Supplementary Material of "Escaping from the Barren Plateau via Gaussian Initializations in Deep Variational Quantum Circuits"

Anonymous Author(s)
Affiliation
Address
`email`

This supplementary material contains four parts:

- Section A provides some additional experiment results.
- Section B provides some technical lemmas which are useful for proving main theorems in this work.
- Section C provides the proof of Theorem 4.1.
- Section D provides the proof of Theorem 4.2.

A Additional Experiments

A.1 Heisenberg model

In this section, we introduce additional experiment results toward finding the ground state energy of the Heisenberg model with different circuit depths and optimizers. We follow simulation details in the main text.

First, we consider the effect of different circuit depths and Gaussian initializations with different variances. The loss function has the formulation Eq. (10) with the number of qubits $N = 15$. We adopt the ansatz circuit 1 with $L_1 \in \{8, 10, 12\}$ layers of $R_Y R_X CZ$ blocks, which correspond with $L \in \{14, 18, 22\}$ case of Theorem 4.1, respectively. In the experiment, we train VQAs using gradient descent with the learning rate 0.01. Since the estimation of gradients on real quantum computers could be perturbed by statistical measurement noise, we compare optimizations using accurate and noisy gradients. For the latter case, we set the variance of measurement noises to be 0.01. We train different Gaussian initialized VQAs with variances $\{0.01\gamma, 0.1\gamma, \gamma, 10\gamma, 100\gamma\}$, where the value γ follows the formulation in Theorem 4.1.

We illustrate results in Figures 1 and 2, which correspond to the noiseless and the noisy case, respectively. As show in figures of the loss during optimizations, the Gaussian initialization with the variance γ outperforms other Gaussian initializations with faster convergence rates. Gaussian initializations with small variances $\{0.01\gamma, 0.1\gamma\}$ have similar performances with the zero initialization for the noisy training case, and Gaussian initializations with large variances $\{10\gamma, 100\gamma\}$ behave similarly with the uniform initialization presented in the main text. Moreover, circuits initialized with larger variances $\{10\gamma, 100\gamma\}$ need more iterations to converge when the depth increases, while circuits with variances $\{0.01\gamma, 0.1\gamma, \gamma\}$ show similar convergence rates for different depths.

Next, we compare different initializations with other optimizers, i.e., the gradient descent with momentum [1], the Nesterov accelerated gradient (NAG) [2], and the adaptive gradient (AdaGrad) [3]. We follow the loss function (10) with $(N, L) = (15, 18)$ and $(N, L) = (18.38)$. The learning rate and the noise are the same as that in the experiment considering different Gaussian variances. We illustrate results in Figures 3 and 4. As shown in Figure 3 and Figure 1 in the main text, the performance of

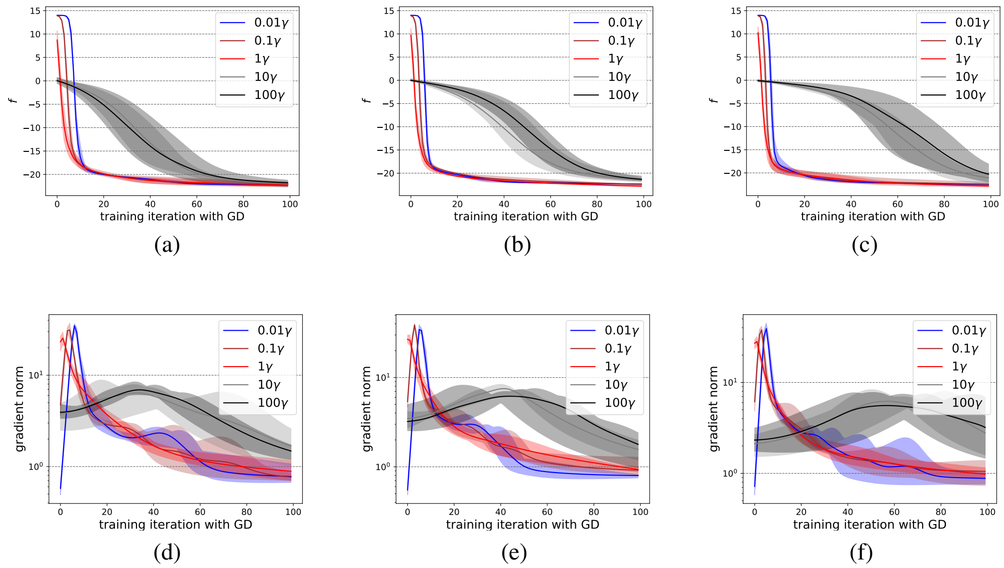

Figure 1: Numerical results of finding the ground state energy of the Heisenberg model using the noiseless gradient descent. Figures 1(a)-1(c) show the loss during optimizations for different $L \in \{14, 18, 22\}$ using the circuit 1 in the main text. For each L, we adopt Gaussian initializaions with different variances $0.01\gamma, 0.1\gamma, \gamma, 10\gamma, 100\gamma$, where the value γ follows the formulation in Theorem 4.1. Figures 1(d)-1(f) show the ℓ_2 norm of corresponding gradients during the optimization. Each line illustrates the average of 5 rounds of independent experiments.

GD with momentum and the NAG is similar to that of the Adam optimizer, while the performance of the AdaGrad is similar to the GD optimizer. By comparing Figures 3 and 4, we notice that uniformly initialized circuits converge slower when the qubit number and the circuit depth increase.

A.2 Quantum chemistry

In this section, we introduce additional experiment results toward finding the ground state energy of the Heisenberg model with different circuit depths. We repeat the LiH task in the main text with the depth $L \in \{24, 48, 72\}$ by stacking the circuit V_{Givens} in Eq. (11). The noise setting follows the adaptive noise with the variance in Eq. (12). We adopt gradient descent and the Adam optimizer with learning rates 0.1 and 0.01, respectively. The result is shown in Figure 5. For the gradient descent case, the convergence rate of the loss function increases when the circuit depth grows. For the Adam case, circuits with different depths show similar convergence speeds.

B Technical Lemmas

In this section, we provide some technical lemmas.

Lemma B.1. *Let θ be a variable with Gaussian distribution $\mathcal{N}(0, \gamma^2)$. Let $\rho = \sum_k c_k \rho_k$ be the linear combination of density matrices $\{\rho_k\}$ with real coefficients $\{c_k\}$. Let G be a hermitian unitary and $V = e^{-i\theta G}$. Let O be an arbitrary hermitian quantum observable that anti-commutes with G. Then*

$$\mathop{\mathbb{E}}_{\theta \sim \mathcal{N}(0,\gamma^2)} \text{Tr}\left[OV\rho V^\dagger\right]^2 \geq (1 - 4\gamma^2)\text{Tr}\left[O\rho\right]^2 + 4\gamma^2(1 - 4\gamma^2)\text{Tr}\left[iGO\rho\right]^2. \tag{1}$$

Proof. By replacing the term

$$V = e^{-i\theta G} = I\cos\theta - iG\sin\theta,$$

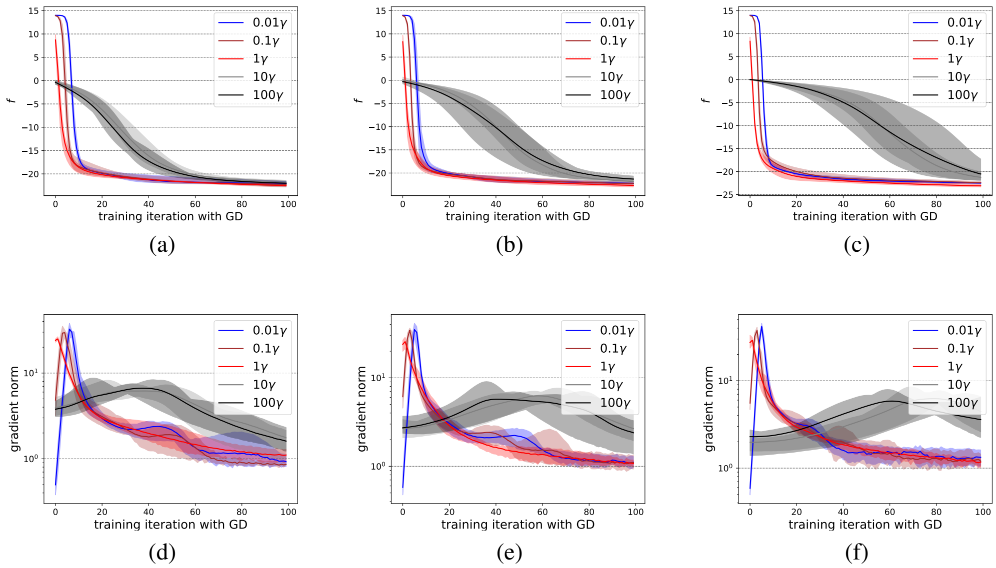

Figure 2: Numerical results of finding the ground state energy of the Heisenberg model using the noisy gradient descent. Figures 2(a)-2(c) show the loss during optimizations for different $L \in \{14, 18, 22\}$ using the circuit 1 in the main text. For each L, we adopt Gaussian initializaions with different variances $0.01\gamma, 0.1\gamma, \gamma, 10\gamma, 100\gamma$, where the value γ follows the formulation in Theorem 4.1. Figures 2(d)-2(f) show the ℓ_2 norm of corresponding gradients during the optimization. Each line illustrates the average of 5 rounds of independent experiments.

we have

$$\begin{aligned}\text{Tr}\left[OV\rho V^\dagger\right] &= \text{Tr}\left[O(I\cos\theta - iG\sin\theta)\rho(I\cos\theta + iG\sin\theta)\right] \\ &= \cos 2\theta \text{Tr}\left[O\rho\right] + \sin 2\theta \text{Tr}\left[iGO\rho\right], \end{aligned} \tag{2}$$

where Eq. (2) follows from the condition $OG + GO = 0$. Since O anti-commutes with G, iGO could be served as a hermitian observable. Based on Eq. (2), we have

$$\begin{aligned}\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \text{Tr}\left[OV\rho V^\dagger\right]^2 &= \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \Big(\cos 2\theta \text{Tr}\left[O\rho\right] + \sin 2\theta \text{Tr}\left[iGO\rho\right]\Big)^2 \\ &= \frac{1+e^{-8\gamma^2}}{2}\text{Tr}\left[O\rho\right]^2 + \frac{1-e^{-8\gamma^2}}{2}\text{Tr}\left[iGO\rho\right]^2 \qquad (3) \\ &\geq (1-4\gamma^2)\text{Tr}\left[O\rho\right]^2 + 4\gamma^2(1-4\gamma^2)\text{Tr}\left[iGO\rho\right]^2, \qquad (4)\end{aligned}$$

where Eq. (3) is obtained by calculating expectation terms. InEq. (4) holds since $1 - 8\gamma^2 \leq e^{-8\gamma^2} \leq 1 - 8\gamma^2 + 32\gamma^4$. Thus, we have proved Eq. (1).

□

Lemma B.2. *Let θ be a variable with Gaussian distribution $\mathcal{N}(0, \gamma^2)$. Let ρ be the density matrix of a quantum state. Let G be a hermitian unitary and $V = e^{-i\theta G}$. Let O be an arbitrary hermitian quantum observable that anti-commutes with G. Then*

$$\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(\frac{\partial}{\partial\theta}\text{Tr}\left[OV\rho V^\dagger\right]\right)^2 \geq (1-4\gamma^2)\left(\frac{\partial}{\partial\theta}\text{Tr}\left[OV\rho V^\dagger\right]\right)^2\Bigg|_{\theta=0} + 16\gamma^2(1-4\gamma^2)\text{Tr}\left[O\rho\right]^2. \tag{5}$$

Proof. By calculating the gradient for both sides of Eq. (2), we obtain

$$\frac{\partial}{\partial\theta}\text{Tr}\left[OV\rho V^\dagger\right] = -2\sin 2\theta \text{Tr}\left[O\rho\right] + 2\cos 2\theta \text{Tr}\left[iGO\rho\right]. \tag{6}$$

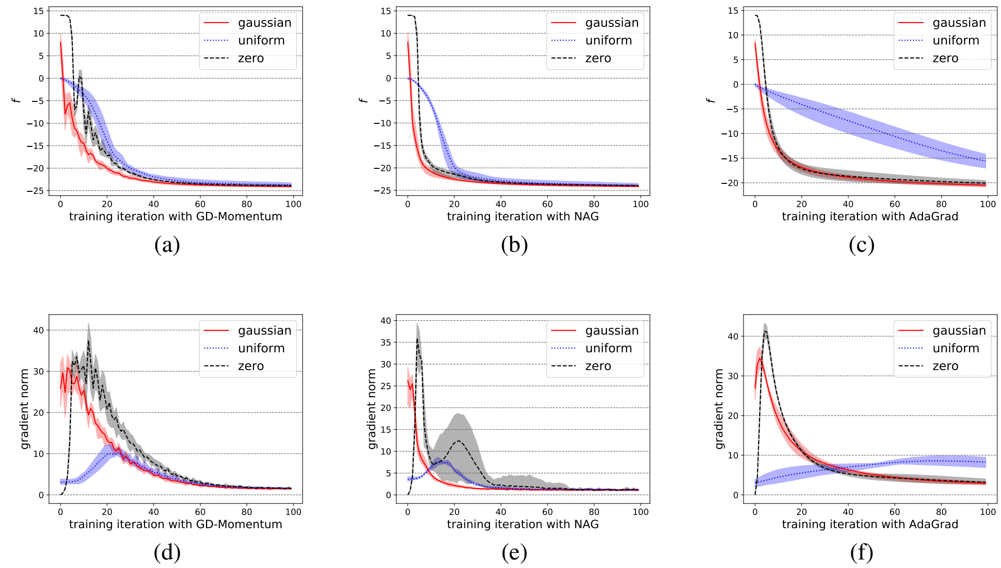

Figure 3: Numerical results of finding the ground state energy of the Heisenberg model with qubits $N = 15$ (noisy case). Figures 3(a)-3(c) show the loss during optimizations using the gradient descent with momentum, the Nesterov accelerated gradient (NAG), and the adaptive gradient (AdaGrad), respectively. Figures 3(d)-3(f) show the ℓ_2 norm of gradients during the optimization. Each line illustrates the average of 5 rounds of independent experiments.

