Vol.2022-QS-5 No.29
2022/3/25

Cost function gradient for general ansatz in variational quantum algorithm

Ryunosuke Okubo[1,a)] Lento Nagano[2,b)]

Abstract: Variational quantum algorithms (VQAs) are expected to be promising strategies to achieve quantum advantages in the near future. However, gradients of some VQA cost functions vanish exponentially with the number of qubits, which requires exponentially large resources for optimizing them. This phenomenon is the so-called barren plateau problem and has been studied in previous works for certain types of ansatzes. We extend the previous works to a more general type of ansatz. Specifically, we calculate the second moment of a cost function gradient for a general ansatz, assuming that it is an unitary 2-design. We also evaluate the second moment without this assumption, which leads to a relation between a metric to quantify ansatz expressibilities and the second moment. This relation implies cost function landscapes for more expressive ansatzes become flatter. Our results hold independently of ansatz structures, so they are applicable to analysis of scalabilities of various VQAs.

Keywords: Quantum Computing, NISQ, Variational Quantum Algorithm, Barren Plateaus

1. Introduction

Real quantum devices in the near future, called noisy intermediate-scale quantum (NISQ) era [1], have a limited number of qubits and suffer from quantum noises. Variational quantum algorithm (VQA) is one of candidates that may achieve quantum advantages in the NISQ era. The first step of a VQA is to define a cost function $C(\Theta)$ whose minimum point gives a solution for a problem, by a Θ-parametrized quantum circuit. Then a classical optimizer trains the parameter Θ to minimize the cost function. VQAs are applicable to a variety of fields such as quantum chemistry [2], [3], combinatorial optimization [4], machine learning [5], [6], quantum circuit compiling [7], [8], [9] and dynamics simulation [10], [11], [12], [13], [14].

While such various applications of VQAs have been proposed, it has been found that some VQAs suffer from an infamous problem called barren plateau phenomena, where gradients of cost functions vanish exponentially with the number of qubits increasing [15], [16], [17], [18], [19]. This vanishing gradient problem makes training process harder even if we use a higher-derivative method [20] or a gradient-free method [21].

Previous works evaluated magnitudes of cost function gradients, assuming that ansatzes are highly expressive to be unitary 2-designs [15], [16], [22], [23]. Furthermore, they gave upper bounds of the magnitudes without the assumption and proved that more expressive ansatzes make cost function landscapes flatter [22], [23]. However their results hold only for specific types of ansatzes such as the hardware efficeint ansatz (HEA) and the Hamiltonian variational ansatz (HVA). For expample, the particle number preserving ansatz (PNPA) [24] is not categorized as ansatzes considered there.

In this paper, we extend the previous results to a more general type of ansatz. We first explicitly calculate the gradient magnitude under the assumption that the ansatz is a unitary 2-design. We also evaluate the gradient without the assumption to give an inequality for its magnitude. We do not assume any detailed structure of the ansatz except for layer-wise unitarity, so our analytical results hold for various types of ansatzes including the HEA, the HVA and the PNPA.

This paper is organized as follows. In section 2, we introduce a general form of ansazes including the HEA, the HVA and the PNPA, a metric to quantify expressibilities of ansatzes and a formal definition of barren plateau in VQA cost function landscapes. Section 3 provides our analytical results that hold for the general ansatz. In section 4, we apply the general results to the HEA and the HVA, and see that our results reproduce the previous works. We also give novel results for the PNPA. Section 5 verifies our analytical results with numerical simulations. In section 6, we discuss our results and conclude this paper.

2. Preliminaries

VQAs solve optimization problems by minimizing a cost function evaluated on a quantum circuit. Throughout this paper, we consider cost functions of the form

$$C(\Theta) = \mathrm{tr}_{\mathcal{H}}\left[OU(\Theta)\rho U^{\dagger}(\Theta)\right], \tag{1}$$

where ρ is an input state on an n-qubit system, O is a Hermitian operator acting on a d-dimensional Hilbert space $\mathcal{H}$ with $d = 2^n$, and $U(\Theta)$ is a unitary operator parametrized by $\Theta \in \mathcal{D} \subset \mathbb{R}^{N_p}$

1 Department of Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
2 International Center for Elementary Particle Physics (ICEPP), The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
a) okubo@icepp.s.u-tokyo.ac.jp
b) lento@icepp.s.u-tokyo.ac.jp

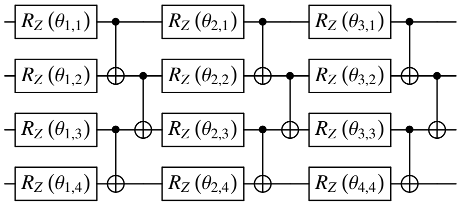

Fig. 1 An example of a four-qubit HEA on a quantum hardware with nearest-neighbor qubit connectivity.

where $\mathcal{D}$ is a parameter space of Θ and N_p is its dimension.

2.1 General form of ansatz

We consider a general ansatz with the following layered structure:

$$U\left(\Theta = (\boldsymbol{\theta}_0, \cdots, \boldsymbol{\theta}_{L-1})\right) = \prod_{l=L-1}^{0} U_l(\boldsymbol{\theta}_l), \tag{2}$$

where $U_l(\boldsymbol{\theta}_l)$ is a unitary operator with real parameters $\boldsymbol{\theta}_l := (\theta_{l,0}, \theta_{l,1}, \ldots, \theta_{l,K_l-1})$. We remark that there is no assumption on a detailed structure of each U_l.

Choosing an appropriate U_l is an important issue in a VQA. We should construct an ansatz to well approximate a solution state of the VQA and be implemented efficiently on a quantum device. If we do not have much information about the solution, then we choose U_l so that the ansatz can express as many unitary operators as possible. One example of such problem-agnostic ansatzes is the HEA. On the other hand, if we have some information about the solution, then the HEA can be inefficient. In such case we should use problem-inspired ansatzes like the HVA and the PNPA. Below we will explain the HEA, the HVA and the PNPA and see that the general ansatz with the structure Eq. (2) includes all of them.

2.1.1 Hardware efficient ansatz

The HEA is constructed as a sequence of gates taken from a native gate set of a given quantum hardware. While its structure is constrained by the architecture of the quantum hardware, a heavy transpilation process of decomposing an ansatz into the native gates is not needed. The HEA was used for finding a ground state energy of a quantum system [3].

A general form of the HEA is given by

$$U_{\mathrm{HEA}}(\Theta) = \prod_{l=L-1}^{0} \left(\prod_{k=K_l-1}^{0} e^{-i\theta_{l,k} V_{l,k}} \right) W_l, \tag{3}$$

where $V_{l,k}$ is a Hermitian and unitary operator, W_l is a unitary operator that is independent of any parameter and each of the unitary gates $e^{-i\theta_{l,k}V_{l,k}}$ and W_l is a native gate of a given quantum hardware. **Fig. 1** shows an example of four-qubit HEAs on a quantum hardware that has a native gate set $\{R_Z, X, \sqrt{X}, \mathrm{CNOT}\}$ and nearest-neighbor qubit connectivity.

