Universal noise-precision relations in variational hybrid quantum-classical algorithms

Kosuke Ito[1,a)] Wataru Mizukami[1,2,3,b)] Keisuke Fujii[1,2,4,c)]

Abstract: Variational quantum algorithms (VQAs) are expected to become a practical application of near-term noisy quantum computers. Although the effect of the noise crucially determines whether a VQA works or not, the heuristic nature of VQAs makes it difficult to establish analytic theories. Analytic estimations of the impact of the noise are urgent for searching for quantum advantages, as numerical simulations of noisy quantum computers on classical computers are heavy and quite limited to small scale problems. In this work, we establish an analytic estimation of the error in the cost function of VQAs due to the noise. The estimation is applicable to any typical VQAs under the Gaussian noise, which is equivalent to a class of stochastic noise models. Notably, the depolarizing noise is included in this model. As a result, we obtain an estimation of the noise level to guarantee a required precision. Our formulae show how the Hessian of the cost function affects the sensitivity to the noise. This insight implies a trade-off relation between the trainability and the noise resilience of the cost function. As a highlight of the applications of the formula, we propose a quantum error mitigation method which is different from the extrapolation and the probabilistic error cancellation.

Keywords: Quantum computing, NISQ, Variational quantum algorithms, Hybrid quantum-classical algorithms, Quantum error mitigation, Effects of the noise

1. Introduction

To make use of noisy intermediate-scale quantum (NISQ) devices in the near future [37], we have to seek a classically intractable task that hundreds of qubits can resolve under the lack of the error correction. A promising framework to realize it is hybrid quantum-classical algorithms, where most of the processes are done on a classical computer, receiving the output from a quantum circuit which computes some classically intractable functions. Especially, variational quantum algorithms (VQAs) have attracted much attention, where the cost function of a variational problem is computed by utilizing low-depth quantum circuits and the optimization of the variational parameters is done on a classical computer. For example, the variational quantum eigensolver (VQE) [1], [20], [36] is a VQA to obtain an approximation of the ground state of a Hamiltonian, and beyond [5], [15], [16], [18], [26], [27], [31], [33], [35], [38], [41]. The quantum approximate optimization algorithm (QAOA) [8], [9], [34] is another attracting VQA for combinatorial optimization problems. The quantum machine learning algorithms [2], [3] for NISQ devices have also been proposed in various settings [6], [13], [21], [24], [28], [44].

The noise is one of the most crucial obstacles to overcome toward achieving quantum advantage via VQAs. The heuristic nature of VQAs makes it difficult to establish analytic theories on the effects of the noise on the performance of VQAs. As numerical simulations of noisy quantum computers on classical computers are heavy and limited to small scale problems, analytic estimations of the impact of the noise are urgent for obtaining knowledge about intermediate scale problems with potential quantum advantage. In fact, this issue has been actively studied in recent years, and some analytic results have been obtained, for example, on the noise resilience of the optimization results [11], [39], noise-induced barren plateaus [42], noise-induced breaking of symmetries [10], effects of the noise on the convergence property of the optimizations in VQAs [12].

In this work, we establish an analytic estimation formula on the error in the cost function of VQAs due to the noise. Especially, we focus on the the effect of the noise on the expectation value in order to investigate ultimately achievable and unachievable precision, aside from the statistical error due to the finiteness of the number of measurements. The estimation is applicable to any typical VQAs under the Gaussian noise, which is equivalent to a class of stochastic noise models. Notably, the depolarizing noise is included in this model. As a result, we obtain an estimation of the or-

1 Center for Quantum Information and Quantum Biology, International Advanced Research Institute, Osaka University, Osaka 560-8531, Japan
2 Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
3 JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan
4 RIKEN Center for Quantum Computing (RQC), Hirosawa 2-1, Wako, Saitama 351-0198, Japan
a) kosuke.ito@qc.ee.es.osaka-u.ac.jp
b) wataru.mizukami.857@qiqb.osaka-u.ac.jp
c) fujii@qc.ee.es.osaka-u.ac.jp

der of magnitude of the noise level to guarantee a required precision. Our formula shows how the Hessian of the cost function affects the sensitivity to the noise. This insight implies trade-off relations between the trainability and the noise resilience of the cost function.

The correspondence from the stochastic noise model to the Gaussian model is given by introducing virtual parametric gates associated with the noise. Our formula essentially comes from the expansion of the cost function with respect to the fluctuations in the parameters due to the noise. This fact implies that a picture of the noise based on fluctuations of parameters of virtual parametric gates can serve as a powerful tool for performance analysis of VQAs. In fact, we propose a quantum error mitigation method based on this expansion including the virtual parameters, which is different from existing error mitigation methods such as the extrapolation [23], [40] and the probabilistic error cancellation [7], [40].

2. Setup

2.1 Gaussian noise model of the parameterized quantum circuit

We consider the following parameterized quantum circuit

$$U(\vec{\theta}) = \prod_{i=1}^{M} U_i(\theta_i) W_i, \tag{1}$$

where $U_i(\theta_i) = \exp[-i\theta_i A_i/2]$ with $A_i^2 = I$, and W_i is a generic non-parametric gate. Typical parameterized quantum circuits such as the hardware efficient ansatz [13], [20] satisfy the above requirements. We focus on a VQA to minimize the cost function $C(\vec{\theta})$ given by the sum of the expectation values of the target Hermitian operators H_l $(l = 1, 2, \cdots, L)$ as

$$C(\vec{\theta}) = \sum_{l=1}^{L} \langle\phi_l| U(\vec{\theta})^\dagger H_l U(\vec{\theta}) |\phi_l\rangle, \tag{2}$$

where $|\phi_l\rangle$ $(l = 1, 2, \cdots, L)$ are the input states.

