

# Variational quantum algorithms to estimate rank, quantum entropies, fidelity, and Fisher information via purity minimization

Kok Chuan Tan<sup>1,\*</sup> and Tyler Volkoff<sup>2,†</sup>

<sup>1</sup>*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Republic of Singapore*

<sup>2</sup>*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*



(Received 15 April 2021; accepted 4 August 2021; published 15 September 2021)

Variational quantum algorithms (VQAs) that estimate values of widely used physical quantities such as the rank, the quantum entropies, the Bures fidelity, and the quantum Fisher information of mixed quantum states are developed. In addition, variations of these VQAs are also adapted to perform other useful functions such as quantum state learning and approximate fractional inverses. The common theme shared by the proposed algorithms is that their cost functions are all based on minimizing the quantum purity of a quantum state. Strategies to mitigate or avoid the problem of exponentially vanishing cost function gradients are also discussed.

DOI: [10.1103/PhysRevResearch.3.033251](https://doi.org/10.1103/PhysRevResearch.3.033251)

## I. INTRODUCTION

In recent years, there has been growing interest and rapid developments in a class of early quantum computing devices collectively referred to as noisy intermediate-scale quantum (NISQ) devices [1]. NISQ devices are essentially quantum computing hardware that lacks full quantum error correction. Due to the lack of error correction, the gate fidelities and total coherence times available on such devices are limited. This means that NISQ devices are limited to applications that can be performed using short-depth quantum circuits, beyond which the measurement output is unreliable. While NISQ devices cannot perform universal quantum computing, it is widely expected that they are sufficient to provide a demonstrable computational advantage over classical computers in the near term [2,3].

One approach to developing algorithms suitable for NISQ devices is to consider hybrid quantum-classical algorithms [4]. Such algorithms seek to lower the overall quantum circuit depth requirements by offloading a portion of the computation to a classical computer. Among these, a class of algorithms called variational quantum algorithms (VQAs) are arguably among the most widely used and promising strategies for designing a NISQ algorithm. In VQAs, a cost function  $C(\theta)$  which is difficult to compute classically is estimated using NISQ hardware, while a classical optimization strategy is used to find the optimal parameter  $\theta^*$  that minimizes the cost function. Such strategies have been used in applications such as finding approximate eigenstates of Hamiltonians [5,6], recompiling quantum circuits [7], and solving linear algebraic problems [8–10]. As a class of quantum algorithms, VQAs

will likely continue to be relevant even when fully quantum error-corrected devices are widely available, as the solution to many classical and quantum problems is naturally phrased in terms of the minimization of some cost function.

In this paper, we develop VQAs to estimate a variety of physical quantities such as the rank of a quantum state, the Rényi [11] and Tsallis [12] quantum entropies, the quantum fidelity [13], and the quantum Fisher information [14]. Variations of these algorithms are also adapted to perform tasks such as quantum state learning [15,16] and approximating fractional powers and inverses of density matrices. The cost functions of these VQAs are based on minimizing the quantum purity of the system, defined as the quantity  $\text{tr}(\sigma^2)$  for any normalized density matrix  $\sigma$ . The quantum purity has previously been considered as a quantum resource [17] and is known to bound the maximum entanglement and quantum coherence generated by a quantum circuit [18]. The algorithms developed here suggest that the quantum purity is also an operationally useful concept in the sense that it can be a useful primitive for constructing cost functions for a variety of VQAs.

The Barren Plateau Landscape (BPL) problem [19] in relation to our proposed VQAs is also discussed. The BPL problem is said to occur when the expected magnitude of the gradient of the cost function  $C(\theta)$  vanishes exponentially with system size, which in turn suggests that exponential resources are required to optimize the parameters. We provide numerical evidence that the BPL problem can be avoided with sufficiently good *Ansatz* design for low-rank input states and also discuss possible strategies to mitigate the BPL problem by adding local terms to the cost function or by adopting correlated *Ansatz* parameters.

## II. PRELIMINARY EQUATIONS: QUANTUM SWAP TEST

The VQAs that we introduce in Secs. III and IV utilize a single algorithmic primitive to calculate the cost function, viz., computation of the trace of a product of quantum states. Although our VQAs are agnostic to the details of how this

\*[bbtankc@gmail.com](mailto:bbtankc@gmail.com)

†[volkoff@lanl.gov](mailto:volkoff@lanl.gov)

computation is carried out, here we briefly describe a method based on the quantum SWAP test [20], which can be generalized for  $k$  systems [21,22]. Given  $k$  input states  $\rho_i$ , where  $i = 1, \dots, k$ , the SWAP test evaluates the value of  $\text{tr}(\prod_{i=1}^k \rho_i)$ .

Let us first consider the pure input state  $|\psi_1\rangle \cdots |\psi_k\rangle$ . In this case, the goal is to evaluate  $\text{tr}(\prod_{i=1}^k |\psi_i\rangle\langle\psi_i|) = \langle\psi_k|\psi_1\rangle\langle\psi_1|\psi_2\rangle \cdots \langle\psi_{k-1}|\psi_k\rangle$ . We define the cyclic permutation operator  $P_k$  acting on  $k$  systems, which performs the operation

$$P_k |\psi_1\rangle \cdots |\psi_k\rangle = |\psi_k\rangle |\psi_1\rangle \cdots |\psi_{k-1}\rangle. \quad (1)$$

One may verify that the expectation value of  $P_k$  for the input state  $|\psi_1\rangle \cdots |\psi_k\rangle$  gives

$$\langle P_k \rangle = \langle\psi_k|\psi_1\rangle\langle\psi_1|\psi_2\rangle \cdots \langle\psi_{k-1}|\psi_k\rangle, \quad (2)$$

which is the required expression.

In order to evaluate this expectation value, we append an ancilla initialized in the state  $|+\rangle$ , where  $|\pm\rangle := (|0\rangle \pm |1\rangle)/\sqrt{2}$ . We then apply a controlled unitary  $U$  which performs  $P_k$  when the ancilla is in state  $|1\rangle$  and the identity operation  $\mathbb{1}_d$  when it is in state  $|0\rangle$ . Finally, we measure the ancilla qubit by projecting it onto the  $P_\pm := |\pm\rangle\langle\pm|$  basis. The corresponding probabilities are given by

$$\text{Prob}(\pm) = \|(\mathbb{1}_d \pm P_k)|\psi_1\rangle \cdots |\psi_k\rangle\|^2/4 \quad (3)$$

$$= \frac{1 \pm \langle P_k \rangle}{2}, \quad (4)$$

such that  $\langle P_k \rangle = \text{Prob}(+) - \text{Prob}(-)$ . This provides a resource-efficient method to evaluate  $\langle P_k \rangle$ .

We now consider mixed quantum states. Suppose that the input state has the form  $\rho_1 \otimes \cdots \otimes \rho_k$ , where  $\rho_i, i = 1, \dots, k$  may be mixed states in general. Let  $\{p_{i,j_i}, |e_{i,j_i}\rangle\}$  denote the  $j$ th eigenvalues and eigenstates of  $\rho_i$ . We can then write

$$\begin{aligned} \text{tr}\left(\prod_{i=1}^k \rho_i\right) &= \sum_{j_1 \cdots j_k} p_{1,j_1} \cdots p_{k,j_k} \\ &\times \text{tr}(P_k |e_{1,j_1}\rangle\langle e_{1,j_1}| \otimes \cdots \otimes |e_{k,j_k}\rangle\langle e_{k,j_k}|), \end{aligned} \quad (5)$$

which is a linear sum of expectation values of  $P_k$  acting on pure states. The SWAP test therefore also applies to mixed states.

### III. VARIATIONAL ALGORITHMS VIA PURITY MINIMIZATION

We utilize the quantum purity as a basic building block to construct efficiently computable cost functions for our VQAs. First, note that the purity is only nontrivial when optimizing over mixed quantum states, since it is always equal to unity over pure states. As such, a basic assumption we will make in the ensuing discussion is that the input state of the algorithm  $\rho$  has  $\text{rank}(\rho) \geq 2$ , which implies that  $\rho$  is a mixed quantum state.