Let $\theta = 0$ in Eq. (6), we obtain

$$\frac{\partial}{\partial \theta} \mathrm{Tr}\left[OV\rho V^\dagger\right]\Big|_{\theta=0} = 2\mathrm{Tr}\left[iGO\rho\right]. \tag{7}$$

Now we proceed to prove Lemma B.2.

$$\begin{aligned}
\text{The left part of Eq. (5)} &= \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(-2\sin 2\theta \mathrm{Tr}\left[O\rho\right] + 2\cos 2\theta \mathrm{Tr}\left[iGO\rho\right]\right)^2 \\
&= 2(1-e^{-8\gamma^2})\mathrm{Tr}\left[O\rho\right]^2 + 2(1+e^{-8\gamma^2})\mathrm{Tr}\left[iGO\rho\right]^2 && (8) \\
&\geq 16\gamma^2(1-4\gamma^2)\mathrm{Tr}\left[O\rho\right]^2 + 4(1-4\gamma^2)\mathrm{Tr}\left[iGO\rho\right]^2 && (9) \\
&= (1-4\gamma^2)\left(\frac{\partial}{\partial\theta}\mathrm{Tr}\left[OV\rho V^\dagger\right]\right)^2\Big|_{\theta=0} + 16\gamma^2(1-4\gamma^2)\mathrm{Tr}\left[O\rho\right]^2. && (10)
\end{aligned}$$

Eq. (8) is obtained by calculating expectation terms. InEq. (9) is obtained by using $1-8\gamma^2 \leq e^{-8\gamma^2} \leq 1-8\gamma^2+32\gamma^4$. Eq. (10) follows from Eq. (7). Thus, we have proved Eq. (5).

□

Lemma B.3. *Denote by $\rho = \sum_k c_k\rho_k$ the linear combination of density matrices $\{\rho_k\}$ with real coefficients $\{c_k\}$. Let $V_h(\theta) = W_1 e^{-i\theta G_1} W_2 \cdots W_h e^{-i\theta G_h}$, where $\{G_n\}_{n=1}^h$ is a list of hermitian unitaries and $\{W_n\}_{n=1}^h$ is a list of unitary matrices. Denote by O an arbitrary hermitian quantum observable. Then*

$$\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \mathrm{Tr}\left[OV_h(\theta)\rho V_h(\theta)^\dagger\right]^2 \geq \mathrm{Tr}\left[OV_h(0)\rho V_h(0)^\dagger\right]^2 - \left[12h(h-1)+4\right]\gamma^2\|\mathbf{c}\|_1^2\|O\|_2^2, \tag{11}$$

where $\|\mathbf{c}\|_1 = \sum_k |c_k|$ denotes the ℓ_1 norm of $\mathbf{c}$, $\|O\|_2$ denotes the spectral norm of O, and the variance $\gamma^2 \leq \frac{1}{12h^2}$.

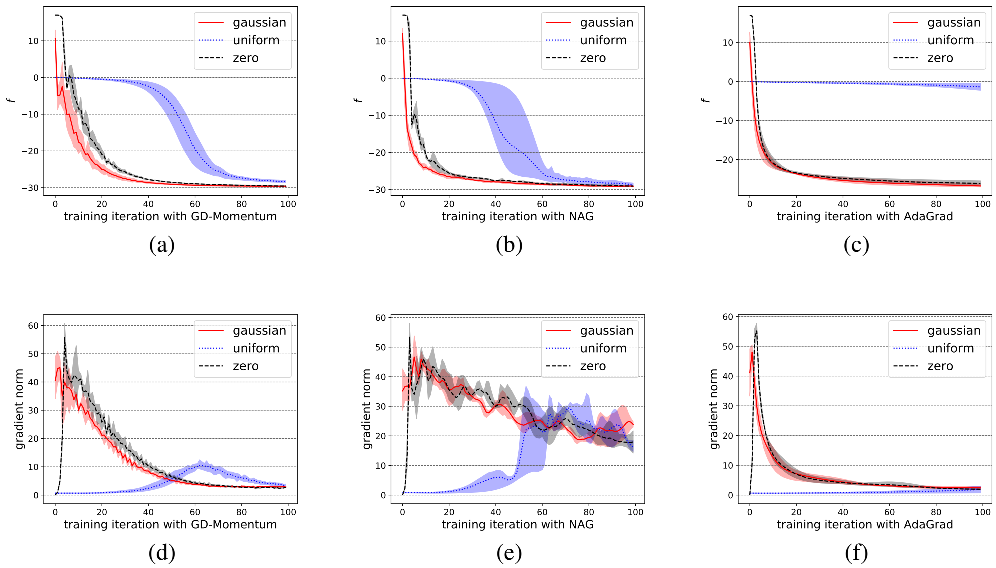

Figure 4: Numerical results of finding the ground state energy of the Heisenberg model with qubits $N = 18$ (noisy case). Figures 4(a)-4(c) show the loss during optimizations using the gradient descent with momentum, the Nesterov accelerated gradient (NAG), and the adaptive gradient (AdaGrad), respectively. Figures 4(d)-4(f) show the ℓ_2 norm of gradients during the optimization. Each line illustrates the average of 3 rounds of independent experiments.

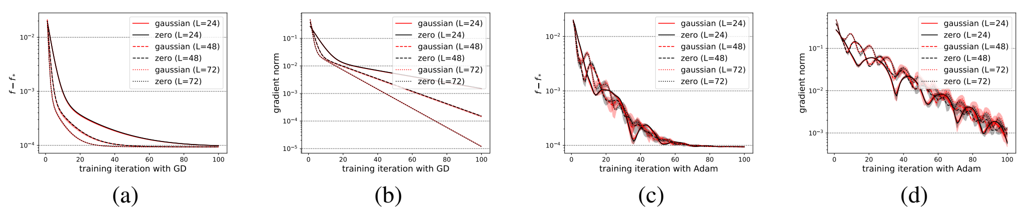

Figure 5: Numerical results of finding the ground state energy of the LiH molecule using noisy gradients. Figures 5(a) and 5(c) show the loss during optimizations for different $L \in \{24, 48, 72\}$ with the gradient descent and the Adam optimizer, respectively. Figures 5(b) and 5(d) show the ℓ_2 norm of gradients during the optimization. Each line illustrates the average of 3 rounds of independent experiments.

Proof. Before the proof, we define several notations for convenience. We define $V_0 = I$ and

$$V_j(\theta) = W_{h+1-j} e^{-i\theta G_{h+1-j}} \cdots W_h e^{-i\theta G_h}, \forall j \in \{1, \cdots, h\}. \tag{12}$$

We denote $\mathbf{0}_k$, $\mathbf{1}_k$, and $\mathbf{2}_k$ as k-dimensional vectors with components 0, 1, and 2, respectively. We define $O^{j_1,\cdots,j_k}_{i_1,\cdots,i_k} = O$ for the $k = 0$ case and

$$O^{j_1,\cdots,j_k}_{i_1,\cdots,i_k} = \begin{cases} {W_k}^\dagger O^{j_1,\cdots,j_{k-1}}_{i_1,\cdots,i_{k-1}} W_k, & \text{if } i_k = 0,\ j_k = 0, \\ \frac{1}{2} G_k \left\{ G_k, {W_k}^\dagger O^{j_1,\cdots,j_{k-1}}_{i_1,\cdots,i_{k-1}} W_k \right\}, & \text{if } i_k = 1,\ j_k = 0, \\ \frac{1}{2} G_k \left[G_k, {W_k}^\dagger O^{j_1,\cdots,j_{k-1}}_{i_1,\cdots,i_{k-1}} W_k \right], & \text{if } i_k = 2,\ j_k = 0, \\ i G_k O^{j_1,\cdots,j_{k-1},0}_{i_1,\cdots,i_{k-1},i_k}, & \text{if } j_k = 1, \end{cases} \tag{13}$$

for increasing $k \in \{1, \cdots, h\}$, where $i_k \in \{0, 1, 2\}$ and $j_k \in \{0, 1\}$.

For all $1 \le k \le \ell \le h$, the definition (13) provides the commuting and anti-commuting parts of $O_{i_1,\cdots,i_{k-1},0,i_{k+1},\cdots,i_\ell}^{j_1,\cdots,j_{k-1},0,j_{k+1},\cdots,j_\ell}$ with respect to G_k, respectively, i.e.,

$$\begin{aligned}
O_{i_1,\cdots,i_{k-1},0,i_{k+1},\cdots,i_\ell}^{j_1,\cdots,j_{k-1},0,j_{k+1},\cdots,j_\ell} &= O_{i_1,\cdots,i_{k-1},1,i_{k+1},\cdots,i_\ell}^{j_1,\cdots,j_{k-1},0,j_{k+1},\cdots,j_\ell} + O_{i_1,\cdots,i_{k-1},2,i_{k+1},\cdots,i_\ell}^{j_1,\cdots,j_{k-1},0,j_{k+1},\cdots,j_\ell}, \\
G_k O_{i_1,\cdots,i_{k-1},1,i_{k+1},\cdots,i_\ell}^{j_1,\cdots,j_{k-1},0,j_{k+1},\cdots,j_\ell} &= O_{i_1,\cdots,i_{k-1},1,i_{k+1},\cdots,i_\ell}^{j_1,\cdots,j_{k-1},0,j_{k+1},\cdots,j_\ell} G_k, \\
G_k O_{i_1,\cdots,i_{k-1},2,i_{k+1},\cdots,i_\ell}^{j_1,\cdots,j_{k-1},0,j_{k+1},\cdots,j_\ell} &= -O_{i_1,\cdots,i_{k-1},2,i_{k+1},\cdots,i_\ell}^{j_1,\cdots,j_{k-1},0,j_{k+1},\cdots,j_\ell} G_k.
\end{aligned}$$

Since for all $k \in [L]$, G_k is a unitary matrix, $O_{\boldsymbol{i}}^{\boldsymbol{j}}$ is a hermitian observable for all $\boldsymbol{i} \in \{0,1,2\}^\ell$, $\boldsymbol{j} \in \{0,1\}^\ell$, and $\ell \in [L]$. Meanwhile, the spectral norm of $O_{\boldsymbol{i}}^{\boldsymbol{j}}$ is bounded,

$$\begin{aligned}
\left\| O_{i_1,\cdots,i_{h-1},i_h}^{j_1,\cdots,j_{h-1},j_h} \right\|_2 \le \left\| O_{i_1,\cdots,i_{h-1},i_h}^{j_1,\cdots,j_{h-1},0} \right\|_2 &\le \frac{1}{2} \left\| O_{i_1,\cdots,i_{h-1},0}^{j_1,\cdots,j_{h-1},0} \right\|_2 + \frac{1}{2} \left\| O_{i_1,\cdots,i_{h-1},0}^{j_1,\cdots,j_{h-1},0} \right\|_2 \\
&= \left\| O_{i_1,\cdots,i_{h-1},0}^{j_1,\cdots,j_{h-1},0} \right\|_2 = \left\| O_{i_1,\cdots,i_{h-1}}^{j_1,\cdots,j_{h-1}} \right\|_2 \le \|O\|_2, \qquad (14)
\end{aligned}$$

where $\|A\|_2$ denotes the spectral norm of the matrix A. Moreover, for all $k, \ell \ge 0$ such that $k+\ell \le h$, the observable $O_{i_1,\cdots,i_k,\mathbf{0}_{h-k-\ell},i_{h-\ell+1},\cdots,i_h}^{j_1,\cdots,j_k,\mathbf{0}_{h-k-\ell},j_{h-\ell+1},\cdots,j_h}$ could be recovered by

$$\sum_{n=k+1}^{h-\ell} \sum_{i_n=1}^{2} O_{i_1,\cdots,i_k,i_{k+1},\cdots,i_{h-\ell},i_{h-\ell+1},\cdots,i_h}^{j_1,\cdots,j_k,\mathbf{0}_{h-k-\ell},j_{h-\ell+1},\cdots,j_h} = O_{i_1,\cdots,i_k,\mathbf{0}_{h-k-\ell},i_{h-\ell+1},\cdots,i_h}^{j_1,\cdots,j_k,\mathbf{0}_{h-k-\ell},j_{h-\ell+1},\cdots,j_h}. \qquad (15)$$

Now we begin the proof. To analyze the expectation with respect to the parameter θ, we need the detailed formulation of $\mathrm{Tr}\left[OV_h\rho V_h{}^\dagger\right]$ as the function of θ. In fact, for all $h' \in \{0,1,\cdots,h\}$ and all

$$\boldsymbol{i} \in \{0,1,2\}^{h-h'},\ \boldsymbol{j} \in \{0,1\}^{h-h'},$$

we have

$$\mathrm{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}} V_{h'}\rho V_{h'}{}^\dagger\right] = \sum_{\boldsymbol{j}'=\mathbf{0}_{h'}}^{\mathbf{1}_{h'}} \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_{h'}}^{\mathbf{2}_{h'}} (\cos 2\theta)^{\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - h'} (\sin 2\theta)^{\|\boldsymbol{j}'\|_1} \mathrm{Tr}\left[O_{\boldsymbol{i},\boldsymbol{i}'}^{\boldsymbol{j},\boldsymbol{j}'}\rho\right], \qquad (16)$$

where $\|\boldsymbol{i}'\|_1 \equiv \sum_{k=1}^{\dim(\boldsymbol{i}')} |i'_k|$ denotes the ℓ_1 norm of the vector $\boldsymbol{i}$.