We consider the Gaussian noise in the parameter, where the cost function $C_{\rm noisy}(\vec{\theta})$ obtained by the noisy circuit is given as

$$\begin{aligned} C_{\rm noisy}(\vec{\theta}) &= \int d\mu(\vec{\Delta}) \ \langle\phi| U(\vec{\theta}+\vec{\Delta})^\dagger H U(\vec{\theta}+\vec{\Delta}) |\phi\rangle \\ &= \int d\mu(\vec{\Delta}) \ C(\vec{\theta}+\vec{\Delta}), \end{aligned} \tag{3}$$

where the components of the noise $\vec{\Delta} = (\Delta_1, \Delta_2, \cdots, \Delta_M)$ are independent random variables following the Gaussian distribution with the mean 0 and the variance σ_i^2 for the i-th component, and μ denotes the measure. The Gaussian noise model represents not only the fluctuation in the parameter but also the stochastic noise in general as shown in Sec. 2.2.

In this paper, we only focus on the effect of the noise on the expectation value in order to investigate ultimately achievable and unachievable precision, aside from the statistical error due to the finiteness of the number of measurements.

2.2 Correspondence to the stochastic noise model

Here, we show the correspondence relation between the Gaussian and stochastic noise models along the same lines with Nielsen and Chuang's textbook [32]. We consider the case where M' stochastic noise channels

$$\mathcal{E}_{A'_j,p_j}(\rho) := (1-p_j)\rho + p_j A'_j \rho A'_j \tag{4}$$

with respect to operators A'_j $(j = 1, 2, \cdots, M')$, $A'^2_j = I$ are inserted in the circuit, where ρ denotes a density operator and $0 < p_j < 1/2$ is the error probability. Hereafter, the prime denotes the operator or parameter associated with the stochastic noise to distinguish from that of the noiseless circuit. We define a map

$$\mathcal{U}_{A'_j,\Delta}(\rho) := e^{-i\frac{\Delta}{2}A'_j} \rho e^{i\frac{\Delta}{2}A'_j}. \tag{5}$$

Using the relation

$$e^{-i\frac{\Delta}{2}A'_j} = \mathbb{1}\cos\frac{\Delta}{2} - iA'_j \sin\frac{\Delta}{2}, \tag{6}$$

we have

$$\mathcal{U}_{A'_j,\Delta}(\rho) + \mathcal{U}_{A'_j,-\Delta}(\rho) = 2\rho\cos^2\frac{\Delta}{2} + 2A'_j\rho A'_j \sin^2\frac{\Delta}{2}. \tag{7}$$

From Eq. (7), we obtain the equivalence between the Gaussian noise channel $\mathcal{G}_{A'_j,\sigma'_j}$ with respect to A'_j with the variance

$$\sigma'^2_j = -2\log(1-2p_j) \tag{8}$$

and the given stochastic noise channel $\mathcal{E}_{A'_j,p_j}$ as follows:

$$\begin{aligned} &\mathcal{G}_{A'_j,\sigma'_j}(\rho) \\ =& \int_{-\infty}^{\infty} \mathcal{U}_{A'_j,\Delta}(\rho) \frac{e^{-\frac{\Delta^2}{2\sigma'^2_j}}}{\sqrt{2\pi}\sigma'_j} d\Delta \\ =& 2\int_0^{\infty} \left(2\rho\cos^2\frac{\Delta}{2} + 2A'_j\rho A'_j \sin^2\frac{\Delta}{2}\right) \frac{e^{-\frac{\Delta^2}{2\sigma'^2_j}}}{\sqrt{2\pi}\sigma'_j} d\Delta \\ =& (1-p_j)\rho + p_j A'_j\rho A'_j \\ =& \mathcal{E}_{A'_j,p_j}(\rho). \end{aligned} \tag{9}$$

Hence, if we consider the stochastic A'_j-noise $(j = 1, 2, \cdots, M')$, it can be treated as the Gaussian noise with respect to the virtually inserted parametric gate $U'_j(\theta'_j) = \exp\left[-i\theta'_j A'_j/2\right]$ at the place where the noise occurs, where $\theta'_j \equiv 0$ throughout the optimization. In the following, $\vec{\theta}$ denotes the abbreviation of the total parameters $(\vec{\theta}, \vec{\theta'}) = (\vec{\theta}, 0)$ when the virtual parameters for the stochastic noises are considered. Especially, the partial derivative of the cost function with respect to a virtual parameter θ'_j with $\vec{\theta'} = 0$ is denoted by $\frac{\partial}{\partial\theta'_j} C(\vec{\theta})$. We remark that Eq. (8) implies that

$$\sigma_j'^2 = 4p_j + O\left(p_j^2\right) \tag{10}$$

for small error probability p_j from the Taylor expansion $-\log(1-x) = x + O(x^2)$.

Especially, the depolarizing noise is one of the most basic and serious error sources for noisy quantum computers. A key feature of the Gaussian noise model is its capability of treating the depolarizing noise via the above correspondence. The depolarizing noise is described by the depolarizing channel $\mathcal{D}_{k,q}(\rho) = (1-q)\rho + q(4^k-1)^{-1}\sum_{i=1}^{4^k-1} P_i\rho P_i$, where P_i runs over all k-qubit Pauli operators except for the identity $I =: P_0$, and q is the error probability. Since we can decompose the depolarizing channel into multiple stochastic noise channels with respect to each single Pauli operator, the above correspondence works.

Lemma 1. *The k-qubit depolarizing channel $\mathcal{D}_{k,q}$ can be decomposed as $\mathcal{D}_{k,q} = \prod_i \mathcal{G}_{P_i,\sigma_{\mathrm{dep},k}}$ into $4^k - 1$ Gaussian noise channels with respect to k-qubit Pauli operators P_i with the common variance*

$$\sigma_{\mathrm{dep},k}^2 = -\frac{1}{4^{k-1}}\log\left(1 - \frac{4^k}{4^k-1}q\right). \tag{11}$$

Proof. We consider the vector space $\mathcal{V}_k$ consisting of the operators acting on k-qubit. Then, the k-qubit Pauli operators $\{P_i | i = 0, 1, \cdots, 4^k-1\}$ is a basis of $\mathcal{V}_k$. It is convenient to consider the matrix representation of quantum channels with respect to this basis. Since the Pauli channels $\mathcal{U}_{P_i}(\rho) = P_i\rho P_i$ are mutually commutative, they are simultaneously diagonalized in the Pauli basis $\{P_i | i = 0, 1, \cdots, 4^k - 1\}$. Then, the calculation of the product $\prod_{i=1}^{4^k-1}[(1-p)\mathcal{I} + p\mathcal{U}_{P_i}]$ is reduced to the calculation of each diagonal component. The (j,j)-component of $(1-p)\mathcal{I} + p\mathcal{U}_{P_i}$ is $1-2p$ if P_i anticommutes with P_j, otherwise 1 (i.e. if P_i commutes with P_j). The number of the generators of the Pauli group which anticommute to each element P_i is calculated as

$$2^k \sum_{r\le k, r:\mathrm{odd}} \binom{k}{r} = 2^k 2^{k-1} = 2\cdot 4^{k-1}. \tag{12}$$

