PAPER

Variational quantum algorithms for trace distance and fidelity estimation

Quantum Science and Technology

PAPER

Variational quantum algorithms for trace distance and fidelity estimation

RECEIVED
5 September 2021

REVISED
4 November 2021

ACCEPTED FOR PUBLICATION
11 November 2021

PUBLISHED
22 December 2021

Ranyiliu Chen, Zhixin Song, Xuanqiang Zhao and Xin Wang*
Institute for Quantum Computing, Baidu Research, Beijing 100193, People's Republic of China
* Author to whom any correspondence should be addressed.

E-mail: wangxin73@baidu.com

Keywords: fidelity, near-term quantum computing, variational quantum algorithms, trace distance, quantum machine learning

Supplementary material for this article is available online

Abstract

Estimating the difference between quantum data is crucial in quantum computing. However, as typical characterizations of quantum data similarity, the trace distance and quantum fidelity are believed to be exponentially-hard to evaluate in general. In this work, we introduce hybrid quantum–classical algorithms for these two distance measures on near-term quantum devices where no assumption of input state is required. First, we introduce the variational trace distance estimation (VTDE) algorithm. We in particular provide the technique to extract the desired spectrum information of any Hermitian matrix by local measurement. A novel variational algorithm for trace distance estimation is then derived from this technique, with the assistance of a single ancillary qubit. Notably, VTDE could avoid the barren plateau issue with logarithmic depth circuits due to a local cost function. Second, we introduce the variational fidelity estimation algorithm. We combine Uhlmann's theorem and the freedom in purification to translate the estimation task into an optimization problem over a unitary on an ancillary system with fixed purified inputs. We then provide a purification subroutine to complete the translation. Both algorithms are verified by numerical simulations and experimental implementations, exhibiting high accuracy for randomly generated mixed states.

1. Introduction

With surging advances in material, manufacturing, and quantum control, quantum computing has been driven into the noisy intermediate-scale quantum (NISQ) era [1], which requires novel algorithms running on a limited number of qubits with unwanted interference of the environment. The hybrid quantum–classical computation framework [2] is regarded as well-suited for execution on NISQ devices and is expected to show practical near-term applications in quantum chemistry and quantum machine learning [3–5]. Specifically, hybrid quantum–classical algorithms utilize the parameterized quantum circuits (PQCs) [6] and classical optimization to solve problems. Such idea was applied to many key areas including Hamiltonian ground and excited states preparation [7, 8], quantum compiling [9], quantum classification [10–12], Gibbs state preparation [13–16], entanglement manipulation [17], and quantum linear algebra [18–22]. We refer to [23–25] for a detailed review.

Applications mentioned above and many other quantum information tasks suffer from unwanted interactions with the environment when implemented by real-world quantum systems, leading to errors in their working qubits. Thus, metric estimation for quantum states is vital to benchmark the tasks' implementation. As scalable quantum computers and quantum error correction are still on their way [26], estimating metric on a near-term device is essential to the verification of quantum information processing tasks and quantify how well quantum information has been preserved. Moreover, metric estimation is also an integral part of quantum machine learning [3–5]. For example, metrics could play the role of the loss function in state learning tasks [27, 28].

The trace distance and fidelity are two typical metrics to quantify how close two quantum states are [29–31]. Given two quantum states ρ and σ, the trace distance $D(\rho,\sigma)$ and the fidelity $F(\rho,\sigma)$ are defined as follows:

$$D(\rho,\sigma) := \frac{1}{2}\|\rho-\sigma\|_1, \tag{1}$$

$$F(\rho,\sigma) := \mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} = \|\sqrt{\rho}\sqrt{\sigma}\|_1, \tag{2}$$

where $\|\cdot\|_1$ denotes the trace norm. When at least one of the states is pure, the task of fidelity estimation reduces to the simple case of calculating the square root of the state overlap $F(\rho,\sigma)=\sqrt{\mathrm{Tr}\ \rho\sigma}$ which can be obtained by the swap-test [32]. Thus the trace distance in this case is bounded by $1-F(\rho,\sigma)\leqslant D(\rho,\sigma)\leqslant\sqrt{1-F(\rho,\sigma)^2}$. However, evaluating these two metrics for mixed states is hard in general. One might attempt to classically compute the two metric from the matrix description of quantum states ρ and σ obtained via quantum state tomography. Nevertheless, this approach is infeasible due to the exponential growth of the matrix dimension with the number of qubits. On quantum computers, evaluating the trace distances is probably hard since even judging whether ρ and σ have large or small trace distance is known to be quantum statistical zero-knowledge (QSZK)-complete [33], where QSZK is a complexity class that includes BQP (bounded-error quantum polynomial time). Hence estimating fidelity and trace distance could probably be hard even for quantum computers.

Several approaches have been proposed for the trace distance and fidelity estimation, and here we focus on the case where the metric is between two unknown general quantum states. In [34], the authors variationally estimate the truncated fidelity via a hybrid classical–quantum algorithm. The truncated fidelity bounds the exact fidelity and is a good approximation of it when one of the states is known to have a low rank. For the estimation of trace distance, several methods are proposed but only applicable in specific quantum environments [35, 36]. To the best of our knowledge, no methods for estimating the trace distance on general NISQ devices has been proposed yet.

To overcome these challenges, we propose the variational trace distance estimation (VTDE) algorithm as well as the variational fidelity estimation (VFE) algorithm. First, we propose a method to estimate the trace norm of an arbitrary Hermitian matrix H, and apply it to trace distance estimation by specifying $H=\frac{1}{2}(\rho-\sigma)$. In particular, we prove that local measurements on an ancillary single qubit can extract the desired spectrum information of any Hermitian H conjugated by a unitary. By optimizing over all unitaries, we can obtain the trace norm of H. Our method notably only employs a local observable in the loss function evaluation, which saves us from the gradient vanishing issue with shallow circuits. Second, we introduce a method to estimate the fidelity of general quantum states, which utilizes the Uhlmann's theorem. We observe that the fidelity can be estimated by optimizing over all unitaries on an ancillary system via the freedom in purification. We also introduce a subroutine that works on NISQ devices to purify quantum states. With the performance analysis of the purification subroutine, we show that only few ancillary qubits are required if the unknown states are low-rank.

This paper is organized as follows. In section 2, we introduce the variational quantum algorithms for trace norm and trace distance estimation. In section 3, we introduce the variational quantum algorithms for quantum state fidelity estimation and its purification subroutine for mixed state learning. Numerical experiments and experimental implementations on IBM superconducting device are provided to show the validity of our methods. We finally deliver concluding remarks in section 4.

2. Variational trace distance estimation

This section introduces a VQA for estimating the trace norm of an arbitrary Hermitian matrix H, which could be easily applied to trace distance estimation. Our method employs the optimization of a PQC, requiring only one ancillary qubit initialized in an arbitrary pure state (typically $|0\rangle\langle 0|$ in practice) and single-qubit measurements. In this sense, our algorithm is practical and efficient for NISQ devices.

We will frequently use symbols such as $\mathcal{H}_A$ and $\mathcal{H}_B$ to denote Hilbert spaces associated with quantum systems A and B, respectively. We use d_A to denote the dimension of system A. The set of linear operators acting on A is denoted by $\mathcal{L}(\mathcal{H}_A)$. We usually write an operator with a subscript indicating the system that the operator acts on, such as M_{AB}, and write $M_A := \mathrm{Tr}_B M_{AB}$. Note that for a linear operator $X\in\mathcal{L}(\mathcal{H}_A)$, we define $|X|=\sqrt{X^\dagger X}$, and the trace norm of X is given by $\|X\|_1 = \mathrm{Tr}|X|$.