The following is the basic cost function that we will consider:

$$C(\theta) = \text{tr}\{[\eta(\theta)^k \rho \eta(\theta)^k / \text{tr}(\eta(\theta)^k \rho \eta(\theta)^k)]^2\}, \quad (6)$$

where  $k$  is any positive integer,  $\rho$  is the density matrix of a possibly un-normalized mixed state, and  $\eta(\theta)$  is some normal-

ized *Ansatz* state. The goal is to minimize the cost function  $C(\theta)$  by varying  $\theta$  in order to find the optimal *Ansatz*  $\eta(\theta^*)$ . If we let  $\sigma = \eta(\theta)^k \rho \eta(\theta)^k / \text{tr}(\eta(\theta)^k \rho \eta(\theta)^k)$ , we see that  $\sigma$  is a normalized density matrix and the optimization reduces to finding  $\min_\sigma \text{tr}(\sigma^2)$ , i.e., it is minimizing the purity of  $\sigma$ . The purity is minimized by the unique solution  $\sigma^* = \mathbb{1}_d/d$ , where  $d \geq 2$  is the rank of the input matrix  $\rho$  and  $\mathbb{1}_d$  is the identity of the subspace spanned by the support of  $\rho$ . As a result, minimizing the cost function is equivalent to finding the optimal state  $\eta(\theta^*) = \rho^{-1/(2k)} / \text{tr}(\rho^{-1/(2k)})$ , given that the parametrized *Ansatz*  $\eta(\theta)$  includes this state. This can be verified by substituting this expression for  $\eta(\theta^*)$  into Eq. (6) to obtain  $C(\theta^*) = \text{tr}[\mathbb{1}_d / \text{tr}(\mathbb{1}_d)^2] = 1/d$ .

We need to further verify that the cost function  $C(\theta)$  is efficiently computable on a quantum computer. Evaluating  $C(\theta)$  requires one to compute the quantities (i)  $\text{tr}\{[\eta(\theta)^k \rho \eta(\theta)^k]^2\}$  and (ii)  $\text{tr}[\eta(\theta)^k \rho \eta(\theta)^k] = \text{tr}[\rho \eta(\theta)^{2k}]$ . Both quantities (i) and (ii) may be efficiently computed via the SWAP test using (i) two copies of  $\rho$  and  $4k$  copies of  $\eta(\theta)$  and (ii) a single copy of  $\rho$  and  $2k$  copies of  $\eta(\theta)$ , respectively. In general, the controlled cyclic permutation operator  $P_k$  from Eq. (1) acting on  $k$  subsystems can be decomposed into a series of  $(k-1)$  controlled SWAP operations. Assuming that each of the  $k$  subsystems is composed of  $n = \log_2 d$  qubits, each of the  $k$  controlled SWAPs on the level of subsystems can be implemented using  $n$  controlled SWAP operations on the level of individual qubits. Combining these two facts, the circuit complexity of the SWAP test is  $O(k \log_2 d)$ , or just  $O(n)$  assuming the value of  $k$  is fixed. This shows that the cost function  $C(\theta)$  can be efficiently sampled in linear time.

We also observe that one may estimate the normalization factor  $\text{tr}(\rho^{-1/(2k)})$  by utilizing the optimal state via the expression  $\text{tr}[\rho \eta(\theta^*)^{2k}] = d / \text{tr}(\rho^{-1/(2k)})^{2k}$ , which results in the identity

$$\begin{aligned} \text{tr}(\rho^{-1/(2k)})^{2k} &= \{\text{tr}[\rho \eta(\theta^*)^{2k}] / d\}^{-1} \\ &= \{\text{tr}[\rho \eta(\theta^*)^{2k}] C(\theta^*)\}^{-1}. \end{aligned} \quad (7)$$

Away from the critical point  $\theta^*$ , the product  $\{\text{tr}[\rho \eta(\theta)^{2k}] C(\theta)\}^{-1}$  provides an approximation of the normalization factor.

#### A. Geometric interpretation of purity minimization

We highlight that quantum purity has a direct geometric interpretation. Let us consider the Hilbert-Schmidt norm  $\|A\| := \sqrt{\text{tr}(A^\dagger A)}$  and the corresponding norm-induced distance  $d(A, B) := \|A - B\|$ . We compute the square distance between the maximally mixed state and a normalized quantum state  $\sigma$ :

$$d\left(\frac{\mathbb{1}_d}{d}, \sigma\right)^2 = \text{tr}\left[\left(\frac{\mathbb{1}_d}{d} - \sigma\right)^2\right] \quad (8)$$

$$= \text{tr}\left(\frac{\mathbb{1}_d}{d^2} + \sigma^2 - 2\frac{\sigma}{d}\right) \quad (9)$$

$$= \text{tr}(\sigma^2) - \frac{1}{d}. \quad (10)$$

From this, it is observed that the quantum purity directly quantifies how far away  $\sigma$  is from the maximally mixed

state, as quantified by the Hilbert-Schmidt distance  $d(A, B)$ . By defining  $\sigma(\theta) := \eta(\theta)^k \rho \eta(\theta)^k / \text{tr}[\eta(\theta)^k \rho \eta(\theta)^k]$ , we have  $\text{tr}[\sigma(\theta)^2] = C(\theta)$ , the cost function in Eq. (6). Minimizing  $C(\theta)$  therefore has a direct interpretation of finding the geometrically closest state  $\sigma(\theta)$  on the *Ansatz* manifold from the the maximally mixed state  $\mathbb{1}_d/d$ .

#### IV. APPLICATIONS OF PURITY MINIMIZATION

It turns out that purity minimization is a useful primitive to solve a range of problems that are relevant in quantum physics. We will demonstrate this by describing several possible applications based on the cost function described in Eq. (6).

##### A. Rank estimation

One immediate application is to estimate the rank of the input state  $\rho$ . This can be useful in, for instance, quantum state tomography, where efficient measurement strategies can be devised for low-rank states [23]. Purity optimization allows one to find a lower bound estimate of the rank. This comes directly from the fact that  $d = \text{rank}(\rho) \approx 1/C(\theta^*)$  and that in general for any  $\theta$ , we have  $d \geq 1/C(\theta)$ . For rank estimation, the size of the quantum circuit required scales with  $5n + 2$ , assuming that the input states are given and the dimension of  $\rho$  is  $d = 2^n$ . The overall quantum circuit complexity is  $O(n)$ .

##### B. Approximating fractional powers and fractional inverses of density matrices

Suppose we would like to raise the power of the density matrix to some power  $\rho^\alpha$ , for some real value  $\alpha \in [-1, 1]$ . We show that this state can be approximately prepared via purity minimization.

We assume that  $\alpha$  is approximated by some rational number  $\pm p/q$ , where  $p, q$  are both positive integers and  $p \leq q$ . The case where  $\alpha = 1$  is trivial since we assumed that  $\rho$  is a given input state. The other trivial case is when  $\alpha = 0$ , where the solution is just the maximally mixed state regardless of the input state  $\rho$ .

For the case  $\alpha = -1$ , we note that this is approximated by minimizing the purity of the (un-normalized) state  $\eta(\theta_1)\rho^2\eta(\theta_1)$ . This finds a solution  $\eta(\theta_1^*)$  such that  $\eta(\theta_1^*)\rho^2\eta(\theta_1^*)$  is proportional to the identity. This implies that  $\eta(\theta_1^*) = \rho^{-1}/\text{tr}(\rho^{-1})$ . Recall that the denominator can be estimated using Eq. (7).

Next, we consider the case where  $\alpha > 0$  and  $p < q$ . By choosing  $k = q$  in Eq. (6), we find a state  $\eta(\theta_1^*) \approx \rho^{-1/(2q)}/\text{tr}(\rho^{-1/(2q)})$ . The next step is to perform another round of purity minimization for the (un-normalized) input state  $\mu(\theta_2)\eta(\theta_1^*)^{4p}\mu(\theta_2)$ , which finds some optimal state  $\mu(\theta_2^*)$ . Since for a well-chosen *Ansatz*,  $\eta(\theta_1^*)^{4p}$  is approximately  $[\rho^{-1/(2q)}]^{4p} = \rho^{-(2p)/q}$  up to a normalization factor, we have  $\mu(\theta_2^*) \approx \rho^{p/q}/\text{tr}(\rho^{p/q})$  when  $\mu(\theta_2^*)\eta(\theta_1^*)^{4p}\mu(\theta_2^*)$  is proportional to the identity.