Therefore, the matrix expression of $\prod_i[(1-p)\mathcal{I} + p\mathcal{U}_{P_i}]$ in the Pauli basis is

$$\mathrm{diag}(1, (1-2p)^{2\cdot 4^{k-1}}, \cdots, (1-2p)^{2\cdot 4^{k-1}}). \tag{13}$$

On the other hand, the matrix expression of the k-qubit depolarizing channel is

$$\mathrm{diag}\left(1, 1 - \frac{4^k}{4^k-1}q, \cdots, 1 - \frac{4^k}{4^k-1}q\right). \tag{14}$$

Thus, $\prod_i[(1-p)\mathcal{I} + p\mathcal{U}_{P_i}]$ is equal to the depolarizing channel with the error probability q if p satisfies

$$2\log(1-2p) = \frac{1}{4^{k-1}}\log\left(1 - \frac{4^k}{4^k-1}q\right). \tag{15}$$

Hence, the variance of the corresponding Gaussian noise (8) reads

$$\sigma_{\mathrm{dep},k}^2 = -2\log(1-2p) = -\frac{1}{4^{k-1}}\log\left(1 - \frac{4^k}{4^k-1}q\right). \tag{16}$$

□

We remark that Eq. (11) implies that

$$\sigma_{\mathrm{dep},k}^2 = \frac{4}{4^k-1}q + O\left(q^2\right) \tag{17}$$

for small error probability q in the same way as Eq. (10).

3. Universal Error Estimation

Our first main result is the following estimation of the error $\epsilon(\vec{\theta}) := C_{\mathrm{noisy}}(\vec{\theta}) - C(\vec{\theta})$ in the cost function due to the noise.

Theorem 1. *We have the following estimation of the deviation of the cost function due to the noise:*

$$\left|\epsilon(\vec{\theta}) - \frac{1}{2}\sum_{i=1}^{M}\frac{\partial^2}{\partial\theta_i^2}C(\vec{\theta})\sigma_i^2\right| \le \frac{\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})}{16}\left(\sum_{i=1}^{M}\sigma_i^2\right)^2, \tag{18}$$

where $E_{0,l}$, $E_{\max,l}$ are the minimum and the largest eigenvalues of H_l, respectively.

Proof. Let us introduce the multi-index notation for $\alpha \in \mathbb{N}^M$ and $\vec{\theta} \in \mathbb{R}^M$ as follows:

$$\vec{\theta}^{\alpha} := \prod_{i=1}^{M}\theta_i^{\alpha_i}, \quad \alpha! := \prod_{i=1}^{M}\alpha_i!, \quad |\alpha| := \sum_{i=1}^{M}\alpha_i. \tag{19}$$

The partial derivatives of a function f are denoted as

$$D^{\alpha}f := \frac{\partial^{|\alpha|}}{\partial\theta_1^{\alpha_1}\partial\theta_2^{\alpha_2}\cdots\partial\theta_M^{\alpha_M}}f. \tag{20}$$

We also consider the scalar multiplication $k\alpha = (k\alpha_1, k\alpha_2, \cdots, k\alpha_M)$ for $k \in \mathbb{N}$, $\alpha \in \mathbb{N}^M$. By applying the Taylor's theorem to the integrand $C(\vec{\theta} + \vec{\Delta})$, the definition of the noisy cost function (3) reads

$$\begin{aligned}
&C_{\mathrm{noisy}}(\vec{\theta}) \\
=&C(\vec{\theta}) + \sum_{|\alpha|=1}^{3}\frac{1}{\alpha!}D^{\alpha}C(\vec{\theta})\int\vec{\Delta}^{\alpha}d\mu(\vec{\Delta}) \\
&+\sum_{|\alpha'|=4}\frac{1}{\alpha'!}\int D^{\alpha'}C(\vec{\theta} + s(\vec{\theta},\vec{\Delta})\vec{\Delta})\vec{\Delta}^{\alpha'}d\mu(\vec{\Delta}) \\
=&C(\vec{\theta}) + \frac{1}{2}\sum_{i=1}^{M}\frac{\partial^2}{\partial\theta_i^2}C(\vec{\theta})\sigma_i^2 \\
&+\sum_{|\alpha|=2}\frac{1}{(2\alpha)!}\int D^{2\alpha}C(\vec{\theta} + s(\vec{\theta},\vec{\Delta})\vec{\Delta})\vec{\Delta}^{2\alpha}d\mu(\vec{\Delta})
\end{aligned} \tag{21}$$

with $0 < s(\vec{\theta},\vec{\Delta}) < 1$, where we have used the fact that all odd moments of the Gaussian random variable vanish. Because of $A_i^2 = 1$ the second derivatives read

$$\frac{\partial^2}{\partial\theta_i^2}C(\vec{\theta}) = -\frac{1}{2}\left[C(\vec{\theta}) - C(\vec{\theta} + \pi\vec{e}_i)\right], \tag{22}$$

where $\vec{e}_i$ denotes the vector whose i-th component is 1 and the other components are 0. Similar relation is used in Refs. [4], [17], [25], [29]. By recursively applying the relation (22), it turns out that the derivatives $D^{2\alpha}C(\vec{\theta})$ have the form

$$D^{2\alpha}C(\vec{\theta}) = \frac{1}{2}\left[\frac{1}{2^{|\alpha|-1}}\sum_{i=1}^{2^{|\alpha|-1}}\left(C(\vec{\theta}_{i,1}) - C(\vec{\theta}_{i,2})\right)\right] \quad (23)$$

with some parameters $\vec{\theta}_{i,1(2)}$. Since $\sum_{l=1}^{L} E_{0,l} \leq C(\vec{\theta}) \leq \sum_{l=1}^{L} E_{\max,l}$ holds for any parameter $\vec{\theta}$, we obtain [22]

$$|D^{2\alpha}C(\vec{\theta})| \leq \frac{\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})}{2}. \quad (24)$$

