Quantum approximation of normalized Schatten norms and applications to learning

Yiyou Chen ,[1,*] Hideyuki Miyahara ,[2,†] Louis-S. Bouchard ,[3,4,5,6,7,‡] and Vwani Roychowdhury [2,5,§]

[1]*Department of Computer Science, University of California, Los Angeles, California 90095, USA*
[2]*Department of Electrical and Computer Engineering, University of California, Los Angeles, California 90095, USA*
[3]*Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA*
[4]*Department of Bioengineering, University of California, Los Angeles, California 90095, USA*
[5]*Center for Quantum Science and Engineering, University of California, Los Angeles, California 90095, USA*
[6]*California NanoSystems Institute, University of California, Los Angeles, California 90095, USA*
[7]*The Molecular Biology Institute, University of California, Los Angeles, California 90095, USA*

(Received 23 June 2022; accepted 21 October 2022; published 9 November 2022)

Efficient measures to determine the similarity of quantum states, such as the fidelity metric, have been widely studied. In this paper, we address the problem of defining a similarity measure for quantum operations that can be *efficiently estimated*. Given two quantum operations, U_1 and U_2, represented in their circuit forms, we first develop a quantum sampling circuit to estimate the normalized Schatten 2-norm of their difference ($\|U_1 - U_2\|_{S_2}$) with precision ϵ, using *only one clean qubit and one classical random variable*. We prove a Poly($\frac{1}{\epsilon}$) upper bound on the sample complexity, which is independent of the size of the quantum system. We then show that such a similarity metric is directly related to a functional definition of similarity of unitary operations using the conventional fidelity metric of quantum states ($\mathcal{F}$): If $\|U_1 - U_2\|_{S_2}$ is sufficiently small (e.g., $\leqslant \frac{\epsilon}{1+\sqrt{2(1/\delta-1)}}$) then the fidelity of states obtained by processing the same randomly and uniformly picked pure state $|\psi\rangle$ is as high as needed [$\mathcal{F}(U_1|\psi\rangle, U_2|\psi\rangle) \geqslant 1-\epsilon$] with probability exceeding $1-\delta$. We provide example applications of this efficient similarity metric estimation framework to quantum circuit learning tasks, such as finding the square root of a given unitary operation.

DOI: 10.1103/PhysRevA.106.052409

I. INTRODUCTION

Recent advances in quantum approximate optimization algorithms (QAOA, [1,2]), the variational quantum eigensolver (VQE, [3]), and the promise of implementing such algorithms using noisy intermediate-scale quantum (NISQ) devices [4] rekindled the prospect of a new era in quantum computing. Researchers started experimenting with quantum machine learning algorithms such as quantum neural networks (QNN) [5–9] and quantum circuit learning [10–14] that are based on variational quantum algorithms [15,16]; a recent work studied the price of the ansatz used in such variational methods [17]. These algorithms assumed a hybrid model which takes advantage of both classical and quantum computations: loss functions are obtained by summing the outputs of a quantum machine whereas the variational parameters of the model (circuit) are learned using a classical optimizer.

A critical factor in the formulation of a learning algorithm is the design of its loss functions, which often involves computing a similarity measure between a target objective and the output of the parameterized model (Fig. 1). In VQE, for instance, the objective is to determine the lowest-energy eigenstate of a given Hermitian operator H. The learning framework assumes an ansatz comprising a quantum circuit with a fixed topology, but where each gate is parameterized to generate a candidate eigenstate vector $|\psi(\xi)\rangle = U(\xi)|\mathbf{0}\rangle$, where $U(\xi)$ is the unitary operator determined by the parameters ξ. Such a pure state vector $|\psi(\xi)\rangle$ is an eigenstate, if $H|\psi(\xi)\rangle = \lambda|\psi(\xi)\rangle$, where λ is to be minimized when searching for the ground state. Thus, the objective function of VQE can be interpreted as minimizing the cosine similarity between $|\psi(\xi)\rangle$ and $H|\psi(\xi)\rangle$ or the expectation value $\langle\psi(\xi)|H|\psi(\xi)\rangle$. This loss term can be physically estimated by performing measurements corresponding to the observables used to define H. This similarity metric is related to the well-known measure of fidelity used for determining the similarity of quantum states, [18–20] and was extensively applied to distinguishing quantum states [21–29].

In contrast, consider the problem where one wants to learn a given quantum operation V, which is also available when controlled by a clean qubit (see Table I). That is, the task is to learn ξ such that $U(\xi) \approx V$. This problem of learning quantum operations is much less studied, in spite of its applications to quantum circuit synthesis [30–33] (where low-depth approximations of quantum circuits are needed) and to distinguish quantum operations and channels [29,34–58]. The main difficulty of such learning tasks is the design of a similarity metric between $U(\xi)$ and V that can be *efficiently estimated*. In Gilchrist *et al.*'s work [59], for example, a similarity metric that was blind to input unitary operations was

*gerry99@ucla.edu
†miyahara@g.ucla.edu; hmiyahara512@gmail.com
‡louis.bouchard@gmail.com
§vwani@g.ucla.edu

TABLE I. Hadamard test circuit. Given an arbitrary quantum operation V controlled by a clean qubit, the above circuit computes $\mathrm{Re}\langle\psi(\xi)|V|\psi(\xi)\rangle$, where $|\psi(\xi)\rangle = U(\xi)|\mathbf{0}\rangle$ and $U(\xi)$ is a quantum circuit parameterized by ξ.

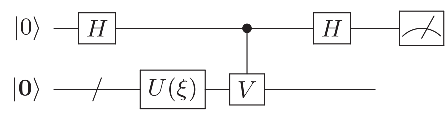
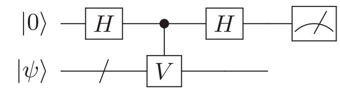


TABLE II. Hadamard test: $\mathrm{Re}\{\langle\psi|V|\psi\rangle\}$.



studied and validated, but the estimation of the metric was inefficient because it required an exponential number of quantum states. Other metrics such as the diamond norm [60,61] were conceptualized to distinguish quantum operations, but they heavily relied on the classical information of the input unitaries such as their eigendecompositions [62,63].

The Schatten norm, studied and explored from an information theory perspective [64–66], is another candidate, and the approximation of which was proven to be DQC1-complete [67,68]. These approximation schemes [67,68], however, required an exponential classical sample complexity when a clean qubit (e.g., the control bit of the Hadamard test circuit in Table I) was provided. Herein we present a random sampling method, using few samples [e.g., $O(\frac{\ln(2/\delta)}{2\epsilon^2})$ sample complexity] and an efficient sampling circuit design [e.g., $O(1)$ in depth], to estimate the normalized trace and normalized Schatten 2-norm of any given quantum operation. We then formulate a similarity metric for quantum operations using the notion of fidelity of quantum states and show how such a metric is closely related to the normalized Schatten 2-norm. As a consequence, we can use the normalized Schatten 2-norm of the difference between $U(\xi)$ and V as a loss function to learn a target quantum operation V.

The paper is organized as follows. In Sec. II, we provide the background concepts and notations. In Sec. III, we present a sampling method to approximate the normalized trace of matrices that are unitarily similar to diagonal matrices, as well as to approximate the normalized Schatten 2-norm of arbitrary $N \times M$ matrices. We prove an upper bound of the sample complexity of such a sampling method. In Sec. IV, we introduce the normalized Schatten 2-norm of mixed quantum operations, which can be estimated efficiently using the sampling method from the previous section and the Hadamard test circuits shown in Tables I. We also present an optimized circuit design for the approximation. In Sec. V, we relate the normalized Schatten 2-norm to a similarity metric of quantum operations. Finally, in Sec. VI, we present an application of the efficient approximation of the normalized Schatten 2-norm to quantum circuit learning.

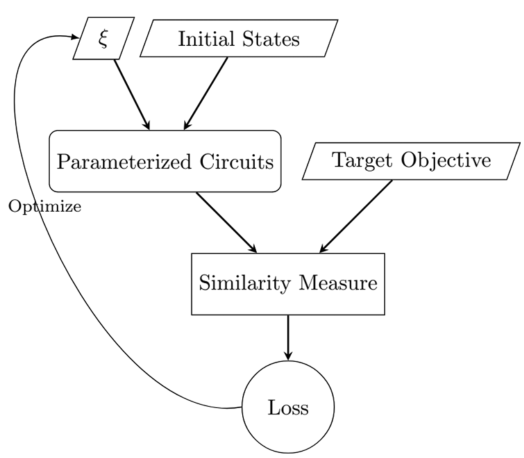
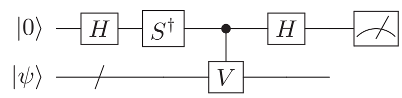


FIG. 1. A schematic illustration of variational quantum algorithms. In the case of VQE, the initial state is $|\mathbf{0}\rangle$, the parameterized circuit is U, the target objective is a Hamiltonian H, the similarity measure is the cosine similarity, and the loss term is $\langle\mathbf{0}|U^\dagger(\xi)HU(\xi)|\mathbf{0}\rangle$.