2.1.2 Hamiltonian variational ansatz

Let us consider a problem to obtain the ground state energy of a Hamiltonian $H = \sum_{k=0}^{K-1} H_k$ where $[H_k, H_{k'}] \neq 0$ for $k \neq k'$. Then a HVA can be constructed as

$$U_{\mathrm{HVA}}(\Theta) = \prod_{l=L-1}^{0} \prod_{k=K-1}^{0} e^{-i\theta_{l,k} H_k}, \tag{4}$$

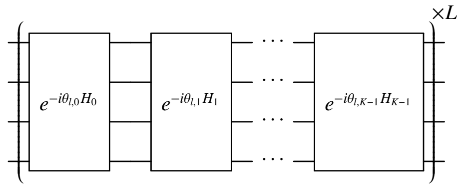

Fig. 2 A HVA for a Hamiltonian $H = \sum_{k=0}^{K-1} H_k$.

as shown in **Fig. 2**. The HVA is inspired by Trotterized adiabatic time evolution operator and used for solving combinatorial optimization [4], [25] and analyzing physics models [14], [26], [27], [28].

2.1.3 Particle number preserving ansatz

Let us define an m-particle subspace $\mathcal{H}_{n,m}$ ($m = 0, 1, \ldots, n$) of $\mathcal{H}$ as

$$\mathcal{H}_{n,m} = \{ |\psi\rangle \in \mathcal{H} \mid Q_{\mathrm{PN}} |\psi\rangle = m |\psi\rangle \}, \tag{5}$$

where Q_{PN} is a particle number operator given by

$$Q_{\mathrm{PN}} = \sum_{j=0}^{n-1} \frac{1 - Z_j}{2}. \tag{6}$$

An orthonormal basis of the subspace $\mathcal{H}_{n,m}$ is given in a computational basis by a set of quantum states with m qubits in the $|1\rangle$ state and $n - m$ qubits in the $|0\rangle$ state, so the dimension $d_{n,m}$ of the subspace $\mathcal{H}_{n,m}$ is $\binom{n}{m}$. We call an ansatz that conserves a particle number of an input state as a PNPA.

An explicit form of the PNPA for a two-qubit system is a gate $A(\theta, \phi)$ ($\theta \in [0, 2\pi)$, $\phi \in [0, 2\pi)$) given by

$$A(\theta, \phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sin\theta & e^{i\phi}\cos\theta & 0 \\ 0 & e^{-i\phi}\cos\theta & -\sin\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{7}$$

in the orthonromal basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. This gate is implemented by a circuit as shown in **Fig. 3**. The PNPA for an n-qubit system can be constructed by a sequence of $A(\theta, \phi)$ with the form

$$U_{\mathrm{PNPA}}(\Theta) = \prod_{l=L-1}^{0} A(\theta_l, \phi_l) \otimes I_{\bar{A}_l}, \tag{8}$$

where $I_{\bar{A}_l}$ is an identity operator acting on all qubits except the two qubits on which the gate $A(\theta_l, \phi_l)$ acts. A quantum circuit of a four-qubit PNPA is shown in **Fig. 4**.

The PNPA was employed for simulating dynamics and finding a ground state energy of quantum systems that conserve the particle number Q_{PN} [29], [30]. These works demonstrated that the PNPA reduced effect of quantum noises.

2.2 Ansatz expressibility

An important property of an ansatz is expressibility, i.e., a degree to which it expresses unitary operators in $\mathrm{U}(d)$. Here we introduce a metric to quantify ansatz expressibilities according to [22], [31], [32]

Let us first define a superoperator $\mathcal{A}_U^{(t)} : \mathcal{L}(\mathcal{H}^{\otimes t}) \to \mathcal{L}(\mathcal{H}^{\otimes t})$ as

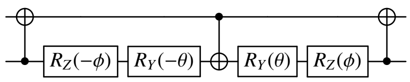

Fig. 3 Construction of the 2-qubit gate $A(\theta, \phi)$.

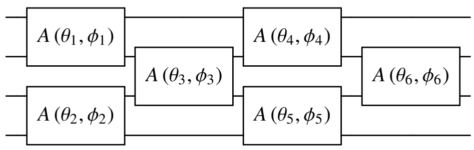

Fig. 4 An example of a 4-qubit PNPA.

$$\mathcal{A}_U^{(t)}(\bullet) := \int_{\mathrm{U}(d)} \mu_{\mathrm{Haar}}(\mathrm{d}V)\, V^{\otimes t}(\bullet)(V^\dagger)^{\otimes t} - \int_{\mathcal{D}} P_\Theta(\mathrm{d}\Theta)\, U(\Theta)^{\otimes t}(\bullet)(U^\dagger(\Theta))^{\otimes t} \tag{9}$$

where t is a positive integer, $\mathcal{L}(\mathcal{H}^{\otimes t})$ is a set of linear operators $\mathcal{H}^{\otimes t} \to \mathcal{H}^{\otimes t}$, μ_{Haar} is the Haar distribution on $\mathrm{U}(d)$, and P_Θ is the uniform distribution of the $\mathcal{D}$-valued Θ. Since the Haar distribution corresponds to the uniform distribution on $\mathrm{U}(d)$, an $\mathrm{U}(d)$-valued random variable V following the Haar distribution can be considered as a maximally expressive ansatz. Therefore, the superoperator $\mathcal{A}_U^{(t)}$ computes the difference between the ansatz U and the maximally expressive ansatz V. Then a quantity defined by

$$\epsilon_U^{(t)}(X) := \left\| \mathcal{A}_U^{(t)}(X^{\otimes t}) \right\|_2 \tag{10}$$

measures an expressibility of the ansatz U for $X \in \mathcal{L}(\mathcal{H})$. Note that smaller values of $\epsilon_U^{(t)}(X)$ indicates that the ansatz U is more expressible. The ansatz U is called a unitary t-design on $\mathrm{U}(d)$ if $\epsilon_U^{(t)}(X) = 0$ for all $X \in \mathcal{L}(\mathcal{H}^{\otimes t})$ [33].

2.3 Barren plateaus

Formally, barren plateau phenomena in VQA cost function landscapes are defined as follows [34] .

Definition 1. A VQA cost function $C\colon \mathcal{D} \to \mathbb{R}$ exhibits a barren plateau landscape with respect to α if for all $\delta > 0$, there exists $b > 1$ such that

$$P_\Theta(|\partial_\alpha C| \geq \delta) \in O(b^{-n}), \tag{11}$$

where P_Θ is the uniform distribution of the $\mathcal{D}$-valued Θ.