Then, applying the formula $\int \Delta_i^{2\alpha_i} d\mu = \sigma_i^{2\alpha_i}(2\alpha_i - 1)!!$ of the even moments of the Gaussian random variable, where $(2k-1)!! := (2k-1)(2k-3)\cdots 3\cdot 1$ denotes the double factorial, we obtain

$$\begin{aligned}
&\left|\epsilon(\vec{\theta}) - \frac{1}{2}\sum_{i=1}^{M}\frac{\partial^2}{\partial\theta_i^2}C(\vec{\theta})\sigma_i^2\right| \\
=&\left|\sum_{|\alpha|=2}\frac{1}{(2\alpha)!}\int D^{2\alpha}C(\vec{\theta}+s(\vec{\theta},\vec{\Delta})\vec{\Delta})\vec{\Delta}^{2\alpha}d\mu(\vec{\Delta})\right| \\
\leq&\frac{\sum_{l=1}^{L}(E_{\max,l}-E_{0,l})}{2}\sum_{|\alpha|=2}\frac{(2\alpha-1)!!}{(2\alpha)!}\vec{\sigma}^{2\alpha} \\
=&\frac{\sum_{l=1}^{L}(E_{\max,l}-E_{0,l})}{16}\left(\sum_{i=1}^{M}\sigma_i^2\right)^2, \quad (25)
\end{aligned}$$

where $2\alpha - 1 := (2\alpha_1 - 1, 2\alpha_2 - 1, \cdots, 2\alpha_M - 1)$ and $\vec{\sigma} := (\sigma_1, \sigma_2, \cdots, \sigma_M)$. □

Theorem 1 implies that the error $\epsilon(\vec{\theta})$ is approximated as

$$\epsilon(\vec{\theta}) \approx \frac{1}{2}\sum_{i=1}^{M}\frac{\partial^2}{\partial\theta_i^2}C(\vec{\theta})\sigma_i^2 \quad (26)$$

if the variances σ_i^2 $(i = 1, 2, \cdots, M)$ are small enough so that $\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})\left(\sum_{i=1}^{M}\sigma_i^2\right)^2$ is sufficiently small. For typical problems, $\sum_{l=1}^{L}(E_{\max,l} - E_{0,l}) \leq 2\sum_{l=1}^{L}\|H_l\|$ is in polynomial order of the number of qubit n, i. e., $\sum_{l=1}^{L}(E_{\max,l} - E_{0,l}) = O(n^r)$ with a positive number r (e. g. $r = 1$ for locally interacting spin systems, $r = 4$ for the Jordan-Wigner transformed full configuration interaction Hamiltonian of molecules [14], [19], [36]). Then, if all the variances are in the same order $\sigma_i^2 = O(\sigma^2)$ $(i = 1, 2, \cdots, M)$, this approximation is valid when $\sigma^2 = o\left(n^{-\frac{r}{2}}M^{-1}\right)$ in the sense that $\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})\left(\sum_{i=1}^{M}\sigma_i^2\right)^2 = o(1)$.

Applying Theorem 1 to the stochastic noise through the correspondence to the Gaussian model shown in Sec. 2.2 with the relations (9) and (10), we obtain the following corollary:

Corollary 1. *If the stochastic noise channels $\mathcal{E}_{A'_j,p_j}(\rho) = (1-p_j)\rho + p_j A'_j\rho A'_j$ $(j = 1, 2, \cdots, M')$ with the error probability $0 < p_j < 1/2$ are inserted in the circuit, we have the following approximation of the error:*

$$\begin{aligned}
\epsilon(\vec{\theta}) =& \frac{1}{2}\sum_{i=1}^{M}\frac{\partial^2}{\partial\theta_i^2}C(\vec{\theta})\sigma_i^2 + 2\sum_{j=1}^{M'}\frac{\partial^2}{\partial\theta_j'^2}C(\vec{\theta})p_j \\
&+ O\left(\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})\left[\left(\sum_{i=1}^{M}\sigma_i^2\right)^2 + \left(\sum_{j=1}^{M'}p_j\right)^2\right]\right), \quad (27)
\end{aligned}$$

where θ'_j is the virtual parameter associated with $\mathcal{E}_{A'_j}$ introduced in Sec. 2.2 to give the correspondence between the stochastic noise and the Gaussian noise models.

Especially, as a typical model, we consider a model such that the depolarizing noise $\mathcal{D}_{k,q_k}$ is inserted after each k-qubit gate, where we set $q_k = (4^{k-1} - 4^{-1})c_k q$ with q being the scaling of the error probability, and c_k being the constant factor characterizing the difference in the error rates between different number-qubit gates. Then, we can apply Corollary 1 to this model in combination with Lemma 1 as follows:

$$\epsilon(\vec{\theta}) = \frac{1}{2}\sum_{i=1}^{M'}\frac{\partial^2}{\partial\theta_i'^2}C(\vec{\theta})c_{k'_i}q + O\left(\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})M^2q^2\right), \quad (28)$$

where θ'_i denotes the virtual parameter associated with each stochastic Pauli noise channel in the decomposition of the k'_i-qubit depolarizing channels, and the total number of the stochastic Pauli noise channels M' satisfies $M' = O(M)$. Since the second derivatives are bounded as $\left|\frac{\partial^2}{\partial\theta_i'^2}C(\vec{\theta})\right| \leq \sum_{l=1}^{L}(E_{\max,l} - E_{0,l})/2$ from Eq. (24), the estimation

$$\sum_{i=1}^{M'}\frac{\partial^2}{\partial\theta_i'^2}C(\vec{\theta})c_{k'_i} = O\left(M\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})\right) = O(Mn^r) \quad (29)$$

holds, where θ'_i denotes the virtual parameter associated with each stochastic Pauli noise channel in the decomposition of the k'_i-qubit depolarizing channels. Hence, when the error probability has the scaling

$$q = O\left(\frac{\epsilon}{n^r M}\right), \quad (30)$$

we can achieve the precision $\sim \epsilon$ as

$$\begin{aligned}
\epsilon(\vec{\theta}) =& \frac{1}{2}\sum_{i=1}^{M'}\frac{\partial^2}{\partial\theta_i'^2}C(\vec{\theta})c_{k'_i}q + O\left(\frac{\epsilon^2}{n^r}\right) \\
=& O(\epsilon). \quad (31)
\end{aligned}$$