II. BACKGROUND AND NOTATIONS

Given an n-qubit quantum system, a quantum state is specified by a density matrix $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, where $|\psi_i\rangle \in \mathbb{CP}^{N-1}$ ($\mathbb{CP}$ denotes the complex projective space) are pure state vectors and $N = 2^n$ is the dimension of the Hilbert space representing the quantum system. To deal with the equivalence class on $\mathbb{CP}^{N-1}$ we adopt the convention of normalization that a pure state $|\psi\rangle$ is a point on the boundary of a unit ball centered at the origin, i.e., $|\psi\rangle \in \partial B_{\mathbb{C}^N}(0,1)$ and thus has a unit norm, i.e., $\langle\psi||\psi\rangle = 1$. We use the term "quantum operation" to refer to a unitary map of a density matrix $\rho \to U\rho U^\dagger$. This linear operator is a type of a Liouville space superoperator. A unitary operator $U \in \mathbb{C}^{N\times N}$ working on n qubits can generally be decomposed as a product of unitary operators, where each U_i is a unitary operator acting on only a reduced number of qubits. For example, $U = \Pi_{i=1}^{L} U_i$, $U_i = U_i^{j,k} \otimes I^{s-\{j,k\}}$ where s is the set of qubits and $U_i^{j,k} \in \mathbb{C}^{4\times 4}$ acts on the jth and kth qubits. Such few-qubit operations (e.g., $U_i^{j,k}$) are referred to as gates, and a quantum circuit is a visual representation of a sequence of gates used to represent a quantum operation. Some well-known quantum gates include $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $H = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$, $\mathrm{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$, $S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$, $R_x(\theta) = e^{-i\frac{\theta}{2}\sigma_x}$, $R_y(\theta) = e^{-i\frac{\theta}{2}\sigma_y}$, and $R_z(\theta) = e^{-i\frac{\theta}{2}\sigma_z}$. In the simplest instance we can define a parameterized quantum circuit $U(\theta)$ as a circuit with learnable parameters θ for its rotational gates. The parameters θ are called variational parameters. Given an arbitrary quantum unitary V and a state $|\psi\rangle$, the expectation of V, $\langle\psi|V|\psi\rangle$, can be estimated using a method called the Hadamard test [69], as shown in Tables II and III. Letting $\Pr(1)$ be the probability of observing $|1\rangle$ from measuring the control qubit in Tables II and III, then $1 - 2\Pr(1)$ evaluates $\mathrm{Re}\{\langle\psi|V|\psi\rangle\}$ and $\mathrm{Im}\{\langle\psi|V|\psi\rangle\}$,

TABLE III. Hadamard test: $\mathrm{Im}\{\langle\psi|\,V\,|\psi\rangle\}$.



respectively. The control bit in the Hadamard test is called a clean qubit.

In this paper, we also consider a generalization of the definition of quantum operations comprising finite linear combinations of unitary quantum operations. We use $\tilde{U}$ to denote such mixed quantum operations $\tilde{U} = \sum_{\kappa=1}^{K} \alpha_\kappa U_\kappa$, where $U_\kappa \in \mathbb{C}^{N\times N}$ are unitary quantum operations and $\sum_{\kappa=1}^{K} |\alpha_\kappa| \leqslant 1$. For simplicity, we only consider the case when K is of $O(1)$. In the case when $\alpha_\kappa \geqslant 0$ and $\sum_{\kappa=1}^{K} \alpha_\kappa = 1$, a mixed quantum operation $\sum_{\kappa=1}^{K} \alpha_\kappa U_\kappa$ represents a quantum computing mixture model that applies the unitary operation U_κ to any state ρ with probability α_κ and yields a density matrix output of $\sum_{\kappa=1}^{K} \alpha_\kappa U_\kappa \rho U_\kappa^\dagger$. Such a mixed quantum system is a special case of a larger class, completely positive trace-preserving (CPTP) maps [18], and could, in principle, be used to model quantum errors. In general, one could ask the following question: Given a mixed quantum system as an oracle, can one design a variational quantum algorithm to approximate it to a high accuracy? This is one of the learning problems we address in Sec. VI.

For any matrix $A \in \mathbb{C}^{N\times M}$, let $A = W\Sigma T^\dagger$ be the singular value decomposition (SVD), where $\Sigma = \mathrm{diag}(\sigma_i)$ and $\sigma_i \geqslant 0$ are the singular values. Note that $AA^\dagger$ is always diagonalizable with eigenvalues σ_i^2 and $AA^\dagger = W\Sigma T^\dagger T\Sigma^\dagger W^\dagger = W(\Sigma\Sigma^\dagger)W^\dagger = W\hat{\Sigma}W^\dagger$, where $\hat{\Sigma} = \mathrm{diag}(\sigma_i^2)$. Moreover, let $|w_i\rangle$ be the column vectors of W (also called left-singular vectors) under bra-ket notation $AA^\dagger = \sum_{i=1}^{N} \sigma_i^2 |w_i\rangle\langle w_i|$.

A square matrix $A \in \mathbb{C}^{N\times N}$ is unitarily similar to a diagonal matrix D if $A = WDW^\dagger$ where W is unitary and D is diagonal in $\mathbb{C}^{N\times N}$. In particular, for any matrix $A \in \mathbb{C}^{N\times M}$, $AA^\dagger$ is unitarily similar to $\mathrm{diag}(\sigma_i^2)$. Another result we use is that all unitary matrices are also unitarily similar to the diagonal matrices [70].

Given two quantum states' density matrices ρ_1, ρ_2, the fidelity is customarily defined as $\mathcal{F}_\rho(\rho_1,\rho_2) = [\mathrm{Tr}(\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}})]^2$ [18]. Given two pure states $\rho_1 = |\psi_1\rangle\langle\psi_1|$ and $\rho_2 = |\psi_2\rangle\langle\psi_2|$, it can be shown that $\mathcal{F}_\rho(\rho_1,\rho_2) = |\langle\psi_1||\psi_2\rangle|^2$. In this paper we work with pure states, and this simplified version of fidelity between wave functions will be used and denoted as $\mathcal{F}(|\psi_1\rangle, |\psi_2\rangle)$. In general, fidelity measures how similar two quantum states are and $\psi_1 = \psi_2$ if and only if $\mathcal{F}(\psi_1,\psi_2) = 1$.

Definition 1. The normalized Schatten p-norm [64–68] for arbitrary matrix $A \in \mathbb{C}^{N\times M}$ and $p \in [1,\infty)$ is defined as

$$\|A\|_{S_p} = \left(\frac{\sum_i {\sigma_i}^p}{N}\right)^{\frac{1}{p}},$$

where σ_i are the singular values of A. Note that the Schatten 2-norm is related to the Frobenius norm via $\|A\|_{S_2} = \frac{\|A\|_F}{\sqrt{N}}$.

We can relate the normalized Schatten norm of the difference of two unitary operations to a functional definition of similarity using the fidelity of states.

Definition 2. Let $|\psi\rangle$ be a random variable defined on a distribution $\mathcal{J} = \mathrm{Uni}[\partial B_{\mathbb{C}^N}(0,1)]$, i.e., $|\psi\rangle$ is uniformly random over all pure quantum states, we define two unitary operations U_1, U_2 to be pure-state (δ,ϵ)-similar if

$$\mathbb{P}_{\psi\sim\mathcal{J}}(\mathcal{F}(U_1|\psi\rangle, U_2|\psi\rangle) \geqslant 1-\epsilon) \geqslant 1-\delta.$$

Let X_1, X_2,..., X_m be independent random variables with $X_i \in [a_i, b_i] \subset \mathbb{R}$ almost surely and define $S_m = \sum_{i=1}^{m} X_i$, the Chernoff-Hoeffding inequality [71] states

$$\mathbb{P}(|S_m - \mathbb{E}[S_m]| > \epsilon) < 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^{m}(b_i-a_i)^2}}.$$

A special case is when X_1,..., X_m are independent and identically distributed random variables (iidrv) on [0,1] almost surely. Setting $\overline{X} = \frac{S_m}{m}$, we obtain the following inequality:

$$\mathbb{P}(|\overline{X} - \mathbb{E}[X_1]| > \epsilon) < 2e^{-2\epsilon^2 m}.$$

Note that when $m(\epsilon,\delta) = O(\frac{\ln(2/\delta)}{2\epsilon^2})$, $\mathbb{P}(|\overline{X} - \mathbb{E}[X_1]| \leqslant \epsilon) \geqslant 1-\delta$.

III. EFFICIENT SAMPLING ALGORITHMS FOR APPROXIMATING NORMALIZED TRACE AND SCHATTEN NORMS

We first consider any matrix $A \in \mathbb{C}^{N\times N}$ that is unitarily similar to a diagonal matrix (e.g., unitary matrices, Hermitian matrices, etc.) and present an efficient sampling technique to approximate its normalized trace. Recall that A can be decomposed as

$$A = WDW^\dagger = \sum_{i=1}^{N} d_i |w_i\rangle\langle w_i|, \; w_i \in W,$$

where W is unitary and D is diagonal. The goal is to find a distribution $\mathcal{D}$ and a random vector $x \sim \mathcal{D}$ such that

$$\mathbb{E}_{x\sim\mathcal{D}}\langle x|A|x\rangle = \frac{\mathrm{Tr}(A)}{N}.$$

It suffices to show $\mathbb{E}_{x\sim\mathcal{D}}\langle w||x\rangle = \frac{1}{N}$ for all unit vectors $w \in \mathbb{C}^N$. As discussed in [67] and [68], this is equivalent to the uniform sampling of $|x\rangle$ (with replacement) from the standard basis $\{e_1,\ldots,e_N\}$, which, in general, requires $\Omega(N)$ sampling complexity. We show in Lemma 1 that if we construct $x(\theta)$ using a *continuous* classical random variable θ, then $\mathbb{E}_\theta\langle w||x(\theta)\rangle = \frac{1}{N}$ holds as desired. Under such a construction we can efficiently approximate the normalized trace $\frac{\mathrm{Tr}(A)}{N}$.

