\frac{1}{2}(\alpha_i+1))}{\Gamma(\sum_i \frac{1}{2}(\alpha_i+1))} & \text{if all } \alpha_i \text{ are even} \end{cases} \tag{38}$$

where Γ is gamma function, $\Gamma(\frac{1}{2}(\alpha + 1)) = \sqrt{\pi}2^\alpha(\alpha - 1)!!$. This gives a polynomial time algorithm for evaluating p_{norm} when d is fixed. This is functionally equivalent to the $O(n^d)$ algorithm for permanents described in [33]: Barvinok describes the more general form that applies to any matrices (not just PSD) of rank d . This fact puts the (exact) permanent calculation in the class XP, or slicewise polynomial time, when parameterized by rank.

4.4 Relationship Between Estimation Problems

Since QBU is not of particular interest to actual tomography tasks, we show it is equivalent (under polynomial many-one reductions) to the more realistic tasks 2 and 3 above, of estimating observables or the state itself. We can show that these are just as difficult (or, just as easy) as the Bayesian update step.

4.4.1 Computing ρ_{Avg}

Given that there will always be room for uncertainty, we cannot meaningfully ask for a single pure state as an answer, but we can ask for ρ_{Avg} : the mixed state representing the correctly updated mixture over all the possible true states, given by $\int P(\rho)\rho d\rho$. The impure ρ_{Avg} reflects the expectation of all observables given our current information. We parameterize the space of density matrices by a single vector $\psi \in S^{2d-1}$, and given some completed observations \mathcal{O} , the Bayesian expected state is

$$\begin{aligned} \rho_{Avg} &= \int_{\psi \in S^{2d-1}} P(|\psi\rangle\langle\psi| | \mathcal{O}) |\psi\rangle\langle\psi| d\psi \\ &= \int_{\psi \in S^{2d-1}} p_{norm}^{-1}(|\psi\rangle\langle\psi|) \prod_{O \in \mathcal{O}} \langle\psi|O|\psi\rangle d\psi \end{aligned}$$

whose individual matrix elements are

$$\langle i | \rho_{Avg} | j \rangle = p_{norm}^{-1} \int_{\psi \in S^{2d-1}} \langle i | \psi \rangle \langle \psi | j \rangle \prod_{O \in \mathcal{O}} \langle \psi | O | \psi \rangle d\psi$$

We have already discussed computing p_{norm} , as a spherical integral of a polynomial. For any given i and j , the remaining integral is also a spherical integral of a polynomial, and can be computed in the same fashion as 4.3. Since each of the d^2 matrix elements can be computed in $O((2n)^{(2d)})$ time, this the Bayesian average state ρ_{Avg} is also computable in polynomial time for a fixed d .

On the other hand, a diagonal element $\langle i | \rho_{avg} | i \rangle$ gives

$$\langle i | \psi \rangle \langle \psi | i \rangle \prod_{O \in \mathcal{O}} \langle \psi | O | \psi \rangle = \langle \psi | (|i\rangle\langle i|) | \psi \rangle \prod_{O \in \mathcal{O}} \langle \psi | O | \psi \rangle = \prod_{O \in (\mathcal{O} \cup \{|i\rangle\langle i|})} \langle \psi | O | \psi \rangle$$

which is the same integrand as for p_{norm} , only with one additional observation $|i\rangle\langle i|$ added.

If we had an algorithm compute ρ_{Avg} efficiently, we could use it to solve the Bayesian update problem on a set of observations \mathcal{O} , as follows. First, discard the last observation O_{last} , compute ρ_{Avg} on the $n - 1$ other observations, and compute $p_{norm,n-1}$ for those $n - 1$ as well. This calls the procedure recursively on one less observation. Then the desired answer is

$$p_{norm} = p_{norm,n-1} \text{Tr}[O_{last} \rho_{Avg}]$$

That is to say, the probability of the n observations is just the probability of the first $n - 1$ observations multiplied by the last probability of the last (conditioned on the first). This recursive approach makes n calls to computing ρ_{Avg} , which shows that the QBU problem is at most n times as hard as state estimation.