For example, when $r = 1$, to achieve $\epsilon(\vec{\theta}) \sim 10^{-3}$ with $n \sim 100$ qubits and the number of gates $M \sim 100$, the error probability $q \sim 10^{-7}$ is sufficient, according to this order estimation. As we will show in Sec. 4, a simple error mitigation method utilizing Theorem 1 can relax this stringent error estimation. We also remark that this order estimation does *not* mean that Eq. (30) is required to achieve the precision ϵ, but it only shows that Eq. (30) is *sufficient* for that.

Hence, larger error probability than this estimation might be acceptable in practice.

From another point of view, the coefficients $\frac{\partial^2}{\partial\theta_i^2}C(\vec{\theta})$ in Eq. (26) give the sensitivity to the noise. Thus, low sensitivity to the noise requires small diagonal components of the Hessian of the cost function. Especially for a minimal point $\vec{\theta}^*$, this means that the trace norm of the Hessian should be small for the low sensitivity to the noise since the Hessian is positive, which implies the flat landscape of the cost function around the minima. However, the optimization in a flat landscape tends to be hard, e. g. due to the required precision of the gradient in gradient descent methods, which increases the required measurement number. Hence, Eq. (26) implies a trade-off relation between the sensitivity to the noise and the trainability of the cost function.

The above argument can also be extended to stochastic noise models where the coefficients includes the derivatives with respect to the virtual parameters as follows. To do so, we have to separate the derivatives with respect to the virtual parameters from those with respect to the real parameters since the virtual parameters have nothing to do with the optimization landscape. We define the noiseless precision δ of the minimization as $\delta := C(\vec{\theta}^*) - E_0$ which attributes to poor expression power of the parameterized quantum circuit $U(\vec{\theta})$ and to the non-globality of the minimization (i. e. $\vec{\theta}^*$ may be a local minimum). We assume that the parameters giving the minima of the noisy cost function does not significantly deviate from the noiseless ones [39]. To estimate the second derivatives of $C(\vec{\theta})$ with respect to the virtual parameters, let us consider a single variable function $C_j(\theta'_j) := C((\vec{\theta}^*, \theta'_j\vec{e}_j))$ of one virtual parameter θ'_j, where $\vec{e}_j$ is the M'-dimensional vector whose j-th component is 1 and the others are 0. Since $C_j(\theta'_j) = a_j\cos\bigl(\theta'_j + b_j\bigr) + c_j$ holds [30], where $a_j \geq 0$ and b_j and c_j are real numbers, we have $\frac{\partial^2}{\partial\theta_i'^2}C(\vec{\theta}^*) = \frac{d^2}{d\theta_j'^2}C_j(0) = -a_j\cos(b_j)$. Because the virtual parameters are fixed to 0 and not optimized, $\frac{\partial^2}{\partial\theta_j'^2}C(\vec{\theta}^*)$ may be negative. In this case, $a_j\cos(b_j) \geq 0$ holds. Since $C((\vec{\theta}, \vec{\theta}')) \geq E_0$ for any value of the parameters, we have $C_j(\theta'_j) \geq C_j(0) - \delta$, which implies

$$a_j\cos\bigl(\theta'_j + b_j\bigr) \geq a_j\cos(b_j) - \delta \geq -\delta. \tag{32}$$

Thus, we obtain $a_j \leq \delta$, and hence

$$\frac{\partial^2}{\partial\theta_i'^2}C(\vec{\theta}^*) = -a_j\cos(b_j) \geq -a_j \geq -\delta. \tag{33}$$

We consider the case where all A_j are Pauli operators, and the depolarizing noise $\mathcal{D}_{k,q_k}$ acting on the same qubit number k as A_j is inserted after each $U_j(\theta_j)$. We again set $q_k = (4^{k-1} - 4^{-1})c_k q$ with the scale q and the constant factor c_k depending on k. In this case, one of the stochastic Pauli noise channels composing the depolarizing noise is the stochastic A_j-channel, and hence the derivative with respect to the virtual parameter associated with this channel is equivalent to the derivative with respect to the real parameter θ_j. Such virtual parameters are not included in $\vec{\theta}' = (\theta'_1, \theta'_2, \cdots, \theta'_{M'})$, where the total number M' of the virtual parameter is again in the order $O(M)$. Let k_i be the number of qubits A_i acting on. For convenience, we rescale the parameter as $\theta_i = \sqrt{c_{k_i}}\tilde{\theta}_i$. We also define k'_i in the same way as in Eq. (28). Then, Eqs. (28) and (33) yield

$$\begin{aligned}
&\epsilon(\vec{\theta}^*)\\
=&\frac{1}{2}\left[\sum_{i=1}^{M}\frac{\partial^2}{\partial\tilde{\theta}_i^2}C(\vec{\theta}^*) + \sum_{i=1}^{M'}\frac{\partial^2}{\partial\tilde{\theta}_i'^2}C(\vec{\theta}^*)c_{k'_i}\right] q\\
&+ O\left(\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})M^2q^2\right)\\
\geq&\frac{1}{2}\left[\sum_{i=1}^{M}\frac{\partial^2}{\partial\tilde{\theta}_i^2}C(\vec{\theta}^*) - M'\delta\right] q + O\left(\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})M^2q^2\right),
\end{aligned} \tag{34}$$

and hence

$$\begin{aligned}
&\frac{\epsilon(\vec{\theta}^*)}{q} + O(M\delta) + O\left(\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})M^2q\right)\\
\geq&\frac{1}{2}\sum_{i=1}^{M}\frac{\partial^2}{\partial\tilde{\theta}_i^2}C(\vec{\theta}^*) = \frac{1}{2}\,\mathrm{Tr}\left|\left(\frac{\partial^2 C}{\partial\tilde{\theta}_i\partial\tilde{\theta}_j}(\vec{\theta}^*)\right)\right|
\end{aligned} \tag{35}$$