We now consider the case  $\alpha < 0$  and  $p < q$ . For this, we will use the state  $\mu(\theta_2^*)$  from the preceding paragraph as the input. We then perform another round of purity minimization for the (un-normalized) state  $\nu(\theta_3)\mu(\theta_2^*)^2\nu(\theta_3)$ . This finds the optimal state  $\nu(\theta_3^*)$  so that  $\nu(\theta_3^*)\mu(\theta_2^*)^2\nu(\theta_3^*)$  is proportional

to the identity, which implies  $\nu(\theta_3^*) \approx \rho^{-p/q}/\text{tr}(\rho^{-p/q})$ . This covers all the important cases and is sufficient to show that purity optimization is able to approximate the state  $\rho^\alpha$ , for any  $\alpha \in [-1, 1]$ .

For approximating fractional powers and fractional inverses, the size of the largest quantum circuit required scales with  $\max((8p + 3)n + 2, (2q + 1)n + 2)$ , assuming that the input states are given and the dimension of  $\rho$  is  $d = 2^n$ . The overall quantum complexity is therefore  $O(n)$  for fixed  $p, q$ .

##### C. Quantum state learning

The goal of quantum state learning [15,16] is to learn a quantum circuit that produces an approximation of a given input state  $\rho$ . Since  $\eta(\theta) = \text{tr}_b[U(\theta)|0\rangle\langle 0|_{ab}U^\dagger(\theta)]$ , where  $U(\theta)$  represents some parametrized quantum circuit, this is equivalent to finding some optimal  $\theta^*$  such that  $\eta(\theta^*) \approx \rho$  for any target mixed state  $\rho$ .

In the preceding section, we described how to find fractional inverses of  $\rho$ . Consider the special case where  $\alpha = -1$ , and the optimal solution is  $\eta(\theta_1^*) \approx \rho^{-1}/\text{tr}(\rho^{-1})$ . We then perform another purity minimization over  $\theta_2$  for the state

$$\frac{\nu(\theta_2)\eta(\theta_1^*)^2\nu(\theta_2)}{\text{tr}[\nu(\theta_2)\eta(\theta_1^*)^2\nu(\theta_2)]}. \quad (11)$$

We see that since  $\eta(\theta_1^*)^2$  approximates  $\rho^{-2}$  up to a normalization factor, the optimal solution is achieved when  $\nu(\theta_2^*) \approx \rho$ , which is the required state. The size of the largest quantum circuit required to perform quantum state learning is  $7n + 2$ , assuming that the input states are given and the dimension of  $\rho$  is  $d = 2^n$ . The overall circuit complexity is  $O(n)$ .

In Ref. [16], a similar technique using the SWAP test was used to compute a different cost function. This leads to a similar circuit complexity of  $O(n)$ . However, we note that the purity minimization approach incurs significant additional overhead because two parameters need to be optimized in order to find  $\eta(\theta_1^*)$  and  $\nu(\theta_2^*)$  while only one parameter needs to be optimized in the method of Ref. [16].

##### D. Estimating Rényi and Tsallis entropies

We recall that the Rényi entropies [20] are defined as the quantity

$$S_\alpha^R(\rho) = \frac{1}{1-\alpha} \log_2 \text{tr}(\rho^\alpha), \quad (12)$$

while Tsallis entropies [12] are defined as

$$S_\alpha^T(\rho) = \frac{1}{1-\alpha} [\text{tr}(\rho^\alpha) - 1], \quad (13)$$

where  $\alpha \in (0, 1) \cup (1, \infty)$ . The widely used von Neumann entropy  $S(\rho) := -\text{tr}(\rho \log_2 \rho)$  is retrieved as the limiting case of both quantities when  $\alpha \rightarrow 1$ , i.e.,  $\lim_{\alpha \rightarrow 1} S_\alpha^R(\rho) = \lim_{\alpha \rightarrow 1} S_\alpha^T(\rho) = S(\rho)$ . Note that both quantities require some estimate of  $\text{tr}(\rho^\alpha)$ . For integer  $\alpha$ , nonvariational quantum algorithms have been proposed [21,24].

In the preceding section, we have discussed how to obtain an approximation of fractional powers of a density matrix, i.e., prepare a quantum approximation  $\eta(\theta^*) \approx \rho^{p/q}/\text{tr}(\rho^{p/q})$ . Observe that any  $\alpha \in (0, 1) \cup (1, \infty)$  has a

rational approximation  $l + p/q$ , where  $l = \lfloor \alpha \rfloor$ ,  $p$ , and  $q$  are positive integers. This allows us to write  $\text{tr}(\rho^\alpha) \approx \text{tr}(\rho^l \rho^{p/q}) \approx \text{tr}[\rho^l \eta(\theta^*)] \text{tr}(\rho^{p/q})$ , where the first approximation is from the rational approximation to  $\alpha$  and the second approximation is due to the form of the *Ansatz*.  $\text{tr}[\rho^l \eta(\theta^*)]$  can be found by performing the SWAP test using  $l$  copies of  $\rho$  and one copy of  $\eta(\theta^*)$  as input. The normalization factor  $\text{tr}(\rho^{p/q})$  can be obtained using information that was already collected during the purity minimization process [see Eq. (7)]. The size of the largest quantum circuit required for this task is  $\max(ln + 2, (8p + 3)n + 2, (2q + 1)n + 2)$ , assuming that the input states are given and the dimension of  $\rho$  is  $d = 2^n$ . The overall circuit complexity is  $O(n)$ . This enables us to estimate both Rényi and Tsallis entropies for fractional  $\alpha$ . By comparison, the algorithms of Refs. [21,24] only apply to integer  $\alpha$ , with a similar circuit complexity of  $O(n)$ .

### E. Quantum fidelity and quantum Fisher information estimation

The Bures fidelity, also called the quantum fidelity, is frequently used to benchmark the quality of a quantum state preparation [13]. It is also applied in fundamental studies that use geometric quantifiers to detect quantum phase transitions in many-body quantum systems [25,26]. A closely related quantity is the quantum Fisher information (QFI). The QFI is most commonly used as a quantifier of the minimal achievable error in a quantum metrology protocol for a quantum probe [14,27] but can also be applied in fundamental studies of quantum nonclassicality [28–30]. We recall the expression for the Bures fidelity:

$$F(\rho, \sigma) = \text{tr}(\sqrt{\sqrt{\sigma} \rho \sqrt{\sigma}}). \quad (14)$$

From this expression, we expect that if the square root of a density matrix can be estimated variationally, then the quantum fidelity can also be estimated. We will describe how to do this via purity minimization.

First, we choose  $k = 1$  to find  $\eta(\theta_1^*) \approx \sigma^{-1/2} / \text{tr}(\sigma^{-1/2})$ . One then uses Eq. (7) to obtain an estimate of  $\sqrt{\sigma}$  given by the expression

$$\sqrt{\sigma} \approx \sigma \eta(\theta_1) K_1, \quad (15)$$

where  $K_1 := \{\text{tr}[\sigma \eta(\theta_1^*)^2] C_1(\theta_1^*)\}^{-1/2}$ , with  $C_1(\theta_1^*)$  being the value of the cost function evaluated at the optimal *Ansatz* state  $\eta(\theta_1^*)$ .