4.4.2 Computing Observable Expectations

We could try to only find the expectation of a particular observable A , and not the whole state ρ_{Avg} , conditioned on our observations. We can write this as $E[A|\mathcal{O}]$. This is also just as hard: density matrices as a $d^2 - 1$ linear space, and expectations of observables are linear in ρ , so by computing the exact expectation of $d^2 - 1$ independent observables, we can find ρ_{Avg} exactly, by solving a linear system. This is of course precisely the idea behind least-squares quantum state estimation, and it shows that computing expectation values is as hard as ρ_{Avg} .

Finally, if we could compute a Bayesian update, we could compute the expectation values of observables. Take the eigendecomposition of our operator A : write $A = \sum \lambda_i |i\rangle\langle i|$, and evaluate

$$E[A|\mathcal{O}] = \sum \lambda_i E[\langle i|\psi\rangle\langle\psi|i\rangle|\mathcal{O}] = \sum \lambda_i p_{norm}^{-1} \int_{\psi \in S^{2d-1}} \langle i|\psi\rangle\langle\psi|i\rangle \prod_{O \in \mathcal{O}} \langle\psi|O|\psi\rangle d\psi$$

$$\sum \lambda_i p_{norm}^{-1} \int_{\psi \in S^{2d-1}} \prod_{O \in (\mathcal{O} \cup \{|i\rangle\langle i|})} \langle\psi|O|\psi\rangle d\psi$$

Computing p_{norm} and each of the d many spherical integrals is a Bayesian update problem. We have reductions (Bayesian update) \rightarrow (Compute ρ_{Avg}) \rightarrow (Compute $E[A|\mathcal{O}]$) \rightarrow (Bayesian update), so these are equivalent in difficulty. Note that these are many-one reductions, which is unavoidable as ρ_{Avg} is a matrix-valued function problem while the two are scalar-valued.

4.5 NP-Hardness of QBU and ρ_{Avg}

We now state the main hardness results on quantum tomography.

Theorem 7 *For any $C < 1$, it is NP-hard to compute the value p_{norm} for Quantum-Bayesian-Update with an approximation factor of at most 2^{n^C} .*

Proof When \mathcal{O}_i are all rank-1 operators, the numerator in Eq. (37) is of the form in Theorem 3, and the denominator in Eq. (37) can be efficiently computed by direct calculation. Thus any PSD permanent can be efficiently reduced to a problem of computing p_{norm} with an approximation-preserving reduction, and QBU is NP-Hard to approximate to the same degree. \square

Theorem 8 *For any $C < 1$, it is NP-hard to compute a diagonal matrix entry of ρ_{Avg} , in any basis, with an approximation factor of at most 2^{n^C} . It is also NP-hard to compute the expectation of a positive semidefinite operator \mathcal{O} with an approximation factor of at most 2^{n^C} .*

Proof A diagonal element of ρ_{Avg} is the expectation value of the rank-1 PSD operator projecting onto that element, so the first statement is a special case of the second. As described above, both of these quantities then also take the form of a PSD permanent, and any PSD permanent can be turned into these problem by taking the desired matrix element (in the first case) or observable \mathcal{O} (in the second case) to be the first vector $V_1^\dagger V_1$. These are also approximation preserving reductions, so these are also NP-hard to approximate. \square

4.6 NP-Completeness of Maximum Likelihood Estimation

In the case of MLE state tomography, we are not so demanding that we require knowledge of the full average state, and we are content with just finding one good explanatory state $|\psi\rangle$. Accordingly, we do not consider a permanent $\int_x G_V(x)$ (a problem of counting solutions to 3-SAT), but just the question of maximizing $G_V(x)$ (a problem of finding a solution to 3-SAT). This allows to show that the problem actually lies in NP, while this is unlikely to be true for the other problems in this paper unless $BPP^{NP} = NP$.

Formulating the MLE problem as a decision problem:

Definition 4 (*C-Approximate-Quantum-MLE*) Given a collection of observations \mathcal{O}_i of an unknown quantum state $|\psi\rangle$, and a real number p , decide whether there is a $|\psi\rangle$ whose likelihood $L(\psi) = \prod_i \langle \psi | \mathcal{O}_i | \psi \rangle$ is at least p , or if $L(\psi) < p/C$ for all ψ , being promised that one of these is the case.