Eq. (11) means that the probability that the gradient $\partial_\alpha C\mid_{\Theta=\Theta_0}$ is greater than or equal to δ if we choose a set of parameters Θ_0 uniformly from the entire parameter space $\mathcal{D}$ decreases exponentially with the number of qubits.

One way to show that a cost function C exhibits a barren plateau landscape is to evaluate the second moment of the gradient $\partial_\alpha C$,

$$\mathbb{E}_\Theta\left[(\partial_\alpha C)^2\right] = \int_{\mathcal{D}} P_\Theta(\mathrm{d}\Theta)\,(\partial_\alpha C)^2 . \tag{12}$$

Once the second moment is obtained, we can bound the probability $P_\Theta(|\partial_\alpha C| \geq \delta)$ from Chebyshev's inequality (see e.g., [35]) as

$$P_\Theta(|\partial_\alpha C| \geq \delta) \leq \frac{1}{\delta^2}\mathbb{E}_\Theta\left[(\partial_\alpha C)^2\right] . \tag{13}$$

Thus if the second moment scales as $O(b^{-n})$ with some $b > 1$, we can conclude that the cost function C exhibits a barren plateau landscape.

3. Main results

We study the second moment of a gradient of the cost function C defined in Eq. (1) with respect to a chosen parameter $\alpha := \theta_{l_0,k_0}$. To this end, we divide the general ansatz U into three parts as

$$U(\Theta) = U_L(\Theta_L)U_M(\Theta_M)U_R(\Theta_R), \tag{14}$$

where

$$\Theta_L = (\boldsymbol{\theta}_{l_0+1}, \cdots, \boldsymbol{\theta}_L), \tag{15}$$
$$\Theta_M = (\boldsymbol{\theta}_{l_0}), \tag{16}$$
$$\Theta_R = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{l_0-1}), \tag{17}$$

and

$$U_L(\Theta_L) := \prod_{l=L-1}^{l_0+1} U_l(\boldsymbol{\theta}_l), \tag{18}$$
$$U_M(\Theta_M) := U_{l_0}(\boldsymbol{\theta}_{l_0}), \tag{19}$$
$$U_R(\Theta_R) := \prod_{l=l_0-1}^{0} U_l(\boldsymbol{\theta}_l). \tag{20}$$

3.1 Gradient for unitary 2-design general ansatz

Assuming that U_L and U_R are highly expressible to be unitary 2-designs, we can exactly calculate the second moment of the gradient as follows.

Theorem 2. Suppose that U_L and U_R are unitary 2-designs on $\mathrm{U}(d)$. Then the second moment of $\partial_\alpha C$ is given by

$$\mathbb{E}_\Theta\left[(\partial_\alpha C)^2\right] = \frac{2d\Delta_d(\rho)\Delta_d(O)}{(d^2-1)^2}\int P_{\Theta_M}(\mathrm{d}\Theta_M)\, f_1(\Theta_M), \tag{21}$$

where $\Delta_d^{(2)}(X) = \mathrm{tr}\left[X^2\right] - \frac{1}{d}\mathrm{tr}\,[X]^2$, P_{Θ_M} denotes the uniform distribution of Θ_M, and

$$f_1(\Theta_M) := \mathrm{tr}\left[U_{M,\alpha}U_{M,\alpha}^\dagger\right] - \frac{1}{d}\left|\mathrm{tr}\left[U_{M,\alpha}U_M^\dagger\right]\right|^2, \tag{22}$$

with $U_{M,\alpha} = \partial_\alpha U_M$.

Eq. (21) says that the second moment of the gradient depends on four factors: the input sate ρ, the observable O, the ansatz U_M and the dimension d of the Hilbert space. We can obtain a scaling of a cost function gradient for a VQA we wish to solve, by substituting the corresponding quantities ρ, O, U_M and d into Eq. (21). If we have $\Delta_d(\rho)\Delta_d(O)\int P_{\Theta_M}(\mathrm{d}\Theta_M)\, f_1(\Theta_M) \in O(d^2)$, the second moment scales as $O(d^{-1}) = O(2^{-n})$, which means that the cost function exhibits a barren plateau landscape.

3.2 Relation between gradient and expressibility

In Section 3.1, we assumed that U_L and U_R are unitary 2-designs. Without this assumption, we can derive the second moment of the gradient as follows.

Proposition 3. The second moment of $\partial_\alpha C$ is given by

$$\begin{aligned}
&\mathbb{E}_\Theta[(\partial_\alpha C)^2] \\
&= \frac{2d\Delta_d(\rho)\Delta_d(O)}{(d^2-1)^2}\int P_{\Theta_M}(\mathrm{d}\Theta_M)\, f_1(\Theta_M) \\
&\quad - \frac{2\Delta_d(\rho)}{d^2-1}\int P_{\Theta_M}(\mathrm{d}\Theta_M)\,\mathrm{tr}\left[\mathcal{A}^{(2)}_{U_L^\dagger}\left(O^{\otimes 2}\right) S\,\Omega_1\right] \\
&\quad - \frac{2\Delta_d(O)}{d^2-1}\int P_{\Theta_M}(\mathrm{d}\Theta_M)\,\mathrm{tr}\left[\mathcal{A}^{(2)}_{U_R}\left(\rho^{\otimes 2}\right) S\,\Omega_2\right] \\
&\quad + \int P_{\Theta_M}(\mathrm{d}\Theta_M)\,\mathrm{tr}\left[\left(\mathcal{J}\circ\mathcal{A}^{(2)}_{U_L^\dagger}\right)\left(O^{\otimes 2}\right)\mathcal{A}^{(2)}_{U_R}\left(\rho^{\otimes 2}\right)\right], \quad (23)
\end{aligned}$$

where

$$\Omega_1 := I\otimes U_{M,\alpha}U_{M,\alpha}^\dagger - U_{M,\alpha}U_M^\dagger \otimes U_M U_{M,\alpha}^\dagger, \quad (24)$$

$$\Omega_2 := U_{M,\alpha}^\dagger U_{M,\alpha}\otimes I - U_M^\dagger U_{M,\alpha}\otimes U_{M,\alpha}^\dagger U_M, \quad (25)$$

$$\begin{aligned}
\mathcal{J}(\bullet) := (U_M^\dagger)^{\otimes 2}(\bullet)\,U_{M,\alpha}^{\otimes 2} + (U_{M,\alpha}^\dagger)^{\otimes 2}(\bullet)\,U_M^{\otimes 2} \\
+ 2\left(U_M^\dagger\otimes U_{M,\alpha}^\dagger\right)(\bullet)\,(U_{M,\alpha}\otimes U_M), \quad (26)
\end{aligned}$$

and S is a subsystem SWAP operation defined by $S : |i\rangle\otimes|j\rangle \mapsto |j\rangle\otimes|i\rangle$.

Proposition 3 gives the following upper bound on the second moment of the gradient using $\epsilon_L^{(2)} := \epsilon_{U_L^\dagger}^{(2)}(O)$ and $\epsilon_R^{(2)} = \epsilon_{U_R}^{(2)}(\rho)$.