Lemma 1. Let $A \in \mathbb{C}^{N\times N}$ be unitarily similar to a diagonal matrix and let $n = \lceil \log_2 N\rceil$. Define a geometric sequence $(\omega)_{i=1}^{n}$ with $w_i = 2^i$. Let random variable $\theta \sim \mathcal{D} = \mathrm{Uni}[-\pi,\pi]$, and define a random vector $x(\theta) \in \mathbb{R}^N$ with N entries $x_0,\ldots,x_{N-1}$ and $x_i(\theta) = \sqrt{\frac{2^n}{N}}\Pi_{j=1}^{n}\cos^{b_{i_j}}(\omega_j\theta)\sin^{1-b_{i_j}}(\omega_j\theta)$, where $b_{i_1}\ldots b_{i_n}$ is the n-bit binary representation of i. For example, when $N = 2^n - 1$, we obtain $x(\theta) \in \mathbb{R}^N$ where the only missing entry is

$\sin(\omega_1\theta)\sin(\omega_2\theta)\ldots\sin(\omega_{n-1}\theta)\sin(\omega_n\theta)$,

$$x(\theta)=\sqrt{\frac{2^n}{N}}\begin{pmatrix}\cos(\omega_1\theta)\ldots\cos(\omega_{n-1}\theta)\cos(\omega_n\theta)\\ \cos(\omega_1\theta)\ldots\cos(\omega_{n-1}\theta)\sin(\omega_n\theta)\\ \cos(\omega_1\theta)\ldots\sin(\omega_{n-1}\theta)\cos(\omega_n\theta)\\ \cos(\omega_1\theta)\ldots\sin(\omega_{n-1}\theta)\sin(\omega_n\theta)\\ \ldots\\ \sin(\omega_1\theta)\ldots\cos(\omega_{n-1}\theta)\cos(\omega_n\theta)\\ \sin(\omega_1\theta)\ldots\cos(\omega_{n-1}\theta)\sin(\omega_n\theta)\\ \sin(\omega_1\theta)\ldots\sin(\omega_{n-1}\theta)\cos(\omega_n\theta)\end{pmatrix}.$$

Then

$$\mathbb{E}_{\theta\sim\mathcal{D}}\,\langle x(\theta)|\,A\,|x(\theta)\rangle=\frac{\mathrm{Tr}(A)}{N}.$$

Proof. We first note that any signed sum of any subset of ω_i's is a *nonzero* integer, i.e. $\forall S\subset\{1,\ldots,n\}:\sum_{s\in S}\pm\omega_s\in\mathbb{Z}\backslash\{0\}$. This statement can be proved by induction, and we omit the proof here. For any such nonempty $S\subset\{1,\ldots,n\}$, it follows that

$$\mathbb{E}_{\theta\sim\mathcal{D}}\Pi_{s\in S}e^{\pm i2\omega_s\theta}=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{2\sum_{s\in S}\pm i\omega_s\theta}d\theta=0. \quad (1)$$

Next, we show that for all $j\neq k$, $\mathbb{E}_{\theta\sim\mathcal{D}}x_jx_k=0$, and $\mathbb{E}_{\theta\sim\mathcal{D}}{x_i}^2=\frac{1}{N}$. For any pair (j,k), let $c_1c_2\ldots c_n$ be the binary representation of $j\oplus k$. We define $S_0=\{p\in\mathbb{N}\cap[1,n]:c_p=0\}$, $S_1=\{q\in\mathbb{N}\cap[1,n]:c_q=1\}$, and 2^{S_0}, 2^{S_1} to be the corresponding power sets. Note that $S_0\cap S_1=\emptyset$,

$$\begin{aligned}&\mathbb{E}_{\theta\sim\mathcal{D}}x_jx_k\\ &=\mathbb{E}_{\theta\sim\mathcal{D}}\frac{2^n}{N}\Pi_{p\in S_0}\frac{1\pm\cos(2\omega_p\theta)}{2}\Pi_{q\in S_1}\frac{\sin(2\omega_q\theta)}{2}\\ &=\mathbb{E}_{\theta}\frac{2^n}{N}\sum_{S\in 2^{S_0}}\frac{\pm1}{2^{|S_0|+|S_1|}}\Pi_{q\in S_1}\sin(2\omega_q\theta)\Pi_{p\in S}\cos(2\omega_p\theta)\\ &=\mathbb{E}_{\theta\sim\mathcal{D}}\frac{2^n}{N}\frac{1}{2^{|S_0|+|S_1|}}\sum_{S\in 2^{S_0}}\frac{\pm1}{2^{|S|}(2i)^{|S_1|}}\Pi_{q\in S_1}(e^{i2\omega_q\theta}\\ &\quad-e^{-i2\omega_q\theta})\Pi_{p\in S}(e^{i2\omega_p\theta}+e^{-i2\omega_p\theta})\\ &=\frac{1}{N}\frac{1}{(2i)^{|S_1|}}\sum_{S\in 2^{S_0}}\frac{\pm1}{2^{|S|}}\mathbb{E}_{\theta\sim\mathcal{D}}\Pi_{q\in S_1}(e^{i2\omega_q\theta}-e^{-i2\omega_q\theta})\\ &\Pi_{p\in S}(e^{i2\omega_p\theta}+e^{-i2\omega_p\theta}).\end{aligned}$$

When $j\neq k$, $j\oplus k\neq 0$ and $S_1\neq\emptyset$. It follows from Eq. (1) that all $\mathbb{E}_{\theta\sim\mathcal{D}}\Pi_{q\in S_1}(e^{i2\omega_q\theta}-e^{-i2\omega_q\theta})\Pi_{p\in S}(e^{i2\omega_p\theta}+e^{-i2\omega_p\theta})$ evaluates to 0. Therefore, $\mathbb{E}_{\theta\sim\mathcal{D}}x_jx_k=0$.

Analogously when $j=k$, $S_0=\{1,\ldots,n\}$ and $S_1=\emptyset$,

$$\begin{aligned}\mathbb{E}_{\theta\sim\mathcal{D}}x_j^2&=\frac{1}{N}+\frac{1}{N}\sum_{S\in 2^{S_0}\backslash\emptyset}\frac{\pm1}{2^{|S|}}\mathbb{E}_{\theta\sim\mathcal{D}}\Pi_{p\in S}(e^{i2\omega_p\theta}\\ &\quad+e^{-i2\omega_p\theta})=\frac{1}{N}.\end{aligned}$$

We thus showed that x is unbiased under the standard basis $\{e_1,e_2,\ldots,e_N\}$ [72]. It remains to show that for arbitrary unit vector w, $\mathbb{E}_{\theta\sim D}|\langle x|\,|w_j\rangle|^2=\frac{1}{N}$. We decompose w in the standard basis, i.e., $w=\sum_{j=1}^N w_je_j$,

$$\begin{aligned}\mathbb{E}_{\theta\sim D}|\langle x|\,|w\rangle|^2&=\sum_{j,k=1}^N\mathbb{E}_{\theta\sim D}x_jx_kw_j^*w_k\\ &=\sum_{j=1}^N\mathbb{E}_{\theta\sim D}|x_j|^2|w_j|^2=\frac{1}{N}.\end{aligned}$$

The main claim $\mathbb{E}_{\theta\sim\mathcal{D}}\,\langle x|\,A\,|x\rangle=\frac{\mathrm{Tr}(A)}{N}$ follows. ■

By Lemma 1, for an arbitrary matrix $A\in\mathbb{C}^{N\times N}$ that is unitarily similar to a diagonal matrix, we can approximate $\frac{\mathrm{Tr}(A)}{N}$ using m random samples. Namely, we randomly sample $\theta_1,\ldots,\theta_m\sim\mathcal{D}=\mathrm{Uni}[-\pi,\pi]$ and use $x(\theta_i)$ as defined in Lemma 1 to approximate

$$\frac{\widehat{\mathrm{Tr}(A)}}{N}=\frac{1}{m}\sum_{i=1}^m\langle x(\theta_i)|\,A\,|x(\theta_i)\rangle\,. \quad (2)$$

We study the sample complexity for such an approximation to achieve a low error rate with high success probability.

Theorem 1. Let $A\in\mathbb{C}^{N\times N}$ be unitarily similar to a diagonal matrix. For any δ, $\epsilon>0$, with sample complexity $m(\epsilon,\delta)=O(\frac{\ln(2/\delta)}{4\epsilon^2})$, samples $\theta_1,\ldots,\theta_{m(\epsilon,\delta)}\sim\mathrm{Uni}[-\pi,\pi]$, and $x(\theta_i)$ as defined in Lemma 1, the following holds for the classical approximation of $\frac{\mathrm{Tr}(A)}{N}$ using (2):

$$\mathbb{P}\left(|\frac{\widehat{\mathrm{Tr}(A)}}{N}-\frac{\mathrm{Tr}(A)}{N}|<\epsilon\right)>1-\delta.$$

Proof. The theorem follows from the Chernoff-Hoeffding bound for complex numbers where we bound the precision for both real and imaginary parts to be within $\frac{\epsilon}{\sqrt{2}}$. ■

The sampling method in Eq. (2) can be generalized to estimate the normalized Schatten p-norms when p is even, but in this paper we are particularly interested in the case when $p=2$.