We first formulate the theory for the trace norm estimation, then we describe in detail the process of our algorithm, followed by the numerical experiments.

2.1. Estimating trace norm via one-qubit overlap maximization

Suppose on system $\mathcal{H}_A \otimes \mathcal{H}_B$ we have a Hermitian matrix H_{AB} that has spectral decomposition

$$H_{AB} = \sum_{j=1}^{d_A d_B} h_{AB}^j |\psi_j\rangle\langle\psi_j|, \tag{3}$$

with decreasing spectrum $\{h_{AB}^j\}_{j=1}^{d_A d_B}$ and orthonormal basis $\{|\psi_j\rangle\}$. Here, for technical convenience, we simply pad the spectrum with 0 s to ensure the expression in equation (3).

Before showing the main algorithm, we introduce an optimization method to obtain some information of the spectrum of a given Hermitian matrix as follows.

Proposition 1. *For any Hermitian matrix $H_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ with spectral decomposition as equation (3), respectively denote the dimension of $\mathcal{H}_A$, $\mathcal{H}_B$ by d_A, d_B. It holds that*

$$\max_U \operatorname{Tr} |0\rangle\langle 0|_A \widetilde{H}_A = \sum_{j=1}^{d_B} h_{AB}^j, \tag{4}$$

where $\widetilde{H}_A = \operatorname{Tr}_B \widetilde{H}_{AB}$, $\widetilde{H}_{AB} = U H_{AB} U^\dagger$, and the optimization is over all unitaries.

This proposition indicates that the optimization over unitaries can estimate the sum of several largest eigenvalues of H_{AB}.

The proof of proposition 1 can be found in appendix IA (https://stacks.iop.org/QST/7/015019/mmedia).

To apply proposition 1 to estimate the trace norm, we fix subsystem A to be single-qubit. In this case, for arbitrary Hermitian H_{AB} on $d = 2^n$-dimensional quantum system $\mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_A = \mathbb{C}^2$ (hence $\mathcal{H}_B = \mathbb{C}^{2^{n-1}}$), the maximal expectation of measurements on system A gives us the sum of the first half eigenvalues of H_{AB}. It is now quite close to the trace norm of H_{AB}, which is the sum of absolute values of its eigenvalues. We take the final step to estimate the trace norm by appending a one-qubit pure state and optimizing twice. Concretely, we derive the following theorem to ensure the validity of our algorithm.

Theorem 2 *For any Hermitian H_A on n-qubit system A, and any single-qubit pure state $|r\rangle$ on system R, it holds that*

$$\|H_A\|_1 = \max_{U^+} \operatorname{Tr} |0\rangle\langle 0|_R Q_R^+ + \max_{U^-} \operatorname{Tr} |0\rangle\langle 0|_R Q_R^-, \tag{5}$$

where $Q_R^\pm = \operatorname{Tr}_A Q_{AR}^\pm$, $Q_{AR}^\pm = U^\pm(\pm H_A \otimes |r\rangle\langle r|_R)U^{\pm\dagger}$, and each optimization is over unitaries on system AR.

This theorem shows how to generally evaluate the trace norm of arbitrary H by optimization. The proof of theorem 2 can be found in appendix IB.

We note that, since the target Hermitian matrix is in many cases (and can always be) written as the linear combination of density matrix with real coefficients c^j:

$$H_A = \sum_j c^j \rho_A^j, \tag{6}$$

we have $\operatorname{Tr} H_A = \sum_j c^j$. Employing the information of $\operatorname{Tr} H_A$ we can save one optimization as the following Corollary shows:

Corollary 3 *For any Hermitian H_A (written as equation (6)) on n-qubit system A, and any single-qubit pure state $|r\rangle$ on system R, it holds that*

$$\begin{aligned} \|H_A\|_1 &= \sum_{h_A^j > 0} h_A^j + \sum_{h_A^j < 0} -h_A^j \\ &= 2 \sum_{h_A^j > 0} h_A^j - \operatorname{Tr} H_A \\ &= 2 \max_U \operatorname{Tr} |0\rangle\langle 0|_R Q_R - \sum_j c^j, \end{aligned} \tag{7}$$

where $Q_R = \operatorname{Tr}_A Q_{AR}$, $Q_{AR} = U(H_A \otimes |r\rangle\langle r|_R)U^\dagger$, and the optimization is over all unitaries on system AR.

Also in this case, $\max_U \operatorname{Tr} |0\rangle\langle 0|_R Q_R$ in practice can be evaluated by post-processing:

$$\operatorname{Tr} |0\rangle\langle 0|_R Q_R = \sum_j c^j \operatorname{Tr} |0\rangle\langle 0|_R (U \rho_A^j \otimes |r\rangle\langle r|_R U^\dagger)_R. \tag{8}$$

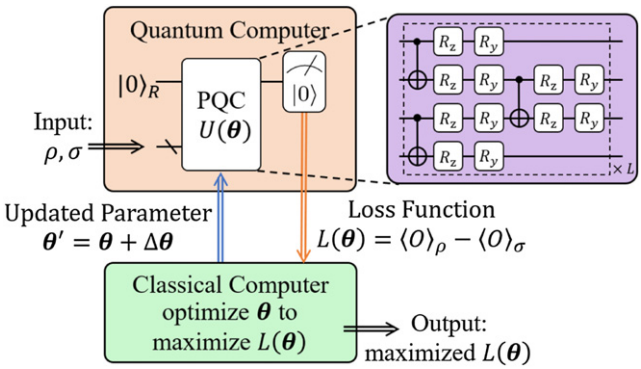

Figure 1. Diagram of VTDE. VTDE employs a single-qubit ancilla for arbitrary number of qubits of ρ and σ. The outputs $\langle O\rangle_\rho, \langle O\rangle_\sigma$ are obtained by locally measuring the overlap with state $|0\rangle_R$. Then the optimization of the measurement outcome is undertaken by a classical computer (the green box). The unitary evolution on the coupled system is simulated by a hardware-efficient ansatz, of which the single qubit rotation gates (R_y, R_z) are controlled by classical parameters $\boldsymbol{\theta}$.

Algorithm 1. Variational trace distance estimation (VTDE).

Input: quantum states ρ_A and σ_A, circuit ansatz of unitary $U_{AR}(\boldsymbol{\theta})$, number of iterations ITR;
Output: an estimate of trace distance $D(\rho_A, \sigma_A)$.
Initialize parameters $\boldsymbol{\theta}$.
Append ρ_A and σ_A with single-qubit state $|0\rangle_R$, respectively.
for itr $= 1, \ldots,$ ITR **do**
Apply $U_{AR}(\boldsymbol{\theta})$ to $\rho_A \otimes |0\rangle\langle 0|_R$ and $\sigma_A \otimes |0\rangle\langle 0|_R$, obtain the states $\widetilde{\rho}_{AR} = U_{AR}(\boldsymbol{\theta})\rho_A \otimes |0\rangle\langle 0|_R U_{AR}(\boldsymbol{\theta})^\dagger$ and $\widetilde{\sigma}_{AR} = U_{AR}(\boldsymbol{\theta})\sigma_A \otimes |0\rangle\langle 0|_R U_{AR}(\boldsymbol{\theta})^\dagger$, respectively.
Evaluate $O_\rho = \mathrm{Tr}\,|0\rangle\langle 0|_R\widetilde{\rho}_R$ and $O_\sigma = \mathrm{Tr}\,|0\rangle\langle 0|_R\widetilde{\sigma}_R$ by measurement on system R.
Compute the loss function $\mathcal{L}_1 := O_\rho - O_\sigma$.
Maximize the loss function $\mathcal{L}_1$ and update parameters $\boldsymbol{\theta}$.
end for
Output the optimized $\mathcal{L}_1$ as the trace distance estimate.