Theorem 4. The second moment of $\partial_\alpha C$ is upper bounded as

$$\begin{aligned}
&\mathbb{E}_\Theta\left[(\partial_\alpha C)^2\right] \\
&\le \frac{2d\Delta_d(\rho)\Delta_d(O)}{(d^2-1)^2}\int P_{\Theta_M}(\mathrm{d}\Theta_M)\, f_1(\Theta_M) \\
&\quad + 2\epsilon_R^{(2)}\epsilon_L^{(2)}\int P_{\Theta_M}(\mathrm{d}\Theta_M)\, f_2(\Theta_M) \\
&\quad + \frac{2\left(\epsilon_L^{(2)}\Delta_d(\rho)+\epsilon_R^{(2)}\Delta_d(O)\right)}{d^2-1}\int P_{\Theta_M}(\mathrm{d}\Theta_M)\, f_3(\Theta_M), \quad (27)
\end{aligned}$$

where

$$f_2(\Theta_M) := \left\|U_{M,\alpha}\right\|_2^2 + \sqrt{d}\left\|U_{M,\alpha}^\dagger U_{M,\alpha}\right\|_2, \quad (28)$$

$$\begin{aligned}
f_3(\Theta_M) := \Big(d\left\|U_{M,\alpha}^\dagger U_{M,\alpha}\right\|_2^2 + \left\|U_{M,\alpha}\right\|_2^4 \\
- 2\,\mathrm{tr}\left[U_{M,\alpha}U_M^\dagger U_{M,\alpha}U_{M,\alpha}^\dagger\right]\mathrm{tr}\left[U_M U_{M,\alpha}^\dagger\right]\Big)^{\frac{1}{2}}. \quad (29)
\end{aligned}$$

The right hand side of Eq. (27) takes the smaller value for smaller $\epsilon_L^{(2)}$ and $\epsilon_R^{(2)}$. We thus find that more expressive ansatzes induce flatter landscapes of cost functions. We remark that this statement holds for the general ansatz including the HEA, the HVA and the PNPA.

4. Ansatz-specific results

In the previous section we studied the gradient of the VQA cost function for the ansatz of the form Eq. (2). In this section, we see that results of the previous works on the HEA [15], [22] and the HVA [23] can be obtained as corollaries of our results. We also give novel results for the PNPA as corollaries of Theorem 2 and Proposition 3.

4.1 Hardware efficient ansatz

We consider VQA cost functions of the form

$$C_{\mathrm{HEA}}(\Theta) = \mathrm{tr}_{\mathcal{H}}\left[O U_{\mathrm{HEA}}(\Theta)\rho U_{\mathrm{HEA}}^\dagger(\Theta)\right], \quad (30)$$

where U_{HEA} is the HEA defined in Eq. (3) and ρ is an input state on an n-qubit Hilbert space $\mathcal{H}$. We study the gradient $\partial_\alpha C_{\mathrm{HEA}}$ with respect to $\alpha = \theta_{l_0}$. The first step to do this is to decompose the ansatz U_{HEA} into three blocks as

$$U_{\mathrm{HEA}}(\Theta) = U_L(\Theta_L)U_M(\Theta_M)U_R(\Theta_R), \quad (31)$$

where

$$\Theta_L = (\theta_{l_0+1}, \theta_{l_0+2}, \ldots, \theta_L), \quad (32)$$

$$\Theta_M = (\theta_{l_0}), \quad (33)$$

$$\Theta_R = (\theta_1, \theta_2, \ldots, \theta_{l_0-1}), \quad (34)$$

and

$$U_L(\Theta_L) = \prod_{l=L-1}^{l_0+1} e^{-i\theta_l V_l} W_l, \quad (35)$$

$$U_M(\Theta_M) = e^{-i\theta_{l_0}V_{l_0}} W_{l_0}, \quad (36)$$

$$U_R(\Theta_L) = \prod_{l=l_0-1}^{1} e^{-i\theta_l V_l} W_l. \quad (37)$$

Assuming that U_L and U_R are unitary 2-designs, we can easily calculate the second moment of the gradient $\partial_\alpha C_{\mathrm{HEA}}$ from Theorem 2 as

$$\mathbb{E}_\Theta\left[(\partial_\alpha C_{\mathrm{HEA}})^2\right] = \frac{2^{n+1}\Delta_{2^n}(\rho)\Delta_{2^n}(O)\Delta_{2^n}(V_{l_0})}{(4^n-1)^2}, \quad (38)$$

which reproduces the result of previous works [15], [22].

We next consider the second moment of $\partial_\alpha C_{\mathrm{HEA}}$ without an assumption that U_L and U_R are unitary 2-designs. Proposition 3 with the fact that $U_{M,\alpha} = -iV_{l_0}U_M$ is unitary leads to the following upper bound on the second moment:

$$\begin{aligned}
&\mathbb{E}_\Theta\left[(\partial_\alpha C_{\mathrm{HEA}})^2\right] \\
&\le \frac{2^{n+1}\Delta_{2^n}(\rho)\Delta_{2^n}(O)\Delta_{2^n}(V_{l_0})}{(4^n-1)^2} \\
&\quad + 4\epsilon_R^{(2)}\epsilon_L^{(2)} + \frac{\sqrt{2^{n+3}\Delta_{2^n}(V_{l_0})}}{4^n-1}\left(\epsilon_R^{(2)}\Delta_{2^n}(O)+\epsilon_L^{(2)}\Delta_{2^n}(\rho)\right), \\
&\qquad (39)
\end{aligned}$$

where $\epsilon_L^{(2)} := \epsilon_{U_L^\dagger}^{(2)}(O)$ and $\epsilon_R^{(2)} := \epsilon_{U_R}^{(2)}(\rho)$. This inequality is tighter than the one derived directly from Theorem 4. Eq. (39) reproduces the upper bound on the second moment in a previous work [18].

4.2 Hamiltonian variational ansatz

We consider VQA cost functions of the form

$$C_{\mathrm{HVA}}(\Theta) = \mathrm{tr}_{\mathcal{H}}\left[O U_{\mathrm{HVA}}(\Theta)\rho U_{\mathrm{HVA}}^\dagger(\Theta)\right], \quad (40)$$

where U_{HVA} is the HVA for the Hamiltonan $H = \sum_{k=0}^{K-1} H_k$ defined in Eq. (4) and ρ is an input state on an n-qubit Hilbert space $\mathcal{H}$. Here we make the following two assumptions:

(i) There is a Hermitian operator Q that has G distinct eigenvalues q_g $(g = 0, 1, \ldots, G-1)$ and commutes with each term H_k of the Hamiltonian $H = \sum_{k=0}^{K-1} H_k$. Then the n-qubit Hilbert space $\mathcal{H}$ is decomposed in a direct sum form $\mathcal{H} = \bigoplus_{g=0}^{G-1} \mathcal{H}_g$, where $\mathcal{H}_g$ is the eigenspace associated with the eigenvalue q_g.