For arbitrary matrix $A\in\mathbb{C}^{N\times M}$, $AA^\dagger$ is unitarily similar to a diagonal matrix in $\mathbb{C}^{N\times N}$ with σ_i^2 along its diagonal. We observe that $\|A\|_{S_2}=\sqrt{\frac{\mathrm{Tr}(AA^\dagger)}{N}}$, based on which we approximate $\|A\|_{S_2}$. Namely,

$$\widehat{\|A\|_{S_2}}=\sqrt{\frac{\widehat{\mathrm{Tr}(AA^\dagger)}}{N}}. \quad (3)$$

Theorem 2. Let $A\in\mathbb{C}^{N\times M}$ be an arbitrary matrix. For any δ, $\epsilon>0$, with sample complexity $m(\epsilon,\delta)=O(\frac{\ln(2/\delta)}{2\epsilon^2}\min\{\epsilon^{-2},\|A\|_{S_2}^{-2}\})$, samples $\theta_1,\ldots,\theta_{m(\epsilon,\delta)}\sim\mathrm{Uni}[-\pi,\pi]$, and $x(\theta_i)$ as defined in Lemma 1, the following holds for the classical approximation of $\|A\|_{S_2}$ using Eq. (3):

$$\mathbb{P}(|\widehat{\|A\|_{S_2}}-\|A\|_{S_2}|<\epsilon)>1-\delta.$$

Proof. It suffices, see Appendix, to show

$$\mathbb{P}\left(|\frac{\widehat{\mathrm{Tr}(AA^\dagger)}}{N}-\|A\|_{S_2}^2|<\epsilon\max\{\epsilon,\|A\|_{S_2}\}\right)>1-\delta,$$

which follows from the result of normalized trace approximation in Theorem 1 and the Chernoff-Hoeffding bound. ■

Since all sampled state vectors $x(\theta_i)$ have dimension N, classically evaluating each $\langle x(\theta_i)|\,AA^\dagger\,|x(\theta_i)\rangle$ requires

Poly(N) arithmetic operations. Next, we show how we can take advantage of quantum computing to speedup the evaluations and thus make the approximation of the normalized Schatten 2-norms more efficient.

IV. QUANTUM APPROXIMATION OF NORMALIZED SCHATTEN 2-NORMS FOR MIXED QUANTUM OPERATIONS

We apply the general results from the previous section to quantum operations. Let n be the number of qubits in a quantum system and $N = 2^n$. A quantum operation (usually denoted as U) can be defined by a unitary matrix in $\mathbb{C}^{N\times N}$.

Recall that we define a mixed quantum operation to be a linear combination of a finite $K \sim O(1)$ quantum operations with coefficients $(\alpha)_{\kappa=1}^{K}$ satisfying $\sum_{\kappa=1}^{K} |\alpha_\kappa| \leqslant 1$:

$$\tilde{U} = \sum_{\kappa=1}^{K} \alpha_\kappa U_\kappa .$$

While all U_κ are unitary quantum operations which are unitarily similar to diagonal matrices, their linear combinations $\tilde{U}$ may not be. Therefore, the approximation of the normalized trace from Theorem 1 does not generalize to mixed quantum operations. Nevertheless, the approximation of the normalized Schatten 2-norms does generalize and we present explicit quantum circuit constructions to approximate $\|\tilde{U}\|_{S_2}$. We start with a toy example when $\tilde{U} = \frac{1}{\sqrt{2}}U_1 - \frac{1}{\sqrt{2}}U_2$. For an arbitrary pure state $|x\rangle \in \partial B_{\mathbb{C}^N}(0, 1)$,

$$\begin{aligned}\langle x|\,\tilde{U}\tilde{U}^\dagger\,|x\rangle &= \tfrac{1}{2}\,\langle x|\,(U_1 - U_2)(U_1^\dagger - U_2^\dagger)\,|x\rangle \\ &= \tfrac{1}{2}(2 - 2\mathrm{Re}\{\langle x|\,U_1U_2^\dagger\,|x\rangle\}) = 1 - \mathrm{Re}\{\langle x|\,U_1U_2^\dagger\,|x\rangle\}.\end{aligned}$$

Similarly, we can generalize such result to mixed quantum operations $\tilde{U} = \sum_{\kappa=1}^{K} \alpha_\kappa U_\kappa$, where $\sum_{\kappa=1}^{K} |\alpha_\kappa| \leqslant 1$,

$$\begin{aligned}\langle x|\,\tilde{U}\tilde{U}^\dagger\,|x\rangle &= \langle x|\sum_{\kappa_1,\kappa_2=1}^{K}\alpha_{\kappa_1}\alpha_{\kappa 2}^{*}U_{\kappa_1}U_{\kappa_2}^\dagger\,|x\rangle \\ &= \sum_{\kappa=1}^{K}|\alpha_\kappa|^2 + \sum_{\kappa_1<\kappa_2} 2\mathrm{Re}\{\alpha_{\kappa_1}\alpha_{\kappa 2}^{*}\,\langle x|\,U_{\kappa_1}U_{\kappa_2}^\dagger\,|x\rangle\} \\ &= \sum_{\kappa=1}^{K}|\alpha_\kappa|^2 + \sum_{\kappa_1<\kappa_2} 2\mathrm{Re}\{\alpha_{\kappa_1}\alpha_{\kappa 2}^{*}\}\mathrm{Re}\{\langle x|\,U_{\kappa_1}U_{\kappa_2}^\dagger\,|x\rangle\} \\ &\quad - \sum_{\kappa_1<\kappa_2} 2\mathrm{Im}\{\alpha_{\kappa_1}\alpha_{\kappa 2}^{*}\}\mathrm{Im}\{\langle x|\,U_{\kappa_1}U_{\kappa_2}^\dagger\,|x\rangle\}.\end{aligned} \quad (4)$$

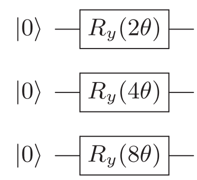
Since $U_{\kappa_1}, U_{\kappa_2}$ are quantum unitaries, the adjoint of them can be efficiently constructed by reversing the order of the gates.

We show a simple construction of the quantum sampling circuit $S(\theta)$ (illustrated in Table IV). Let n be the number of qubits, we define a geometric sequence $(\omega)_{i=1}^{n}$ with $w_i = 2^i$.

$$S(\theta) = \bigotimes_{i=1}^{n} R_y(2\omega_i\theta). \quad (5)$$

Note that $S(\theta)\,|\mathbf{0}\rangle$ creates a quantum state $|x(\theta)\rangle$ with a state vector matching the one defined in Lemma 1.

TABLE IV. $S(\theta)$: sampling circuit for three qubits.



To apply the Hadamard test to evaluate $\mathrm{Re}\{\alpha_{\kappa_1}\alpha_{\kappa 2}^{*}\,\langle x(\theta)|\,U_{\kappa_1}U_{\kappa_2}^\dagger\,|x(\theta)\rangle\}$, measurement circuits such as which in Table V are used.

Lemma 2. $O(\frac{2\ln(2/\delta)}{\epsilon^2})$ measurements suffice to bound the error from a Hadamard test to be within ϵ with probability $1-\delta$.

Proof. Assume m measurements are performed on the control bit of the Hadamard test with outcomes $M_1, M_2, \ldots, M_m \in \{0, 1\}$, we can approximate Pr(1) using $\widehat{\Pr(1)} = \frac{1}{m}\sum_{i=1}^{m} M_i$. Applying the Chernoff-Hoeffding bound

$$\begin{aligned}&\mathbb{P}(|[1 - 2\widehat{\Pr(1)}] - [1 - 2\Pr(1)]| > \epsilon) \\ &\quad = \mathbb{P}\left(|\widehat{\Pr(1)} - \Pr(1)| > \frac{\epsilon}{2}\right) < 2e^{-\epsilon^2 m/2}.\end{aligned}$$

When $m \geqslant \frac{2\ln(2/\delta)}{\epsilon^2}$, $2e^{-\epsilon^2 m/2} \leqslant \delta$, and this completes the proof. ■

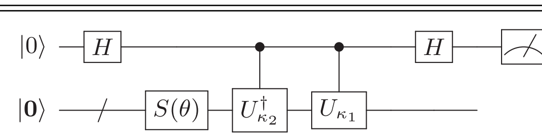
Lemma 3. Given an arbitrary pure state $|x\rangle \in \partial B_{\mathbb{C}^N}(0, 1)$ and a mixed quantum operation $\tilde{U} = \sum_{\kappa=1}^{K}\alpha_\kappa U_\kappa$ with $\sum_{\kappa=1}^{K}|\alpha_\kappa| \leqslant 1$ and $K \sim O(1)$, $O(\frac{32K^4\ln(4K^2/\delta)}{\epsilon^2})$ measurements suffice to estimate $\langle x|\,\tilde{U}\tilde{U}^\dagger\,|x\rangle$ to an error within ϵ with probability at least $1-\delta$.

Proof. According to Eq. (4), $\langle x|\,\tilde{U}\tilde{U}^\dagger\,|x\rangle$ can be estimated with a summation of $O(2K^2)$ measurements from the Hadamard tests. By the assumption $\sum_{\kappa=1}^{K}|\alpha_\kappa| \leqslant 1$, $2\mathrm{Re}\{\alpha_{\kappa_1}\alpha_{\kappa 2}^{*}\}$ and $2\mathrm{Im}\{\alpha_{\kappa_1}\alpha_{\kappa 2}^{*}\}$ are both $\leqslant 2$. Applying the triangle inequality and the union bound, it suffices to bound the error of each Hadamard test to be within $\frac{\epsilon}{4K^2}$ with probability at least $1 - \frac{\delta}{2K^2}$. We then apply Lemma 2 and obtain an upper bound on the sample complexity, $m = O(\frac{32K^4\ln(4K^2/\delta)}{\epsilon^2})$.

With the assumption $K \sim O(1)$, Lemma 3 implies a low-order polynomial measurement complexity to estimate the measurement outcome of the Hadamard test to a marginal error. Thus, we make an assumption that *all measurements are error-free* from now on.