We will show that even the approximate problem is NP-complete, for any C .

Theorem 9 *For any $C > 1$, the C-Approximate-Quantum-MLE problem is NP-complete.*

Proof Containment in NP is straightforward, as one can supply a description of the state $|\psi\rangle$, which requires only $2d$ many real numbers, and then $L(\psi)$ can be directly evaluated. To be contained in NP, we need to show that only $poly(n)$ bits of precision are needed, equivalently a $1 + 2^{-poly(n)}$ approximation ratio to the exact number. Every factor in $L(\psi)$, a single $\langle \psi | \mathcal{O}_i | \psi \rangle$, is a sum of products of numbers in our witness ψ and the given \mathcal{O}_i . Each individual product has a $1 + 2^{-poly(n)}$ ratio, but when we add them there may be a catastrophic cancellation, and $\langle \psi | \mathcal{O}_i | \psi \rangle$ is very small, so that we

get a large multiplicative error. But we know this cannot be optimal: this would mean that \mathcal{O}_i projects onto a subspace that $|\psi\rangle$ is exponentially close to (within an angle of $2^{-poly(n)}$). We can choose any other point $|\psi\rangle$ that is *not* exponentially close to any projected subspace, and get at least a better score. Such a point is guaranteed to exist, because n hyperplanes can only divide the d -sphere into $2^{poly(n,d)}$ many sections; by taking a larger $poly$ in our precision, we avoid catastrophic cancellation and compute $\langle\psi|\mathcal{O}_i|\psi\rangle$ to within $1 + 2^{-poly(n)}$. Finally, the multiplicative accuracies in each factor $\langle\psi|\mathcal{O}_i|\psi\rangle$ multiply together, so we can guarantee the C constant approximation ratio necessary to distinguish the two sides of the promise.

We also need to check that the witness is normalized, $\langle\psi|\psi\rangle \approx 1$. As long as this is accurate to within $1 + 2^{-n}$, rescaling $|\psi\rangle$ by a factor f just scales $L(\psi)$ by f^{2n} , so also preserves the good multiplicative approximation. This shows containment in NP.

To show hardness, we use the same NAE3SAT construction as in Theorem 4. As was shown in the proof of that theorem, any good point (thus, a solution to the underlying NAE3SAT problem) has

$$L(\psi) = G_0(x) \geq p \left(1 - \frac{\ln^2 d}{\sqrt{d}}\right).$$

We also show in that proof that, if there are no good points (and thus no solutions) then

$$L(\psi) = G_0(x) \leq p/d^{d^2}$$

for all points. Thus, the existence of a high likelihood point even within $C < d^{d^2}$ implies the existence of a solution. \square

4.7 Practical Difficulty of Tomography

Although the above results imply that several approaches to quantum state tomography may be difficult to compute exactly, these difficult instances are somewhat artificial and unlikely to occur in practice. Additionally, difficult instances such as the one constructed in the above proofs could be readily addressed in practice by the addition of measurements in e.g. the X measurement basis, which would directly probe the relative signs in the state vector and allow relatively efficient readout of the state. Additionally, the constraint that we only search for pure states—while a useful prior that could be relevant once high-fidelity quantum computer exists—makes a highly nonconvex search space. If we relax this and take a prior with uniform measure over the space of density matrices, then the resulting likelihood function is logarithmically convex and the resulting MLE problem can be solved in polynomial time in d . Thus, these results should not be taken as a statement that quantum state tomography is actually exponentially hard in the Hilbert space dimension d . Rather, any analysis of quantum state tomography procedures will need at least one of: careful choice of measurement basis, only probabilistic guarantees on convergence, or (if doing MLE) a convex prior.

This work has little implication for the *learnability* of quantum states, which asks how many (or what kind of) samples we need in order to constrain a state. The optimal *sample complexity* of states is already known [34, 35] to scale as $O(d)$ for pure states

and $O(d^2)$ for mixed states. It is also known that simple reconstruction algorithms, this quantity of data, converge with high probability. Our work, in contrast, shows that the optimal answer *given a fixed set of data* is hard to compute in general.