(ii) The input state ρ lives only on $\mathcal{H}_g$, i.e., $\rho = \rho^{[g]} \oplus \left(\bigoplus_{g'\neq g} 0\right)$. Here we introduced the notation $X^{[g]} = P_g X P_g^\dagger$ where P_g is a projector from the n-qubit Hilbert space $\mathcal{H}$ onto the eigenspace $\mathcal{H}_g$.

Under these assumptions, we study the second moment of the gradient $\partial_\alpha C_{\text{HVA}}$ with respct to $\alpha = \theta_{l_0,k_0}$. The HVA can be decomposed as

$$U_{\text{HVA}}(\Theta) = U_L(\Theta_L) U_M(\Theta_M) U_R(\Theta_R)\,, \tag{41}$$

where

$$\boldsymbol{\theta}_l = (\theta_{l,0}, \theta_{l,1}, \ldots, \theta_{l,K-1})\,, \tag{42}$$
$$\Theta_L = (\boldsymbol{\theta}_{l_0+1}, \boldsymbol{\theta}_{l_0+2}, \ldots, \boldsymbol{\theta}_{L-1})\,, \tag{43}$$
$$\Theta_M = (\boldsymbol{\theta}_{l_0})\,, \tag{44}$$
$$\Theta_R = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{l_0-1})\,, \tag{45}$$

and

$$U_L(\Theta_L) = \prod_{l=L-1}^{l_0+1} \prod_{k=K-1}^{0} e^{-i\theta_{l,k} H_k}\,, \tag{46}$$
$$U_M(\Theta_M) = \prod_{k=K-1}^{0} e^{-i\theta_{l_0,k} H_k}\,, \tag{47}$$
$$U_R(\Theta_L) = \prod_{l=l_0-1}^{0} \prod_{k=K-1}^{0} e^{-i\theta_{l,k} H_k}\,. \tag{48}$$

Each $e^{-i\theta_{l,k} H_k}$ that commutes with Q is block diagonal as

$$e^{-i\theta_{l,k} H_k} = \bigoplus_{g'=0}^{G-1} e^{-i\theta_{l,k} H_k^{[g']}}\,, \tag{49}$$

so U_L, U_M, and U_R are also block diagonal as

$$U_L(\Theta_L) = \bigoplus_{g'=0}^{G-1} U_L^{[g']}(\Theta_L),\ U_L^{[g']}(\Theta_L) = \prod_{l=L-1}^{l_0+1} \prod_{k=K-1}^{0} e^{-i\theta_{l,k} H_k^{[g']}}\,, \tag{50}$$

$$U_M(\Theta_M) = \bigoplus_{g'=0}^{G-1} U_M^{[g']}(\Theta_M),\ U_M^{[g']}(\Theta_M) = \prod_{k=K-1}^{0} e^{-i\theta_{l_0,k} H_k^{[g']}}\,, \tag{51}$$

$$U_R(\Theta_R) = \bigoplus_{g'=0}^{G-1} U_R^{[g']}(\Theta_R),\ U_R^{[g']}(\Theta_R) = \prod_{l=l_0-1}^{0} \prod_{k=K-1}^{0} e^{-i\theta_{l,k} H_k^{[g']}}\,. \tag{52}$$

This implies that U_L, U_M, and U_R act non-trivially only on $\rho^{[g]}$ if the input state ρ satisfies the second assumption (ii). Thus we should focus only on the subspace $\mathcal{H}_g$ rather than the entier n-qubit Hilbert space $\mathcal{H}$, and can rewrite the cost function C_{HVA} as

$$C_{\text{HVA}}(\Theta) = \text{tr}_{\mathcal{H}_g}\Big[O^{[g]} U_L^{[g]} U_M^{[g]} U_R^{[g]} \rho^{[g]} U_R^{[g]\dagger} U_M^{[g]\dagger} U_L^{[g]\dagger}\Big], \tag{53}$$

by tracing out the other sectors $\mathcal{H}_{g'}$ $(g' \neq g)$.

Assuming that $U_L^{[g]}$ and $U_R^{[g]}$ are unitary 2-designs on $\mathrm{U}(d_g)$, where d_g denotes the dimension of $\mathcal{H}_g$, Theorem 2 gives

$$\mathbb{E}_\Theta\left[(\partial_\alpha C_{\text{HVA}})^2\right] = \frac{2 d_g \Delta_{d_g}(\rho^{[g]}) \Delta_{d_g}(O^{[g]}) \Delta_{d_g}(H_{k_0}^{[g]})}{(d_g^2-1)^2}\,, \tag{54}$$

which reproduces the result of a previous work [23].

Without the assumption that U_L and U_R are unitary 2-designs, we can evaluate the second moment of the gradient from Theorem 4 as

$$\begin{aligned}
&\mathbb{E}_\Theta\left[(\partial_\alpha C_{\text{HVA}})^2\right] \\
&\le \frac{2 d_g \Delta_{d_g}(\rho^{[g]}) \Delta_{d_g}(O^{[g]}) \Delta_{d_g}(V_{k_0}^{[g]})}{(d_g^2-1)^2} \\
&\quad + 2\epsilon_R^{(2)}\epsilon_L^{(2)}\left(\text{tr}\left[H_{k_0}^{[g]2}\right] + \sqrt{d_g \text{tr}\left[H_{k_0}^{[g]4}\right]}\right) \\
&\quad + \frac{2\left(\epsilon_R^{(2)}\Delta_{d_g}(O^{[g]}) + \epsilon_L^{(2)}\Delta_{d_g}(\rho^{[g]})\right)}{d_g^2-1} \\
&\qquad \times \sqrt{d_g \text{tr}\left[H_{k_0}^{[g]4}\right] + \text{tr}\left[H_{k_0}^{[g]2}\right]^2 - 2\text{tr}\left[H_{k_0}^{[g]3}\right]\text{tr}\left[H_{k_0}^{[g]}\right]}\,,
\end{aligned} \tag{55}$$

where $\epsilon_L^{(2)} := \epsilon_{U_L^{[g]\dagger}}^{(2)}(O^{[g]})$ and $\epsilon_R^{(2)} := \epsilon_{U_R^{[g]}}^{(2)}(\rho^{[g]})$. Another upper bound on the second moment for the HVA is given by [23].

4.3 Particle number preserving ansatz

We consider VQA cost functions of the form

$$C_{\text{PNPA}}(\Theta) = \text{tr}_{\mathcal{H}}\left[O U_{\text{PNPA}}(\Theta) \rho U_{\text{PNPA}}^\dagger(\Theta)\right], \tag{56}$$

where U_{PNPA} is the PNPA defined in Eq. (8) and ρ is an m-particle state ($m = 1, 2, \ldots, n-1$) on an n-qubit Hilbert space $\mathcal{H}$.