This means that we could use PQC, taking ρ_A^j as inputs, to perform the optimization for the trace norm estimation. In next section we will adopt this strategy to estimate the trace distance, for which $H = \frac{1}{2}(\rho - \sigma)$.

2.2. VTDE algorithm

Based on corollary 3, we are now ready to show our trace distance estimation (VTDE) algorithm for arbitrary quantum states ρ and σ. Figure 1 demonstrates the diagram of our algorithm.

Specifically, with $H_A = \frac{1}{2}(\rho_A - \sigma_A)$ and $\mathrm{Tr}H_A = 0$, equation (7) could be written as

$$\begin{aligned} D(\rho, \sigma) &= \frac{1}{2}\|\rho_A - \sigma_A\|_1 \\ &= \max_U \left(\mathrm{Tr}\,|0\rangle\langle 0|_R\widetilde{\rho}_R - \mathrm{Tr}\,|0\rangle\langle 0|_R\widetilde{\sigma}_R\right), \end{aligned} \tag{9}$$

where $\widetilde{\rho}_{AR} = U(\rho_A \otimes |r\rangle\langle r|_R)U^\dagger$, $\rho_R = \mathrm{Tr}_A\rho_{AR}$ and $\widetilde{\sigma}_{AR} = U(\sigma_A \otimes |r\rangle\langle r|_R)U^\dagger$, $\sigma_R = \mathrm{Tr}_A\sigma_{AR}$. This observation finally enables us to estimate the trace distance by optimization over unitaries, and the algorithm is given in algorithm 1 in detail. In our VTDE algorithm, $\mathrm{Tr}\,|0\rangle\langle 0|_R\widetilde{\rho}_R$ and $\mathrm{Tr}\,|0\rangle\langle 0|_R\widetilde{\sigma}_R$ are evaluated successively, then the difference between them are maximized by a classical computer. In practice, the ancillary pure state can be initially set as $|r\rangle = |0\rangle$. We adopt a hardware efficient ansatz [37, 38] consisting of parameterized single-qubit R_y and R_z rotations, along with CNOT gates on adjacent qubits as entanglement gates (also see figure 1) to implement the tunable unitary evolution.

From algorithm 1 one can tell that our VTDE estimates the trace distance for arbitrary quantum states without requiring any pre-knowledge. Furthermore, algorithm 1 and figure 1 imply that VTDE employs a single-qubit ancillary qubit for any input size n of ρ and σ, and perform local (single-qubit) measurement in each iteration. In this sense, VTDE is general and efficient for NISQ devices. Recall that algorithm 1 can be adapted to trace norm estimation for any H with decomposition $H = \sum_j c^j\rho^j$ based on corollary 3.

Given the VTDE, we briefly discuss analytic gradient and the gradient vanishing (barren plateau (BP) [39]) issues. Analytic gradient enables us to perform gradient descent to optimize the parameters. As our VTDE employs an hardware-efficient PQC, the parameter shift rule [40, 41] is capable in our case to obtain

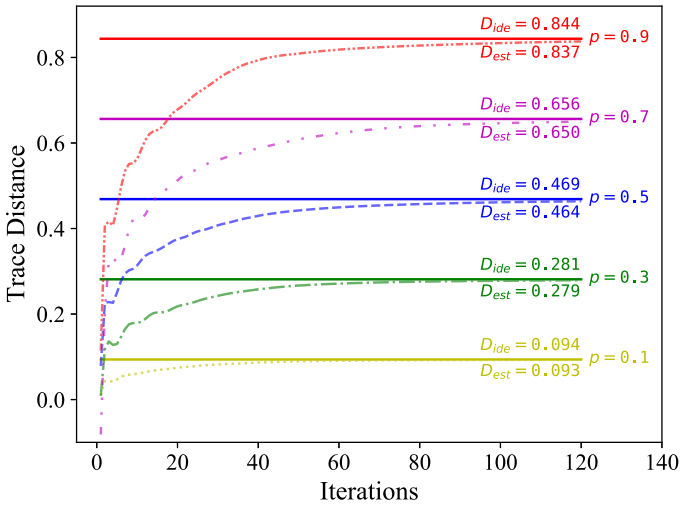

Figure 2. Learning processes of VTDE, where the input states ρ and σ are four-qubit GHZ state and its output after the depolarizing channel. The different colors correspond to considered depolarizing channel parameters $p = \{0.1, 0.3, 0.5, 0.7, 0.9\}$, with $p = 0.9$ being the top two red lines. We denote the ideal trace distances by D_{ide} and solid lines, and denote the estimated ones by D_{est} and dashed/dotted lines.

analytic gradient. As for the BP from which many variational quantum algorithms may suffer. We remark that our VTDE performs single-qubit measurements and takes the result as the loss function. This is essentially equivalent to a local observable, which has been proved to have at worst a polynomially vanishing gradient with a shallow PQC [42]. In this sense, our VTDE could avoid the BP issue when the number of layers $L \in \mathcal{O}(\log n)$.

We would like to remark that, the Naimark extension (see example 2.2 of [43]) together with the property of trace distance that

$$D(\rho, \sigma) = \max_{0 \leqslant P \leqslant 1} \text{Tr } P(\rho - \sigma), \tag{10}$$

also explain VTDE. Indeed our idea is to study the capability of PQCs in extracting spectrum information from local measurements. Nevertheless, VTDE can be generalized for trace norm estimation of arbitrary H (theorem 2 and corollary 3), where trace norm estimation in equation (10) does not hold if we simply replace $(\rho - \sigma)$ by a non-traceless H. This generalization can be lead to wider applications in quantum information like the estimation of log negativity. Our intermediate product (proposition 1) also indicates that the partial sum of the spectrum can be extracted by local measurements.

2.3. Numerical experiments

Numerical experiments are undertaken to demonstrate the validity and advantage of VTDE. All simulations including optimization loops are implemented via Paddle quantum [44] on the PaddlePaddle deep learning platform [45, 46].

We firstly estimate the trace distance between the four-qubit GHZ state $|\psi\rangle = \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle)$ and its output after the depolarizing channel:

$$\text{Dep}_p(\rho) = p \text{ Tr } \rho \frac{I}{2^4} + (1 - p)\rho. \tag{11}$$

In the experiment setting, we let $\rho = |\psi\rangle\langle\psi|$ and $\sigma = \text{Dep}_p(\rho)$, where the channel parameters are $p = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. For the hyper-parameters, we set ITR $= 120$ and learning rate LR $= 0.02$. Figure 2 shows the trace distance learned by VTDE versus the numbers of iterations. As one can see, for all considered channel parameters, the trace distance between ρ and σ can be estimated accurately within feasible iterations (< 120).