since the Hessian of the cost function is positive at $\vec{\theta}^*$. For a successful minimization, δ should be small, and hence $O(M\delta)$ term is negligible for not so large scale M. $O\left(\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})M^2q\right)$ term is also negligible for sufficiently small error probability q. Then, Eq. (35) implies that the trace norm of the Hessian of the cost function should be small if the error probability q is not sufficiently small compared to the required level of the error $\epsilon(\vec{\theta}^*) < \epsilon$ due to the noise. This fact implies the hardness of the optimization due to the flat landscape of the vicinity of the minima. Moreover, in this case, optimization algorithms utilizing the Hessian become hard since high precision of the estimation of the Hessian is required if the Hessian is small. Oppositely, at least we need $q = O(\epsilon)$ to achieve $\epsilon > \epsilon(\vec{\theta}^*)$ avoiding such hardness.

As another remark, recent result [43] has shown that the trace of the Hessian of the loss function of the overparameterized networks decreases on average during the optimization steps of the stochastic gradient descent (SGD). If this result is extended to VQA, it implies that SGD may decrease the effect of the noise during the optimization steps, which may hint on when and to what extent we should use the error mitigation techniques in optimizations.

4. An error mitigation method

We can apply Theorem 1 to derive a error mitigation method. We can cancel the error by subtracting the leading term of the error $\frac{1}{2}\sum_{i=1}^{M}\frac{\partial^2}{\partial\theta_i^2}C(\vec{\theta})\sigma_i^2$ given that we know the error model, and σ_i^2 is small enough so that the sub-leading order terms of $O\left(\sum_{l=1}^{L}(E_{\max,l} - E_{0,l})\left(\sum_{i=1}^{M}\sigma_i^2\right)^2\right)$ are negligible. An advantage of this method is that we only use the noisy estimation of the derivatives of the cost function to mitigate the error, and we do not need to change the noise strength as in the extrapolation method [23], [40], nor to sample various circuits as in the probabilistic error cancellation [7], [40]. Using the shift rule (22), we can calculate the second derivatives from noisy evaluations of the cost function. The effect of the noise in this noisy estimation $D_i^2 C_{\mathrm{noisy}}(\theta)$ of the second derivative is estimated by applying Theorem 1 again, which reads

$$D_i^2 C_{\mathrm{noisy}}(\theta) = \frac{\partial^2}{\partial \theta_i^2} C(\vec{\theta}) + O\left(\sum_{l=1}^{L} (E_{\mathrm{max},l} - E_{0,l}) \sum_{i=1}^{M} \sigma_i^2\right). \tag{36}$$

Therefore, we have

$$\begin{aligned} &C_{\mathrm{noisy}}(\theta) - \frac{1}{2}\sum_{i=1}^{M} D_i^2 C_{\mathrm{noisy}}(\theta)\sigma_i^2 \\ =&C(\theta) + \frac{1}{2}\sum_{i=1}^{M} O\left(\sum_{l=1}^{L}(E_{\mathrm{max},l} - E_{0,l})\sum_{i=1}^{M}\sigma_i^2\right)\sigma_i^2 \\ &+ O\left(\sum_{l=1}^{L}(E_{\mathrm{max},l} - E_{0,l})\left(\sum_{i=1}^{M}\sigma_i^2\right)^2\right) \\ =&C(\theta) + O\left(\sum_{l=1}^{L}(E_{\mathrm{max},l} - E_{0,l})\left(\sum_{i=1}^{M}\sigma_i^2\right)^2\right). \end{aligned} \tag{37}$$

In this way, we can mitigate the error up to the sub-leading order $O\left(\sum_{l=1}^{L}(E_{\mathrm{max},l} - E_{0,l})\left(\sum_{i=1}^{M}\sigma_i^2\right)^2\right)$.

This method is also applicable to the stochastic noise including the depolarizing noise by applying Corollary 1. The overhead of this protocol is the evaluations of the noisy cost function at the π-shift of every parameter including the virtual parameters. In that case, the Pauli rotation gate of each virtual parameter is inserted to calculate the derivative with respect to the virtual parameter, and the noise is also added along with this gate. However, the order estimation is not affected, since at most a single gate is added for each evaluation. We again consider the same depolarizing noise model with the scaling of the error probability q as the one to obtain Eq. (28). We also assume that $\sum_{l=1}^{L}(E_{\mathrm{max},l} - E_{0,l}) = O(n^r)$. Then, in order to achieve a given precision ϵ, it is sufficient to have

$$q = O\left(\frac{\epsilon^{\frac{1}{2}}}{n^{\frac{r}{2}} M}\right) \tag{38}$$

by applying this error mitigation. In comparison to Eq. (30), the order estimation of the sufficient noise level is relaxed by $\sqrt{\epsilon/n^r}$ via this error mitigation. For example, when $r = 1$, to achieve the precision $\sim 10^{-3}$ with $n \sim 100$ qubits and the number of gates $M \sim 100$, the error probability $q \sim 3 \times 10^{-5}$ is sufficient, which is about 10^2 times larger in comparison with the one without the error mitigation shown below Eq. (31), although it is still stringent. However, we again remark that this estimation is only the sufficient order of the error probability to achieve a given precision, but not necessary. Moreover, we can take into account the next-leading order in expansion (21) to improve the error mitigation if the overhead is acceptable. Further analysis on the practical effectiveness of this error mitigation method including the finiteness of the sampling and the comparison with different error mitigation techniques will be done in a successive work.

5. Conclusion

We have established an analytic formula for estimating the error in the cost function of VQAs due to the Gaussian noise. We can also apply our formula to a wide class of stochastic noise including the depolarizing noise model via their equivalence with the Gaussian noise. The first main result Theorem 1 gives the leading-order approximation of the error $\epsilon(\vec{\theta})$ in the cost function due to the noise. The Hessian of the cost function as the coefficients of the noise effect implies a trade-off relation between the hardness of the optimization of the parameters and the noise resilience of the cost function. We have derived an order estimation of the sufficient error probability to achieve a given precision based on this formula. This estimation offers stringently small error probability if no error mitigation is taken into account. This is partially because, the estimation is nothing but a sufficient condition to achieve the given precision.