In order to study the gradient $\partial_\alpha C_{\text{PNPA}}$ with respect to $\alpha \in \{\theta_{l_0}, \phi_{l_0}\}$, let us decompose U_{PNPA} as

$$U_{\text{PNPA}}(\Theta) = U_L(\Theta_L) U_M(\Theta_M) U_R(\Theta_R)\,, \tag{57}$$

where

$$\Theta_L = ((\theta_{l_0+1}, \phi_{l_0+1}), (\theta_{l_0+2}, \phi_{l_0+2}), \ldots, (\theta_L, \phi_L))\,, \tag{58}$$
$$\Theta_M = (\theta_{l_0}, \phi_{l_0})\,, \tag{59}$$
$$\Theta_R = ((\theta_1, \phi_1), (\theta_2, \phi_2), \ldots, (\theta_{l_0-1}, \phi_{l_0-1}))\,, \tag{60}$$

and

$$U_L(\Theta_L) = \prod_{l=L-1}^{l_0+1}\left(A(\theta_l, \phi_l) \otimes I_{\bar{A}_l}\right)\,, \tag{61}$$
$$U_M(\Theta_M) = A(\theta_{l_0}, \phi_{l_0}) \otimes I_{\bar{A}_{l_0}}\,, \tag{62}$$
$$U_R(\Theta_L) = \prod_{l=l_0-1}^{0}\left(A(\theta_l, \phi_l) \otimes I_{\bar{A}_l}\right)\,. \tag{63}$$

The decomposed ansatzes U_L, U_M, and U_R conserve the particle number since they are sequences of the particle number preserving gates $A(\theta, \phi)$. Thus they are block diagonal as

$$U_L(\Theta_L) = \bigoplus_{m'=0}^{n} U_L^{(m')}(\Theta_L)\,, \tag{64}$$
$$U_M(\Theta_M) = \bigoplus_{m'=0}^{n} U_M^{(m')}(\Theta_M)\,, \tag{65}$$
$$U_R(\Theta_R) = \bigoplus_{m'=0}^{n} U_R^{(m')}(\Theta_R)\,. \tag{66}$$

Here we introduced the notation $X^{(m)} = P_{n,m} X P_{n,m}^{\dagger}$ where $P_{n,m}$ is a projector from the n-qubit Hilbert space $\mathcal{H}$ onto the m-particle subspace $\mathcal{H}_{n,m}$. As a result, the cost function C_{PNPA} can be expressed as

$$C_{\mathrm{PNPA}}(\Theta) = \mathrm{tr}_{\mathcal{H}_{n,m}}\left[O^{(m)} U_L^{(m)} U_M^{(m)} U_R^{(m)} \rho^{(m)} U_R^{(m)\dagger} U_M^{(m)\dagger} U_L^{(m)\dagger}\right], \tag{67}$$

since the initial state ρ is an m-particle state. Theorem 2 gives a scaling of the gradient $\partial_\alpha C_{\mathrm{PNPA}}$ if $U_L^{(m)}$ and $U_R^{(m)}$ are unitary 2-designs as the following corollary.

Corollary 5. We consider the gradient of the cost function C_{PNPA} defined in Eq. (8) with respect to $\alpha \in \{\theta_{l_0}, \phi_{l_0}\}$. Suppose that the input state ρ is an n-qubit and m-particle state ($m = 1, 2, \ldots, n-1$) and $U_L^{(m)}$ and $U_R^{(m)}$ are unitary 2-designs on $\mathrm{U}(d_{n,m})$, then the second moment of the gradient is given by

$$\mathbb{E}_\Theta\left[(\partial_\alpha C_{\mathrm{PNPA}})^2\right] = \frac{4 b_\alpha d_{n,m} d_{n-2,m-1} \Delta_{d_{n,m}}(\rho^{(m)}) \Delta_{d_{n,m}}(O^{(m)})}{(d_{n,m}^2 - 1)^2}, \tag{68}$$

where

$$b_\alpha := \begin{cases} 1 & (\alpha = \theta_{l_0}) \\ \frac{1}{2} & (\alpha = \phi_{l_0}) \end{cases}. \tag{69}$$

Corollary 5 shows that the second moment depends on the parameter α differentiating C_{PNPA} and the particle number m of the initial state ρ. For example, suppose that the input state ρ is a m-particle pure state $|\underbrace{01\cdots01}_{2m}\underbrace{00\cdots00}_{n-2m}\rangle\langle\underbrace{01\cdots01}_{2m}\underbrace{00\cdots00}_{n-2m}|$ and the observable $O = -|\underbrace{01\cdots01}_{2m}\underbrace{00\cdots00}_{n-2m}\rangle\langle\underbrace{01\cdots01}_{2m}\underbrace{00\cdots00}_{n-2m}|$, then the second moment is proportional to $d_{n-2,m-1} d_{n,m}^{-1}(d_{n,m}+1)^{-2}$. Therefore it scales as $O(n^{-5})$ for $m = 2$ while $O(n4^{-n})$ for $m = n/2$. This implies that whether or not the C_{PNPA} exhibits a barren plateau landscape depends on the particle number of the input state.

As with Eq. (39), Proposition 3 leads to a relation between the second moment and $\epsilon_L^{(2)} := \epsilon_{U_L^{(m)\dagger}}^{(2)}(O^{(m)})$ and $\epsilon_R^{(2)} := \epsilon_{U_R^{(m)}}^{(2)}(\rho^{(m)})$.

Corollary 6. We consider the gradient of the cost function C_{PNPA} defined in Eq. (8) with respect to $\alpha \in \{\theta_{l_0}, \phi_{l_0}\}$. Suppose that the input state ρ is an n-qubit and m-particle state ($m = 1, 2, \ldots, n-1$), then the second moment of the gradient is bounded as

$$\begin{aligned}&\mathbb{E}_\Theta\left[(\partial_\alpha C_{\mathrm{PNPA}})^2\right] \\ &\le \frac{4 b_\alpha d_{n,m} d_{n-2,m-1} \Delta_{d_{n,m}}(\rho^{(m)}) \Delta_{d_{n,m}}(O^{(m)})}{(d_{n,m}^2 - 1)^2} + 4 b_\alpha \epsilon_R^{(2)} \epsilon_L^{(2)} \\ &+ \frac{2 b_\alpha d_{n-2,m-1}}{d_{n,m}^2 - 1} \sqrt{4 + \frac{2n(n-1)}{m(n-m)}} \left(\epsilon_R^{(2)} \Delta_{d_{n,m}}(O^{(m)}) + \epsilon_L^{(2)} \Delta_{d_{n,m}}(\rho^{(m)})\right).\end{aligned} \tag{70}$$

In the next section, we will numerically evaluate how tight this bound is.