Eq. (16) can be proved inductively. First, for the case $h'=0$, Eq. (16) holds trivially. Next, we assume that Eq. (16) holds for the $h'=k$ case. Then for all

$$\boldsymbol{i} \in \{0,1,2\}^{h-k-1},\ \boldsymbol{j} \in \{0,1\}^{h-k-1},$$

we have

$$\begin{aligned}
\mathrm{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}} V_{k+1}\rho V_{k+1}{}^\dagger\right] =& \mathrm{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}} W_{h-k}(I\cos\theta - iG_{h-k}\sin\theta)V_k\rho V_k{}^\dagger (I\cos\theta + iG_{h-k}\sin\theta)W_{h-k}{}^\dagger\right] \\
=& \cos^2\theta \mathrm{Tr}\left[O_{\boldsymbol{i},0}^{\boldsymbol{j},0} V_k\rho V_k{}^\dagger\right] + \sin^2\theta \mathrm{Tr}\left[G_{h-k} O_{\boldsymbol{i},0}^{\boldsymbol{j},0} G_{h-k} V_k\rho V_k{}^\dagger\right] \\
&+ \sin\theta\cos\theta \mathrm{Tr}\left[iG_{h-k} O_{\boldsymbol{i},0}^{\boldsymbol{j},0} V_k\rho V_k{}^\dagger\right] - \sin\theta\cos\theta \mathrm{Tr}\left[O_{\boldsymbol{i},0}^{\boldsymbol{j},0} iG_{h-k} V_k\rho V_k{}^\dagger\right] \qquad (17) \\
=& \mathrm{Tr}\left[O_{\boldsymbol{i},1}^{\boldsymbol{j},0} V_k\rho V_k{}^\dagger\right] + \cos 2\theta \mathrm{Tr}\left[O_{\boldsymbol{i},2}^{\boldsymbol{j},0} V_k\rho V_k{}^\dagger\right] + \sin 2\theta \mathrm{Tr}\left[O_{\boldsymbol{i},2}^{\boldsymbol{j},1} V_k\rho V_k{}^\dagger\right], \qquad (18)
\end{aligned}$$

where Eqs. (17) and (18) are derived by using the definition (13). We proceed by employing the $h'=k$ case of Eq. (16), such that

$$\begin{aligned}
\text{Eq. (18)} =& \sum_{\boldsymbol{j}'=\mathbf{0}_k}^{\mathbf{1}_k} \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} (\cos 2\theta)^{\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k} (\sin 2\theta)^{\|\boldsymbol{j}'\|_1} \mathrm{Tr}\left[O_{\boldsymbol{i},1,\boldsymbol{i}'}^{\boldsymbol{j},0,\boldsymbol{j}'}\rho\right] \\
&+ \cos 2\theta \sum_{\boldsymbol{j}'=\mathbf{0}_k}^{\mathbf{1}_k} \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} (\cos 2\theta)^{\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k} (\sin 2\theta)^{\|\boldsymbol{j}'\|_1} \mathrm{Tr}\left[O_{\boldsymbol{i},2,\boldsymbol{i}'}^{\boldsymbol{j},0,\boldsymbol{j}'}\rho\right]
\end{aligned}$$

$$
\begin{aligned}
&+ \sin 2\theta \sum_{\boldsymbol{j}'=\boldsymbol{0}_k}^{\boldsymbol{1}_k} \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_k}^{\boldsymbol{2}_k} (\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-k} (\sin 2\theta)^{\|\boldsymbol{j}'\|_1} \operatorname{Tr}\left[O_{\boldsymbol{i},2,\boldsymbol{i}'}^{\boldsymbol{j},1,\boldsymbol{j}'}\rho\right] \\
=& \sum_{\boldsymbol{j}'=\boldsymbol{0}_{k+1}}^{\boldsymbol{1}_{k+1}} \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_{k+1}}^{\boldsymbol{2}_{k+1}} (\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-k-1} (\sin 2\theta)^{\|\boldsymbol{j}'\|_1} \operatorname{Tr}\left[O_{\boldsymbol{i},\boldsymbol{i}'}^{\boldsymbol{j},\boldsymbol{j}'}\rho\right], \qquad (19)
\end{aligned}
$$

which matches the formulation of the $h' = k+1$ case of Eq. (16). Thus, Eq. (16) has been proved. Now we begin to prove Eq. (11). Employing the $h' = h$ case of Eq. (16) could yield

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(\operatorname{Tr}\left[O V_h \rho V_h{}^\dagger\right]\right)^2 \\
=& \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(\sum_{\boldsymbol{j}=\boldsymbol{0}_h}^{\boldsymbol{1}_h} \sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h} (\sin 2\theta)^{\|\boldsymbol{j}\|_1} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \right)^2 \qquad (20) \\
=& \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(\sum_{\boldsymbol{i}=\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}\|_1-h} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{0}_h}\rho\right] + \sum_{\boldsymbol{j}>\boldsymbol{0}_h}^{\boldsymbol{1}_h} \sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h} (\sin 2\theta)^{\|\boldsymbol{j}\|_1} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \right)^2 \\
& \qquad (21) \\
\geq& \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(\sum_{\boldsymbol{i}=\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}\|_1-h} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{0}_h}\rho\right] \right)^2 \\
&+ 2 \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\boldsymbol{j}>\boldsymbol{0}_h}^{\boldsymbol{1}_h} \sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h} (\sin 2\theta)^{\|\boldsymbol{j}\|_1} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \sum_{\boldsymbol{i}'=\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}'\|_1-h} \operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{0}_h}\rho\right]. \\
& \qquad (22)
\end{aligned}
$$

InEq. (22) is obtained by discarding the square of the latter term in the bracket of Eq. (21). We remark that if Eqs. (23) and (24) hold, we can prove Eq. (11) by using Eqs. (20-22).

$$
\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(\sum_{\boldsymbol{i}=\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}\|_1-h} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{0}_h}\rho\right] \right)^2 - \left(\operatorname{Tr}\left[O V_h(0) \rho V_h(0)^\dagger\right]\right)^2 \geq -(6h-2)\gamma^2 \|\boldsymbol{c}\|_1^2 \|O\|_2^2, \qquad (23)
$$

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\boldsymbol{j}>\boldsymbol{0}_h}^{\boldsymbol{1}_h} \sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h} (\sin 2\theta)^{\|\boldsymbol{j}\|_1} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \sum_{\boldsymbol{i}'=\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}'\|_1-h} \operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{0}_h}\rho\right] \\
\geq& -\left(6h^2-9h+3\right)\gamma^2 \|\boldsymbol{c}\|_1^2 \|O\|_2^2. \qquad (24)
\end{aligned}
$$

In the following proof, we would derive Eqs. (23) and (24). We focus on the Eq. (23) first. In fact, the left side of Eq. (23) is bounded by

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(\sum_{\boldsymbol{i}=\boldsymbol{1}_h}^{\boldsymbol{2}_h} \left[1-(\cos 2\theta)^{\|\boldsymbol{i}\|_1-h}-1\right] \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{0}_h}\rho\right] \right)^2 - \left(\operatorname{Tr}\left[O_{\boldsymbol{0}_h}^{\boldsymbol{0}_h}\rho\right]\right)^2 \qquad (25) \\
=& \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(\sum_{\boldsymbol{i}=\boldsymbol{1}_h}^{\boldsymbol{2}_h} \left[1-(\cos 2\theta)^{\|\boldsymbol{i}\|_1-h}-1\right] \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{0}_h}\rho\right] \right)^2 - \left(\sum_{\boldsymbol{i}=\boldsymbol{1}_h}^{\boldsymbol{2}_h} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{0}_h}\rho\right] \right)^2 \qquad (26) \\
\geq& -2 \left| \sum_{\boldsymbol{i}=\boldsymbol{1}_h}^{\boldsymbol{2}_h} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{0}_h}\rho\right] \right| \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left| \sum_{\boldsymbol{i}=\boldsymbol{1}_h}^{\boldsymbol{2}_h} \left[1-(\cos 2\theta)^{\|\boldsymbol{i}\|_1-h}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{0}_h}\rho\right] \right| \qquad (27) \\
=& -2 \left|\operatorname{Tr}\left[O_{\boldsymbol{0}_h}^{\boldsymbol{0}_h}\rho\right]\right| \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left| \sum_{\boldsymbol{i}=\boldsymbol{1}_h}^{\boldsymbol{2}_h} \left[1-(\cos 2\theta)^{\|\boldsymbol{i}\|_1-h}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{0}_h}\rho\right] \right| \qquad (28)
\end{aligned}
$$

$$
\begin{aligned}
&\geq -2\|\boldsymbol{c}\|_1\|O\|_2 \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left|\sum_{\boldsymbol{i}=\boldsymbol{1}_h}^{\boldsymbol{2}_h}\left[1-(\cos 2\theta)^{\|\boldsymbol{i}\|_1-h}\right]\mathrm{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{0}_h}\rho\right]\right| &(29)\\
&\geq -2\|\boldsymbol{c}\|_1^2\|O\|_2^2 \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left[(2-\cos 2\theta)^h-1\right]. &(30)
\end{aligned}
$$

Eq. (25) is obtained by using the definition (13). Eq. (26) is derived by using Eq. (15). InEq. (27) is obtained by using $(a-b)^2-b^2 \geq -2|a|\cdot|b|$. Eq. (28) yields from Eq. (15). InEq. (29) is derived by using

$$
\left|\mathrm{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right| = \left|\sum_k c_k \mathrm{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho_k\right]\right| \leq \sum_k |c_k| \left|\mathrm{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho_k\right]\right| \leq \sum_k |c_k| \left\|O_{\boldsymbol{i}}^{\boldsymbol{j}}\right\|_2 \leq \|\boldsymbol{c}\|_1\|O\|_2. \quad (31)
$$

InEq. (30) is obtained by using the $h'=h$ case of InEq. (32), i.e.,

$$
\left|\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_{h'}}^{\boldsymbol{2}_{h'}}\left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-h'}\right]\mathrm{Tr}\left[O_{\boldsymbol{i}',\boldsymbol{i}}^{\boldsymbol{j}',\boldsymbol{j}}\rho\right]\right| \leq \left[(2-\cos 2\theta)^{h'-\|\boldsymbol{j}'\|_1}-1\right]\|\boldsymbol{c}\|_1\|O\|_2 \quad (32)
$$

for all $h' \in \{0,1,\cdots,h\}$, $\boldsymbol{j}' \in \{0,1\}^{h'}$, $\boldsymbol{i} \in \{0,1,2\}^{h-h'}$, and $\boldsymbol{j} \in \{0,1\}^{h-h'}$.