Next, we explore the required number of layers of the ansatz for input states with distinct ranks. Specifically, we set the number of qubits $n = 3$ and $\text{Rank}(\sigma) = 2$, while $\text{Rank}(\rho)$ ranges from 1 to $2^n = 8$. We randomly sample 100 states for each number of ranks of σ, and compute the accuracy for circuits with one, two and four layers. Here the accuracy is defined as Acc. $= D_{\text{est}}/D_{\text{ide}}$, where $D_{\text{est,ide}}$ are the estimated and ideal trace distances, respectively. Hyper-parameters are the same as the previous experiment. Figure 3 summarizes the result, telling that the behavior of a four-layer circuit is more accurate and stable, while

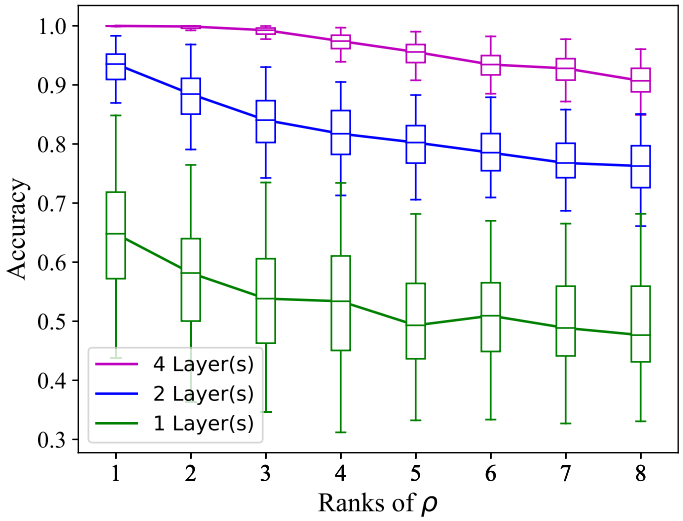

Figure 3. Trace distance learned by VTDE with different number of layers versus the rank of ρ. The curves connect the median values in the boxplots displaying the distribution of the estimation of D between 100 sampled state pairs. Each curve corresponds to the one-, two-, or four-layer case.

Table 1. Error analysis for VTDE between $\rho = |+\rangle$ and $\sigma = \text{Deph}_p(\rho)$. We choose $p = 0.7$ and repeat the experiment 10 times independently.

Backends	Mean	Variance	Error rate
Theoretical value	0.7	—	—
Simulator	0.703 22	0.000 10	0.46%
ibmq_quito	0.643 12	0.000 25	8.13%

circuits with fewer layers perform badly and unstably for any rank of ρ, possibly due to the lack in expressibility [47].

2.4. Implementation on superconducting quantum processor

We also apply our VTDE to estimate the trace distance between $|+\rangle$ state and itself affected by a dephasing channel:

$$\text{Deph}_p(\rho) = pZ\rho Z + (1-p)\rho. \tag{12}$$

The result with comparison to the values achieved from simulation are listed in table 1. The experiments are performed on the *IBM quantum* platform, loading the quantum device *ibmq_quito* containing five qubits.

The input state we choose here are $\rho = |+\rangle$ and $\sigma = \text{Deph}_{0.7}(\rho)$. On the quantum device, the input state σ is prepared by applying two Ry gates on the working and ancilla qubits with parameters $\pi/2$ and $2\arcsin\sqrt{0.7} \approx 1.982$ respectively, followed by a control-Z gate, then discarding the ancilla. Next, we operate PQC on qubits, and use sequential minimal optimization [48] to optimize the parameters until the cost converge to its maximum. We repeat ten independent experiments with same input states and randomly initialized parameters.

As demonstrated in table 1, the estimates converges stably for the simulator and the quantum device. Correspondingly, the achieved trace distance estimates by simulator (0.7032) is closer to the theoretical values (0.7) than that of quantum device (0.643 12), which may caused by the quantum device noises.

3. Variational fidelity estimation

In this section, we introduce VFE as a hybrid quantum–classical algorithm for estimating fidelity in the most general case where two mixed states are provided. The intuition of VFE lies in Uhlmann's theorem and the freedom in purification, based on which we prove that optimization over unitaries can obtain the fidelity. The purification of each quantum state is required by the optimization, for which we design a variational quantum state learning (VQSL) algorithm as a subroutine. Both VFE and VQSL employs PQC to implement the optimization over unitaries. We give the theory behind and the process of VFE and VQSL,

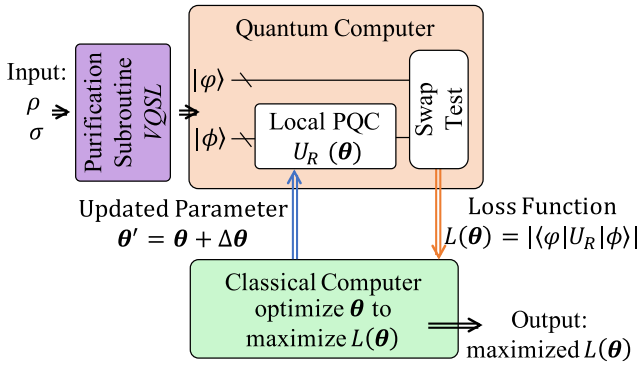

Figure 4. Workflow of VFE. VFE employs a purification subroutine to purify ρ and σ. The outputs $\langle\psi|U_R|\phi\rangle$ are obtained by the swap test. Then the optimization of the measurement outcome is undertaken by a classical computer (the green box). The unitary evolution on the ancillary system is realized by a hardware efficient ansatz.

respectively, followed by experimental results from a simulator and a five-qubit IBMQ superconducting processor. The workflow of VFE is shown in figure 4.

3.1. Theory and algorithm for fidelity estimation

Suppose we have two unknown quantum states, ρ_A and σ_A, on system A. With arbitrary purification of ρ_A and σ_A (denoted by $|\psi\rangle_{AR}$ and $|\phi\rangle_{AR}$, respectively), corollary 6 allows us to estimate the fidelity of ρ_A and σ_A. As the prerequisite of corollary 6, we first introduce Uhlmann's theorem and the freedom in purification:

Theorem 4 (Uhlmann's theorem [29]). *Suppose ρ and σ are states of quantum system A. Then, for given quantum purification $|\psi\rangle_{AR}$ of ρ_A on system AR,*

$$F(\rho_A, \sigma_A) = \max_{|\phi\rangle_{AR}} |\langle\psi|\phi\rangle_{AR}|, \tag{13}$$

where the maximization is over all purifications $|\phi\rangle_{AR}$ of σ_A on system AR.

Lemma 5 (Freedom in purification [29]). *For two purifications $|\phi_1\rangle_{AR}, |\phi_2\rangle_{AR}$ of σ_A on system AR, there exists a unitary transformation U_R such that*

$$|\phi_2\rangle_{AR} = (I_A \otimes U_R)|\phi_1\rangle_{AR}, \tag{14}$$

Based on the fundamental properties of quantum fidelity discussed above, we observe that the following optimization problem characterizes the fidelity function.

Corollary 6 *For any quantum states ρ_A and σ_A on system A, and arbitrary purification $|\psi\rangle_{AR}$ and $|\phi\rangle_{AR}$ of ρ_A and σ_A, it holds that*

$$F(\rho, \sigma) = \max_{U_R} |\langle\psi|_{AR}(I_A \otimes U_R)|\phi\rangle_{AR}|, \tag{15}$$

where the optimization is over any unitaries on system R.

Corollary 6 gives us an elegant variational representation of the fidelity function that only requires purification of input states and optimization over the ancillary system. Following this line of reasoning, we design a VQA to estimate the fidelity of quantum states ρ and σ. We note that in algorithm 2 the procedure of purifying ρ and σ is realized by VQSL as a subroutine, which will be presented in section 3.2. Our method formulates the problem of directly calculating the fidelity between two mixed states into an optimization procedure over the ancillary system of two purified states. At the cost of extra subroutines, this approach could deal with arbitrary high-rank states as long as we provide enough ancillary qubits. For numerical performance analysis, we refer to sections 3.3 and 3.4. Besides, we also derive the analytical gradient of this loss function in appendix III, which is required for gradient-based classical optimization.