For an arbitrary mixed quantum operation $\tilde{U}$, we can approximate $\|\tilde{U}\|_{S_2}$ using m random samples. Namely, we randomly sample $\theta_1, \ldots, \theta_m \sim \mathcal{D} = \mathrm{Uni}[-\pi, \pi]$ and use the

TABLE V. Circuit for estimating $\mathrm{Re}\{\langle\mathbf{0}|\,S^\dagger(\theta_i)U_{\kappa_1}U_{\kappa_2}^\dagger S(\theta_i)\,|\mathbf{0}\rangle\}$ using Hadamard test as in Table II.



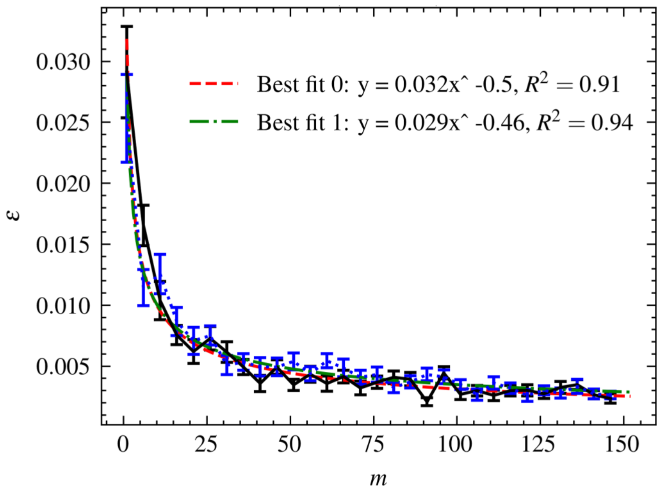


FIG. 2. For each randomly generated mixed quantum operations $\tilde{U} = \frac{1}{\sqrt{2}}U_1 - \frac{1}{\sqrt{2}}U_2$ where U_1, U_2 are randomly generated using the QR decomposition [73], we plot the error of the approximation $\widehat{\|\tilde{U}\|_{S_2}}$ with respect to the sample size m as defined in Theorem 3. A six-qubits system is considered. The means and standard errors ($\pm\sigma$ error bars) are computed over 30 sets of random samples.

sampling circuit S as defined by Eq. (5) to approximate

$$\widehat{\|\tilde{U}\|_{S_2}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\langle\mathbf{0}|\,S^\dagger(\theta_i)\tilde{U}\tilde{U}^\dagger S(\theta_i)\,|\mathbf{0}\rangle}, \tag{6}$$

where each $\langle\mathbf{0}|\,S^\dagger(\theta_i)\tilde{U}\tilde{U}^\dagger S(\theta_i)\,|\mathbf{0}\rangle$ can be measured with negligible error by Lemma 3.

Theorem 3. For an arbitrary mixed quantum operation $\tilde{U}$, we can estimate its normalized Schatten 2-norm efficiently using quantum sampling circuits of depth overhead $O(1)$. Moreover, for any δ, $\epsilon > 0$, with sample complexity $m(\epsilon,\delta) = O(\frac{\ln(2/\delta)}{2\epsilon^2}\min\{\epsilon^{-2}, \|\tilde{U}\|_{S_2}^{-2}\})$, samples $\theta_1\ldots,\theta_{m(\epsilon,\delta)} \sim \mathrm{Uni}[-\pi,\pi]$, and $S(\theta_i)$ as defined by Eq. (5), the following holds for the quantum approximation of $\|\tilde{U}\|_{S_2}$ using Eq. (6):

$$\mathbb{P}(|\widehat{\|\tilde{U}\|_{S_2}} - \|\tilde{U}\|_{S_2}| < \epsilon) > 1-\delta.$$

Proof. The theorem follows from Theorem 2, Table V, and Lemma 3. ∎

Note that the sample complexity for the approximation of the normalized Schatten 2-norm is independent of the number of qubits and is polynomial to $\frac{1}{\epsilon}$, which implies a potential quantum advantage. In Fig. 2, we present simulated results supporting the relation $\epsilon \propto m(\epsilon,\delta)^{-1/2}$ which is in agreement with Theorem 3. In the next section, we build a connection from the normalized Schatten 2-norm of the difference of quantum operations to the similarity metric defined in Sec. II.

V. FROM THE NORMALIZED SCHATTEN 2-NORM TO A FIDELITY-BASED SIMILARITY METRIC

Recall the definition of pure-state (ϵ,δ)-similarity. Let $|\psi\rangle$ be a random state sampled from the distribution $\mathcal{J} = \mathrm{Uni}[\partial B_{\mathbb{C}^N}(0,1)]$, we define two unitary operations U_1, U_2 to be pure-state (δ,ϵ)-similar if

$$\mathbb{P}_{\psi\sim\mathcal{J}}(\mathcal{F}(U_1|\psi\rangle, U_2|\psi\rangle) \geqslant 1-\epsilon) \geqslant 1-\delta.$$

The following lemma is significant as it relates pure-state (ϵ,δ)-similarity to the normalized Schatten 2-norm.

Lemma 4. Let U_1, U_2 be two unitary quantum operations. U_1, U_2 are pure-state (ϵ,δ)-similar if $\|U_1 - U_2\|_{S_2} \leqslant \frac{\epsilon}{1+\sqrt{2(1/\delta-1)}}$.

Proof. We apply statistical analysis to study the expectation and the variance of $\mathcal{F}(U_1|\psi\rangle, U_2|\psi\rangle)$ when $\psi\sim\mathcal{J} = \mathrm{Uni}[\partial B_{\mathbb{C}^N}(0,1)]$. For any pure state $|\psi\rangle$,

$$\begin{aligned}\mathcal{F}(U_1|\psi\rangle, U_2|\psi\rangle) &= |\langle\psi|\,U_1^\dagger U_2\,|\psi\rangle|^2\\ &\geqslant \mathrm{Re}^2\{\langle\psi|\,U_1^\dagger U_2\,|\psi\rangle\} = \left(1 - \frac{\sum_{i=1}^{N}\sigma_i{}^2|\langle w_i|\,|\psi\rangle|^2}{2}\right)^2\\ &\geqslant 1 - \sum_{i=1}^{N}\sigma_i{}^2|\langle w_i|\,|\psi\rangle|^2,\end{aligned}$$

where w_i and σ_i are the left-singular vectors and singular values of $U_1 - U_2$. Let $\hat{\epsilon} = \|U_1 - U_2\|_{S_2}$ and $\psi\sim\mathcal{J}$,

$$\mathbb{E}_{\psi\sim\mathcal{J}}\mathcal{F}(U_1|\psi\rangle, U_2|\psi\rangle) \geqslant 1 - \|U_1-U_2\|_{S_2}^2 \geqslant 1-\hat{\epsilon}^2.$$

We next compute the variance of the fidelity

$$\begin{aligned}&\mathrm{Var}_{\psi\sim\mathcal{J}}[\mathcal{F}(U_1|\psi\rangle, U_2|\psi\rangle)]\\ &\quad= \mathbb{E}_\psi\mathcal{F}^2(U_1|\psi\rangle, U_2|\psi\rangle)\\ &\qquad- (\mathbb{E}_\psi\mathcal{F}(U_1|\psi\rangle, U_2|\psi\rangle))^2 \leqslant 1-(1-\hat{\epsilon}^2)^2 \leqslant 2\hat{\epsilon}^2.\end{aligned}$$

Application of the Chebyshev-Cantelli inequality, for arbitrary $c > 0$,

$$\mathbb{P}_{\psi\sim\mathcal{J}}(\mathcal{F}(U_1|\psi\rangle, U_2|\psi\rangle) \geqslant 1-c) \geqslant 1 - \frac{2\hat{\epsilon}^2}{(c-\hat{\epsilon}^2)^2+2\hat{\epsilon}^2}.$$

Setting $c=\epsilon$ and $\frac{2\hat{\epsilon}^2}{(c-\hat{\epsilon}^2)^2+2\hat{\epsilon}^2} = \delta$, it suffices to have $\hat{\epsilon} \leqslant \frac{\epsilon}{1+\sqrt{2(1/\delta-1)}}$. ∎

Empirical relations between $\mathcal{F}(U_1|\psi\rangle, U_2|\psi\rangle)$ and $\|U_1 - U_2\|_{S_2}$ are illustrated in Figs. 3 and 4. In both experiments, U_1 is a fixed random unitary generated by QR decomposition [73] and U_2 is constructed by applying rotation operators to U_1. Figure 4 supports the bound derived in Lemma 4 as the probability for U_1, U_2 to be pure-state $[(1+\sqrt{8})\|U_1-U_2\|_{S_2}, 0.2]$-similar is much higher than 0.8 for all pairs of U_1 and U_2 used (setting $\delta = 0.2$).