InEq. (32) can be proved inductively. First, for the case $h'=0$, Eq. (32) holds trivially. Next we assume that Eq. (32) holds for the case $h'=k$. Then for all $\boldsymbol{i} \in \{0,1,2\}^{h-k-1}$ and $\boldsymbol{j} \in \{0,1\}^{h-k-1}$, we have

$$
\begin{aligned}
&\left|\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_{k+1}}^{\boldsymbol{2}_{k+1}}\left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-k-1}\right]\mathrm{Tr}\left[O_{\boldsymbol{i}',\boldsymbol{i}}^{\boldsymbol{j}',\boldsymbol{j}}\rho\right]\right|\\
=&\left|\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_k}^{\boldsymbol{2}_k}\sum_{i'_{k+1}=j'_{k+1}+1}^{2}\left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1+i'_{k+1}-\|\boldsymbol{j}'\|_1-j'_{k+1}-k-1}\right]\mathrm{Tr}\left[O_{\boldsymbol{i}',i'_{k+1},\boldsymbol{i}}^{\boldsymbol{j}',j'_{k+1},\boldsymbol{j}}\rho\right]\right|. &(33)
\end{aligned}
$$

For the case $j'_{k+1}=1$,

$$
\begin{aligned}
\text{Eq. (33)} &= \left|\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_k}^{\boldsymbol{2}_k}\left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-k}\right]\mathrm{Tr}\left[O_{\boldsymbol{i}',2,\boldsymbol{i}}^{\boldsymbol{j}',1,\boldsymbol{j}}\rho\right]\right|\\
&\leq \left[(2-\cos 2\theta)^{k-\|\boldsymbol{j}'\|_1}-1\right]\|\boldsymbol{c}\|_1\|O\|_2 &(34)\\
&= \left[(2-\cos 2\theta)^{k+1-\|\boldsymbol{j}'\|_1-j'_{k+1}}-1\right]\|\boldsymbol{c}\|_1\|O\|_2. &(35)
\end{aligned}
$$

InEq. (34) is derived by using the $h'=k$ case of InEq. (32). Eq. (35) is derived by using $j'_{k+1}=1$. For the case $j'_{k+1}=0$,

$$
\begin{aligned}
\text{Eq. (33)} =& \left|\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_k}^{\boldsymbol{2}_k}\left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-k}\right]\mathrm{Tr}\left[O_{\boldsymbol{i}',1,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right]\right.\\
&\left.+\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_k}^{\boldsymbol{2}_k}\left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-k+1}\right]\mathrm{Tr}\left[O_{\boldsymbol{i}',2,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right]\right|\\
=& \left|\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_k}^{\boldsymbol{2}_k}\left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-k}\right]\mathrm{Tr}\left[O_{\boldsymbol{i}',0,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right]\right.\\
&\left.+(1-\cos 2\theta)\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_k}^{\boldsymbol{2}_k}\left[(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-k}-1+1\right]\mathrm{Tr}\left[O_{\boldsymbol{i}',2,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right]\right| &(36)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \left| \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-k}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}',0,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right] \right| \\
&\quad + (1-\cos 2\theta) \left| \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-k}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}',2,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right] \right| \\
&\quad + (1-\cos 2\theta) \left| \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \operatorname{Tr}\left[O_{\boldsymbol{i}',2,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right] \right| \qquad (37) \\
&\leq \left[(2-\cos 2\theta)^{k-\|\boldsymbol{j}'\|_1}-1\right] \|\boldsymbol{c}\|_1 \|O\|_2 + (1-\cos 2\theta)\left[(2-\cos 2\theta)^{k-\|\boldsymbol{j}'\|_1}-1\right] \|\boldsymbol{c}\|_1 \|O\|_2 \\
&\quad + (1-\cos 2\theta)\|\boldsymbol{c}\|_1 \|O\|_2 \qquad (38) \\
&\leq \left[(2-\cos 2\theta)^{k+1-\|\boldsymbol{j}'\|_1-j'_{k+1}}-1\right] \|\boldsymbol{c}\|_1 \|O\|_2 . \qquad (39)
\end{aligned}
$$

Eq. (36) is derived by using Eq. (15). InEq. (37) is obtained since $|a+b| \leq |a|+|b|$. InEq. (38) is obtained using the $h'=k$ case of Eq. (32) and Eq. (15). InEq. (39) is derived by using Eq. (14). Thus we have proved Eq. (32) since Eqs. (35) and (39) match the $h'=k+1$ case.

Since $\cos 2\theta \geq 1-2\theta^2$, we could further bound Eq. (30) by

$$
\begin{aligned}
\text{Eq. (30)} &\geq -2\|\boldsymbol{c}\|_1^2\|O\|_2^2 \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left[(1+2\theta^2)^h-1\right] \qquad (40) \\
&= -2\|\boldsymbol{c}\|_1^2\|O\|_2^2 \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{t=1}^{h}\binom{h}{t}(2\theta^2)^t \\
&= -2\|\boldsymbol{c}\|_1^2\|O\|_2^2 \sum_{t=1}^{h}\binom{h}{t}(2t-1)!!(2\gamma^2)^t \qquad (41) \\
&\geq -2\|\boldsymbol{c}\|_1^2\|O\|_2^2 \sum_{t=1}^{h} h(h-1)^{t-1}2^{t-1}\left(\frac{1}{6h^2}\right)^{t-1}(2\gamma^2) \qquad (42) \\
&= -2\|\boldsymbol{c}\|_1^2\|O\|_2^2 2h\gamma^2 \left[1+\frac{h-1}{3h^2}\sum_{t=0}^{h-2}\left(\frac{h-1}{3h^2}\right)^t\right] \\
&\geq -(6h-2)\|\boldsymbol{c}\|_1^2\|O\|_2^2\gamma^2. \qquad (43)
\end{aligned}
$$

Eq. (41) is derived by calculating expectation terms. InEq. (42) yields from $\frac{(2t-1)!!}{t!} \leq 2^{t-1}$ and the condition $\gamma^2 \leq \frac{1}{12h^2}$. Thus, we have proved InEq. (23).

Next, we focus on the Eq. (24). The left side of Eq. (24) could be bounded by

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\boldsymbol{j}>\mathbf{0}_h,2|\|\boldsymbol{j}\|_1}^{\mathbf{1}_h} \sum_{\boldsymbol{i}=\boldsymbol{j}+\mathbf{1}_h}^{\mathbf{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h}(\sin 2\theta)^{\|\boldsymbol{j}\|_1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \\
&\qquad \cdot \sum_{\boldsymbol{i}'=\mathbf{1}_h}^{\mathbf{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}'\|_1-h}\operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\mathbf{0}_h}\rho\right] \qquad (44) \\
&\geq -\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\boldsymbol{j}>\mathbf{0}_h,2|\|\boldsymbol{j}\|_1}^{\mathbf{1}_h} (\sin 2\theta)^{\|\boldsymbol{j}\|_1} \left(\left| \sum_{\boldsymbol{i}=\boldsymbol{j}+\mathbf{1}_h}^{\mathbf{2}_h} \left[1-(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \right| \right. \\
&\qquad \left. + \left| \sum_{\boldsymbol{i}=\boldsymbol{j}+\mathbf{1}_h}^{\mathbf{2}_h} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \right| \right) \cdot \left(\left| \sum_{\boldsymbol{i}'=\mathbf{1}_h}^{\mathbf{2}_h} \left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-h}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\mathbf{0}_h}\rho\right] \right| + \left| \sum_{\boldsymbol{i}'=\mathbf{1}_h}^{\mathbf{2}_h} \operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\mathbf{0}_h}\rho\right] \right| \right) \\
&\qquad (45) \\
&\geq -\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\boldsymbol{j}>\mathbf{0}_h,2|\|\boldsymbol{j}\|_1}^{\mathbf{1}_h} (\sin 2\theta)^{\|\boldsymbol{j}\|_1} \left(\left[(2-\cos 2\theta)^{h-\|\boldsymbol{j}\|_1}-1\right] \|\boldsymbol{c}\|_1 \left\|O_{\mathbf{0}_h}^{\mathbf{0}_h}\right\|_2 + \left|\operatorname{Tr}\left[O_{2\boldsymbol{j}}^{\boldsymbol{j}}\rho\right]\right| \right)
\end{aligned}
$$

$$\cdot \left(\left[(2-\cos 2\theta)^h - 1 \right] \|\boldsymbol{c}\|_1 \left\| O^{\mathbf{0}_h}_{\mathbf{0}_h} \right\|_2 + \left| \mathrm{Tr}\left[O^{\mathbf{0}_h}_{\mathbf{0}_h} \rho \right] \right| \right) \tag{46}$$

$$\geq - \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\boldsymbol{j}>\mathbf{0}_h, 2|\|\boldsymbol{j}\|_1}^{\mathbf{1}_h} (\sin 2\theta)^{\|\boldsymbol{j}\|_1} (2-\cos 2\theta)^{2h-\|\boldsymbol{j}\|_1} \|\boldsymbol{c}\|_1^2 \left\|O\right\|_2^2 \tag{47}$$

$$\geq - \|\boldsymbol{c}\|_1^2 \|O\|_2^2 \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\boldsymbol{j}>\mathbf{0}_h, 2|\|\boldsymbol{j}\|_1}^{\mathbf{1}_h} (2\theta)^{\|\boldsymbol{j}\|_1} (1+2\theta^2)^{2h-\|\boldsymbol{j}\|_1}. \tag{48}$$

Eq. (44) is obtained by noticing that the expectation of $\sin^a 2\theta \cos^b 2\theta$ equals to zero, if a is odd. InEq. (45) is obtained by using $\sum_{i,j} a_i b_j \geq -(\sum_i |a_i|)(\sum_j |b_j|)$ and $|a+b| \leq |a|+|b|$. InEq. (46) is derived by using the $h' = h$ case of Eq. (32) and Eq. (15). InEq. (47) is obtained by using $\|O^{\mathbf{0}_h}_{\mathbf{0}_h}\| = \|O\|$ and Eq. (31). InEq. (48) is derived by using $(\sin 2\theta)^2 \leq (2\theta)^2$ and $\cos 2\theta \geq 1 - 2\theta^2$.

We proceed from InEq. (48), which could be further bounded by

$$= - \|\boldsymbol{c}\|_1^2 \|O\|_2^2 \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{t=1}^{\lfloor h/2 \rfloor} \binom{h}{2t} (2\theta)^{2t} \sum_{m=0}^{2h-2t} \binom{2h-2t}{m} (2\theta^2)^m \tag{49}$$

$$= - \|\boldsymbol{c}\|_1^2 \|O\|_2^2 \sum_{t=1}^{\lfloor h/2 \rfloor} \binom{h}{2t} \sum_{m=0}^{2h-2t} \binom{2h-2t}{m} 2^{2t+m} (2t+2m-1)!! \gamma^{2t+2m} \tag{50}$$

$$\begin{aligned} \geq & - \|\boldsymbol{c}\|_1^2 \|O\|_2^2 \sum_{t=1}^{\lfloor h/2 \rfloor} \sum_{m=0}^{2h-2t} \frac{h(h-1)^{2t-1}}{2^t t! (2t-1)!!} \frac{(2h-2)^m}{m!} 2^{2t+m} \\ & \cdot (2t-1)!!(2t+1)(2t+3)\cdots(2t+2m-1)\gamma^{2t+2m} \end{aligned} \tag{51}$$

$$\geq - \|\boldsymbol{c}\|_1^2 \|O\|_2^2 \sum_{t=1}^{\lfloor h/2 \rfloor} \sum_{m=0}^{2h-2t} \frac{h(h-1)^{2t-1}}{2^t 2^{t-1}} \frac{(2h-2)^m}{m!} 2^{2t+m} (2h)^m \gamma^{2t+2m} \tag{52}$$

$$\geq - \|\boldsymbol{c}\|_1^2 \|O\|_2^2 \sum_{t=1}^{\lfloor h/2 \rfloor} \left(2h(h-1)^{2t-1}\gamma^{2t} + \sum_{m=1}^{2h-2t} 4h(h-1)^{2t-1}(2h-2)^m (2h)^m \gamma^{2t+2m} \right) \tag{53}$$

$$= - \|\boldsymbol{c}\|_1^2 \|O\|_2^2 \left(\sum_{t=1}^{\lfloor h/2 \rfloor} 2h(h-1)^{2t-1}\gamma^{2t} \right) \cdot \left(1 + 2 \sum_{m=1}^{2h-2t} \left[4h(h-1)\gamma^2 \right]^m \right) \tag{54}$$

$$\geq - \|\boldsymbol{c}\|_1^2 \|O\|_2^2 3h(h-1)\gamma^2 \left(1 + 12h(h-1)\gamma^2 \right) \tag{55}$$

$$\geq - \left(6h^2 - 9h + 3 \right) \gamma^2 \|\boldsymbol{c}\|_1^2 \|O\|_2^2. \tag{56}$$

Here, Eq. (49) is obtained since the summation $\sum_{\boldsymbol{j}>\mathbf{0}_h}^{\mathbf{1}_h}$ contains $\binom{h}{2t}$ terms such that $\|\boldsymbol{j}\|_1 = 2t$, for all $t \in \{1, \cdots, \lfloor \frac{h}{2} \rfloor\}$. Eq. (50) is derived by calculating expectation terms. InEq. (51) is obtained by using

$$\binom{h}{2t} \leq \frac{h(h-1)^{2t-1}}{(2t)!} = \frac{h(h-1)^{2t-1}}{2^t t! (2t-1)!!} \text{ and } \binom{2h-2t}{m} \leq \frac{(2h-2t)^m}{m!}.$$

InEq. (52) is derived by using $t! \geq 2^{t-1}$ and

$$(2t+2k-1)(2t+2m-2k+1) \leq (2t+m)^2 \leq (2h)^2, \forall k \in \{1, \cdots, m-1\}.$$

InEq. (53) is obtained by splitting the summation $\sum_m$ and using $m! \geq 2^{m-1}, \forall m \geq 1$. InEq. (55) is derived by calculating geometric sequences with the condition $\gamma^2 \leq \frac{1}{12h^2}$. InEq. (56) follows from the condition $\gamma^2 \leq \frac{1}{12h^2}$. Thus, we have proved Eq. (24).