3.2. Subroutine—quantum state learning and purification

The task of quantum state learning is to find the correct unitary operation such that one could prepare any target mixed state ρ from an initialized state (usually $|0\rangle\langle 0|$ on each qubit). In reference [49], the method of state learning for a pure state ρ, where $\mathrm{Rank}(\rho) = 1$, is proposed. Our approach further generalizes it to work for a mixed state ρ, where $\mathrm{Rank}(\rho) \geqslant 1$. In order to implement our algorithm in NISQ devices, we

Algorithm 2. Variational fidelity estimation (VFE).

Input: quantum states ρ_A and σ_A, circuit ansatz of unitary $U_R(\boldsymbol{\theta})$, number of iterations ITR;
Output: an estimate of fidelity $F(\rho_A, \sigma_A)$.
Use Purification Subroutine to learn the purified state $|\psi\rangle_{AR}$ of ρ_A and the purified state $|\phi\rangle_{AR}$ of σ_A.
Initialize parameters $\boldsymbol{\theta}$.
for itr $= 1, \ldots,$ ITR **do**
Apply $U_R(\boldsymbol{\theta})$ to $|\phi\rangle_{AR}$ and obtain the resulting state $|\widetilde{\phi}\rangle_{AR} = I_A \otimes U_R(\boldsymbol{\theta})|\phi\rangle_{AR}$.
Compute the loss function $\mathcal{L}_2(\boldsymbol{\theta}) := |\langle\widetilde{\phi}|\psi\rangle|_{AR}$.
Maximize the loss function and update parameters $\boldsymbol{\theta}$;
end for
Output the optimized $\mathcal{L}_2$ as the final fidelity estimation;

consider quantum state learning via PQC and the framework of VQA. The main issue is how to design a faithful loss function $\mathcal{L}_3$ to guide the learning direction. Such a loss function should be able to quantify the closeness between target ρ and prepared state χ. For pure state cases, simply maximize the state overlap $\mathrm{Tr}\,\rho\chi$ could help us learn the target state. But for general mixed states, this is not the case. Consider a counter example of $\rho = I/2$ and χ being a random one-qubit mixed state. The overlap is always $\mathrm{Tr}\,\rho\chi = 1/2$ but this is clearly not the correct distance measure between ρ and χ. On the other hand, the Hilbert–Schmidt norm defined as follows could be a good candidate:

$$\Delta(\rho, \chi) := \|\rho - \chi\|_2^2 = \mathrm{Tr}\,(\rho - \chi)^2. \tag{16}$$

As the unknown state ρ is fixed, we have

$$\begin{aligned} &\mathrm{argmin}_\chi \Delta(\rho, \chi) \\ &= \mathrm{argmin}_\chi\ \mathrm{Tr}\ \rho^2 + \mathrm{Tr}\ \chi^2 - 2\ \mathrm{Tr}\ \rho\chi \\ &= \mathrm{argmin}_\chi\ \mathrm{Tr}\ \chi^2 - 2\ \mathrm{Tr}\ \rho\chi, \end{aligned} \tag{17}$$

where $0 < \mathrm{Tr}\,\rho^2 < 1$ for mixed state ρ is a constant and hence does not influence the optimization direction. Therefore, we choose our loss function to be

$$\min_{\boldsymbol{\theta}} \mathcal{L}_3(\boldsymbol{\theta}) := \mathrm{Tr}\ \chi^2 - 2\ \mathrm{Tr}\ \rho\chi. \tag{18}$$

Note that this loss function can be implemented on near-term devices since the state overlap can be computed via the swap test [32, 50]. As a brief reminder, the swap test evaluates the overlap of two arbitrary states by a single qubit measurement after a combination of Hadamard gates and controlled-swap gates. Evidence has been found that the swap test has a simple physical implementation in quantum optics [51, 52] and can be experimentally implemented on near-term quantum hardware [53–55].

After discussing the general picture of state learning, we introduce its application in learning the purification of quantum state ρ_A on system A. This can be done by providing an ancillary qubit system R with dimension d_R and initializing the complete system with $|\chi_0\rangle = |00\rangle_{AR}$. Then we apply a parametrized unitary operation $U_{AR}(\boldsymbol{\theta})$ to drive the system and use classical optimization methods to minimize equation (18). In the case of learning purification, we need to set

$$\chi_A = \mathrm{Tr}_R \left[U_{AR}(\boldsymbol{\theta})|\chi_0\rangle\langle\chi_0|U_{AR}^\dagger(\boldsymbol{\theta}) \right], \tag{19}$$

where the symbol Tr_R denotes the partial trace operation with respect to the ancillary system R. With the above set up, we introduce a VQA to learn the purification of a quantum state ρ_A as follows.

We want to emphasize that the scope of state learning is much broader than the purification learning task. One could further develop this approach for quantum state preparation and many other applications. For our purpose, learning a purified state to estimate the fidelity is enough. At the cost of introducing ancillary qubits, one could prepare a purification $|\psi\rangle_{AR}$ of target mixed state ρ_A with high fidelity. Then, it will be important to study the performance of VQSL with different ancillary dimensions d_R. The following analysis suggests that only a few ancillary qubits are necessary for low-rank states. For best performance, one should choose the dimension of the ancillary system to be the same as the original system A such that $d_R = d_A$.

Proposition 7 *Suppose the input state ρ_A has the spectral decomposition $\rho_A = \sum_{j=1}^k \lambda_j |\psi_j\rangle\langle\psi_j|$ with decreasing spectrum $\{\lambda_j\}_{j=1}^k$. There exists a quantum circuit U_{AR} that generates χ_A from $|00\rangle_{AR}$ to approximate the target*

Algorithm 3. Variational quantum state learning (VQSL).

Input: quantum states ρ_A, circuit ansatz of unitary $U_{AR}(\boldsymbol{\theta})$, number of iterations ITR;
Output: purification of ρ_A.
Initialize parameters $\boldsymbol{\theta}$.
for itr $= 1, \ldots,$ ITR **do**
Apply $U_{AR}(\boldsymbol{\theta})$ to three equivalent initial states $|\chi_0\rangle$ on system AR and obtain the resulting partial states:
$\chi_A^1 = \chi_A^2 = \chi_A^3 = \mathrm{Tr}_R\left[U_{AR}(\boldsymbol{\theta})|\chi_0\rangle\langle\chi_0|U_{AR}^\dagger(\boldsymbol{\theta})\right]$;
Measure the overlap $\langle O\rangle_1 = \mathrm{Tr}(\chi_A^1\chi_A^2)$ via swap test;
Measure the overlap $\langle O\rangle_2 = \mathrm{Tr}(\rho_A\chi_A^3)$ via swap test;
Compute the loss function $\mathcal{L}_3(\boldsymbol{\theta}) = \langle O\rangle_1 - 2\langle O\rangle_2$;
Perform optimization for $\mathcal{L}_3(\boldsymbol{\theta})$ and update parameters $\boldsymbol{\theta}$;
end for
Output the final purified state $|\psi\rangle_{AR} = U_{AR}(\boldsymbol{\theta}^*)|00\rangle_{AR}$;

state ρ_A with a fidelity that satisfies

$$F(\rho_A, \chi_A) = \begin{cases} 1 & \text{if } k \leqslant d_R, \\ \sqrt{\sum_{j=1}^{d_R}\lambda_j} & \text{otherwise,} \end{cases} \tag{20}$$

where the second case is at least $\sqrt{d_R/k}$.