The lemma can be generalized to mixed quantum operations. Let $\mathcal{J} = \mathrm{Uni}[\partial B_{\mathbb{C}^N}(0,1)]$, we define two mixed quantum operations $\tilde{U}_1, \tilde{U}_2$ to be pure-state (ϵ,δ)-similar if

$$\mathbb{P}_{\psi\sim\mathcal{J}}\left(\mathcal{F}(\tilde{U}_1|\psi\rangle, \tilde{U}_2|\psi\rangle) \geqslant \mathbb{E}_\psi^2\frac{\langle\psi|\,\tilde{U}_1\tilde{U}_1^\dagger + \tilde{U}_2\tilde{U}_2^\dagger\,|\psi\rangle}{2} - \epsilon\right) \geqslant 1-\delta.$$

Lemma 5. Let $\tilde{U}_1, \tilde{U}_2$ be two mixed quantum operations and $\mathcal{J} = \mathrm{Uni}[\partial B_{\mathbb{C}^N}(0,1)]$. $\tilde{U}_1$, $\tilde{U}_2$ are pure-state

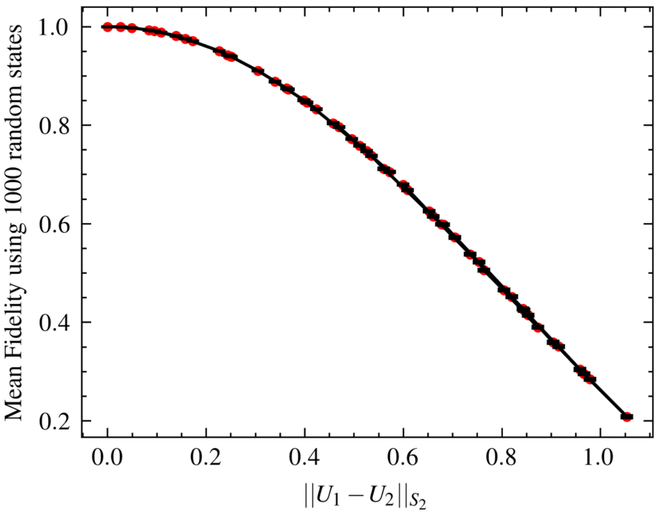


FIG. 3. Empirical relation between approximated mean $\mathbb{E}_{\psi\sim \mathrm{Uni}[\partial B_{\mathbb{C}^N}(0,1)]}\mathcal{F}(U_1\,|\psi\rangle\,,U_2\,|\psi\rangle)$ and $\|U_1-U_2\|_{S_2}$. For each pair of U_1 and U_2, the mean fidelity is computed over 1000 randomly sampled (with replacement) states $\psi\sim\mathrm{Uni}[\partial B_{\mathbb{C}^N}(0,1)]$ and error bars shown ($\pm\sigma$) are smaller than the symbols. A six-qubit system is considered.

(ϵ,δ)-similar if $\|\tilde U_1-\tilde U_2\|_{S_2}\leqslant\sqrt{\frac{\epsilon^2-(1/\delta-1)(\tau-\tau^4)}{2\tau[\epsilon+(1/\delta-1)\tau^2]}}$, where $\tau=\mathbb{E}_{\psi\sim\mathcal{J}}\frac{\langle\psi|\tilde U_1\tilde U_1^\dagger+\tilde U_2\tilde U_2^\dagger|\psi\rangle}{2}$.

Proof. For any given pure state $|\psi\rangle\in\partial B_{\mathbb{C}^N}(0,1)$,

$$\begin{aligned}\mathcal{F}(\tilde U_1\,|\psi\rangle\,,\tilde U_2\,|\psi\rangle)&\geqslant \mathrm{Re}^2\{\langle\psi|\,\tilde U_1^\dagger\tilde U_2\,|\psi\rangle\}\\&=\bigg(\frac{\langle\psi|\,\tilde U_1\tilde U_1^\dagger\,|\psi\rangle+\langle\psi|\,\tilde U_2\tilde U_2^\dagger\,|\psi\rangle}{2}\\&\quad-\frac{\sum_{i=1}^N{\sigma_i}^2|\langle w_i|\,|\psi\rangle|^2}{2}\bigg)^2,\end{aligned}$$

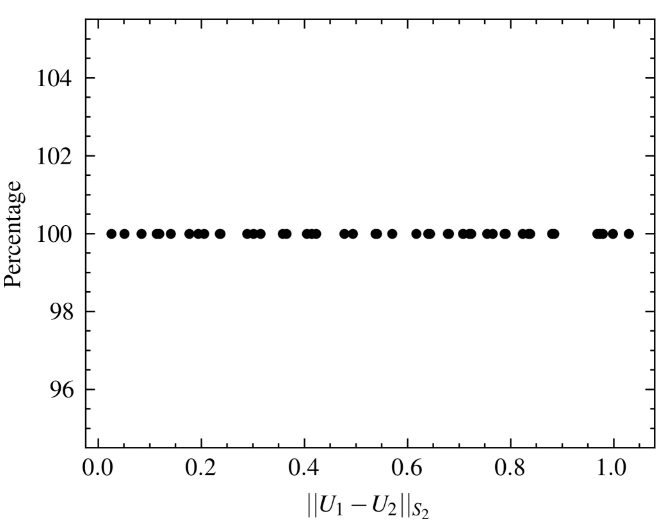


FIG. 4. Verification of Lemma 4 when $\delta=0.2$. For 48 pairs of random U_1 and U_2, we plot the percentage of 1000 randomly sampled (with replacement) $\psi\sim\mathrm{Uni}[\partial B_{\mathbb{C}^N}(0,1)]$ which satisfy $\mathcal{F}(U_1\psi,U_2\psi)\geqslant 1-(1+\sqrt{8.0})\|U_1-U_2\|_{S_2}$. A six-qubits system is considered.

where w_i and σ_i are the left-singular vectors and singular values of $\tilde U_1-\tilde U_2$. Let $\tau=\mathbb{E}_{\psi\sim\mathcal{J}}\frac{\langle\psi|\tilde U_1\tilde U_1^\dagger|\psi\rangle+\langle\psi|\tilde U_2\tilde U_2^\dagger|\psi\rangle}{2}$ and $\hat\epsilon=\|\tilde U_1-\tilde U_2\|_{S_2}$,

$$\begin{aligned}&\mathbb{E}_{\psi\sim\mathcal{J}}\mathcal{F}(\tilde U_1\,|\psi\rangle\,,\tilde U_2\,|\psi\rangle)\\&\quad\geqslant\mathbb{E}_\psi\Big(\frac{\langle\psi|\,\tilde U_1\tilde U_1^\dagger\,|\psi\rangle+\langle\psi|\,\tilde U_2\tilde U_2^\dagger\,|\psi\rangle}{2}\Big)^2-\tau\hat\epsilon^2\\&\quad\geqslant\tau^2-\tau\hat\epsilon^2.\\&\mathrm{Var}_{\psi\sim\mathcal{J}}[\mathcal{F}(\tilde U_1\,|\psi\rangle\,,\tilde U_2\,|\psi\rangle)]\\&\quad=\mathbb{E}_\psi\mathcal{F}^2(\tilde U_1\,|\psi\rangle\,,\tilde U_2\,|\psi\rangle)-(\mathbb{E}_\psi\mathcal{F}(\tilde U_1\,|\psi\rangle\,,\tilde U_2\,|\psi\rangle))^2\\&\quad\leqslant\tau-(\tau^2-\tau\hat\epsilon^2)^2\leqslant\tau-\tau^4+2\tau^3\hat\epsilon^2.\end{aligned}$$

Applying the Chebyshev-Cantelli inequality, for arbitrary $c>0$,

$$\begin{aligned}&\mathbb{P}_{\psi\sim\mathcal{J}}\big(\mathcal{F}(\tilde U_1\,|\psi\rangle\,,\tilde U_2\,|\psi\rangle)\geqslant\tau^2-c\big)\\&\quad\geqslant 1-\frac{\tau-\tau^4+2\tau^3\hat\epsilon^2}{(c-\tau\hat\epsilon^2)^2+\tau-\tau^4+2\tau^3\hat\epsilon^2}.\end{aligned}\qquad(7)$$

Setting $c=\epsilon$ and $\frac{\tau-\tau^4+2\tau^3\hat\epsilon^2}{(c-\tau\hat\epsilon^2)^2+\tau-\tau^4+2\tau^3\hat\epsilon^2}=\delta$, it suffices to have $\hat\epsilon\leqslant\sqrt{\frac{\epsilon^2-(1/\delta-1)(\tau-\tau^4)}{2\tau[\epsilon+(1/\delta-1)\tau^2]}}$. ■

Unlike for unitary quantum operations, the error bound for mixed quantum operations depends on τ, which incorporates the difference of $\tilde U_1$ and $\tilde U_2$ in "size." Combining Lemma 4 and Theorem 3, we obtain the following theorem.

Theorem 4. Consider arbitrary $\epsilon,\delta,\hat\delta>0$ and input unitary quantum operations U_1, U_2. Consider m independent samples $\theta_1,\dots,\theta_m\sim\mathrm{Uni}[-\pi,\pi]$. Let $S(\theta_i)$ be as defined by Eq. (5) and the corresponding quantum approximations be as defined by Eq. (6). Then $\mathbb{P}[U_1,U_2$ are (ϵ,δ)-similar$]\geqslant 1-\hat\delta$ if the following inequality holds:

$$\begin{aligned}&\|\widehat{U_1-U_2}\|_{S_2}+\min\{\sqrt[4]{\frac{2\ln(2/\hat\delta)}{m}},\sqrt{\frac{2\ln(2/\hat\delta)}{m}}\\&\|U_1-U_2\|_{S_2}^{-1}\}\leqslant\frac{\epsilon}{1+\sqrt{2(1/\delta-1)}}.\end{aligned}$$

Proof. Following from Lemma 4, it suffices to show that $\|U_1-U_2\|_{S_2}\leqslant\frac{\epsilon}{1+\sqrt{2(1/\delta-1)}}$ with probability at least $1-\hat\delta$. By Theorem 3, with m samples,

$$\begin{aligned}&\mathbb{P}\bigg(\frac{|\|\widehat{U_1-U_2}\|_{S_2}-\|U_1-U_2\|_{S_2}|}{\sqrt2}\leqslant\min\{\sqrt[4]{\frac{\ln(2/\hat\delta)}{2m}},\\&\sqrt{\frac{\ln(2/\hat\delta)}{2m}}\sqrt2\|U_1-U_2\|_{S_2}^{-1}\}\bigg)\geqslant 1-\hat\delta.\end{aligned}$$

TABLE VI. Sample based quantum circuit learning

Input: Target unitary V.