$\square$

Lemma B.4. *Let ρ be the density matrix of a quantum state. Let $V_h = W_1 e^{-i\theta G_1} W_2 \cdots W_h e^{-i\theta G_h}$, where $\{G_n\}_{n=1}^h$ is a list of hermitian unitaries and $\{W_n\}_{n=1}^h$ is a list of unitary matrices. Denote by O an arbitrary hermitian quantum observable. Then*

$$\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(\frac{\partial}{\partial\theta} \mathrm{Tr}\left[O V_h \rho {V_h}^\dagger \right] \right)^2 \geq (1-4\gamma^2) \left(\frac{\partial}{\partial\theta} \mathrm{Tr}\left[O V_h \rho {V_h}^\dagger \right] \Big|_{\theta=0} \right)^2$$

$$- 96h^2(h-1)\gamma^2\|O\|_2^2 - 20h^2(h-1)(h-2)\gamma^2\|O\|_2^2, \tag{57}$$

where $\|O\|_2$ denotes the spectral norm of O and the variance $\gamma^2 \leq \frac{1}{16h^3}$.

Proof. For convenience, we follow the notation $O_{\boldsymbol{i}}^{\boldsymbol{j}}$ in Eq. (13). We can obtain the detailed formulation of $\frac{\partial}{\partial\theta}\operatorname{Tr}\left[OV_h\rho V_h{}^\dagger\right]$ by using the $h'=h$ case of Eq. (16),

$$\frac{\partial}{\partial\theta}\operatorname{Tr}\left[OV_h\rho V_h{}^\dagger\right] = \frac{\partial}{\partial\theta}\sum_{\boldsymbol{j}=\boldsymbol{0}_h}^{\boldsymbol{1}_h}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h}(\sin 2\theta)^{\|\boldsymbol{j}\|_1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \tag{58}$$

$$\begin{aligned}
=&2\sum_{\boldsymbol{j}=\boldsymbol{0}_h}^{\boldsymbol{1}_h}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(h+\|\boldsymbol{j}\|_1-\|\boldsymbol{i}\|_1)(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h-1}(\sin 2\theta)^{\|\boldsymbol{j}\|_1+1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\\
&+2\sum_{\boldsymbol{j}=\boldsymbol{0}_h}^{\boldsymbol{1}_h}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\|\boldsymbol{j}\|_1(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h+1}(\sin 2\theta)^{\|\boldsymbol{j}\|_1-1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]
\end{aligned} \tag{59}$$

$$\begin{aligned}
=&2\sum_{\boldsymbol{j}=\boldsymbol{0}_h}^{\boldsymbol{1}_h}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(h+\|\boldsymbol{j}\|_1-\|\boldsymbol{i}\|_1)(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h-1}(\sin 2\theta)^{\|\boldsymbol{j}\|_1+1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\\
&+2\sum_{\|\boldsymbol{j}\|_1=1}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(\cos 2\theta)^{\|\boldsymbol{i}\|_1-h}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\\
&+2\sum_{\|\boldsymbol{j}\|_1\geq 2}^{\boldsymbol{1}_h}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\|\boldsymbol{j}\|_1(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h+1}(\sin 2\theta)^{\|\boldsymbol{j}\|_1-1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right].
\end{aligned} \tag{60}$$

Here, Eq. (58) follows from Eq. (16). Eq. (59) is derived by calculating the gradient of sine and cosine terms. By discarding the square of the sum of the first and the third term in Eq. (60), we obtain

$$\begin{aligned}
\left(\frac{\partial}{\partial\theta}\operatorname{Tr}\left[OV_h\rho V_h{}^\dagger\right]\right)^2 \geq& 4\left(\sum_{\|\boldsymbol{j}\|_1=1}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(\cos 2\theta)^{\|\boldsymbol{i}\|_1-h}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right)^2\\
&+8\left(\sum_{\|\boldsymbol{j}\|_1\geq 2}^{\boldsymbol{1}_h}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\|\boldsymbol{j}\|_1(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h+1}(\sin 2\theta)^{\|\boldsymbol{j}\|_1-1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right)\\
&\cdot\left(\sum_{\|\boldsymbol{j}'\|_1=1}\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-h}\operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{j}'}\rho\right]\right)\\
&+8\left(\sum_{\boldsymbol{j}=\boldsymbol{0}_h}^{\boldsymbol{1}_h}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(h+\|\boldsymbol{j}\|_1-\|\boldsymbol{i}\|_1)(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h-1}(\sin 2\theta)^{\|\boldsymbol{j}\|_1+1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right)\\
&\cdot\left(\sum_{\|\boldsymbol{j}'\|_1=1}\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-h}\operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{j}'}\rho\right]\right).
\end{aligned} \tag{61}$$

Let $\theta=0$ in Eq. (60), we obtain

$$\frac{\partial}{\partial\theta}\operatorname{Tr}\left[OV_h\rho V_h{}^\dagger\right]\Big|_{\theta=0} = 2\sum_{\|\boldsymbol{j}\|_1=1}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]. \tag{62}$$

Thus, we could obtain Eq. (57) if Eqs. (63-65) hold.

$$\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}\left(\sum_{\|\boldsymbol{j}\|_1=1}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(\cos 2\theta)^{\|\boldsymbol{i}\|_1-h}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right)^2 - (1-4\gamma^2)\left(\sum_{\|\boldsymbol{j}\|_1=1}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right)^2$$

$$\geq -\frac{13}{3}h^2(h-1)\gamma^2\|O\|_2^2, \tag{63}$$

$$\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}\left(\sum_{\boldsymbol{j}=\boldsymbol{0}_h}^{\boldsymbol{1}_h}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\left(h+\|\boldsymbol{j}\|_1-\|\boldsymbol{i}\|_1\right)(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h-1}(\sin 2\theta)^{\|\boldsymbol{j}\|_1+1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right)$$
$$\cdot\left(\sum_{\|\boldsymbol{j}'\|_1=1}\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-h}\operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{j}'}\rho\right]\right)\geq -\frac{59}{6}h^2(h-1)\gamma^2\|O\|_2^2, \tag{64}$$

$$\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}\left(\sum_{\|\boldsymbol{j}\|_1\geq 2}^{\boldsymbol{1}_h}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\|\boldsymbol{j}\|_1(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h+1}(\sin 2\theta)^{\|\boldsymbol{j}\|_1-1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right)$$
$$\cdot\left(\sum_{\|\boldsymbol{j}'\|_1=1}\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-h}\operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{j}'}\rho\right]\right)\geq -\frac{5}{2}h^2(h-1)(h-2)\gamma^2\|O\|_2^2. \tag{65}$$

We begin by proving Eq. (63). The left side of Eq. (63) can be lower bounded as

$$\begin{aligned}
\geq\ &\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}\left(\sum_{\|\boldsymbol{j}\|_1=1}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\left[\cos 2\theta-(\cos 2\theta)^{\|\boldsymbol{i}\|_1-h}-\cos 2\theta\right]\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right)^2\\
&-\left(\cos 2\theta\sum_{\|\boldsymbol{j}\|_1=1}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right)^2 &(66)\\
\geq\ &-2\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}(\cos 2\theta)^2\sum_{\|\boldsymbol{j}\|_1=1}\left|\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\left[1-(\cos 2\theta)^{\|\boldsymbol{i}\|_1-1-h}\right]\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right|\\
&\cdot\sum_{\|\boldsymbol{j}'\|_1=1}\left|\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{j}'}\rho\right]\right| &(67)\\
\geq\ &-2\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}(\cos 2\theta)^2\sum_{\|\boldsymbol{j}\|_1=1}\left[(2-\cos 2\theta)^{h-1}-1\right]\|O\|_2\cdot\sum_{\|\boldsymbol{j}\|_1=1}\left|\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right| &(68)\\
\geq\ &-2\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}(\cos 2\theta)^2 h\left[(2-\cos 2\theta)^{h-1}-1\right]\|O\|_2\cdot h\|O\|_2 &(69)\\
\geq\ &-2h^2\|O\|_2^2\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}\left[(1+2\theta^2)^{h-1}-1\right] &(70)\\
\geq\ &-\frac{13}{3}h^2(h-1)\gamma^2\|O\|_2^2. &(71)
\end{aligned}$$

Here, InEq. (66) follows from

$$1-4\gamma^2=\mathbb{E}_\theta[1-4\theta^2]\leq\mathbb{E}_\theta(1-2\theta^2)^2\leq\mathbb{E}_\theta(\cos 2\theta)^2.$$

InEq. (67) is obtained by using $(a-b)^2-b^2\geq -2|a|\cdot|b|$. InEq. (68) follows from the $h'=h$ and $\|\boldsymbol{j}\|=1$ case of Eq. (32). InEq. (69) is derived by using Eq. (15). InEq. (70) is obtained by using $\cos 2\theta\geq 1-2\theta^2$. InEq. (71) follows from the derivation below.

$$\begin{aligned}
\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}(1+2\theta^2)^{h-1}-1 &= \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}\sum_{t=1}^{h-1}\binom{h-1}{t}(2\theta^2)^t\\
&=\sum_{t=1}^{h-1}\binom{h-1}{t}(2t-1)!!(2\gamma^2)^t &(72)\\
&\leq\sum_{t=1}^{h-1}(h-1)(h-2)^{t-1}2^{t-1}(2\gamma^2)^t
\end{aligned}$$

$$\leq 2(h-1)\gamma^2 \sum_{t=1}^{h-1} \left[h^3\gamma^2\right]^{t-1} \tag{73}$$

$$\leq \frac{13}{6}(h-1)\gamma^2, \tag{74}$$

where Eq. (72) is obtained by calculating expectation terms. InEq (73) follows from $h^3 \geq 4(h-2)$ for integer h. InEq. (74) is derived by calculating geometric sequences with the condition $\gamma^2 \leq \frac{1}{16h^3}$.

Next, we prove Eq. (64). The left side of Eq. (64) could be lower bounded by

$$\begin{aligned}
= & \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \left(\sum_{\|\boldsymbol{j}'\|_1=1} \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\cos 2\theta)^{\|\boldsymbol{i}'\|_1-1-h} \operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{j}'}\rho\right] \right) \\
& \cdot \left(\sum_{\substack{\boldsymbol{j}=\boldsymbol{0}_h \\ 2|(\|\boldsymbol{j}\|_1+1)}}^{\boldsymbol{1}_h} \sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h} (h+\|\boldsymbol{j}\|_1-\|\boldsymbol{i}\|_1)(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h}(\sin 2\theta)^{\|\boldsymbol{j}\|_1+1} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \right)
\end{aligned} \tag{75}$$

$$\begin{aligned}
\geq & - \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\|\boldsymbol{j}'\|_1=1} \left(\left| \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_h}^{\boldsymbol{2}_h} \left[1-(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-1-h}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{j}'}\rho\right] \right| + \left| \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_h}^{\boldsymbol{2}_h} \operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{j}'}\rho\right] \right| \right) \\
& \cdot \sum_{\substack{\boldsymbol{j}=\boldsymbol{0}_h \\ 2|(\|\boldsymbol{j}\|_1+1)}}^{\boldsymbol{1}_h} (\sin 2\theta)^{\|\boldsymbol{j}\|_1+1} \left| \sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h} (\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h)(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h} \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \right|
\end{aligned} \tag{76}$$

$$\begin{aligned}
\geq & - \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\|\boldsymbol{j}'\|_1=1} (2-\cos 2\theta)^{h-1}\|O\|_2 \sum_{\substack{\boldsymbol{j}=\boldsymbol{0}_h \\ 2|(\|\boldsymbol{j}\|_1+1)}}^{\boldsymbol{1}_h} (\sin 2\theta)^{\|\boldsymbol{j}\|_1+1} \\
& \left| \sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h} \left[(h-\|\boldsymbol{j}\|_1)-(\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h)(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h}-(h-\|\boldsymbol{j}\|_1)\right] \operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right] \right|
\end{aligned} \tag{77}$$

$$\begin{aligned}
\geq & - h\|O\|_2^2 \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} (2-\cos 2\theta)^{h-1} \\
& \cdot \sum_{\substack{\boldsymbol{j}=\boldsymbol{0}_h \\ 2|(\|\boldsymbol{j}\|_1+1)}}^{\boldsymbol{1}_h} (\sin 2\theta)^{\|\boldsymbol{j}\|_1+1}(h-\|\boldsymbol{j}\|_1)(3-\cos 2\theta)(2-\cos 2\theta)^{h-\|\boldsymbol{j}\|_1-1}.
\end{aligned} \tag{78}$$

Here, Eq. (75) is obtained by noticing that the expectation of $\sin^a 2\theta \cos^b 2\theta$ equals to zero, if a is odd. InEq. (76) is derived by using $|\sum_k a_k| \leq \sum_k |a_k|$. InEq. (77) is obtained by using the $h'=h$, $\|\boldsymbol{j}\|_1 = 1$ case of Eq. (32) and Eq. (15). InEq. (78) follows from Eq. (15) and the $h'=h$ case of Eq. (79), i.e.