Proof. Assume $d_R \leqslant d_A$. The goal is to solve

$$\max_{U_{AR}} F(\rho_A, \chi_A), \tag{21}$$

where $\chi_A = \mathrm{Tr}_R\left[U_{AR}|00\rangle\langle 00|_{AR}U_{AR}^\dagger\right]$ is the reduced state on the subsystem A. For convenience, denote the state $U_{AR}|00\rangle_{AR}$ by $|\psi\rangle$ and hence

$$\chi_A = \mathrm{Tr}_R|\psi\rangle\langle\psi|_{AR}. \tag{22}$$

Since the Schmidt rank of $|\psi\rangle$ is at most d_R, the rank of χ_A is also at most d_R.

If $k \leqslant d_R$, one could choose unitary U_{AR} such that $|\psi\rangle = U_{AR}|00\rangle_{AR} = \sum_{j=1}^k\sqrt{\lambda_j}|\psi_j\rangle_A|\varphi\rangle_{j_R}$, where $\{|\psi_j\rangle_A\}$ and $\{|\varphi_j\rangle_R\}$ are orthonormal state sets. This ensures that $\chi_A = \mathrm{Tr}_R|\psi\rangle\langle\psi| = \sum_j\lambda_j|\psi_j\rangle\langle\psi_j| = \rho_A$ and leads to a fidelity $F(\rho_A, \chi_A) = 1$. On the other hand, if $k > d_R$, there exists an unitary U such that $|\psi\rangle = U_{AR}|00\rangle_{AR} = \sum_{j=1}^{d_R}\sqrt{\xi_j}|\psi_j\rangle_A|\varphi_j\rangle_R$. In particular, one could further choose $\xi_j = \frac{\lambda_j}{\eta}$ for $j = 1, \ldots, d_R$, where the denominator is defined as $\eta = \sum_{\ell=1}^{d_R}\lambda_\ell$. Then, one can calculate the fidelity F as

$$\begin{aligned} F(\rho_A, \chi_A) &= \mathrm{Tr}\sqrt{\chi_A^{1/2}\rho_A\chi_A^{1/2}} \\ &= \sum_{j=1}^{d_R}\sqrt{\lambda_j^2/\eta} \\ &= \sum_{j=1}^{d_R}\lambda_j/\sqrt{\eta} \\ &= \sqrt{\eta} \geqslant \sqrt{\frac{d_R}{k}}. \end{aligned} \tag{23}$$

■

3.3. Numerical experiments and trainability

In this section, we conduct simulations to investigate the performance of VFE for state fidelity estimation and its subroutine VQSL for purification learning. The parametrized quantum circuit $U_{AR}(\boldsymbol{\theta})$ used for algorithm 3 VQSL and $U_R(\boldsymbol{\theta})$ for algorithm 2 are recorded in appendix IV. All simulations including optimization loops are implemented via Paddle quantum [44] on the PaddlePaddle deep learning platform [45, 46].

We firstly generate ten random pairs of full-rank density matrices $\{\rho_A^{(1)}, \sigma_A^{(1)}|\cdots|\rho_A^{(10)}, \sigma_A^{(10)}\}$ for $n_A = \{1, 2, 3\}$ qubits and calculate the deviation ΔF between the target fidelity and estimated fidelity. We choose $d_R = d_A$ for best performance in purification learning. The results are summarized in table 2. The

Table 2. Error analysis for fidelity estimation VFE between randomly generated full-rank density operators. For purification learning, the Adam optimizer [57] is adopted and hyper-parameters are taken to be depth $L = 6$, learning rate LR $= 0.2$, and iteration loops ITR $= 100$ (the latter two hyper-parameters are same for VFE).

Qubit number #	$n_A = 1$	$n_A = 2$	$n_A = 3$
Average error rate $\mathbb{E}[\Delta F]$	0.1694%	0.2476%	0.1960%
Standard deviation $\sigma[\Delta F]$	0.1594%	0.1457%	0.1388%

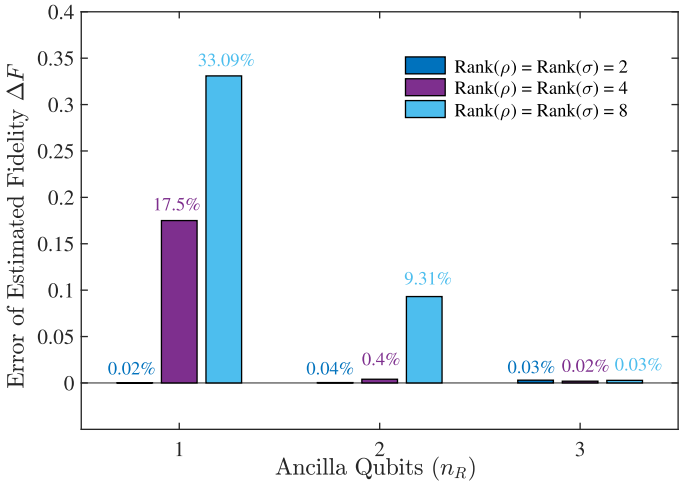

Figure 5. Performance of VFE given two randomly generated three-qubit mixed states ρ, σ with different number of ancilla qubits n_R.

average error rate can be reduced to the level of $<0.5\%$. The maximum average error $\mathbb{E}_{\max}[\Delta F] \approx 0.2476\%$ happens at $n_A = 2$, and no clear scaling phenomenon of average error is observed.

Next, we conduct simulations to qualitatively study the influence of limited ancillary qubits given high-rank states $k > d_R$. Three random pairs of density matrices, $\{\rho_A^{(1)}, \sigma_A^{(1)} | \rho_A^{(2)}, \sigma_A^{(2)} | \rho_A^{(3)}, \sigma_A^{(3)}\}$, are prepared with Rank $= \{2, 4, 8\}$ respectively. Then, we test the performance of VFE with different number of ancillary qubits $n_R = \{1, 2, 3\}$. The results are summarized in figure 5. By providing enough ancilla qubits, VFE could estimate the fidelity between two arbitrary full rank mixed states. If not, the purification subroutine will be restricted to produce pure states with fidelity bounded by equation (20), which is consistent with the numerical simulation. This scalability requirement offers the flexibility of providing few ancilla qubits for low-rank states but can bring challenges when accurately evaluating the fidelity between two high-rank quantum state in large dimensions.

Here we discuss the trainability issues for VFE and possible solutions. With an extended system dimension introduced by the purification, VFE might exhibit a BP [39] due to a global loss function. This would result an exponentially suppressed gradient with respect to the problem dimension, a phenomenon known as the BP. Reference [56] shows BP is independent of the optimization methods, meaning that simply change into a gradient-free optimizer would not help. In order to mitigate this trainability issue, several approaches have been proposed such as the parameter initialization and correlation strategies [57, 58], layer-wise learning [59], and variable structure ansatz [60–62]. We leave the adaptation of these strategies to VFE for future study.

3.4. Implementation on superconducting quantum processor

We also apply our VFE to estimate the fidelity between states which are $|+\rangle$ states affected by dephasing channel Deph_p with different intensity p. The result with comparison to the values achieved from simulation are listed in table 3. The experiments are performed on the *IBM quantum* platform, loading the quantum device *ibmq_quito* containing five qubits.