A highlight of the applications of our formula is the proposal of a quantum error mitigation method shown in Sec. 4. The essence of this error mitigation method is the cancellation of the error based on the expansion of the error with respect to the fluctuations of the parameters including the virtual parameters. An advantage of this method is that we only use the noisy estimation of the derivatives of the cost function to mitigate the error, and we do not need to change the noise strength as in the extrapolation method [23], [40], nor to sample various circuits as in the probabilistic error cancellation [7], [40]. Although the effectiveness of this method is still inconclusive since we have only an estimation of the sufficient order of the error probability for this method to work, there is a possibility of this method to be efficient in some situations. It may also possible to improve this method by taking into account higher order expansions. In a future work, further analysis will be done on this error mitigation method including the finiteness of the sampling and the comparison with other error mitigation methods. To take into account the statistical error due to the finiteness of the sampling, the effect of the noise on the variance of the cost function should also be considered in future works.

Acknowledgments The authors would like to thank Yasunari Suzuki for very helpful comments. This work is supported by MEXT Quantum Leap Flagship Program (MEXT QLEAP) Grant Number JPMXS0118067394 and JPMXS0120319794. K. F. is supported by JSPS KAKENHI Grant No. 16H02211, JST ERATO JPMJER1601, and JST CREST JPMJCR1673. W. M. wishes to thank Japan Society for the Promotion of Science (JSPS) KAKENHI No.

18K14181 and JST PRESTO No. JPMJPR191A. We also acknowledge support from JST COI-NEXT program.

References

[1] Bauer, B., Wecker, D., Millis, A. J., Hastings, M. B. and Troyer, M.: Hybrid Quantum-Classical Approach to Correlated Materials, *Phys. Rev. X*, Vol. 6, p. 031045 (online), DOI: 10.1103/PhysRevX.6.031045 (2016).

[2] Benedetti, M., Lloyd, E., Sack, S. and Fiorentini, M.: Parameterized quantum circuits as machine learning models, *Quantum Science and Technology*, Vol. 4, No. 4, p. 043001 (online), DOI: 10.1088/2058-9565/ab4eb5 (2019).

[3] Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N. and Lloyd, S.: Quantum machine learning, *Nature*, Vol. 549, No. 7671, pp. 195–202 (online), DOI: 10.1038/nature23474 (2017).

[4] Cerezo, M. and Coles, P. J.: Impact of Barren Plateaus on the Hessian and Higher Order Derivatives, arXiv:2008.07454 (2020).

[5] Colless, J. I., Ramasesh, V. V., Dahlen, D., Blok, M. S., Kimchi-Schwartz, M. E., McClean, J. R., Carter, J., de Jong, W. A. and Siddiqi, I.: Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm, *Phys. Rev. X*, Vol. 8, p. 011021 (online), DOI: 10.1103/PhysRevX.8.011021 (2018).

[6] Dallaire-Demers, P.-L. and Killoran, N.: Quantum generative adversarial networks, *Phys. Rev. A*, Vol. 98, p. 012324 (online), DOI: 10.1103/PhysRevA.98.012324 (2018).

[7] Endo, S., Benjamin, S. C. and Li, Y.: Practical Quantum Error Mitigation for Near-Future Applications, *Phys. Rev. X*, Vol. 8, p. 031027 (online), DOI: 10.1103/PhysRevX.8.031027 (2018).

[8] Farhi, E., Goldstone, J. and Gutmann, S.: A Quantum Approximate Optimization Algorithm, arXiv:1411.4028 (2014).

[9] Farhi, E. and Harrow, A. W.: Quantum Supremacy through the Quantum Approximate Optimization Algorithm, arXiv:1602.07674 (2016).

[10] Fontana, E., Cerezo, M., Arrasmith, A., Rungger, I. and Coles, P. J.: Optimizing parametrized quantum circuits via noise-induced breaking of symmetries, arXiv:2011.08763 (2020).

[11] Fontana, E., Fitzpatrick, N., Ramo, D. M., Duncan, R. and Rungger, I.: Evaluating the noise resilience of variational quantum algorithms, arXiv:2011.01125 (2020).

[12] Gentini, L., Cuccoli, A., Pirandola, S., Verrucchi, P. and Banchi, L.: Noise-resilient variational hybrid quantum-classical optimization, *Physical Review A*, Vol. 102, No. 5, p. 052414 (online), DOI: 10.1103/PhysRevA.102.052414 (2020).

[13] Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M. and Gambetta, J. M.: Supervised learning with quantum-enhanced feature spaces, *Nature*, Vol. 567, No. 7747, pp. 209–212 (online), DOI: 10.1038/s41586-019-0980-2 (2019).

[14] Helgaker, T., Jorgensen, P. and Olsen, J.: *Molecular Electronic-Structure Theory*, Wiley, Sussex (2002).

[15] Heya, K., Suzuki, Y., Nakamura, Y. and Fujii, K.: Variational Quantum Gate Optimization, arXiv:1810.12745 (2018).

[16] Higgott, O., Wang, D. and Brierley, S.: Variational Quantum Computation of Excited States, *Quantum*, Vol. 3, p. 156 (online), DOI: 10.22331/q-2019-07-01-156 (2019).

[17] Huembeli, P. and Dauphin, A.: Characterizing the loss landscape of variational quantum circuits, *Quantum Science and Technology*, Vol. 6, No. 2, p. 025011 (online), DOI: 10.1088/2058-9565/abdbc9 (2021).

[18] Jones, T., Endo, S., McArdle, S., Yuan, X. and Benjamin, S. C.: Variational quantum algorithms for discovering Hamiltonian spectra, *Phys. Rev. A*, Vol. 99, p. 062304 (online), DOI: 10.1103/PhysRevA.99.062304 (2019).

[19] Jordan, P. and Wigner, E.: Über das Paulische Äquivalenzverbot, *Zeitschrift fur Physik*, Vol. 47, No. 9-10, pp. 631–651 (1928).