5. Numerical simulation

The previous section analytically evaluates the second moment of the gradient for the HEA, the HVA and the PNPA. In this section, we numerically confirm the analytical results for the PNPA, i.e., Corollary 5 and Corollary 6. Such numerical simulations for the HEA and the HVA were done by [18] and [23] respectively.

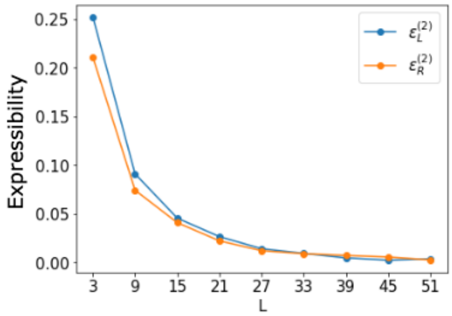

Fig. 5 Expressibilities $\epsilon_L^{(2)}$ and $\epsilon_R^{(2)}$ versus number L of the layers of the eight-qubit PNPA.

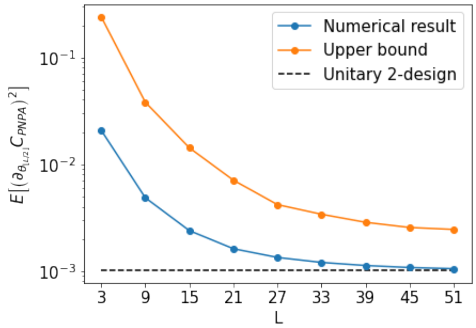

Fig. 6 Comparison between the second moment of the gradient and its upper bound for the eight-qubit PNPA. The blue line denotes the numerically estimated second moment. The orange line is the numerically estimated upper bound on the second moment obtained in Corollary 6. The black dashed line denotes the second moment derived in Corollary 5 assuming that $U_L^{(2)}$ and $U_R^{(2)}$ are unitary 2-designs on $\mathrm{U}(d_{8,2})$.

In our numerical simulation, we consider an eight-qubit system, a two-particle input state $\rho = |10100000\rangle\langle 10100000|$, and a global measurement operator $O = -|10100000\rangle\langle 10100000|$. We then numerically estimate second moments of the gradient of the cost function C_{PNPA} with respect to $\theta_{\lfloor L/2 \rfloor}$ and the quantities $\epsilon_L^{(2)} := \epsilon_{U_L^{(2)\dagger}}^{(2)}(O^{(2)})$ and $\epsilon_R^{(2)} := \epsilon_{U_R^{(2)}}^{(2)}(\rho^{(2)})$. These numerical simulations are implemented with Qulacs [36].

We observe from **Fig. 5** that $\epsilon_L^{(2)}$ and $\epsilon_R^{(2)}$ for the eight-qubit PNPA monotonically decreases and converges to 0 with the number of layers increasing. This implies that sufficiently deep $U_L^{(2)}$ and $U_R^{(2)}$ are approximately unitary 2-designs on $\mathrm{U}(d_{8,2})$. It is thus expected that the value of the second moment of the gradient for such a deep ansatz is near to that derived in Corollary 5. This expectation is verified in the blue line of **Fig. 6**. We also compares, in Fig. 6, the numerically estimated second moment for the eight-qubit PNPA with its upper bound obtained in Corollary 6. We can see that the differences between the second moment and the upper bound are around 0.18 for $L = 3$ and around 0.002 for $L = 51$ respectively, and thus our bound is tighter for a deeper ansatz.

6. Discussion and Conclusion

In this paper, we generalize previous analytical studies on the barren plateaus. Although their results hold only for specific types of ansatzes such as the HEA and the HVA, we extend them to a general type of ansatz including the HEA, the HVA and the PNPA. Theorem 2 shows the second moment of the cost function gradient for ansatzes of the form (2) assuming that the ansatz is a unitary 2-design. Without this assumption, we analytically derive the second moment in Proposition 3. Moreover, Proposition 3 gives an upper bound on the second moment using the ansatz expressibilities in Theorem 4, which shows that more expressive ansatzes make the cost function landscapes flatter. As shown in section 4, our analytical results are useful for understanding scalabilities of various VQAs.

Theorem 4 implies that using a less experssive ansatz is a strategy to avoid the barren plateau problem. Reducing number of layers of an ansatz and restricting ranges of its parameters lower its expressibility [22]. We however remark that we should keep the ansatz expressive enough to express an exact or well approximate solution of a VQA while suppressing its expressibility.

While VQAs are run in practice under noisy situations, our analytical results for a general type of ansatz hold only for noiseless situations. A previous work has analyzed noise-induced cost function landscapes for the HEA and the HVA [37]. A natural question that arises is to understand a cost function landscape for the general type of ansatz under such practical situations, which we leave for future work.

We numerically confirm our analytical results in section 5. We see that our upper bound on the second moment is not sufficiently tight, at least for shallow PNPAs. Similar behavior is also observed for shallow HEAs [22]. Thus further analysis to obtain tighter bounds would be important.

References

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

[2] Peruzzo, A., McClean, J., Shadbolt, P., Yung, M.-H., Zhou, X.-Q., Love, P. J., Aspuru-Guzik, A. and O'Brien, J. L.: A variational eigenvalue solver on a photonic quantum processor, *Nature Communications*, Vol. 5, No. 1, p. 4213 (online), DOI: 10.1038/ncomms5213 (2014).

[3] Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J. M. and Gambetta, J. M.: Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature*, Vol. 549, No. 7671, p. 242–246 (online), DOI: 10.1038/nature23879 (2017).

[4] Farhi, E., Goldstone, J. and Gutmann, S.: A Quantum Approximate Optimization Algorithm (2014).

[5] Farhi, E. and Neven, H.: Classification with Quantum Neural Networks on Near Term Processors (2018).

[6] Mitarai, K., Negoro, M., Kitagawa, M. and Fujii, K.: Quantum circuit learning, *Physical Review A*, Vol. 98, No. 3 (online), DOI: 10.1103/physreva.98.032309 (2018).

[7] Khatri, S., LaRose, R., Poremba, A., Cincio, L., Sornborger, A. T. and Coles, P. J.: Quantum-assisted quantum compiling, *Quantum*, Vol. 3, p. 140 (online), DOI: 10.22331/q-2019-05-13-140 (2019).

[8] Sharma, K., Khatri, S., Cerezo, M. and Coles, P. J.: Noise resilience of variational quantum compiling, *New Journal of Physics*, Vol. 22, No. 4, p. 043006 (online), DOI: 10.1088/1367-2630/ab784c (2020).

[9] Jones, T. and Benjamin, S. C.: Robust quantum compilation and circuit optimisation via energy minimisation, *Quantum*, Vol. 6, p. 628 (online), DOI: 10.22331/q-2022-01-24-628 (2022).