Acknowledgements I would like to thank Eric Vigoda for substantial guidance in submitting this work, and many useful conversations.

Author Contributions AM was the sole author.

Funding No funding was received to assist with the preparation of this manuscript. The authors have no relevant financial or non-financial interests to disclose.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ryser, HJohn: Combinatorial Mathematics, vol. 14. American Mathematical Soc., Providence (1963)
2. Valiant, L.: The complexity of computing the permanent. Theoret. Comput. Sci. **8**(2), 189–201 (1979). [https://doi.org/10.1016/0304-3975\(79\)90044-6](https://doi.org/10.1016/0304-3975(79)90044-6)
3. Ben-Dor, A., Halevi, S.: Zero-one permanent is # p-complete, a simpler proof. In: Proceedings of the 2nd Israel Symposium on the Theory and Computing Systems, (1993)
4. Jerrum, M., Sinclair, A., Vigoda, E.: A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. J. ACM **51**(4), 671–697 (2004). <https://doi.org/10.1145/1008731.1008738>
5. Gurvits, Samorodnitsky: A deterministic algorithm for approximating the mixed discriminant and mixed volume, and a combinatorial corollary. Discrete Comput. Geom. **27**(4), 531–550 (2002). <https://doi.org/10.1007/s00454-001-0083-2>
6. Tamma, V., Laibacher, S.: Multiboson correlation interferometry with multimode thermal sources. Phys. Rev. A **90**, 063836 (2014). <https://doi.org/10.1103/PhysRevA.90.063836>
7. Rahimi-Keshari, S., Lund, A.P., Ralph, T.C.: What can quantum optics say about computational complexity theory? Phys. Rev. Lett. **114**, 060501 (2015). <https://doi.org/10.1103/PhysRevLett.114.060501>
8. Kim, Y., Hong, K.-H., Kim, Y.-H., Huh, J.: Connection between bosonsampling with quantum and classical input states. Opt. Express **28**(5), 6929–6936 (2020). <https://doi.org/10.1364/OE.384973>
9. Grier, D., Schaeffer, L.: New hardness results for the permanent using linear optics. In: Proceedings of the 33rd Computational Complexity Conference, ser. CCC'18, San Diego, California: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, (2018)
10. Stockmeyer, L.: The complexity of approximate counting. In: Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing, ser. STOC'83, New York, NY, USA: Association for Computing Machinery, pp. 118–126. (1983) <https://doi.org/10.1145/800061.808740>
11. Marcus, M.: The permanent analogue of the Hadamard determinant theorem. Bull. Am. Math. Soc. **69**(4), 494–496 (1963). <https://doi.org/10.1090/s0002-9904-1963-10975-1>
12. Anari, N., Gurvits, L., Gharan, S. O., Saberi, A.: Simply exponential approximation of the permanent of positive semidefinite matrices. In: 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), pp. 914–925 (2017) <https://doi.org/10.1109/FOCS.2017.89>