[20] 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, pp. 242–246 (online), DOI: 10.1038/nature23879 (2017).

[21] Khoshaman, A., Vinci, W., Denis, B., Andriyash, E., Sadeghi, H. and Amin, M. H.: Quantum variational autoencoder, *Quantum Science and Technology*, Vol. 4, No. 1, p. 014001 (online), DOI: 10.1088/2058-9565/aada1f (2018).

[22] Kübler, J. M., Arrasmith, A., Cincio, L. and Coles, P. J.: An Adaptive Optimizer for Measurement-Frugal Variational Algorithms, *Quantum*, Vol. 4, p. 263 (online), DOI: 10.22331/q-2020-05-11-263 (2020).

[23] Li, Y. and Benjamin, S. C.: Efficient Variational Quantum Simulator Incorporating Active Error Minimization, *Phys. Rev. X*, Vol. 7, p. 021050 (online), DOI: 10.1103/PhysRevX.7.021050 (2017).

[24] Lloyd, S., Mohseni, M. and Rebentrost, P.: Quantum algorithms for supervised and unsupervised machine learning, arXiv:1307.0411 (2013).

[25] Mari, A., Bromley, T. R. and Killoran, N.: Estimating the gradient and higher-order derivatives on quantum hardware, *Phys. Rev. A*, Vol. 103, p. 012405 (online), DOI: 10.1103/PhysRevA.103.012405 (2021).

[26] 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, p. 75 (online), DOI: 10.1038/s41534-019-0187-2 (2019).

[27] McClean, J. R., Kimchi-Schwartz, M. E., Carter, J. and de Jong, W. A.: Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, *Phys. Rev. A*, Vol. 95, p. 042308 (online), DOI: 10.1103/PhysRevA.95.042308 (2017).

[28] Mitarai, K., Negoro, M., Kitagawa, M. and Fujii, K.: Quantum circuit learning, *Phys. Rev. A*, Vol. 98, p. 032309 (online), DOI: 10.1103/PhysRevA.98.032309 (2018).

[29] Mitarai, K., Nakagawa, Y. O. and Mizukami, W.: Theory of analytical energy derivatives for the variational quantum eigensolver, *Phys. Rev. Research*, Vol. 2, p. 013129 (online), DOI: 10.1103/PhysRevResearch.2.013129 (2020).

[30] Nakanishi, K. M., Fujii, K. and Todo, S.: Sequential minimal optimization for quantum-classical hybrid algorithms, *Physical Review Research*, Vol. 2, No. 4, p. 043158 (online), DOI: 10.1103/PhysRevResearch.2.043158 (2020).

[31] Nakanishi, K. M., Mitarai, K. and Fujii, K.: Subspace-search variational quantum eigensolver for excited states, *Phys. Rev. Research*, Vol. 1, p. 033062 (online), DOI: 10.1103/PhysRevResearch.1.033062 (2019).

[32] Nielsen, M. and Chuang, I.: *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge (2000).

[33] Ollitrault, P. J., Kandala, A., Chen, C.-F., Barkoutsos, P. K., Mezzacapo, A., Pistoia, M., Sheldon, S., Woerner, S., Gambetta, J. M. and Tavernelli, I.: Quantum equation of motion for computing molecular excitation energies on a noisy quantum processor, *Phys. Rev. Research*, Vol. 2, p. 043140 (online), DOI: 10.1103/PhysRevResearch.2.043140 (2020).

[34] Otterbach, J. S., Manenti, R., Alidoust, N., Bestwick, A., Block, M., Bloom, B., Caldwell, S., Didier, N., Fried, E. S., Hong, S., Karalekas, P., Osborn, C. B., Papageorge, A., Peterson, E. C., Prawiroatmodjo, G., Rubin, N., Ryan, C. A., Scarabelli, D., Scheer, M., Sete, E. A., Sivarajah, P., Smith, R. S., Staley, A., Tezak, N., Zeng, W. J., Hudson, A., Johnson, B. R., Reagor, M., da Silva, M. P. and Rigetti, C.: Unsupervised Machine Learning on a Hybrid Quantum Computer, arXiv:1712.05771 (2017).

[35] Parrish, R. M., Hohenstein, E. G., McMahon, P. L. and Martínez, T. J.: Quantum Computation of Electronic Transitions Using a Variational Quantum Eigensolver, *Phys. Rev. Lett.*, Vol. 122, p. 230401 (online), DOI: 10.1103/PhysRevLett.122.230401 (2019).

[36] 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).

[37] Preskill, J.: Quantum Computing in the NISQ era and beyond, *Quantum*, Vol. 2, p. 79 (online), DOI: 10.22331/q-2018-08-06-79 (2018).

[38] Santagati, R., Wang, J., Gentile, A. A., Paesani, S., Wiebe, N., McClean, J. R., Morley-Short, S., Shadbolt, P. J., Bonneau, D., Silverstone, J. W., Tew, D. P., Zhou, X., O'Brien, J. L. and Thompson, M. G.: Witnessing eigenstates for quantum simulation of Hamiltonian spectra, *Science Advances*, Vol. 4, No. 1 (2018).

[39] 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).

[40] Temme, K., Bravyi, S. and Gambetta, J. M.: Error Mitigation for Short-Depth Quantum Circuits, *Phys. Rev. Lett.*, Vol. 119, p. 180509 (online), DOI: 10.1103/PhysRevLett.119.180509 (2017).

[41] Tilly, J., Jones, G., Chen, H., Wossnig, L. and Grant, E.: Computation of molecular excited states on IBM quantum computers using a discriminative variational quantum eigensolver, *Physical Review A*, Vol. 102, No. 6, p. 062425 (2020).

[42] Wang, S., Fontana, E., Cerezo, M., Sharma, K., Sone, A., Cincio, L. and Coles, P. J.: Noise-Induced Barren Plateaus in Variational Quantum Algorithms, arXiv:2007.14384 (2020).

[43] Wei, M. and Schwab, D. J.: How noise affects the Hessian spectrum in overparameterized neural networks, arXiv:1910.00195 (2019).

[44] Wiebe, N., Kapoor, A. and Svore, K. M.: Quantum Deep Learning, arXiv:1412.3489 (2014).