$$\begin{aligned}
& \left| \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_{h'}}^{\boldsymbol{2}_{h'}} \left[(h'-\|\boldsymbol{j}'\|_1)-(\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-h')(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-\|\boldsymbol{j}'\|_1-h'}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}',\boldsymbol{i}}^{\boldsymbol{j}',\boldsymbol{j}}\rho\right] \right| \\
& \leq g(h'-\|\boldsymbol{j}'\|_1)\|O\|_2
\end{aligned} \tag{79}$$

for all $h' \in \{0,1,\cdots,h\}$, $\boldsymbol{i} \in \{0,1,2\}^{h-h'}$, and $\boldsymbol{j} \in \{0,1\}^{h-h'}$, where

$$g(x) = x\left[(3-\cos 2\theta)(2-\cos 2\theta)^{x-1}-1\right]. \tag{80}$$

InEq. (79) can be proved inductively. First, InEq. (79) holds trivially when $h'=0$. Next, we assume that InEq. (79) holds for the $h'=k$ case. Then, for all $\boldsymbol{i} \in \{0,1,2\}^{h-k-1}$ and $\boldsymbol{j} \in \{0,1\}^{h-k-1}$, we have

$$
\left| \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_{k+1}}^{\mathbf{2}_{k+1}} \left[(k+1-\|\boldsymbol{j}'\|_1) - (\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k - 1)(\cos 2\theta)^{\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k - 1}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}',\boldsymbol{i}}^{\boldsymbol{j}',\boldsymbol{j}}\rho\right] \right|
$$

$$
\begin{aligned}
= \Bigg| \sum_{i'_{k+1}=j'_{k+1}+1}^{2} \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} &\Big[(k+1-\|\boldsymbol{j}'\|_1 - j'_{k+1}) \\
&- (\|\boldsymbol{i}'\|_1 + i'_{k+1} - \|\boldsymbol{j}'\|_1 - j'_{k+1} - k - 1)(\cos 2\theta)^{\|\boldsymbol{i}'\|_1 + i'_{k+1} - \|\boldsymbol{j}'\|_1 - j'_{k+1} - k - 1}\Big] \operatorname{Tr}\left[O_{\boldsymbol{i}',i'_{k+1},\boldsymbol{i}}^{\boldsymbol{j}',j'_{k+1},\boldsymbol{j}}\rho\right] \Bigg| .
\end{aligned}
\tag{81}
$$

For the case $j'_{k+1} = 1$, we have

$$
\begin{aligned}
\text{Eq. (81)} &= \left| \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \left[(k - \|\boldsymbol{j}'\|_1) - (\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k)(\cos 2\theta)^{\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}',2,\boldsymbol{i}}^{\boldsymbol{j}',1,\boldsymbol{j}}\rho\right] \right| \\
&\leq g(k - \|\boldsymbol{j}'\|_1) \, \|O\|_2 \qquad (82) \\
&= g(k + 1 - \|\boldsymbol{j}'\|_1 - j'_{k+1}) \, \|O\|_2 \, . \qquad (83)
\end{aligned}
$$

Here, Eq. (82) follows from the $h' = k$ case of InEq. (79). Eq. (83) is obtained by using $j'_{k+1} = 1$. We remark that InEq. (83) matches the $h' = k + 1$ case of InEq. (79).

For the case $j'_{k+1} = 0$, the situation is more complicated. We have

$$
\begin{aligned}
&\text{Eq. (81)} \\
&= \Bigg| \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \left[(k+1-\|\boldsymbol{j}'\|_1) - (\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k)(\cos 2\theta)^{\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}',1,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right] \\
&\quad + \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \left[(k+1-\|\boldsymbol{j}'\|_1) - (\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k + 1)(\cos 2\theta)^{\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k + 1}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}',2,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right] \Bigg| \\
&\leq \Bigg| \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \operatorname{Tr}\left[O_{\boldsymbol{i}',1,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right] \\
&\quad + \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \left[(k-\|\boldsymbol{j}'\|_1) - (\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k)(\cos 2\theta)^{\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}',0,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right] \\
&\quad + (1 - \cos 2\theta)(k + 1 - \|\boldsymbol{j}'\|_1) \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \operatorname{Tr}\left[O_{\boldsymbol{i}',2,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right] \\
&\quad + \cos 2\theta \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \left[1 - (\cos 2\theta)^{\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}',2,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right] \\
&\quad - (1 - \cos 2\theta) \sum_{\boldsymbol{i}'=\boldsymbol{j}'+\mathbf{1}_k}^{\mathbf{2}_k} \left[(k - \|\boldsymbol{j}'\|) - (\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k)(\cos 2\theta)^{\|\boldsymbol{i}'\|_1 - \|\boldsymbol{j}'\|_1 - k}\right] \operatorname{Tr}\left[O_{\boldsymbol{i}',2,\boldsymbol{i}}^{\boldsymbol{j}',0,\boldsymbol{j}}\rho\right] \Bigg| \\
&\leq \left\| O_{\mathbf{0}_k,1,\boldsymbol{i}}^{\mathbf{0}_k,0,\boldsymbol{j}} \right\|_2 + g(k - \|\boldsymbol{j}'\|_1) \, \|O\|_2 + (1 - \cos 2\theta)(k + 1 - \|\boldsymbol{j}'\|_1) \left\| O_{\mathbf{0}_k,2,\boldsymbol{i}}^{\mathbf{0}_k,0,\boldsymbol{j}} \right\|_2 \\
&\quad + |\cos 2\theta| \left[(2 - \cos 2\theta)^{k - \|\boldsymbol{j}'\|_1} - 1\right] \left\| O_{\mathbf{0}_k,2,\boldsymbol{i}}^{\mathbf{0}_k,0,\boldsymbol{j}} \right\|_2 + (1 - \cos 2\theta) g(k - \|\boldsymbol{j}'\|_1) \, \|O\|_2 \qquad (84) \\
&\leq \Bigg[(2 - \cos 2\theta)(k - \|\boldsymbol{j}'\|_1) \left[(3 - \cos 2\theta)(2 - \cos 2\theta)^{k - \|\boldsymbol{j}'\|_1 - 1} - 1\right] \\
&\quad + (2 - \cos 2\theta)^{k - \|\boldsymbol{j}'\|_1} + (1 - \cos 2\theta)(k + 1 - \|\boldsymbol{j}'\|_1)\Bigg] \|O\|_2 \qquad (85)
\end{aligned}
$$

$$\leq g(k+1-\|\boldsymbol{j}'\|-j'_{k+1})\,\|O\|_2\,. \tag{86}$$

Here, InEq. (84) is obtained by using the $h'=k$ case of Eq. (79). InEq. (85) is obtained by using Eqs. (14) and (80). InEq. (86) follows from Eq. (80) and the condition $j'_{k+1}=0$. Since Eqs. (83) and (86) match the formulation of the $h'=k+1$ case of Eq. (79), we have proved Eq. (79) for general cases.

We proceed from Eq. (78), which can be lower bounded by

$$\geq\ -h(h-1)\|O\|_2^2 \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\substack{\boldsymbol{j}=\boldsymbol{0}_h \\ 2|(\|\boldsymbol{j}\|_1+1)}}^{\boldsymbol{1}_h} (2\theta)^{\|\boldsymbol{j}\|_1+1}\, 2(1+2\theta^2)^{2h-1-\|\boldsymbol{j}\|_1} \tag{87}$$

$$\geq\ -2h(h-1)\|O\|_2^2 \mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{t=1}^{\lfloor\frac{h+1}{2}\rfloor} \binom{h}{2t-1}(2\theta)^{2t}\sum_{m=0}^{2h-2t}\binom{2h-2t}{m}(2\theta^2)^m \tag{88}$$

$$=\ -2h(h-1)\|O\|_2^2 \sum_{t=1}^{\lfloor\frac{h+1}{2}\rfloor} \binom{h}{2t-1}\sum_{m=0}^{2h-2t}\binom{2h-2t}{m}2^{2t+m}(2t+2m-1)!!\gamma^{2t+2m} \tag{89}$$

$$\geq\ -\frac{59}{6}h^2(h-1)\gamma^2\|O\|_2^2. \tag{90}$$

Here, InEq. (87) is obtained by using $1\geq\cos 2\theta\geq 1-2\theta^2$. InEq. (88) is obtained since the summation $\sum_{\boldsymbol{j}=\boldsymbol{0}_h}^{\boldsymbol{1}_h}$ contains $\binom{h}{2t-1}$ terms such that $\|\boldsymbol{j}\|_1=2t-1$, for all $t\in\{1,\cdots,\lfloor\frac{h+1}{2}\rfloor\}$. Eq. (89) is derived by calculating expectation terms. InEq. (90) is obtained by bounding the summation terms, i.e.

$$\begin{aligned}
&\sum_{t=1}^{\lfloor\frac{h+1}{2}\rfloor} \binom{h}{2t-1}\sum_{m=0}^{2h-2t}\binom{2h-2t}{m}2^{2t+m}(2t+2m-1)!!\gamma^{2t+2m}\\
\leq\ &\sum_{t=1}^{\lfloor\frac{h+1}{2}\rfloor} \frac{h(h-1)^{2t-2}}{2^{t-1}(t-1)!(2t-1)!!}\sum_{m=0}^{2h-2t}\frac{(2h-2t)^m}{m!}2^{2t+m}\\
&\quad\cdot(2t-1)!!(2t+1)(2t+3)\cdots(2t+2m-1)\gamma^{2t+2m} && (91)\\
\leq\ &\sum_{t=1}^{\lfloor\frac{h+1}{2}\rfloor} h(h-1)^{2t-2}2^{t+1}\gamma^{2t}\sum_{m=0}^{2h-2t}(2h-2)^m2^m(2h)^m\gamma^{2m} && (92)\\
\leq\ &4h\gamma^2\sum_{t=1}^{\lfloor\frac{h+1}{2}\rfloor}\left[h(h-1)^2\gamma^2\right]^{t-1}\sum_{m=0}^{2h-2t}(2h^3\gamma^2)^m && (93)\\
\leq\ &4h\gamma^2\frac{16}{15}\times\frac{8}{7}\leq\frac{59}{12}h\gamma^2. && (94)
\end{aligned}$$

Here, InEq (91) follows from $t\geq 1$ and $(2t-1)!=2^t t!(2t-1)!!$. InEq. (92) is obtained by using $t\geq 1$ and

$$(2t+2k-1)(2t+2m-2k+1)\leq(m+2t)^2\leq(2h)^2,\ \forall k\in\{1,\cdots,m\}.$$

InEq. (93) is derived by using $8h(h-1)\leq 2h^3$ for the integer h. InEq. (94) is obtained by calculating geometric sequences with the condition $\gamma^2\leq\frac{1}{16h^3}$.

Finally, we prove Eq. (65). The left side of Eq. (65) could be lower bounded by

$$\begin{aligned}
=\ &\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)}\left(\sum_{\|\boldsymbol{j}'\|_1=1}\sum_{\boldsymbol{i}'=\boldsymbol{j}'+\boldsymbol{1}_h}^{\boldsymbol{2}_h}(\cos 2\theta)^{\|\boldsymbol{i}'\|_1-h-1}\operatorname{Tr}\left[O_{\boldsymbol{i}'}^{\boldsymbol{j}'}\rho\right]\right)\\
&\cdot\left(\sum_{\substack{\boldsymbol{j}=\boldsymbol{0}_h\\ \|\boldsymbol{j}\|_1\geq 2,2|(\|\boldsymbol{j}\|_1-1)}}^{\boldsymbol{1}_h}\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\|\boldsymbol{j}\|_1(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h+2}(\sin 2\theta)^{\|\boldsymbol{j}\|_1-1}\operatorname{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right)
\end{aligned} \tag{95}$$

$$
\begin{aligned}
\geq\ & -\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} \sum_{\|\boldsymbol{j}'\|_1=1} (2-\cos 2\theta)^{h-1}\|O\|_2 \sum_{\substack{\boldsymbol{j}=\boldsymbol{0}_h\\ \|\boldsymbol{j}\|_1\geq 2,2|(\|\boldsymbol{j}\|_1-1)}}^{\boldsymbol{1}_h} \|\boldsymbol{j}\|_1 \left(\sin 2\theta\right)^{\|\boldsymbol{j}\|_1-1}\left(\cos 2\theta\right)^2 \\
& \left\{\left|\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\left[1-(\cos 2\theta)^{\|\boldsymbol{i}\|_1-\|\boldsymbol{j}\|_1-h}\right]\mathrm{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right| + \left|\sum_{\boldsymbol{i}=\boldsymbol{j}+\boldsymbol{1}_h}^{\boldsymbol{2}_h}\mathrm{Tr}\left[O_{\boldsymbol{i}}^{\boldsymbol{j}}\rho\right]\right|\right\} \qquad (96)\\
\geq\ & -\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} h(2-\cos 2\theta)^{h-1}\|O\|_2 \\
& \cdot \sum_{\substack{\boldsymbol{j}=\boldsymbol{0}_h\\ \|\boldsymbol{j}\|_1\geq 2,2|(\|\boldsymbol{j}\|_1-1)}}^{\boldsymbol{1}_h} \|\boldsymbol{j}\|_1 \left(\sin 2\theta\right)^{\|\boldsymbol{j}\|_1-1}\left(\cos 2\theta\right)^2 \left(2-\cos 2\theta\right)^{h-\|\boldsymbol{j}\|_1}\|O\|_2 \qquad (97)\\
\geq\ & -\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} h\|O\|_2^2 \sum_{\substack{\boldsymbol{j}=\boldsymbol{0}_h\\ \|\boldsymbol{j}\|_1\geq 2,2|(\|\boldsymbol{j}\|_1-1)}}^{\boldsymbol{1}_h} \|\boldsymbol{j}\|_1 \left(2\theta\right)^{\|\boldsymbol{j}\|_1-1}\left(1+2\theta^2\right)^{2h-1-\|\boldsymbol{j}\|_1} \qquad (98)\\
\geq\ & -\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} h\|O\|_2^2 \sum_{t=1}^{\lfloor\frac{h-1}{2}\rfloor}\binom{h}{2t+1}(2t+1)(2\theta)^{2t}\left(1+2\theta^2\right)^{2h-2-2t}. \qquad (99)
\end{aligned}
$$