The next step is to perform another purity minimization to find  $\nu(\theta_2^*) \approx (\sqrt{\sigma} \rho \sqrt{\sigma})^{-1/2} / \text{tr}[(\sqrt{\sigma} \rho \sqrt{\sigma})^{-1/2}]$ . In order to do this, one minimizes the purity of the (un-normalized) state *Ansatz*

$$\nu(\theta_2) \sigma \eta(\theta_1^*) \rho \eta(\theta_1^*) \sigma \nu(\theta_2) \quad (16)$$

by optimizing  $\theta_2$ . We refer to the cost function Eq. (6) applied to *Ansatz*  $\nu(\theta_2)$  as  $C_2(\theta_2)$ . Since  $\sigma \eta(\theta_1^*) \rho \eta(\theta_1^*) \sigma \propto \sqrt{\sigma} \rho \sqrt{\sigma}$ , we must have  $\nu(\theta_2^*) \propto (\sqrt{\sigma} \rho \sqrt{\sigma})^{-1/2}$ . This gives us  $\sigma \eta(\theta_1^*) \rho \eta(\theta_1^*) \sigma \nu(\theta_2^*) \propto \sqrt{\sqrt{\sigma} \rho \sqrt{\sigma}}$ . One may verify the final expression

$$K_1 K_2 \text{tr}[\sigma \eta(\theta_1^*) \rho \eta(\theta_1^*) \sigma \nu(\theta_2^*)] \approx \text{tr}(\sqrt{\sqrt{\sigma} \rho \sqrt{\sigma}}) \quad (17)$$

$$= F(\rho, \sigma), \quad (18)$$

where  $K_2 := \{\text{tr}[\sigma \eta(\theta_1^*) \rho \eta(\theta_1^*) \sigma \nu(\theta_2^*)^2] C_2(\theta_2^*)\}^{-1/2}$ , with  $C_2(\theta_2^*)$  being the cost function value corresponding to the state  $\nu(\theta_2^*)$ . Both  $K_1$  and  $K_2$  can be efficiently estimated using information that was already collected during the purity minimization process [see Eq. (7)]. This allows us to variationally find the quantum fidelity  $F(\rho, \sigma)$  using multiple copies of  $\rho$  and  $\sigma$  as inputs. The space requirements of the protocol are obtained by counting the number of copies of  $\rho$  and  $\sigma$  and variational *Ansätze* needed to compute each factor of the left-hand side of Eq. (17). For example, the purity minimization in  $K_1$  requires 6 input states plus one ancilla qubit to measure, while  $K_2$  requires 14 input states plus one ancilla qubit in total. For fidelity estimation, the size of the largest quantum circuit scales with  $13n + 2$ , assuming that the input states are given and the dimensions of  $\rho$  and  $\sigma$  are  $d = 2^n$ . The overall circuit complexity is  $O(n)$ .

In order to approximate the QFI, one can utilize the above purity-based VQA for the fidelity. Suppose that a smooth quantum trajectory  $\rho_\theta$  in state space is parametrized by the real parameter  $\theta$ . Assume that every point along this trajectory is given and can be used as an input to some quantum algorithm. One may then approximate the Fisher information via the identity

$$I(\theta) = \lim_{\delta \rightarrow 0} 8 \frac{1 - F(\rho_\theta, \rho_{\theta+\delta})}{\delta^2}. \quad (19)$$

Therefore the QFI can be estimated by variationally optimizing for the quantum fidelity  $F(\rho_\theta, \rho_{\theta+\delta})$  between two nearby states  $\rho_\theta$  and  $\rho_{\theta+\delta}$ . The overall circuit complexity of the QFI estimation is similar to fidelity estimation, assuming that  $\rho_\theta$  is given.

We note that there are several other recently proposed algorithms for estimating the quantum fidelity. Reference [16] contains an alternative VQA approach based on Uhlmann’s theorem, which achieves similar overall complexity  $O(n)$ . Similar to our approach, it also requires two separate optimization procedures, but our purity minimization procedure incurs significant overhead due to the fact that estimating the purity of Eq. (16) requires 14 states as input in total. Reference [31] describes a VQA that estimates the fidelity by computing upper and lower bounds in  $\text{poly}(n)$  time, compared with the  $O(n)$  scaling using purity minimization. Reference [32] describes a fully quantum algorithm that estimates the quantum fidelity of low-rank states using block encoding and quantum phase estimation. For low-rank states, this achieves an overall complexity of  $\text{poly}(n)$  but is more difficult to implement on NISQ devices compared with other VQA approaches due to its heavy reliance on quantum phase estimation. A VQA for finding a lower bound estimate of the QFI is also described in Refs. [33,34]. This utilizes the fidelity estimation technique of Ref. [31] as a subroutine and has a circuit complexity of  $\text{poly}(n)$ .

## V. NUMERICAL SIMULATION

In order to demonstrate the proposed algorithms, we performed a numerical simulation to find the rank, Rényi entropy, and quantum fidelity for a single-qubit input state  $\rho = \cos^2(\phi/2)|0\rangle\langle 0| + \sin^2(\phi/2)|1\rangle\langle 1|$ . For this purpose, we chose a simple diagonal *Ansatz* of the form  $\eta(\theta) = \cos^2(\theta/2)|0\rangle\langle 0| + \sin^2(\theta/2)|1\rangle\langle 1|$  initialized at  $\theta = \pi/2$  and

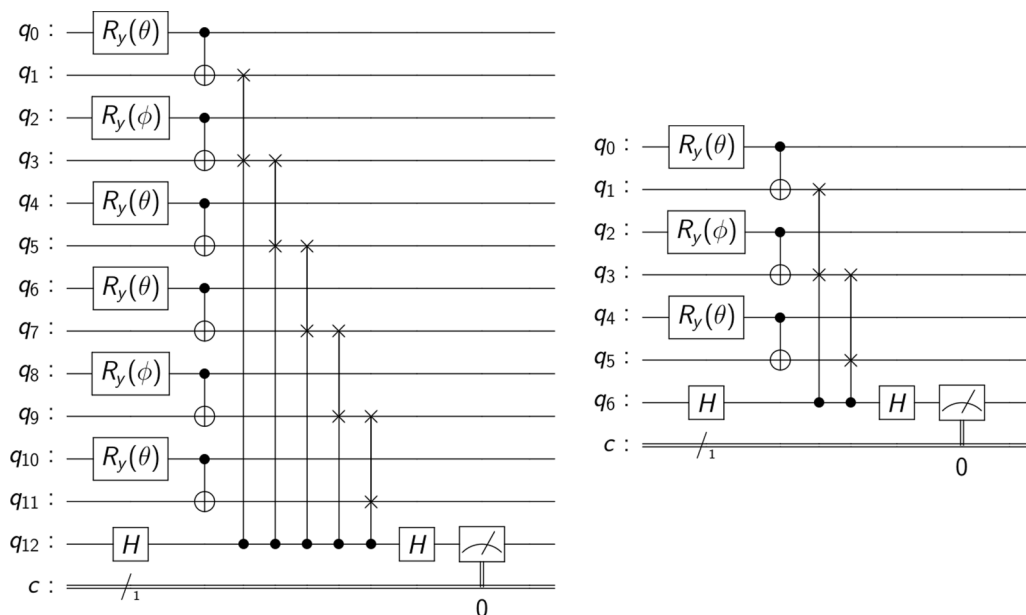


FIG. 1. Example circuits to compute (left) the numerator and (right) the denominator of the cost function  $C(\theta)$  with  $k = 1$  [see Eq. (6)]. In both cases, the output is obtained by measuring a single qubit in the computational basis

minimized the cost function over  $\theta$  using a gradient descent method. The gradient of the cost function is computed via a finite difference method using a step size set to  $\Delta\theta = 0.005$ , while the learning rate of the gradient descent, which determines the rate at which  $\theta$  is updated during gradient descent, is set to  $r = 0.05$ . The algorithm is then executed via simulated quantum circuits using the QISKIT quantum simulator package [35]. Figure 1 shows example circuits for calculating the numerator and denominator of the cost function in Eq. (6) with  $k = 1$ .

We used the proposed algorithms to estimate the rank of  $\rho$  as well as Rényi entropy for  $\alpha = 1/2$ . We also used the algorithm to estimate the quantum fidelity  $F(\rho, \sigma)$ , where the second argument was chosen to be the maximally mixed qubit state, i.e.,  $\sigma = \mathbb{1}_2/2$ . The results are shown in Fig. 2. It is observed that the algorithms were able to accurately reproduce the expected exact results up to the numerical precision of the optimization.