Output: ξ and $U(\xi)$ which approximates V.

1: $\Theta \leftarrow \emptyset$ $\triangleright \Theta$ saves the random samples
2: **for** $i : 1 \rightarrow m$ **do** $\triangleright$Generating m samples
3: $\theta_i \sim \text{Uniform}[-\pi, \pi]$ $\triangleright \mathcal{D} = \text{Uniform}[-\pi, \pi]$
4: $\Theta \leftarrow \Theta \cup \theta_i$
5: **endfor**
6: Randomly initialize ξ.
7: **loop**
8: $f(\xi) = 2 - \frac{1}{m}\sum_{i=1}^{m} 2\Re\{\langle \mathbf{0}| S^\dagger(\theta_i)V^\dagger U(\xi)S(\theta_i)|\mathbf{0}\rangle\}$ $\triangleright$Objective
9: $\xi \leftarrow \xi - \eta\nabla_\xi f(\xi)$ $\triangleright$Gradient Descent
10: **endloop**

For any x, x_0, c,

$$\mathbb{P}(|x - x_0| \leqslant c) = \mathbb{P}(x_0 - c \leqslant x \leqslant x_0 + c) \leqslant \mathbb{P}(x \leqslant x_0 + c).$$

We obtain

$$\mathbb{P}\left(\|U_1 - U_2\|_{S_2} \leqslant \|\widehat{U_1 - U_2}\|_{S_2} + \min\{\sqrt[4]{\frac{2\ln(2/\hat{\delta})}{m}}, \sqrt{\frac{2\ln(2/\hat{\delta})}{m}}\|U_1 - U_2\|_{S_2}^{-1}\} \right) \geqslant 1 - \hat{\delta}.$$

The main claim follows. ■

VI. APPLICATIONS TO QUANTUM CIRCUIT LEARNING

Quantum circuit learning is one of the most natural applications of the similarity metric. A problem setting is as follows: given a target unitary quantum operation V, represented via its clean-qubit controlled circuit, find a parameter set $\hat{\xi}$ of a variational circuit $U(\xi)$ that best approximates V. Theorem 4 inspires a circuit learning algorithm whose cost function $\|\widehat{U(\xi) - V}\|_{S_2}^2$ utilizes the normalized Schatten 2-norm of the difference between $U(\xi)$ and V (see Algorithm VI). We increase the similarity between $U(\xi)$ and V by minimizing the cost function.

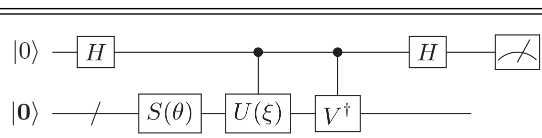
A quantum circuit diagram for approximating $\text{Re}\{\langle \mathbf{0}| S^\dagger(\theta_i)V^\dagger U(\xi)S(\theta_i)|\mathbf{0}\rangle\}$ is shown in Table VII.

To obtain an estimate of the gradient with respect to ξ_i, we could apply a black-box gradient approximation [10,74–77]

$$\frac{\partial f}{\partial \xi_i} \underset{\varepsilon \to 0}{\approx} \frac{f(\xi_i + \varepsilon) - f(\xi_i - \varepsilon)}{2\varepsilon}.$$

A better accuracy of the approximation can be achieved by increasing m, which is consistent with Theorem 4. One possible application of the algorithm is learning the square root of a quantum operation V, where we use $U(\xi)U(\xi)$ to approximate V.

TABLE VII. Circuit for three qubits: approximating $\text{Re}\{\langle \mathbf{0}| S^\dagger(\theta_i)V^\dagger U(\xi)S(\theta_i)|\mathbf{0}\rangle\}$



VII. CONCLUDING REMARKS

In summary, we defined and introduced the normalized Schatten norms and a set of similarity metrics between quantum operations that can be efficiently estimated. We discussed sufficient and necessary conditions for a sampling circuit to estimate the normalized Schatten 2-norm and showed one optimal design of such sampling circuits. We then studied the sample complexity required by the sampling circuit and obtained an upper bound that was polynomial to $\frac{1}{\epsilon}$. With such an efficient sampling method, we were able to estimate the normalized Schatten 2-norms of mixed quantum operations. We next related the similarity of quantum operations based on the normalized Schatten 2-norm to a similarity metric induced by the traditional fidelity metric used for quantum states. Finally, we showed how such a connection could lead to a design of the loss function for circuit learning for tasks such as approximating a given quantum circuit or its square root.

In this paper we emphasized circuit learning applications to the problem of approximating unitary operations. We did not explore learning of mixed operations. A similar circuit learning approach would apply to mixed quantum operations with a corresponding modification of the loss term at the line 8 of Algorithm VI. However, as noted in Lemma 5, the error bound posed on the normalized Schatten 2-norm of the difference of the mixed operations is also a function of $\tau = \mathbb{E}_{\psi\sim\mathcal{J}} \frac{\langle\psi|\tilde{U}_1\tilde{U}_1^\dagger + \tilde{U}_2\tilde{U}_2^\dagger|\psi\rangle}{2}$. When τ is not close 1, there is a weaker correlation between the normalized Schatten 2-norm of the difference and the fidelity-based similarity metric.

APPENDIX: THEOREM 2 SUPPLEMENT

We complete the proof of theorem 2 by showing the following implication:

$$\mathbb{P}\left(\left|\frac{\widehat{\text{Tr}(AA^\dagger)}}{N} - \|A\|_{S_2}^2\right| < \epsilon \max\{\epsilon, \|A\|_{S_2}\}\right) > 1 - \delta$$
$$\Rightarrow \mathbb{P}(|\widehat{\|A\|}_{S_2} - \|A\|_{S_2}| < \epsilon) > 1 - \delta.$$

Proof. Based on our classical approximation algorithm defined by Eq. (3),

$$\widehat{\|A\|}_{S_2} = \sqrt{\frac{\widehat{\text{Tr}(AA^\dagger)}}{N}}.$$

Let $\mathcal{M} = \frac{\widehat{\text{Tr}(AA^\dagger)}}{N} = \widehat{\|A\|}_{S_2}^2$, we divide the proof into two cases.

If $\epsilon < \|A\|_{S_2}$,

$$|\mathcal{M} - \|A\|_{S_2}^2| \leqslant \epsilon\|A\|_{S_2}$$
$$\Rightarrow \|A\|_{S_2}^2 - \epsilon\|A\|_{S_2} \leqslant \mathcal{M} \leqslant \|A\|_{S_2}^2 + \epsilon\|A\|_{S_2}$$
$$\Rightarrow \|A\|_{S_2}^2 - 2\epsilon\|A\|_{S_2} + \epsilon^2 \leqslant \mathcal{M} \leqslant \|A\|_{S_2}^2 + 2\epsilon\|A\|_{S_2} + \epsilon^2$$
$$\Rightarrow (\|A\|_{S_2} - \epsilon)^2 \leqslant \mathcal{M} \leqslant (\|A\|_{S_2} + \epsilon)^2$$
$$\Rightarrow |\sqrt{\mathcal{M}} - \|A\|_{S_2}| \leqslant \epsilon.$$

If $\epsilon \geqslant \|A\|_{S_2}$, $|\mathcal{M} - \|A\|_{S_2}^2| \leqslant \epsilon^2$, which leads to the following series of implications:

$$\Rightarrow \|A\|_{S_2}^2 - \epsilon^2 \leqslant \mathcal{M} \leqslant \|A\|_{S_2}^2 + \epsilon^2$$

$$\Rightarrow 0 \leqslant \mathcal{M} \leqslant \|A\|_{S_2}^2 + 2\epsilon\|A\|_{S_2} + \epsilon^2$$

$$\Rightarrow 0 \leqslant \mathcal{M} \leqslant (\|A\|_{S_2} + \epsilon)^2$$

$$\Rightarrow |\sqrt{\mathcal{M}} - \|A\|_{S_2}| \leqslant \epsilon.$$

[1] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[2] E. Farhi and A. W. Harrow, Quantum supremacy through the quantum approximate optimization algorithm, arXiv:1602.07674.

[3] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[4] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[5] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, Nat. Commun. **11**, 808 (2020).

[6] W. Li, Z. Lu, and D.-L. Deng, Quantum Neural Network Classifiers: A Tutorial, SciPost Phys. Lect. Notes 61 (2022).

[7] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, Nat. Phys. **15**, 1273 (2019).

[8] S. C. Kak, *Quantum Neural Computing* (Elsevier, Amsterdam, 1995), pp. 259–313.

[9] A. A. Ezhov and D. Ventura, Quantum neural networks, in *Future Directions for Intelligent Systems and Information Sciences: The Future of Speech and Image Technologies, Brain Computers, WWW, and Bioinformatics*, edited by N. Kasabov, (Physica-Verlag HD, Heidelberg, 2000), pp. 213–235.

[10] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98**, 032309 (2018).

[11] M. Panella and G. Martinelli, Neural networks with quantum architecture and quantum learning, Int. J. Circuit Theory Appl. **39**, 61 (2011).

[12] R. M. Gingrich and C. P. Williams, Non-unitary probabilistic quantum computing, in *Proceedings of the Winter International Synposium on Information and Communication Technologies*, WISICT '04 (Trinity College, Dublin, 2004), pp. 1–6.

[13] A. W. Schlimgen, K. Head-Marsden, L. M. Sager, P. Narang, and D. A. Mazziotti, Quantum Simulation of Open Quantum Systems Using a Unitary Decomposition of Operators, Phys. Rev. Lett. **127**, 270503 (2021).