Eq. (95) is obtained by noticing that the expectation of $\sin^a 2\theta \cos^b 2\theta$ equals to zero, if a is odd. InEq. (96) follows from the derivation (75-77). InEq. (97) is obtained by using the $h'=h$, $\|\boldsymbol{j}\|_1=1$ case of Eq. (32). InEq. (98) follows from $1\geq\cos 2\theta\geq 1-2\theta^2$ and $(\sin 2\theta)^2\leq(2\theta)^2$. InEq. (99) is obtained since the summation $\sum_{\boldsymbol{j}=\boldsymbol{0}_h}^{\boldsymbol{1}_h}$ contains $\binom{h}{2t+1}$ terms such that $\|\boldsymbol{j}\|_1=2t+1$, for all $t\in\{1,\cdots,\lfloor\frac{h-1}{2}\rfloor\}$. We further bound InEq. (99) by

$$
\begin{aligned}
=\ & -\mathop{\mathbb{E}}_{\theta\sim\mathcal{N}(0,\gamma^2)} h\|O\|_2^2 \sum_{t=1}^{\lfloor\frac{h-1}{2}\rfloor}\binom{h}{2t+1}(2t+1)(2\theta)^{2t}\sum_{m=0}^{2h-2-2t}\binom{2h-2-2t}{m}(2\theta^2)^m \\
\geq\ & -h\|O\|_2^2 \sum_{t=1}^{\lfloor\frac{h-1}{2}\rfloor}\binom{h}{2t+1}(2t+1)\sum_{m=0}^{2h-2-2t}\binom{2h-2-2t}{m}2^{2t+m}(2t+2m-1)!!\gamma^{2t+2m} \\
& \qquad (100)\\
\geq\ & -h\|O\|_2^2 \sum_{t=1}^{\lfloor\frac{h-1}{2}\rfloor}\frac{h(h-1)(h-2)(h-3)^{2t-2}}{(2t)!}2^{2t}\gamma^{2t} \\
& \cdot\sum_{m=0}^{2h-2-2t}(2h-2-2t)^m 2^m(2t+2m-1)!!\gamma^{2m} \qquad (101)\\
=\ & -h^2(h-1)(h-2)\|O\|_2^2 \sum_{t=1}^{\lfloor\frac{h-1}{2}\rfloor}\frac{(h-3)^{2t-2}}{2^t t!(2t-1)!!}2^{2t}\gamma^{2t} \\
& \cdot\sum_{m=0}^{2h-2-2t}(2h-2-2t)^m 2^m(2t+2m-1)!!\gamma^{2m} \qquad (102)\\
=\ & -\frac{5}{2}h^2(h-1)(h-2)\gamma^2\|O\|_2^2. \qquad (103)
\end{aligned}
$$

Here, InEq. (100) is obtained by calculating expectation terms. InEq. (101) is derived by using $t\geq 1$. Eq. (102) follows from $(2t)!=2^t t!(2t-1)!!$. Eq. (103) is obtained by bounding the summation terms, i.e.

$$
\sum_{t=1}^{\lfloor\frac{h-1}{2}\rfloor}\frac{(h-3)^{2t-2}}{2^t t!(2t-1)!!}2^{2t}\gamma^{2t}\sum_{m=0}^{2h-2-2t}(2h-2-2t)^m 2^m(2t+2m-1)!!\gamma^{2m}
$$

$$\leq \sum_{t=1}^{\lfloor \frac{h-1}{2} \rfloor} \frac{(h-3)^{2t-2}}{2^{t-1}} 2^t \gamma^{2t} \sum_{m=0}^{2h-2-2t} (2h-4)^m 2^m (2h-2)^m \gamma^{2m} \tag{104}$$

$$\leq 2\gamma^2 \sum_{t=1}^{\lfloor \frac{h-1}{2} \rfloor} \left[(h-3)^2\gamma^2\right]^{t-1} \sum_{m=0}^{2h-2-2t} \left(h^3\gamma^2\right)^m \tag{105}$$

$$\leq 2\gamma^2 \left(\frac{16}{15}\right)^2 \leq \frac{5}{2}\gamma^2. \tag{106}$$

Here, InEq. (104) is obtained by using $t! \geq 2^{t-1}, \forall t \geq 1$ and

$$(2t+2k-1)(2t+2m-2k+1) \leq (m+2t)^2 \leq (2h-2)^2,\ \forall k \in \{1, \cdots, m\}.$$

InEq. (105) follows from $h^3 \geq 8(h-1)(h-2)$ for integer h. InEq. (106) is obtained by calculating geometric sequences with the condition $\gamma^2 \leq \frac{1}{16h^3}$. Thus, we have proved Eq. (65).

□

C Proof of Theorem 4.1

Proof. Denote by $I_S := \{m | i_m \neq 0, m \in [N]\}$ the set of qubits where the observable acts non-trivially. First, we notice that the norm of the whole gradient is lower bounded by that of particle derivatives summed over a part of parameters, i.e.

$$\mathbb{E}_{\boldsymbol{\theta}} \|\nabla_{\boldsymbol{\theta}} f\|^2 \geq \sum_{q=1}^{L} \sum_{n \in I_S} \mathbb{E}_{\boldsymbol{\theta}} \left(\frac{\partial f}{\partial \theta_{q,n}}\right)^2. \tag{107}$$

Thus, we could obtain the formulation in the theorem if

$$\mathbb{E}_{\boldsymbol{\theta}} \left(\frac{\partial f}{\partial \theta_{q,n}}\right)^2 \geq \frac{1}{S^{S+1}(L+2)^{S+1}} \mathrm{Tr}\left[\sigma_{\boldsymbol{j}} \rho_{\text{in}}\right]^2, \tag{108}$$

holds for any $q \in \{1, \cdots, L\}$ and $n \in I_S$.

Now we begin to prove Eq. (108). Our main idea is to integrate the square of the partial derivative of f with respect to $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{L+2})$ by using Lemma B.1 and Lemma B.2.

We introduce several notations for convenience. Denote the variance $\gamma^2 = \frac{1}{4S(L+2)}$. Denote the ℓ-th single qubit rotations and CZ layer as $R_\ell(\boldsymbol{\theta}_\ell)$ and CZ_ℓ, respectively, where

$$R_\ell(\boldsymbol{\theta}_\ell) = e^{-i\theta_{\ell,1}G_{\ell,1}} \otimes e^{-i\theta_{\ell,2}G_{\ell,2}} \otimes \cdots \otimes e^{-i\theta_{\ell,N}G_{\ell,N}}, \tag{109}$$

and $G_{\ell,j}$ is the Hamiltonian corresponding to the parameter $\theta_{\ell,j}$. Denote by ρ_k the state after the k-th layer, $\forall k \in \{0, 1, \cdots, 2L+2\}$,

$$\rho_k := \begin{cases} \left(\prod_{i=\frac{k}{2}}^{1} \mathrm{CZ}_i R_i(\boldsymbol{\theta}_i)\right) \rho_{\text{in}} \left(\prod_{i=1}^{\frac{k}{2}} R_i(\boldsymbol{\theta}_i)^\dagger \mathrm{CZ}_i^\dagger\right) & (k = 2\ell \leq 2L), \\ R_{\frac{k+1}{2}}(\boldsymbol{\theta}_{\frac{k+1}{2}}) \rho_{k-1} R_{\frac{k+1}{2}}(\boldsymbol{\theta}_{\frac{k+1}{2}})^\dagger & (k = 2\ell+1 \leq 2L+1), \\ R_{L+2}(\boldsymbol{\theta}_{L+2}) \rho_{k-1} R_{L+2}(\boldsymbol{\theta}_{L+2})^\dagger & (k = 2L+2). \end{cases} \tag{110}$$

Thus, ρ_k is parameterized by $\{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_p\}$, where $p = \ell$ if $k = 2\ell \leq 2L$, $p = \ell + 1$ if $k = 2\ell + 1 \leq 2L + 1$, and $p = L + 2$ if $k = 2L + 2$.

Next, rewrite the formulation of Eq. (108) in detail:

$$\begin{aligned} & \mathbb{E}_{\boldsymbol{\theta}_1} \cdots \mathbb{E}_{\boldsymbol{\theta}_{L+2}} \left(\frac{\partial}{\partial \theta_{q,n}} \mathrm{Tr}\left[\sigma_{\boldsymbol{i}} V(\boldsymbol{\theta}) \rho_{\text{in}} V(\boldsymbol{\theta})^\dagger\right]\right)^2 \\ = & \mathbb{E}_{\boldsymbol{\theta}_1} \cdots \mathbb{E}_{\boldsymbol{\theta}_{L+2}} \left(\frac{\partial}{\partial \theta_{q,n}} \mathrm{Tr}\left[\sigma_{\boldsymbol{i}} \rho_{2L+2}\right]\right)^2 \end{aligned} \tag{111}$$

$$\geq \left[4\gamma^2(1-4\gamma^2)\right]^{S_1}\left(1-4\gamma^2\right)^{S_3}\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{L+1}}\left(\frac{\partial}{\partial\theta_{q,n}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i};\boldsymbol{1}}\rho_{2L+1}\right]\right)^2 \tag{112}$$

$$\geq \left[4\gamma^2(1-4\gamma^2)\right]^{S_1+S_2}\left(1-4\gamma^2\right)^{S_1+2S_3}\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{L}}\left(\frac{\partial}{\partial\theta_{q,n}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\rho_{2L}\right]\right)^2 \tag{113}$$

$$\geq \left[4\gamma^2(1-4\gamma^2)\right]^{S}\left(1-4\gamma^2\right)^{S}\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{L}}\left(\frac{\partial}{\partial\theta_{q,n}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\rho_{2L}\right]\right)^2, \tag{114}$$

where $\boldsymbol{3}|\boldsymbol{i}$ denotes the index by replacing non-zero elements of $\boldsymbol{i}=(i_1,\cdots,i_N)$ with 3 and $\boldsymbol{3}|\boldsymbol{i};\boldsymbol{1}$ denotes the index by replacing non-zero elements of $\boldsymbol{i}=(i_1,\cdots,i_N)$ with 3 if the original value is 1. We refer to S_1, S_2, and S_3 as the number of 1, 2, and 3 in the index $\boldsymbol{i}$, respectively. Eq. (111) is obtained by using the notation ρ_{2L+2} defined in (110). We obtain Eqs. (112) and (113) by using Lemma B.1 for the R_Y and R_X gate case, respectively. InEq. (114) follows from $S=S_1+S_2+S_3$.

Then, we proceed from Eq. (114) and take the expectation for parameters in $(\boldsymbol{\theta}_L,\cdots,\boldsymbol{\theta}_{q+1})$.

$$\text{Eq. (114)} = \left[2\gamma(1-4\gamma^2)\right]^{2S}\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{L}}\left(\frac{\partial}{\partial\theta_{q,n}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\mathrm{CZ}_L R_L(\boldsymbol{\theta}_L)\rho_{2L-2}R_L(\boldsymbol{\theta}_L)^\dagger\mathrm{CZ}_L^\dagger\right]\right)^2 \tag{115}$$

$$= \left[2\gamma(1-4\gamma^2)\right]^{2S}\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{L}}\left(\frac{\partial}{\partial\theta_{q,n}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}R_L(\boldsymbol{\theta}_L)\rho_{2L-2}R_L(\boldsymbol{\theta}_L)^\dagger\right]\right)^2 \tag{116}$$

$$\geq \left[2\gamma(1-4\gamma^2)\right]^{2S}\left(1-4\gamma^2\right)^{S}\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{L-1}}\left(\frac{\partial}{\partial\theta_{q,n}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\rho_{2L-2}\right]\right)^2 \tag{117}$$

$$\geq \left[2\gamma(1-4\gamma^2)\right]^{2S}\left(1-4\gamma^2\right)^{(L-q)S}\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{q}}\left(\frac{\partial}{\partial\theta_{q,n}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\rho_{2q}\right]\right)^2. \tag{118}$$

Eq. (115) follows from the definition of ρ_{2L} (110). Eq. (116) is obtained since

$$\mathrm{CZ}(\sigma_j\otimes\sigma_k)\mathrm{CZ}^\dagger = \sigma_j\otimes\sigma_k, \forall j,k\in\{0,3\}.$$

InEq. (117) is derived by using the Lemma B.1. We repeat the derivation in Eqs. (115-117) inductively for parameters $(\boldsymbol{\theta}_L,\cdots,\boldsymbol{\theta}_{q+1})$, which yields InEq. (118).