A limitation of simulating actual quantum circuits is that it is very memory intensive, which limits our ability to scale up to larger system sizes. In order to scale the proposed al-

gorithms, we numerically evaluate the cost function equation (6) directly instead of simulating actual quantum circuits. The gradient descent is otherwise performed just as before. The input states are product states  $\rho = [\cos^2(\phi/2)|0\rangle\langle 0| + \sin^2(\phi/2)|1\rangle\langle 1|]^{\otimes n}$ , where  $n = 3, 6, 9$  is the number of qubits. The Ansatz state is chosen to be  $\eta(\theta) = [\cos^2(\theta/2)|0\rangle\langle 0| + \sin^2(\theta/2)|1\rangle\langle 1|]^{\otimes n}$ , initialized at  $\theta = \pi/2$ . The simulated results are shown in Fig. 3, where it is observed that the proposed VQAs give highly accurate estimates of the rank, Rényi entropy, and fidelity. Over the range of parameters tested, the errors do not appear to grow exponentially with system size, which is a major problem faced by many VQAs. For example, for  $n = 3, 6, 9$ , the mean errors of the rank estimation are 0.0348, 0.0695, and 0.104%, respectively, which approximately follows a linear scaling. However, because the rank of  $\rho$  grows exponentially in this example, we expect errors to be a much more significant issue for larger system sizes. In general, the scaling of errors with system size is highly dependent on both the rank of the input state and the design of the Ansatz. This will be discussed in greater detail in the following section.

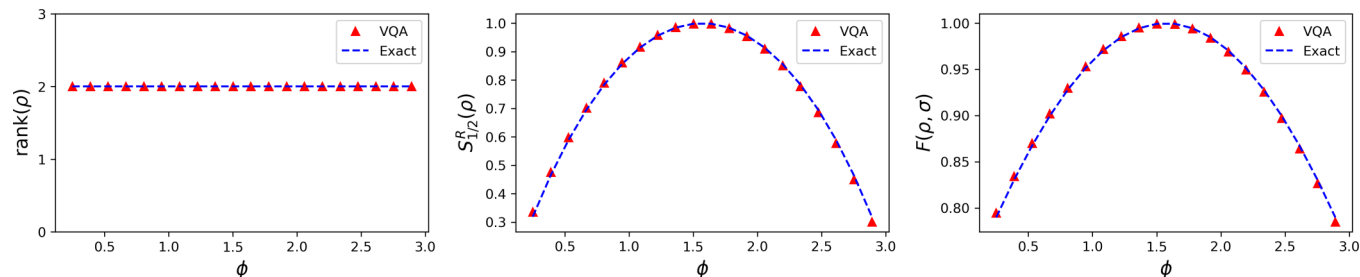


FIG. 2. Exact and estimated values of the (left) rank, (middle) Rényi entropy for  $\alpha = 1/2$ , and (right) fidelity using the proposed VQAs. The input state is parametrized by  $\phi$  and is given by  $\rho = \cos^2(\phi/2)|0\rangle\langle 0| + \sin^2(\phi/2)|1\rangle\langle 1|$ . For the fidelity calculation  $F(\rho, \sigma)$ , the second input state was chosen to be the maximally mixed state  $\sigma = \mathbb{1}_2/2$ .

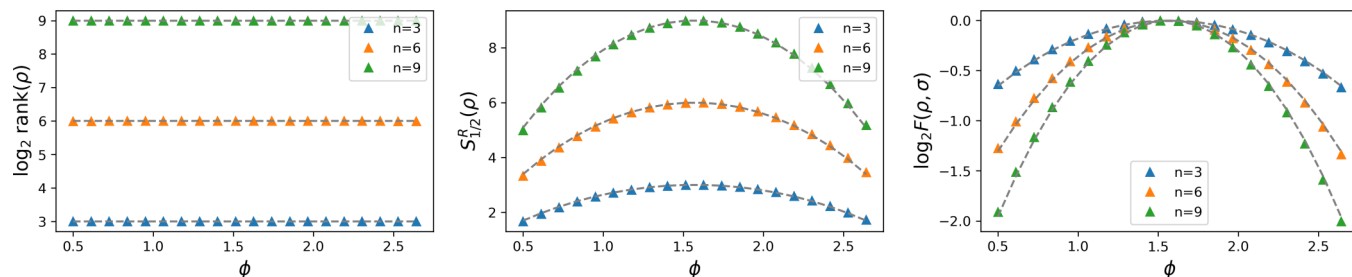


FIG. 3. Exact (dashed lines) and estimated (triangle points) values of the (left) rank, (middle) Rényi entropy for  $\alpha = 1/2$ , and (right) fidelity for  $n = 3, 6, 9$  qubit states. The input state is a product state given by  $\rho = [\cos^2(\phi/2)|0\rangle\langle 0| + \sin^2(\phi/2)|1\rangle\langle 1|]^{\otimes n}$ . For the fidelity calculation  $F(\rho, \sigma)$ , the second input state was chosen to be the maximally mixed state  $\sigma = \mathbb{1}_d/d$ , where  $d = 2^n$ . Here, the cost function is directly computed rather than obtained via simulated quantum circuits as in Fig. 2.

### VI. STRATEGIES FOR MITIGATING THE BARREN PLATEAU LANDSCAPE PROBLEM

The cost function in Eq. (6) involves measurements of global properties, which may lead to the barren plateau landscape (BPL) problem, where the expected magnitude of the gradient of the cost function decreases exponentially with system size [19,36].

For example, suppose the input state is a tensor product state of the form  $\rho = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^{\otimes n}$ ,  $\lambda_1 > \lambda_2$ , and that the Ansatz state has the form  $\eta(\theta) = R(\theta)\rho^{-1/(2k)}/\text{tr}(\rho^{-1/(2k)})R(\theta)^\dagger$ , where  $\theta = (\theta_1, \dots, \theta_n)$  and  $R(\theta) = \bigotimes_{i=1}^n \exp(-i\theta_i\sigma_y/2)$ . Despite the fact that  $\eta(\theta)$  is an SU(2) orbit of the state that achieves the global minimum of Eq. (6), it can be shown that this combination of input state  $\rho$  and Ansatz  $\eta(\theta)$  encounters the BPL problem. To see this, first note that the cost function can be written in the form

$$C(\theta) = \frac{1}{2^n} \prod_{j=1}^n \frac{N(\theta_j)}{D(\theta_j)}, \tag{20}$$

where  $N(\theta_j)$  and  $D(\theta_j)$  are quartic polynomials in  $\tan \frac{\theta_j}{2}$  with coefficients that depend on  $\lambda_1, \lambda_2$ . From the independence of the random variables  $\theta_j$  distributed according to the uniform distribution  $p$  on  $[-\pi, \pi]$ , we note that

$$E_{\theta \sim p^{\otimes n}} (|\partial_{\theta_1} C(\theta)|) = 2^{-n} E_{\theta_1 \sim p} \left( \left| \partial_{\theta_1} \frac{N(\theta_1)}{D(\theta_1)} \right| \right) \left[ E_{x \sim p} \left( \frac{N(x)}{D(x)} \right) \right]^{n-1}, \tag{21}$$

where we considered optimization with respect to  $\theta_1$  without loss of generality. The first factor of Eq. (21) is independent of  $n$ , and when  $p$  is restricted to the neighborhood  $[-\delta, \delta]$  for  $\delta \ll \frac{\lambda_2}{\lambda_1}$ , the second factor is given by

$$\left[ \left( \frac{1}{2} - \frac{\delta^2 \left( \frac{\lambda_2}{\lambda_1} + \frac{\lambda_1}{\lambda_2} \right)}{48} \right) + O(\delta^2) \right]^n, \tag{22}$$

which has an exponential dependence on  $n$ . Therefore, from Chebyshev's inequality in the form  $P(|X| > \epsilon) \leq \epsilon^{-1} E(|X|)$ , it follows that on  $[-\delta, \delta]$ ,  $P(|\partial_{\theta_1} C(\theta)| \geq \epsilon) \leq O(b^n)$ , where  $b$  is taken from Eq. (22). This demonstrates that in some local neighborhood of the optimal solution, the gradient exponentially decays with the system size  $n$ . The exponential decay of the gradient magnitude with respect to  $n$  is the defining property of the BPL.