The input state we choose here are $\rho = \text{Deph}_{p_1}(|+\rangle)$ and $\sigma = \text{Deph}_{p_2}(|+\rangle)$, where $p_1 = 0.2$ and $p_2 = 0.9$. On the quantum device, the input states are prepared similarly to the preparation in section 2.4 and we keep the ancilla as a purification. Next, we operate PQC on the ancilla, and use sequential minimal

Table 3. Error analysis for VFE between $\rho = \mathrm{Deph}_{p_1}(|+\rangle)$ and $\sigma = \mathrm{Deph}_{p_2}(|+\rangle)$. We choose $p_1 = 0.2$ and $p_2 = 0.9$, and repeat the experiment 10 times independently.

Backends	Mean	Variance	Error rate
Theoretical value	0.70710	—	—
Simulator	0.70721	0.00002	0.016%
ibmq_quito	0.71692	0.00003	1.388%

optimization [48] to optimize the parameters until the loss function converges to its maximum. We repeat ten independent experiments with same input states and randomly initialized parameters. As demonstrated in table 3, the estimated fidelity from our method are close to the ideal value with stable performance due to a very small variance.

4. Conclusion and outlook

In this work, we have introduced near-term quantum algorithms VTDE and VFE to estimate trace distance and quantum fidelity. A strength of our algorithms is that they estimate the metrics directly rather than estimating their bounds. Our algorithms also do not require any assumption on the unknown input states. These algorithms are executable on near-term quantum devices equipped with PQCs. In particular, VTDE could be easily generalized for trace norm estimation for any Hermitian matrix and could avoid the BP issue with logarithmic depth parameterized circuits.

Beyond benchmarking the quantum algorithms' behavior, our VTDE and VFE could have a wide range of applications in quantum information processing. A direct extension of VTDE might be the estimation for the diamond norm [63–65], which is a widely-used distance measure for quantum channels. The trace distance can also be applied to quantify the Bell non-locality [66], quantum entanglement [67], and the security of quantum cryptography protocols [68], where VTDE could be utilized as a practical subroutine. VFE may be applied to evaluate the conditional quantum mutual information of tripartite quantum states [69]. It is also of great interest to have a further study on the estimation of sandwiched/geometric Rényi relative entropies [70–72] as an extension of VFE. Efficient estimations of these distance measures could be used to further evaluate the resource measures of entanglement, coherence, magic and other resources in quantum information [73–83].

Acknowledgments

We thank Runyao Duan, Yuao Chen, and Mark M Wilde for helpful discussions. RC and ZS contributed equally to this work. This work was done when RC, ZS, and XZ were research interns at Baidu Research.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Xin Wang https://orcid.org/0000-0002-0641-3186

References

[1] Preskill J 2018 Quantum computing in the NISQ era and beyond *Quantum* **2** 79
[2] McClean J R, Romero J, Babbush R and Aspuru-Guzik A 2016 The theory of variational hybrid quantum–classical algorithms *New J. Phys.* **18** 023023
[3] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N and Lloyd S 2017 Quantum machine learning *Nature* **549** 195–202
[4] Schuld M, Sinayskiy I and Petruccione F 2015 An introduction to quantum machine learning *Contemp. Phys.* **56** 172–85
[5] Arunachalam S and de Wolf R 2017 Guest column *ACM SIGACT News* **48** 41–67
[6] Benedetti M, Lloyd E, Sack S and Fiorentini M 2019 Parameterized quantum circuits as machine learning models *Quantum Sci. Technol.* **4** 043001
[7] Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 A variational eigenvalue solver on a photonic quantum processor *Nat. Commun.* **5** 4213

[8] Nakanishi K M, Mitarai K and Fujii K 2019 Subspace-search variational quantum eigensolver for excited states *Phys. Rev. Res.* **1** 033062
[9] Sharma K, Cerezo M, Cincio L and Coles P J 2020 Trainability of dissipative perceptron-based quantum neural networks (arXiv:2005.12458)
[10] Schuld M, Bocharov A, Svore K M and Wiebe N 2020 Circuit-centric quantum classifiers *Phys. Rev.* A **101** 032308
[11] Grant E, Benedetti M, Cao S, Hallam A, Lockhart J, Stojevic V, Green A G and Severini S 2018 Hierarchical quantum classifiers *npj Quantum Inf.* **4** 17–9
[12] Li G, Song Z and Wang X 2021 VSQL: variational shadow quantum learning for classification *Proc. AAAI Conf. Artificial Intelligence* vol 35 pp 8357–65
[13] Yuan X, Endo S, Zhao Q, Li Y and Benjamin S C 2019 Theory of variational quantum simulation *Quantum* **3** 191
[14] Wu J and Hsieh T H 2019 Variational thermal quantum simulation via thermofield double states *Phys. Rev. Lett.* **123** 220502
[15] Wang Y, Li G and Wang X 2020 Variational quantum Gibbs state preparation with a truncated Taylor series (arXiv:2005.08797)
[16] Chowdhury A N, Hao Low G and Wiebe N 2020 A variational quantum algorithm for preparing quantum Gibbs states (arXiv:2002.00055)
[17] Zhao X, Zhao B, Wang Z, Song Z and Wang X 2021 LOCCNet: a machine learning framework for distributed quantum information processing *npj Quantum Inf.* **7** 159
[18] Xu X, Sun J, Endo S, Li Y, Benjamin S C and Yuan X 2019 Variational algorithms for linear algebra (arXiv:1909.03898)
[19] Bravo-Prieto C, LaRose R, Cerezo M, Yigit S, Cincio L and Coles P J 2019 Variational quantum linear solver (arXiv:1909.05820)
[20] Huang H-Y, Bharti K and Rebentrost P 2019 Near-term quantum algorithms for linear systems of equations (arXiv:1909.07344)
[21] Wang X, Song Z and Wang Y 2021 Variational quantum singular value decomposition *Quantum* **5** 483
[22] Bravo-Prieto C, Garcia-Martin D and Latorre J I 2020 Quantum singular value decomposer *Phys. Rev.* A **101** 062310
[23] Endo S, Cai Z, Benjamin S C and Yuan X 2020 Hybrid quantum-classical algorithms and quantum error mitigation (arXiv:2011.01382)
[24] Cerezo M *et al* 2021 Variational quantum algorithms *Nat. Rev. Phys.* **3** 625–44
[25] Bharti K *et al* 2021 Noisy intermediate-scale quantum (NISQ) algorithms (arXiv:2101.08448)
[26] Terhal B M 2015 Quantum error correction for quantum memories *Rev. Mod. Phys.* **87** 307–46
[27] Hu L *et al* 2019 Quantum generative adversarial learning in a superconducting quantum circuit *Sci. Adv.* **5** eaav2761
[28] Braccia P, Caruso F and Banchi L 2020 How to enhance quantum generative adversarial learning of noisy information (arXiv:2012.05996)
[29] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[30] Jozsa R 1994 Fidelity for mixed quantum states *J. Mod. Opt.* **41** 2315–23
[31] Uhlmann A 1976 The 'transition probability' in the state space of a $*$-algebra *Rep. Math. Phys.* **9** 273–9
[32] Buhrman H, Cleve R, Watrous J and de Wolf R 2001 Quantum fingerprinting *Phys. Rev. Lett.* **87** 167902
[33] Watrous J 2008 Quantum computational complexity (arXiv:0804.3401)
[34] Cerezo M, Poremba A, Cincio L and Coles P J 2020 Variational quantum fidelity estimation *Quantum* **4** 248
[35] Zhang J, Ruggiero P and Calabrese P 2019 Subsystem trace distance in quantum field theory *Phys. Rev. Lett.* **122** 141602
[36] Smirne A, Brivio D, Cialdi S, Vacchini B and Paris M G A 2011 Experimental investigation of initial system–environment correlations via trace-distance evolution *Phys. Rev.* A **84** 032112
[37] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets *Nature* **549** 242–6
[38] Commeau B, Cerezo M, Holmes Z, Cincio L, Coles P J and Sornborger A 2020 Variational Hamiltonian diagonalization for dynamical quantum simulation (arXiv:2009.02559)
[39] McClean J R, Boixo S, Smelyanskiy V N, Ryan B and Neven H 2018 Barren plateaus in quantum neural network training landscapes *Nat. Commun.* **9** 4812
[40] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 Quantum circuit learning *Phys. Rev.* A **98** 032309
[41] Schuld M, Bergholm V, Gogolin C, Izaac J and Killoran N 2019 Evaluating analytic gradients on quantum hardware *Phys. Rev.* A **99** 032331
[42] Cerezo M, Sone A, Volkoff T, Cincio L and Coles P J 2020 Cost-function-dependent barren plateaus in shallow quantum neural networks (arXiv:2001.00550)
[43] Wilde M M 2013 Sequential decoding of a general classical–quantum channel *Proc. R. Soc.* A **469** 20130259
[44] 2020 *Paddle Quantum* https://qml.baidu.com/
[45] Parallel distributed deep LEarning: machine learning framework from industrial practice.
[46] Martseniuk V, Milian N, Wu T and Wang H 2019 PaddlePaddle: an open-source deep learning platform from industrial practice *Front. Data Domput.* **1** 105–15
[47] Sim S, Johnson P D and Aspuru-Guzik A 2019 Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum–classical algorithms *Adv Quantum Technol.* **2** 1900070
[48] Nakanishi K M, Fujii K and Todo S 2020 Sequential minimal optimization for quantum–classical hybrid algorithms *Phys. Rev. Res.* **2** 043158
[49] Lee S M, Lee J and Bang J 2018 Learning unknown pure quantum states *Phys. Rev.* A **98** 052302
[50] Gottesman D and Chuang I 2001 Quantum digital signatures (arXiv:quant-ph/0105032)
[51] Ekert A K, Alves C M, Oi D K L, Horodecki M, Horodecki P and Kwek L C 2002 Direct estimations of linear and nonlinear functionals of a quantum state *Phys. Rev. Lett.* **88** 217901
[52] Garcia-Escartin J C and Chamorro-Posada P 2013 Swap test and Hong–Ou–Mandel effect are equivalent *Phys. Rev.* A **87** 052330
[53] Islam R, Ma R, Preiss P M, Eric Tai M, Lukin A, Rispoli M and Greiner M 2015 Measuring entanglement entropy in a quantum many-body system *Nature* **528** 77–83
[54] Patel R B, Ho J, Ferreyrol F, Ralph T C and Pryde G J 2016 A quantum Fredkin gate *Sci. Adv.* **2** e1501531
[55] Linke N M, Johri S, Figgatt C, Landsman K A, Matsuura A Y and Monroe C 2018 Measuring the Rényi entropy of a two-site Fermi–Hubbard model on a trapped ion quantum computer *Phys. Rev.* A **98** 052334
[56] Arrasmith A, Cerezo M, Czarnik P, Cincio L and Coles P J 2020 Effect of barren plateaus on gradient-free optimization (arXiv:2011.12245)
[57] Grant E, Wossnig L, Ostaszewski M and Benedetti M 2019 An initialization strategy for addressing barren plateaus in parametrized quantum circuits *Quantum* **3** 214