Next, we consider the expectation with respect to $\boldsymbol{\theta}_q$. We have

$$\begin{aligned}\text{Eq. (118)} &= \left[2\gamma(1-4\gamma^2)\right]^{2S}\left(1-4\gamma^2\right)^{(L-q)S}\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{q}}\left(\frac{\partial}{\partial\theta_{q,n}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\rho_{2q-1}\right]\right)^2\\ &\geq \left[2\gamma(1-4\gamma^2)\right]^{2S}\left(1-4\gamma^2\right)^{(L-q)S}\left(1-4\gamma^2\right)^{S-1}\left[4\gamma^2(1-4\gamma^2)\right]4\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{q-1}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\rho_{2q-2}\right]^2,\end{aligned} \tag{119}$$

where expectations with respect to parameters $\{\theta_{q,j}\}_{j\in I_S, j\neq n}$ are calculated via Lemma B.1 and the expectation with respect to $\theta_{q,n}$ is calculated via Lemma B.2.

Finally we proceed from Eq. (119) and take the expectation for parameters in $(\boldsymbol{\theta}_{q-1},\cdots,\boldsymbol{\theta}_1)$. We have

$$\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{q-1}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\rho_{2q-2}\right]^2 = \mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{q-1}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\mathrm{CZ}_{q-1}R_{q-1}(\boldsymbol{\theta}_{q-1})\rho_{2q-4}R_{q-1}(\boldsymbol{\theta}_{q-1})^\dagger\mathrm{CZ}_{q-1}^\dagger\right]^2 \tag{120}$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{q-1}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}R_{q-1}(\boldsymbol{\theta}_{q-1})\rho_{2q-4}R_{q-1}(\boldsymbol{\theta}_{q-1})^\dagger\right]^2 \tag{121}$$

$$\geq \left(1-4\gamma^2\right)^{S}\mathop{\mathbb{E}}_{\boldsymbol{\theta}_1}\cdots\mathop{\mathbb{E}}_{\boldsymbol{\theta}_{q-2}}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\rho_{2q-4}\right]^2 \tag{122}$$

$$\geq \left(1-4\gamma^2\right)^{(q-1)S}\mathrm{Tr}\left[\sigma_{\boldsymbol{3}|\boldsymbol{i}}\rho_0\right]^2. \tag{123}$$

Eq. (120) is derived by using the definition of ρ_{2q-2}. Eq. (121) is obtained since

$$\mathrm{CZ}(\sigma_j\otimes\sigma_k)\mathrm{CZ}^\dagger = \sigma_j\otimes\sigma_k, \forall j,k\in\{0,3\}.$$

InEq. (122) is derived by using Lemma B.1. We repeat the derivation in Eqs. (120-122) inductively for parameters $(\boldsymbol{\theta}_{q-1}, \cdots, \boldsymbol{\theta}_1)$, which yields InEq. (123). Employing Eq. (123) to Eq. (119) yields

$$\begin{aligned}
\text{Eq. (118)} &\geq 4\left(4\gamma^2\right)^{S+1}\left(1-4\gamma^2\right)^{S(L+2)}\operatorname{Tr}\left[\sigma_{\mathbf{3}|\boldsymbol{i}}\rho_0\right]^2 && (124)\\
&= 4\left(\frac{1}{S(L+2)}\right)^{S+1}\left(1-\frac{1}{S(L+2)}\right)^{S(L+2)}\operatorname{Tr}\left[\sigma_{\mathbf{3}|\boldsymbol{i}}\rho_0\right]^2 && (125)\\
&\geq 4\left(\frac{1}{S(L+2)}\right)^{S+1}\left(1-\frac{1}{2}\right)^{2}\operatorname{Tr}\left[\sigma_{\mathbf{3}|\boldsymbol{i}}\rho_0\right]^2 && (126)\\
&= \frac{1}{S^{S+1}(L+2)^{S+1}}\operatorname{Tr}\left[\sigma_{\mathbf{3}|\boldsymbol{i}}\rho_0\right]^2. && (127)
\end{aligned}$$

Eq. (125) is derived by using the condition $\gamma^2 = \frac{1}{4S(L+2)}$. Eq. (126) is obtained by noticing that function $g(x) = (1-\frac{1}{x})^x$ is monotonically increasing when $x \geq 2$. Thus, we have proved Eq. (108).

$\square$

D Proof of Theorem 4.2

Proof. To begin with, we define several notations for convenience. Denote by ρ_j the state after the j-th parameterized operator, i.e.

$$\rho_j(\theta_1, \cdots, \theta_j) = \left(\prod_{i=j}^{1} V_i(\theta_i)\right)\rho_{\text{in}}\left(\prod_{i=1}^{j} V_i(\theta_i)^\dagger\right). \tag{128}$$

Denote by O_j the observable, i.e.

$$O_j = V_j(0)^\dagger \cdots V_L(0)^\dagger O V_L(0) \cdots V_j(0),\ \forall j \in \{1, \cdots, L\}. \tag{129}$$

Now we begin to prove the Theorem. First, we remark that $\forall j \in [L]$, the $a_j \neq 1$ case can be converted to the $a_j = 1$ case by using the transformation

$$\theta_j' = \frac{\theta_j}{a_j},$$

where the variance of the new and the old parameter satisfies

$$\operatorname{Var}[\theta_j'] = \frac{1}{a_j^2}\operatorname{Var}[\theta_j].$$

In the following proof, we assume that $a_j = 1, \forall j \in [L]$. By using the parameter-shift rule, $\frac{\partial f}{\partial \theta_\ell}$ could be written as the linear sum of $2h$ expectations on the observable O with coefficients ± 1. Then for the case $\ell \leq L-1$, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{\theta}}\left(\frac{\partial f}{\partial \theta_\ell}\right)^2 &= \mathop{\mathbb{E}}_{\theta_1}\cdots\mathop{\mathbb{E}}_{\theta_L}\left(\frac{\partial}{\partial \theta_\ell}\operatorname{Tr}\left[OV_L(\theta_L)\rho_{L-1}V_L(\theta_L)^\dagger\right]\right)^2 && (130)\\
&\geq \mathop{\mathbb{E}}_{\theta_1}\cdots\mathop{\mathbb{E}}_{\theta_{L-1}}\left(\frac{\partial}{\partial \theta_\ell}\operatorname{Tr}\left[OV_L(0)\rho_{L-1}V_L(0)^\dagger\right]\right)^2 - \left[12h_L(h_L-1)+4\right]4h_L^2\gamma_L^2\|O\|_2^2 && (131)\\
&= \mathop{\mathbb{E}}_{\theta_1}\cdots\mathop{\mathbb{E}}_{\theta_{L-1}}\left(\frac{\partial}{\partial \theta_\ell}\operatorname{Tr}\left[O_L\rho_{L-1}\right]\right)^2 - \left[12h_L(h_L-1)+4\right]4h_L^2\gamma_L^2\|O\|_2^2 && (132)\\
&= \mathop{\mathbb{E}}_{\theta_1}\cdots\mathop{\mathbb{E}}_{\theta_{L-1}}\left(\frac{\partial f}{\partial \theta_\ell}(\theta_1, \cdots, \theta_{L-1}, 0)\right)^2 - \left[12h_L(h_L-1)+4\right]4h_L^2\gamma_L^2\|O\|_2^2, && (133)
\end{aligned}$$

where Eq. (130) follows from the definition of ρ_j in Eq. (128). InEq. (131) is obtained by using Lemma B.3, where $\|\boldsymbol{c}\|_1 = 2h$. Eq. (132) follows from the definition of O_j in Eq. (129). Eq. (133)

follows from the formulation $f(\boldsymbol{\theta}) = \mathrm{Tr}[O\rho(\boldsymbol{\theta})]$. By proceeding the derivation (130-133) for $L-\ell$ times, we have

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\boldsymbol{\theta}}\left(\frac{\partial f}{\partial \theta_\ell}\right)^2 \geq \mathop{\mathbb{E}}_{\theta_1}\cdots\mathop{\mathbb{E}}_{\theta_\ell}\left(\frac{\partial f}{\partial \theta_\ell}(\theta_1,\cdots,\theta_\ell,0,\cdots,0)\right)^2 - \sum_{j=\ell+1}^{L}\left[12h_j(h_j-1)+4\right]4h_j^2\gamma_j^2\|O\|_2^2 \\
&= \mathop{\mathbb{E}}_{\theta_1}\cdots\mathop{\mathbb{E}}_{\theta_\ell}\left(\frac{\partial}{\partial \theta_\ell}\mathrm{Tr}\left[O_{\ell+1}V_\ell(\theta_\ell)\rho_{\ell-1}V_\ell(\theta_\ell)^\dagger\right]\right)^2 - \sum_{j=\ell+1}^{L}\left[12h_j(h_j-1)+4\right]4h_j^2\gamma_j^2\|O\|_2^2 \quad (134)\\
&\geq \mathop{\mathbb{E}}_{\theta_1}\cdots\mathop{\mathbb{E}}_{\theta_{\ell-1}}(1-4\gamma_\ell^2)\left(\frac{\partial}{\partial \theta_\ell}\mathrm{Tr}\left[O_{\ell+1}V_\ell(\theta_\ell)\rho_{\ell-1}V_\ell(\theta_\ell)^\dagger\right]\right)^2\Bigg|_{\theta_\ell=0} - 96h_\ell^2(h_\ell-1)\gamma_\ell^2\|O\|_2^2 \\
&\quad - 20h_\ell^2(h_\ell-1)(h_\ell-2)\gamma_\ell^2\|O\|_2^2 - \sum_{j=\ell+1}^{L}\left[12h_j(h_j-1)+4\right]4h_j^2\gamma_j^2\|O\|_2^2 \quad (135)\\
&\geq \mathop{\mathbb{E}}_{\theta_1}\cdots\mathop{\mathbb{E}}_{\theta_{\ell-1}}\left(\frac{\partial f}{\partial \theta_\ell}(\theta_1,\cdots,\theta_{\ell-1},0,0,\cdots,0)\right)^2 - 4\gamma_\ell^2(2h_\ell)^2\|O\|_2^2 - 96h_\ell^2(h_\ell-1)\gamma_\ell^2\|O\|_2^2 \\
&\quad - 20h_\ell^2(h_\ell-1)(h_\ell-2)\gamma_\ell^2\|O\|_2^2 - \sum_{j=\ell+1}^{L}\left[12h_j(h_j-1)+4\right]4h_j^2\gamma_j^2\|O\|_2^2, \quad (136)
\end{aligned}
$$

where Eq. (134) follows from definitions ρ_j (128) and O_j (129). InEq. (135) is derived by using Lemma B.4. InEq. (136) follows from the parameter-shift rule. We proceed from InEq. (136) by employing the derivation (130-133) for parameters $(\theta_{\ell-1},\cdots,\theta_1)$, which yields

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\boldsymbol{\theta}}\left(\frac{\partial f}{\partial \theta_\ell}\right)^2 \\
&\geq \left(\frac{\partial f}{\partial \theta_\ell}\right)^2\Bigg|_{\mathbf{0}} - \sum_{j=1}^{\ell-1}16h_j^2\left[3h_j(h_j-1)+1\right]\gamma_j^2\|O\|_2^2 - \sum_{j=\ell+1}^{L}16h_j^2\left[3h_j(h_j-1)+1\right]\gamma_j^2\|O\|_2^2 \\
&\quad - 16h_\ell^2\gamma_\ell^2\|O\|_2^2 - 96h_\ell^2(h_\ell-1)\gamma_\ell^2\|O\|_2^2 - 20h_\ell^2(h_\ell-1)(h_\ell-2)\gamma_\ell^2\|O\|_2^2 \quad (137)\\
&\geq \left(\frac{\partial f}{\partial \theta_\ell}\right)^2\Bigg|_{\mathbf{0}} - \sum_{j=1}^{L}16h_j^2\left[3h_j(h_j-1)+1\right]\gamma_j^2\|O\|_2^2 \\
&\geq (1-\epsilon)\left(\frac{\partial f}{\partial \theta_\ell}\right)^2\Bigg|_{\mathbf{0}}. \quad (138)
\end{aligned}
$$

InEq. (137) is obtained by using Lemma B.3, where $\|\boldsymbol{c}\|_1 = 2h$. InEq. (138) follows from the condition $\gamma_j^2 \leq \frac{a_j^2\epsilon}{16h_j^2(3h_j(h_j-1)+1)L\|O\|_2^2}\left(\frac{\partial f}{\partial \theta_\ell}\right)^2\Big|_{\boldsymbol{\theta}=\mathbf{0}}$ and $a_j = 1, \forall j \in [L]$. Thus, we have proved the theorem. □

References

[1] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1139–1147, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

[2] Y Nesterov. A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. In *Sov. Math. Dokl*, volume 27.

[3] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(61):2121–2159, 2011.