One possible way to mitigate the BPL problem is the introduction of additional local terms to the cost function [37]. We demonstrate this possibility for at least some choices of the state  $\rho$ . Let us now denote the cost function from Eq. (6) as  $C_G(\theta)$  to reflect the global property of the cost function. We consider additional local terms of the form

$$C_L(\theta) = \frac{1}{n} \sum_i \text{tr}(\sigma_i^2), \tag{23}$$

where  $n$  is the number of subsystems of  $\rho$  and  $\sigma_i = \eta_i(\theta)^k \rho_i \eta_i(\theta)^k / \text{tr}[\eta_i(\theta)^k \rho_i \eta_i(\theta)^k]$ . The states  $\eta_i(\theta) = R_i(\theta_i) \rho_i^{-1/(2k)} / \text{tr}(\rho_i^{-1/(2k)}) R_i(\theta_i)^\dagger$  and  $\rho_i = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$  are taken to be the reduced density matrices of the  $i$ th subsystem of  $\eta(\theta)$  and  $\rho$ , respectively. Observe that unlike  $C_G(\theta)$ , each term in  $C_L(\theta)$  measures a local property of the state and scales at most polynomially with system size.

We then observe that in the example we have considered,  $\text{tr}(\sigma_i^2)$ , and hence  $C_L(\theta)$ , is minimized when  $\theta_i = 0$ . This is the same critical point as the optimal solution for the global cost function  $C_G(\theta)$ . This allows us to circumvent the BPL problem. More generally, however, it is not always true that the solutions to the local and the global cost functions coincide. As a possible BPL mitigation strategy, one approach is to consider a linear combination of both local and global cost functions of the form  $C_{G+L}(\theta) = (1 - \epsilon)C_G(\theta) + \epsilon C_L(\theta)$ , where  $\epsilon \ll 1$  is a small constant. By choosing small  $\epsilon$ , we ensure that we are close to the ideal solution in the worst case, while eliminating the BPL problem in cases such as the previously considered examples. To see that we will always be close to the actual solution for small  $\epsilon$ , we observe that  $\|\mathbb{1}_d/d - \sigma\|^2 < C_G(\theta) < C_{G+L}(\theta) \lesssim \epsilon$  [see Eq. (10)]. In the last inequality, we assumed the approximation  $C_G(\theta) \approx 0$  when the dimension  $d$  is sufficiently large and the inequality  $C_L(\theta) \leq 1$ .

Another possible strategy to mitigate the BPL problem is to consider correlated Ansatz parameters. Because  $\rho$  is an exchangeable state [13], one expects that efficient optimization of the cost function  $C_G(\theta)$  can be achieved by correlation of the Ansatz parameters in  $R(\theta)$ , e.g., by taking  $\theta_j = \theta$  for all  $j$  [38]. However, we observe that the  $2^{-n}$  factor in Eq. (20) cannot be overcome by straightforwardly correlating the parameters  $\theta_j$ , so an alternative strategy is required. In Appendix A, we show that the BPL still exists for this scenario. In the case that  $\rho$  is an un-normalized state in Eq. (6), e.g.,

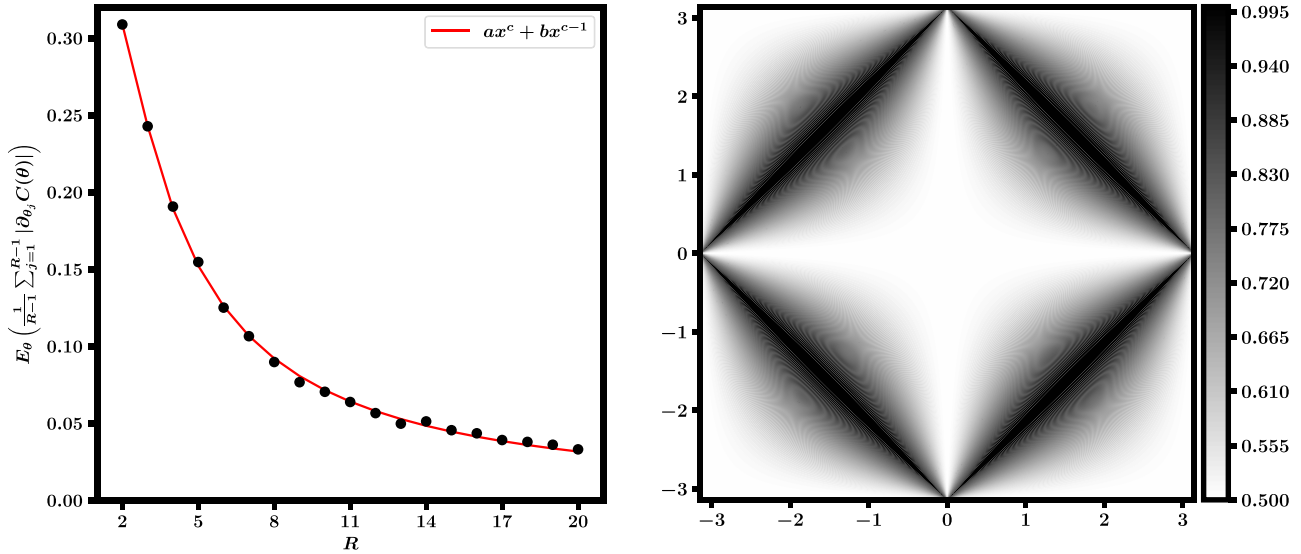


FIG. 4. Left: Average magnitude of the gradient components of cost function (6) for  $k = 4$  and for  $R = 2, 3, \dots, 20$ , with  $\rho$  being the completely mixed state of rank  $R$ . Error bars are smaller than the dot radius. The red curve is a least-squares fit of the function  $f(x) = ax^c + bx^{c-1}$  to the data; optimal parameters are  $a \approx 1.41$ ,  $b \approx -1.35$ ,  $c \approx -1.25$ . Right: Grayscale plot of  $C(\theta)$  defined by (26) and (27) with  $\lambda = 1/2$ ,  $\mu = \lambda^{1/2k}/[\lambda^{1/2k} + (1-\lambda)^{1/2k}]$ , and  $n = 2$ . The gradient vanishes with  $n$  in the center region near the solution.

$\rho = \sigma^\ell$  for quantum state  $\sigma$ , the purity minimization results in a state proportional to  $\sigma^{-\ell/2k}$ . Then the cost function for a correlated  $R(\theta)$ ,  $\sigma = \text{diag}(\lambda_1, \lambda_2)$ , and  $\eta(\theta) \propto R(\theta)\sigma^{-\ell/2k}R(\theta)^\dagger$  is a function of  $\theta$  and  $\ell$ . For any  $\ell$ , one can demonstrate a BPL in a neighborhood of the cost function minimum by using the same method that is used to show the existence of a local BPL for  $C_G(\theta)$  in Appendix A.

We now provide evidence that for low-rank states, the BPL problem can be avoided with a sufficiently good *Ansatz*. We show that if an  $n$ -qubit state  $\rho$  has rank  $R$ , then there exists an *Ansatz*  $\eta(\theta)$  defined by a parameter manifold of dimension  $R - 1$  (specifically, the  $R$  sphere) for which BPL does not occur for the cost function equation (6) if  $R = \text{poly}(n)$ . Consider the *Ansatz*

$$\eta(\theta) = \cos^2 \frac{\theta_1}{2} |\psi_1\rangle\langle\psi_1| + \sin^2 \frac{\theta_1}{2} \cos^2 \frac{\theta_2}{2} |\psi_2\rangle\langle\psi_2| + \dots + \prod_{j=1}^{R-1} \sin^2 \frac{\theta_j}{2} |\psi_{R-1}\rangle\langle\psi_{R-1}|, \quad (24)$$

where  $\{|\psi_i\rangle\}_{i=1}^R$  are the eigenvectors of  $\rho$ . To show that the resulting cost function equation (6) does not exhibit a BPL if  $R = \text{poly}(n)$ , it is sufficient to show that the expected average magnitude of the gradient components  $E(\frac{1}{R-1} \sum_{j=1}^{R-1} |\partial_{\theta_j} C(\theta)|)$  goes to 0 polynomially in  $R$ . We demonstrate this numerically in Fig. 4 by fitting the mean of Monte Carlo estimates of the expected gradient components to a polynomially decreasing function.