[58] Volkoff T and Coles P J 2021 Large gradients via correlation in random parameterized quantum circuits *Quantum Sci. Technol.* **6** 025008
[59] Skolik A, McClean J R, Mohseni M, van der Smagt P and Leib M 2021 Layerwise learning for quantum neural networks *Quantum Mach. Intell.* **3** 1–11
[60] Grimsley H R, Economou S E, Barnes E and Mayhall N J 2019 An adaptive variational algorithm for exact molecular simulations on a quantum computer *Nat. Commun.* **10** 1–9
[61] Rattew A G, Hu S, Pistoia M, Chen R and Wood S 2019 A domain-agnostic, noise-resistant, hardware-efficient evolutionary variational quantum eigensolver (arXiv:1910.09694)
[62] Bilkis M, Cerezo M, Verdon G, Coles P J and Cincio L 2021 A semi-agnostic ansatz with variable structure for quantum machine learning (arXiv:2103.06712)
[63] Kitaev A and Watrous J 2000 Parallelization, amplification, and exponential time simulation of quantum interactive proof systems *Proc. 32nd annual ACM Symp. Theory of computing—STOC '00* (New York: ACM Press) pp 608–17
[64] Watrous J 2009 Semidefinite programs for completely bounded norms *Theory Comput.* **5** 217–38
[65] Watrous J 2012 Simpler semidefinite programs for completely bounded norms (arXiv:1207.5726)
[66] Brito S G A, Amaral B and Chaves R 2018 Quantifying bell nonlocality with the trace distance *Phys. Rev.* A **97** 022111
[67] Vidal G and Werner R F 2002 Computable measure of entanglement *Phys. Rev.* A **65** 032314
[68] Yuen H 2014 What the trace distance security criterion in quantum key distribution does and does not guarantee (arXiv:1410.6945)
[69] Berta M and Tomamichel M 2016 The fidelity of recovery is multiplicative *IEEE Trans. Inf. Theory* **62** 1758–63
[70] Müller-Lennert M, Dupuis F, Szehr O, Fehr S and Tomamichel M 2013 On quantum Rényi entropies: a new generalization and some properties *J. Math. Phys.* **54** 122203
[71] Wilde M M, Winter A and Yang D 2014 Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy *Commun. Math. Phys.* **331** 593–622
[72] Matsumoto K 2018 A new quantum version of f-divergence Reality and Measurement in Algebraic Quantum Theory (Nagoya, Japan, 9–13 March 2018) pp 229–73
[73] Chitambar E and Gour G 2019 Quantum resource theories *Rev. Mod. Phys.* **91** 025001
[74] Plenio M B and Virmani S S 2007 An introduction to entanglement measures *Quantum Inf. Comput.* **7** 1–51
[75] Wang X and Duan R 2017 Irreversibility of asymptotic entanglement manipulation under quantum operations completely preserving positivity of partial transpose *Phys. Rev. Lett.* **119** 180506
[76] Vedral V 2002 The role of relative entropy in quantum information theory *Rev. Mod. Phys.* **74** 197
[77] Wang X and Wilde M M 2020 Cost of quantum entanglement simplified *Phys. Rev. Lett.* **125** 040502
[78] Datta N 2009 Min- and max-relative entropies and a new entanglement monotone *IEEE Trans. Inf. Theory* **55** 2816–26
[79] Wang K, Wang X and Wilde M M 2019 Quantifying the unextendibility of entanglement (arXiv:1911.07433)
[80] Yuan X, Liu Y, Zhao Q, Regula B, Thompson J and Gu M 2019 Universal and operational benchmarking of quantum memories (arXiv:1907.02521)
[81] Streltsov A, Adesso G and Plenio M B 2017 Colloquium: quantum coherence as a resource *Rev. Mod. Phys.* **89** 041003
[82] Wang X, Wilde M M and Su Y 2020 Efficiently computable bounds for magic state distillation *Phys. Rev. Lett.* **124** 090505
[83] Veitch V, Hamed Mousavian S A, Gottesman D and Emerson J 2014 The resource theory of stabilizer quantum computation *New J. Phys.* **16** 013009