[14] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum Algorithm for Linear Systems of Equations, Phys. Rev. Lett. **103**, 150502 (2009).

[15] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[16] M. Lubasch, J. Joo, P. Moinier, M. Kiffner, and D. Jaksch, Variational quantum algorithms for nonlinear problems, Phys. Rev. A **101**, 010301(R) (2020).

[17] H. Miyahara and V. Roychowdhury, Ansatz-independent variational quantum classifier, arXiv:2102.01759.

[18] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, England, 2010).

[19] R. Jozsa, Fidelity for mixed quantum states, J. Mod. Opt. **41**, 2315 (1994).

[20] L. H. Pedersen, N. M. Møller, and K. Mølmer, Fidelity of quantum operations, Phys. Lett. A **367**, 47 (2007).

[21] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, Discriminating States: The Quantum Chernoff Bound, Phys. Rev. Lett. **98**, 160501 (2007).

[22] C. Fuchs and J. van de Graaf, Cryptographic distinguishability measures for quantum-mechanical states, IEEE Trans. Inf. Theory **45**, 1216 (1999).

[23] M. Raginsky, A fidelity measure for quantum channels, Phys. Lett. A **290**, 11 (2001).

[24] D. Aharonov, A. Kitaev, and N. Nisan, Quantum circuits with mixed states, in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98 (Association for Computing Machinery, New York, 1998), pp. 20–30.

[25] J. Bae and L.-C. Kwek, Quantum state discrimination and its applications, J. Phys. A: Math. Theor. **48**, 083001 (2015).

[26] J. A. Bergou, Discrimination of quantum states, J. Mod. Opt. **57**, 160 (2010).

[27] W. Matthews, S. Wehner, and A. Winter, Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding, Commun. Math. Phys. **291**, 813 (2009).

[28] D. Markham, J. A. Miszczak, Z. Puchała, and K. Życzkowski, Quantum state discrimination: A geometric approach, Phys. Rev. A **77**, 042111 (2008).

[29] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, England, 2018).

[30] T. Patel, E. Younis, C. Iancu, W. de Jong, and D. Tiwari, Quest: Systematically approximating quantum circuits for higher output fidelity, in *Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems* (Association for Computing Machinery, New York, 2022), pp. 514–528.

[31] D. Camps and R. Van. Beeumen, Approximate quantum circuit synthesis using block encodings, Phys. Rev. A **102**, 052411 (2020).

[32] M. Amy, D. Maslov, M. Mosca, and M. Roetteler, A meet-in-the-middle algorithm for fast synthesis of depth-optimal quantum circuits, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **32**, 818 (2013).

[33] E. Younis, K. Sen, K. Yelick, and C. Iancu, Qfast: Conflating search and numerical optimization for scalable quantum circuit synthesis, in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, New York, 2021), pp. 232–243.

[34] Y. Li, R. Duan, and M. Ying, Local unambiguous discrimination with remaining entanglement, Phys. Rev. A **82**, 032339 (2010).
[35] R. Duan, Y. Feng, Z. Ji, and M. Ying, Distinguishing Arbitrary Multipartite Basis Unambiguously Using Local Operations and Classical Communication, Phys. Rev. Lett. **98**, 230502 (2007).
[36] X. Wu and R. Duan, Exact quantum search by parallel unitary discrimination schemes, Phys. Rev. A **78**, 012303 (2008).
[37] R. Duan, Y. Feng, Y. Xin, and M. Ying, Distinguishability of quantum states by separable operations, IEEE Trans. Inf. Theory **55**, 1320 (2009).
[38] R. Duan, Y. Feng, and M. Ying, Perfect Distinguishability of Quantum Operations, Phys. Rev. Lett. **103**, 210501 (2009).
[39] R. Duan, Y. Xin, and M. Ying, Locally indistinguishable subspaces spanned by three-qubit unextendible product bases, Phys. Rev. A **81**, 032329 (2010).
[40] C. Lu, J. Chen, and R. Duan, Optimal perfect distinguishability between unitaries and quantum operations, arXiv:1010.2298.
[41] C. Lu, J. Chen, and R. Duan, Some bounds on the minimum number of queries required for quantum channel perfect discrimination, Quantum Info. Comput. **12**, 138 (2012).
[42] N. Yu, R. Duan, and M. Ying, Four Locally Indistinguishable Ququad-Ququad Orthogonal Maximally Entangled States, Phys. Rev. Lett. **109**, 020506 (2012).
[43] N. Yu, R. Duan, and M. Ying, Distinguishability of quantum states by positive operator-valued measures with positive partial transpose, IEEE Trans. Inf. Theory **60**, 2069 (2014).
[44] R. Duan, C. Guo, C.-K. Li, and Y. Li, Parallel distinguishability of quantum operations, in *2016 IEEE International Symposium on Information Theory (ISIT)* (IEEE, New York, 2016) pp. 2259–2263.
[45] E. Chitambar, R. Duan, and M.-H. Hsieh, When do local operations and classical communication suffice for two-qubit state discrimination?, IEEE Trans. Inf. Theory **60**, 1549 (2014).
[46] J. Watrous, Distinguishing quantum operations having few kraus operators, Quantum Inf. Comput. **8**, (2007).
[47] M. Piani and J. Watrous, All Entangled States are Useful for Channel Discrimination, Phys. Rev. Lett. **102**, 250501 (2009).
[48] A. Shimbo, A. Soeda, and M. Murao, Equivalence determination of unitary operations, arXiv:1803.11414.
[49] A. Acín, Statistical Distinguishability between Unitary Operations, Phys. Rev. Lett. **87**, 177901 (2001).
[50] D. Yang, Distinguishability, Classical information of quantum operations, arXiv:quant-ph/0504073.
[51] M. F. Sacchi, Optimal discrimination of quantum operations, Phys. Rev. A **71**, 062340 (2005).
[52] G. Wang and M. Ying, Unambiguous discrimination among quantum operations, Phys. Rev. A **73**, 042301 (2006).
[53] R. Duan, Y. Feng, and M. Ying, Entanglement is Not Necessary for Perfect Discrimination between Unitary Operations, Phys. Rev. Lett. **98**, 100503 (2007).
[54] L. Li, A bound on discrimination of quantum operations, Applied Mechanics and Materials **556-562**, 4293 (2014).
[55] A. Chefles, A. Kitagawa, M. Takeoka, M. Sasaki, and J. Twamley, Unambiguous discrimination among oracle operators, J. Phys. A: Math. Theor. **40**, 10183 (2007).
[56] L.-J. Li, Ambiguous discrimination of general quantum operations, Commun. Theor. Phys. **62**, 813 (2014).
[57] J. Chen and M. Ying, Ancilla-assisted discrimination of quantum gates, Quantum Info. Comput. **10**, 160 (2010).
[58] B. Rosgen and J. Watrous, On the hardness of distinguishing mixed-state quantum computations, arXiv:cs/0407056.
[59] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, Phys. Rev. A **71**, 062310 (2005).
[60] D. Aharonov, A. Kitaev, and N. Nisan, Quantum circuits with mixed states, , in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC '98* (Association for Computing Machinery, New York, NY, USA, 1998), pp. 20–30.
[61] B. Regula, R. Takagi, and M. Gu, Operational applications of the diamond norm and related measures in quantifying the non-physicality of quantum maps, Quantum **5**, 522 (2021).
[62] A. Ben-Aroya and A. Ta-Shma, On the complexity of approximating the diamond norm, Quantum Info. Comput. **10**, 77 (2010).
[63] J. Watrous, Semidefinite programs for completely bounded norms, arXiv:quant-ph/9806029.
[64] D. Pé rez-García, M. M. Wolf, D. Petz, and M. B. Ruskai, Contractivity of positive and trace-preserving maps under lp norms, J. Math. Phys. **47**, 083506 (2006).
[65] A. Ben-Aroya, O. Regev, and R. de Wolf, A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs, in *2008 49th Annual IEEE Symposium on Foundations of Computer Science* (IEEE, New York, 2008).
[66] P. Hayden and A. Winter, Counterexamples to the maximal p-norm multiplicativity conjecture for all $p > 1$, Commun. Math. Phys. **284**, 263 (2008).
[67] P. W. Shor and S. P. Jordan, Estimating jones polynomials is a complete problem for one clean qubit, Quantum Info. Comput. **8**, 681 (2008).
[68] C. Cade and A. Montanaro, The quantum complexity of computing schatten p-norms, arXiv.1706.09279.
[69] D. Aharonov, V. Jones, and Z. Landau, A polynomial quantum algorithm for approximating the jones polynomial, in *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '06 (Association for Computing Machinery, New York, 2006) pp. 427–436.
[70] J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics*, 2nd ed., (Cambridge University Press, San Francisco, 2017).
[71] W. Hoeffding, Probability inequalities for sums of bounded random variables, in *The collected works of Wassily Hoeffding* (Springer, New York, 1994), pp. 409–426.
[72] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury, and F. Vatan, A new proof for the existence of mutually unbiased bases, Algorithmica **34**, 512 (2002).
[73] F. Mezzadri, How to generate random matrices from the classical compact groups, Notices of the AMS **54**, 592 (2007).
[74] N. Yamamoto, On the natural gradient for variational quantum eigensolver, arXiv.1909.05074.
[75] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, Quantum **4**, 269 (2020).
[76] A. W. Harrow and J. C. Napp, Low-Depth Gradient Measurements Can Improve Convergence in Variational Hybrid Quantum-Classical Algorithms, Phys. Rev. Lett. **126**, 140502 (2021).
[77] G. E. Crooks, Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition, arXiv.1905.13311.