More generally, low-rank states combined with low-rank variational *Ansätze*  $\eta(\theta)$  do not guarantee the absence of a BPL in a given problem. Consider cost function (6) with rank-2 state  $\rho = \lambda|0\rangle\langle 0|^{\otimes n} + (1-\lambda)|1\rangle\langle 1|^{\otimes n}$  and with *Ansatz*  $\eta(\theta) := R(\theta)\sigma R(\theta)^\dagger$ , where the normalized state  $\sigma$  is defined by

$$\sigma = (1-\mu)|0\rangle\langle 0|^{\otimes n} + \mu|1\rangle\langle 1|^{\otimes n}, \quad (25)$$

$\mu \in (0, 1)$ . For simplicity, we consider fixed  $\lambda$  and  $\mu$ , viz., the spectra of the target  $\rho^{-1/2k}$  and the *Ansatz*  $\eta(\theta)$  do not vary with  $n$ . Unlike the *Ansatz* (24),  $\eta(\theta) = R(\theta)\sigma R(\theta)^\dagger$  does not commute with  $\rho$  pointwise over  $\theta$ . Considering  $C(\theta)$  as a function of  $\lambda$ ,  $\mu$ , and  $\theta$ , one has that  $C(\theta) \geq \frac{1}{2}$ , as expected for a state  $\rho$  in a two-dimensional subspace. The cost function is given by  $C(\theta) = N(\theta)/D(\theta)$  with

$$\begin{aligned} N(\theta) = & \lambda^2[(1-\mu)^{2k}a(\theta) + \mu^{2k}b(\theta)] \\ & + 2\lambda(1-\lambda)[(1-\mu)^{2k} + \mu^{2k}]^2a(\theta)b(\theta) \\ & + (1-\lambda)^2[(1-\mu)^{2k}b(\theta) + \mu^{2k}a(\theta)] \end{aligned} \quad (26)$$

and

$$\begin{aligned} D(\theta) = & \{ \lambda[(1-\mu)^{2k}a(\theta) + \mu^{2k}b(\theta)] \\ & + (1-\lambda)[(1-\mu)^{2k}b(\theta) + \mu^{2k}a(\theta)] \}^2, \end{aligned} \quad (27)$$

where  $a(\theta) := \prod_{j=1}^n \cos^2 \frac{\theta_j}{2}$ ,  $b(\theta) := \prod_{j=1}^n \sin^2 \frac{\theta_j}{2}$ . The cost function landscape is shown in Fig. 4, with  $\lambda = 1/2$  and  $\mu = \lambda^{1/2k}/[\lambda^{1/2k} + (1-\lambda)^{1/2k}]$  so that  $C(\theta = 0) = 1/2$ , i.e., the *Ansatz* is fully expressive. One finds that  $|\partial_{\theta_j} C(\theta)| = F(\lambda, \mu) \frac{|\theta_j|}{4^{n-1}} \prod_{j=2}^n \theta_j^2 + O(\delta^{4n})$  with constant  $F(\lambda, \mu)$  in a sufficiently small neighborhood  $(-\delta, \delta)^{\times n}$ ,  $\delta < 1$ , of  $\theta = 0$ . It follows that a BPL is present in a neighborhood of a global minimum. In fact, for the fully expressive *Ansatz* in Fig. 4 we find that a BPL exists on sufficiently small constant-volume neighborhoods, e.g.,  $(-2, 2)^{\times n}$ . However, similarly to the above analysis of a correlated *Ansatz* for a full-rank state, for a given  $\lambda$ , there are  $\mu$  for which a BPL is absent on the global domain  $[-\pi, \pi]^{\times n}$ . The presence of a local BPL even though there is no global BPL present suggests that this low-rank *Ansatz* was only able to partially mitigate the BPL problem. Therefore, although it is possible that a randomly initialized gradient descent algorithm could rapidly obtain a cost function value close to the global minimum, the *Ansatz*  $\eta(\theta)$  will

require exponential resources to completely converge to the solution  $\rho^{-1/2k}/\text{tr}\rho^{-1/2k}$  in terms of trace distance, even if the *Ansatz* is fully expressive.

VII. CONCLUSION

We proposed VQAs to estimate physical quantities such as the rank, the Rényi and Tsallis quantum entropies, the quantum fidelity, and the quantum Fisher information for mixed quantum states. These VQAs share a common theme of minimizing a cost function related to the quantum purity of a normalized quantum state. In general, these cost functions are efficiently computable on quantum computers via a quantum SWAP test, while it may be inefficient on a classical computer. We also apply similar cost functions to perform tasks such as preparing fractional powers and fractional inverses of quantum states, as well as quantum state learning. It is noteworthy that already a rather wide variety of problems can be solved via the purity minimization approach, which may suggest further applications beyond those we have considered thus far. As a proof of concept, numerical simulations of the VQAs were performed, which demonstrates that the proposed algorithms are able to retrieve the exact values for the rank, Rényi entropy, and quantum fidelity for up to  $n = 9$  qubit states.

We also studied the BPL problem in relation to the basic cost function that we proposed in Eq. (6). Explicit examples exhibiting the BPL problem were discussed, and strategies such as adding local terms to the cost function or correlating the *Ansatz* parameters were considered in order to mitigate this issue. Our numerical evidence further suggests that for low-rank input states, there exist low-dimension *Ansatz* manifolds for which a BPL does not occur. This supports the recent results in Ref. [32], which suggests that for high-rank states, computing the quantum fidelity may be hard even for quantum computers.

We are hopeful that the methods proposed here will inspire new applications for VQAs and for quantum computing in general. In particular, we are optimistic that some of the proposed VQAs can be applied on already available NISQ computing devices to probe nonclassical properties of quantum states.

ACKNOWLEDGMENTS

K.C.T. was supported by the NTU Presidential Postdoctoral Fellowship program funded by Nanyang Technological University. T.V. acknowledges support from the LDRD program at LANL.

APPENDIX: BPL FOR ANSATZ  $\eta(\theta)$  WITH CORRELATED PARAMETERS

Correlating the angles in  $R(\theta_1, \dots, \theta_n)$  via  $\theta_j = \theta$  for all  $j$  gives the variational *Ansatz*

$$\eta(\theta) = (e^{-i\frac{\theta}{2}\sigma_y})^{\otimes n} \frac{\rho^{-1/(2k)}}{\text{tr}(\rho^{-1/(2k)})} (e^{i\frac{\theta}{2}\sigma_y})^{\otimes n}. \tag{A1}$$

The cost function (20) then takes the form

$$\begin{aligned} C(\theta) &= \frac{1}{2^n} \left( \frac{N(\theta)}{D(\theta)} \right)^n, \\ N(\theta) &= 1 + b \tan^2 \frac{\theta}{2} + a \tan^4 \frac{\theta}{2}, \\ D(\theta) &= \left( 1 + c \tan^2 \frac{\theta}{2} \right)^2, \end{aligned} \tag{A2}$$

where  $a := \frac{1}{2}(\frac{\lambda_1^2}{\lambda_2^2} + \frac{\lambda_2^2}{\lambda_1^2})$ ,  $b := 2(\frac{\lambda_2}{\lambda_1} + \frac{\lambda_1}{\lambda_2} - 1)$ ,  $c := \frac{1}{2}(\frac{\lambda_2}{\lambda_1} + \frac{\lambda_1}{\lambda_2})$ . We seek the  $n \rightarrow \infty$  asymptotics of  $E_{\theta \sim p}(|\partial_\theta C(\theta)|)$ , in particular, an upper bound that vanishes exponentially. We restrict the uniform measure  $p$  to the uniform measure on  $A_\delta = [-\delta, \delta]$ , where  $\delta \ll \sqrt{\frac{\lambda_2}{\lambda_1}}$ , and prove that BPL occurs in this neighborhood of the optimum. We first obtain the expression

$$E_{\theta \sim p}(|\partial_\theta C(\theta)|) = \frac{n}{2^n} E_{\theta \sim p} \left[ \left( \frac{N(\theta)}{D(\theta)} \right)^{n-1} \left| \partial_\theta \frac{N(\theta)}{D(\theta)} \right| \right]. \tag{A3}$$