13. Yuan, C., Parrilo, P.A.: Maximizing products of linear forms, and the permanent of positive semidefinite matrices. *Math. Program.* (2021). <https://doi.org/10.1007/s10107-021-01616-3>
14. Chakhmakchyan, L., Cerf, N.J., Garcia-Patron, R.: Quantum-inspired algorithm for estimating the permanent of positive semidefinite matrices. *Phys. Rev. A* **96**, 022329 (2017). <https://doi.org/10.1103/PhysRevA.96.022329>
15. Barvinok, A.: A remark on approximating permanents of positive definite matrices, 2020. [arXiv: 2005.06344](https://arxiv.org/abs/2005.06344) [cs.DS]
16. Impagliazzo, R., Paturi, R.: Complexity of k-sat. In: Proceedings of Fourteenth Annual IEEE Conference on Computational Complexity (Formerly: Structure in Complexity Theory Conference) (Cat. No.99CB36317), pp. 237–240 (1999). <https://doi.org/10.1109/CCC.1999.766282>
17. Barvinok, A.: *Combinatorics and Complexity of Partition Functions*, vol. 1. Springer, Berlin (2017)
18. Zvonkin, A.: Matrix integrals and map enumeration: an accessible introduction. *Math. Comput. Model.* **26**(8), 281–304 (1997). [https://doi.org/10.1016/S0895-7177\(97\)00210-0](https://doi.org/10.1016/S0895-7177(97)00210-0)
19. Folland, G.B.: How to integrate a polynomial over a sphere. *Am. Math. Mon.* **108**(5), 446–448 (2001). <https://doi.org/10.1080/00029890.2001.11919774>
20. Schaefer, T. J.: The complexity of satisfiability problems. In: Proceedings of the Tenth Annual ACM Symposium on Theory of Computing, ser. STOC'78, San Diego, California, USA: Association for Computing Machinery, pp. 216–226 (1978). <https://doi.org/10.1145/800133.804350>
21. Huszár, F., Houlisby, N.M.T.: Adaptive Bayesian quantum tomography. *Phys. Rev. A* **85**, 052120 (2012). <https://doi.org/10.1103/PhysRevA.85.052120>
22. Quek, Y., Fort, S., Ng, H.K.: Adaptive quantum state tomography with neural networks. *npj Quantum Inf.* **7**(1), 105 (2021). <https://doi.org/10.1038/s41534-021-00436-9>
23. Qi, B., Hou, Z., Li, L., Dong, D., Xiang, G., Guo, G.: Quantum state tomography via linear regression estimation. *Sci. Rep.* **3**(1), 3496 (2013). <https://doi.org/10.1038/srep03496>
24. D'Auria, V., Fornaro, S., Porzio, A., Solimeno, S., Olivares, S., Paris, M.G.A.: Full characterization of gaussian bipartite entangled states by a single homodyne detector. *Phys. Rev. Lett.* **102**, 020502 (2009). <https://doi.org/10.1103/PhysRevLett.102.020502>
25. Lvovsky, Alexander I.: Iterative maximum-likelihood reconstruction in quantum homodyne tomography. *J. Opt. B: Quantum Semiclass. Opt.* **6**, 556 (2004). <https://doi.org/10.1088/1464-4266/6/6/014>
26. Řeháček, J., Hradil, Z., Ježek, M.: Iterative algorithm for reconstruction of entangled states. *Phys. Rev. A* **63**, 040303 (2001). <https://doi.org/10.1103/PhysRevA.63.040303>
27. Torlai, G., Mazzola, G., Carrasquilla, J., Troyer, M., Melko, R., Carleo, G.: Neural-network quantum state tomography. *Nat. Phys.* **14**(5), 447–450 (2018). <https://doi.org/10.1038/s41567-018-0048-5>
28. Blume-Kohout, Robin: Optimal, reliable estimation of quantum states. *New J. Phys.* **12**(4), 043034 (2010). <https://doi.org/10.1088/1367-2630/12/4/043034>
29. Smith, A.W.R., Gray, J., Kim, M.S.: Efficient quantum state sample tomography with basisdependent neural networks. *PRX Quantum* **2**, 020348 (2021)
30. Aaronson, S.: The learnability of quantum states. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **463**(2088), 3089–3114 (2007). <https://doi.org/10.1098/rspa.2007.0113>
31. Aaronson, S.: Shadow tomography of quantum states. In: Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, ser. STOC 2018, Los Angeles, CA, USA: Association for Computing Machinery, pp. 325–338 (2018). <https://doi.org/10.1145/3188745.3188802>
32. Huang, H.-Y., Kueng, R., Preskill, J.: Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**(10), 1050–1057 (2020). <https://doi.org/10.1038/s41567-020-0932-7>
33. Barvinok, A.I.: Two algorithmic results for the traveling salesman problem. *Math. Oper. Res.* **21**(1), 65–84 (1996)
34. O'Donnell, R., Wright, J.: Efficient quantum tomography. In: Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing, ser. STOC'16, Cambridge, MA, USA: Association for Computing Machinery, pp. 899–912 (2016). <https://doi.org/10.1145/2897518.2897544>
35. Haah, J., Harrow, A.W., Ji, Z., Wu, X., Yu, N.: Sample-optimal tomography of quantum states. *IEEE Trans. Inf. Theory* **63**(9), 5628–5641 (2017). <https://doi.org/10.1109/TIT.2017.2719044>