[10] McArdle, S., Jones, T., Endo, S., Li, Y., Benjamin, S. C. and Yuan, X.: Variational ansatz-based quantum simulation of imaginary time evolution, *npj Quantum Information*, Vol. 5, No. 1 (online), DOI: 10.1038/s41534-019-0187-2 (2019).

[11] Otten, M., Cortes, C. L. and Gray, S. K.: Noise-Resilient Quantum Dynamics Using Symmetry-Preserving Ansatzes (2019).

[12] Cirstoiu, C., Holmes, Z., Iosue, J., Cincio, L., Coles, P. J. and Sornborger, A.: Variational fast forwarding for quantum simulation beyond the coherence time, *npj Quantum Information*, Vol. 6, No. 1 (online), DOI: 10.1038/s41534-020-00302-0 (2020).

[13] Commeau, B., Cerezo, M., Holmes, Z., Cincio, L., Coles, P. J. and Sornborger, A.: Variational Hamiltonian Diagonalization for Dynamical Quantum Simulation (2020).

[14] Berthusen, N. F., Trevisan, T. V., Iadecola, T. and Orth, P. P.: Quantum dynamics simulations beyond the coherence time on NISQ hardware by variational Trotter compression (2021).

[15] McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R. and Neven, H.: Barren plateaus in quantum neural network training landscapes, *Nature communications*, Vol. 9, No. 1, pp. 1–6 (2018).

[16] Cerezo, M., Sone, A., Volkoff, T., Cincio, L. and Coles, P. J.: Cost function dependent barren plateaus in shallow parametrized quantum circuits, *Nature Communications*, Vol. 12, p. 1791 (online), DOI: 10.1038/s41467-021-21728-w (2021).

[17] Marrero, C. O., Kieferová, M. and Wiebe, N.: Entanglement Induced Barren Plateaus (2021).

[18] Holmes, Z., Arrasmith, A., Yan, B., Coles, P. J., Albrecht, A. and Sornborger, A. T.: Barren Plateaus Preclude Learning Scramblers, *Phys. Rev. Lett.*, Vol. 126, p. 190501 (online), DOI: 10.1103/PhysRevLett.126.190501 (2021).

[19] Thanasilp, S., Wang, S., Nghiem, N. A., Coles, P. J. and Cerezo, M.: Subtleties in the trainability of quantum machine learning models (2021).

[20] Cerezo, M. and Coles, P. J.: Higher order derivatives of quantum neural networks with barren plateaus, *Quantum Science and Technology*, Vol. 6, No. 3, p. 035006 (2021).

[21] Arrasmith, A., Cerezo, M., Czarnik, P., Cincio, L. and Coles, P. J.: Effect of barren plateaus on gradient-free optimization, *Quantum*, Vol. 5, p. 558 (2021).

[22] Holmes, Z., Sharma, K., Cerezo, M. and Coles, P. J.: Connecting ansatz expressibility to gradient magnitudes and barren plateaus, *arXiv preprint arXiv:2101.02138* (2021).

[23] Larocca, M., Czarnik, P., Sharma, K., Muraleedharan, G., Coles, P. J. and Cerezo, M.: Diagnosing barren plateaus with tools from quantum optimal control, *arXiv preprint arXiv:2105.14377* (2021).

[24] Gard, B. T., Zhu, L., Barron, G. S., Mayhall, N. J., Economou, S. E. and Barnes, E.: Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm, *npj Quantum Information*, Vol. 6, No. 1 (online), DOI: 10.1038/s41534-019-0240-1 (2020).

[25] Hadfield, S., Wang, Z., O'gorman, B., Rieffel, E. G., Venturelli, D. and Biswas, R.: From the quantum approximate optimization algorithm to a quantum alternating operator ansatz, *Algorithms*, Vol. 12, No. 2, p. 34 (2019).

[26] Wecker, D., Hastings, M. B. and Troyer, M.: Progress towards practical quantum variational algorithms, *Physical Review A*, Vol. 92, No. 4, p. 042303 (2015).

[27] Ho, W. W. and Hsieh, T. H.: Efficient variational simulation of nontrivial quantum states, *SciPost Physics*, Vol. 6, No. 3, p. 029 (2019).

[28] Cade, C., Mineh, L., Montanaro, A. and Stanisic, S.: Strategies for solving the Fermi-Hubbard model on near-term quantum computers, *Physical Review B*, Vol. 102, No. 23, p. 235122 (2020).

[29] Otten, M., Cortes, C. L. and Gray, S. K.: Noise-Resilient Quantum Dynamics Using Symmetry-Preserving Ansatzes (2019).

[30] Barron, G. S., Gard, B. T., Altman, O. J., Mayhall, N. J., Barnes, E. and Economou, S. E.: Preserving Symmetries for Variational Quantum Eigensolvers in the Presence of Noise, *Phys. Rev. Applied*, Vol. 16, p. 034003 (online), DOI: 10.1103/PhysRevApplied.16.034003 (2021).

[31] Sim, S., Johnson, P. D. and Aspuru-Guzik, A.: Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms, *Advanced Quantum Technologies*, Vol. 2, No. 12, p. 1900070 (online), DOI: 10.1002/qute.201900070 (2019).

[32] Nakaji, K. and Yamamoto, N.: Expressibility of the alternating layered ansatz for quantum computation, *Quantum*, Vol. 5, p. 434 (2021).

[33] Brandão, F. G. S. L., Harrow, A. W. and Horodecki, M.: Local Random Quantum Circuits are Approximate Polynomial-Designs, *Communications in Mathematical Physics*, Vol. 346, No. 2, p. 397–434 (online), DOI: 10.1007/s00220-016-2706-8 (2016).

[34] Volkoff, T. and Coles, P. J.: Large gradients via correlation in random parameterized quantum circuits, *Quantum Science and Technology*, Vol. 6, No. 2, p. 025008 (online), DOI: 10.1088/2058-9565/abd891 (2021).

[35] Klenke, A.: *Probability Theory: A Comprehensive Course*, Springer (2008).

[36] Suzuki, Y., Kawase, Y., Masumura, Y., Hiraga, Y., Nakadai, M., Chen, J., Nakanishi, K. M., Mitarai, K., Imai, R., Tamiya, S., Yamamoto, T., Yan, T., Kawakubo, T., Nakagawa, Y. O., Ibe, Y., Zhang, Y., Ya-

mashita, H., Yoshimura, H., Hayashi, A. and Fujii, K.: Qulacs: a fast and versatile quantum circuit simulator for research purpose, *Quantum*, Vol. 5, p. 559 (online), DOI: 10.22331/q-2021-10-06-559 (2021).
[37] Wang, S., Fontana, E., Cerezo, M., Sharma, K., Sone, A., Cincio, L. and Coles, P. J.: Noise-induced barren plateaus in variational quantum algorithms, *Nature Communications*, Vol. 12, No. 1 (online), DOI: 10.1038/s41467-021-27045-6 (2021).