It follows that

$$\begin{aligned} \left| \partial_\theta \frac{N(\theta)}{D(\theta)} \right| &= \left| \frac{[(2a - bc) \tan^2 \frac{\theta}{2} + b - 2c] \tan \frac{\theta}{2}}{(1 + c \tan^2 \frac{\theta}{2})^3 \cos^2 \frac{\theta}{2}} \right| \\ &\sim \left| \frac{(b - 2c) \frac{\theta}{2}}{(1 + c \frac{\theta^2}{4})^3} \right| + O(\theta^3) \quad (\text{for } \theta \rightarrow 0). \end{aligned} \tag{A4}$$

So we calculate

$$\begin{aligned} E_{\theta \sim p}(|\partial_\theta C(\theta)|) &= \frac{|b - 2c|n}{2^{n+1}} \frac{1}{2\delta} \int_{-\delta}^{\delta} dx \frac{|x|}{(1 + \frac{cx^2}{4})^{n+2}} + O(\delta^3) \\ &= \frac{|b - 2c|n}{c(n+1)2^n} \left[ 1 - \left( 1 + \frac{c\delta^2}{4} \right)^{-(n+1)} \right]. \end{aligned} \tag{A5}$$

Again using the Chebyshev inequality, the probability of an arbitrarily small gradient magnitude scales as  $O(2^{-n})$ , so there is a barren plateau even when an *Ansatz* with correlated parameters is used.

[1] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).  
 [2] J. Preskill, Quantum computing and the entanglement frontier, [arXiv:1203.5813](https://arxiv.org/abs/1203.5813).  
 [3] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A.

Fowler *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).  
 [4] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, *J. Phys. Soc. Jpn.* **90**, 032001 (2021).  
 [5] 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).
- [6] K. M. Nakanishi, K. Mitarai, and K. Fujii, Subspace-search variational quantum eigensolver for excited states, *Phys. Rev. Res.* **1**, 033062 (2019).
- [7] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum-assisted quantum compiling, *Quantum* **3**, 140 (2019).
- [8] H.-Y. Huang, K. Bharti, and P. Rebentrost, Near-term quantum algorithms for linear systems of equations, [arXiv:1909.07344](https://arxiv.org/abs/1909.07344).
- [9] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles, Variational quantum linear solver: A hybrid algorithm for linear systems, [arXiv:1909.05820](https://arxiv.org/abs/1909.05820).
- [10] D. An and L. Lin, Quantum linear system solver based on time-optimal adiabatic quantum computing and quantum approximate optimization algorithm, [arXiv:1909.05500](https://arxiv.org/abs/1909.05500).
- [11] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, On quantum Rényi entropies: A new generalization and some properties, *J. Math. Phys. (Melville, NY)* **54**, 122203 (2013).
- [12] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, *J. Stat. Phys.* **52**, 479 (1988).
- [13] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, New York, 2018).
- [14] K. C. Tan and H. Jeong, Nonclassical light and metrological power: An introductory review, *AVS Quantum Sci.* **1**, 014701 (2019).
- [15] S. M. Lee, J. Lee, and J. Bang, Learning unknown pure quantum states, *Phys. Rev. A* **98**, 052302 (2018).
- [16] R. Chen, Z. Song, X. Zhao, and X. Wang, Variational quantum algorithms for trace distance and fidelity estimation, [arXiv:2012.05768](https://arxiv.org/abs/2012.05768).
- [17] M. Horodecki, P. Horodecki, and J. Oppenheim, Reversible transformations from pure to mixed states and the unique measure of information, *Phys. Rev. A* **67**, 062104 (2003).
- [18] A. Streltsov, H. Kampermann, S. Wölk, M. Gessner, and D. Bruß, Maximal coherence and the resource theory of purity, *New J. Phys.* **20**, 053058 (2018).
- [19] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [20] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, Quantum Fingerprinting, *Phys. Rev. Lett.* **87**, 167902 (2001).
- [21] S. Johri, D. S. Steiger, and M. Troyer, Entanglement spectroscopy on a quantum computer, *Phys. Rev. B* **96**, 195136 (2017).
- [22] J. Yirka and Y. Subasi, Qubit-efficient entanglement spectroscopy using qubit resets, [arXiv:2010.03080](https://arxiv.org/abs/2010.03080).
- [23] R. O'Donnell and J. Wright, Efficient quantum tomography, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing, STOC '16* (Association for Computing Machinery, New York, 2016), pp. 899–912.
- [24] Y. Subasi, L. Cincio, and P. J. Coles, Entanglement spectroscopy with a depth-two quantum circuit, *J. Phys. A: Math. Theor.* **52**, 044001 (2019).
- [25] A. Carollo, D. Valenti, and B. Spagnolo, Geometry of quantum phase transitions, *Phys. Rep.* **838**, 1 (2020).
- [26] K. C. Tan, Identifying quantum phase transitions via geometric measures of nonclassicality, *Phys. Rev. A* **102**, 022421 (2020).
- [27] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
- [28] P. Hyllus, W. Laskowski, R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, L. Pezzé, and A. Smerzi, Fisher information and multiparticle entanglement, *Phys. Rev. A* **85**, 022321 (2012).
- [29] K. C. Tan, S. Choi, H. Kwon, and H. Jeong, Coherence, quantum Fisher information, superradiance, and entanglement as interconvertible resources, *Phys. Rev. A* **97**, 052304 (2018).
- [30] H. Kwon, K. C. Tan, T. Volkoff, and H. Jeong, Nonclassicality as a Quantifiable Resource for Quantum Metrology, *Phys. Rev. Lett.* **122**, 040503 (2019).
- [31] M. Cerezo, A. Poremba, L. Cincio, and P. J. Coles, Variational quantum fidelity estimation, *Quantum* **4**, 248 (2020).
- [32] Q. Wang, Z. Zhang, K. Chen, J. Guan, W. Fang, and M. Ying, Quantum algorithm for fidelity estimation, [arXiv:2103.09076](https://arxiv.org/abs/2103.09076).
- [33] J. L. Beckey, M. Cerezo, A. Sone, and P. J. Coles, Variational quantum algorithm for estimating the quantum Fisher information, [arXiv:2010.10488](https://arxiv.org/abs/2010.10488).
- [34] A. Sone, M. Cerezo, J. L. Beckey, and P. J. Coles, A generalized measure of quantum Fisher information, [arXiv:2010.02904](https://arxiv.org/abs/2010.02904).
- [35] H. Abraham, AduOffei, R. Agarwal, G. Agliardi, M. Aharoni, I. Y. Akhalwaya, G. Aleksandrowicz, T. Alexander, M. Amy, S. Anagolum, E. Arbel, A. Asfaw, A. Athalye, A. Avkhadiev, C. Azaustre, A. Banerjee, S. Banerjee, W. Bang, A. Bansal, P. Barkoutsos *et al.*, Qiskit: An open-source framework for quantum computing, version 0.7.2, 2019, <https://doi.org/10.5281/zenodo.2562111>.
- [36] T. J. Volkoff, Efficient trainability of linear optical modules in quantum optical neural networks, *J. Russ. Laser Res.* **42**, 250 (2021).
- [37] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, *Nat. Commun.* **12**, 1791 (2021).
- [38] T. Volkoff and P. J. Coles, Large gradients via correlation in random parameterized quantum circuits, *Quantum Sci. Technol.* **6**, 025008 (2021).