Optimizing Parameterized Quantum Circuits with Free-Axis Selection

Hiroshi C. Watanabe*,[1,2] Rudy Raymond*,[3,1] Yu-ya Ohnishi,[4,1] Eriko Kaminishi,[1,2] and Michihiko Sugawara[1]

[1] *Quantum Computing Center, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, 223-8522, Japan*
[2] *JST PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan*
[3] *IBM Quantum, IBM Japan 19-21 Nihonbashi Hakozaki-cho, Chuo-ku, Tokyo, 103-8510, Japan*
[4] *Materials Informatics Initiative, RD Technology & Digital Transformation Center, JSR Corporation, 100 Kawajiri-cho, Yokkaichi, Mie, 510-8552, Japan*

We propose a new method to optimize a PQC by continuous parameterization of both the angles and the axes of its single-qubit rotations. The method is based on the observation that when rotational angles are fixed, optimal axes of rotations can be computed by solving a system of linear equations whose coefficients can be determined from the PQC with small computational overhead. The method can be further simplified to select axes freely from continuous parameters with rotational angles fixed to π. We show the simplified free-axis selection method has better expressibility against other structural optimization methods when measured with Kullback-Leibler (KL) divergence. We also demonstrate PQCs with free-axis selection are more effective to search the ground states of Heisenberg models and molecular Hamiltonians. Because free-axis selection allows designing PQCs without specifying their single-qubit rotational axes, it may significantly improve the handiness of PQCs.

I. INTRODUCTION

Parameterized quantum circuit (PQC) is one of the most essential components of hybrid quantum-classical algorithms on near-term quantum devices [1, 2]. With PQC, a quantum state is expressed by a sequence of one- and two-qubit gates, in which the rotation angles and axes are classically controllable. The design of PQC is critical in variational quantum algorithms. Oversimplified PQC cannot express the optimal quantum state even if it could be implemented on noisy quantum devices. On the other hand, a PQC designed with a deep circuit for high expressibility cannot be implemented on currently-available noisy quantum devices. A physics-based ansatz such as unitary coupled-cluster method in VQE is one of the examples of a deep circuit, in which the required number of two-qubit gates is too large to execute on near-term quantum devices [3]. Highly expressible PQC can also be created in hardware efficient (or heuristic) ansatz[4–9] by applying multiple layers. However, it often results in either too complicated landscape of cost function to find the global minimum, or barren plateau in which the gradient vanishes[1, 10–15]. Hardware-efficient ansatz also suffers from initial guess of parameters. Preparing a suitable set of initial parameters for typical sets of ansatz for various problem is challenging and has received a lot of attention.

Nevertheless, while the design of PQCs has been optimized by using rotations of qubit states around predetermined axes, the choice of such axes are arbitrary and the optimal ones in various problem settings, which are obviously influenced by the properties of the Hamiltonian, can be non trivial. For example, the so-called R_y (rotations around the y-axis) gates are popular for molecular Hamiltonians whose ground states are of real-valued probability amplitudes, but other rotational gates may be necessary for different types of Hamiltonians. Moreover, such additional rotations may facilitate shortcuts to quickly find the ground states as demonstrated in [16].

We show it is possible to efficiently optimize over the continuous choices of rotational axes of single-qubit gates in a PQC and thus obtain a significant improvement over structural optimization using limited set of rotations in [16–18]. The proposed method, which we refer to as *Free-Axis Selection* (or, in the hereafter shortened to **Fraxis**), handles the task to find an optimal axis of rotation with a given rotational angle θ on a single-qubit gate of a PQC.

II. METHODS

We consider the problem of optimizing a PQC that consists of a sequence of parameterized unitary gates $U = U_1 \dots U_D$ so that from a given initial state ρ the PQC generates a quantum state $U\rho U^\dagger$ minimizing an objective function encoded in a Hamiltonian M. Namely, the value of $\langle M \rangle \equiv \mathrm{tr}(MU\rho U^\dagger)$ is minimized over all parameter

[*] These authors contributed equally to this work

space of the PQC. We first present simple examples on why our proposed Fraxis can lead to more efficient PQCs and describe the basics of optimizing PQCs by borrowing the techniques in [16]. We then derive Fraxis in its general form, and present its simplified form that turns out to be computationally more efficient than the Rotoselect [16].

A. Motivating Examples

Here is an example why selecting axes of rotations can be better than fixing the axes and optimizing the continuous angles. Let us consider a simple toy Hamiltonian $H = X + Y + Z$ of a single qubit system. The ground state of H can be computed from a single-qubit PQC with R_yR_z ansatz. Namely, a typical PQC consists of a $R_y(\theta_1)$ gate followed by a $R_z(\theta_2)$ gate. Notice that if the initial quantum state is $|0\rangle$, then clearly no PQCs with single R_y or R_z gate can find the ground state $\rho = \frac{1}{2}\left(I - \frac{1}{\sqrt{3}}(X+Y+Z)\right)$.

This can be seen from the Bloch-vector representation of the ground state, which is $\boldsymbol{r}_g = -\frac{1}{\sqrt{3}}(1,1,1)^T$, and that of the initial state, which is $\boldsymbol{r}_0 = (0,0,1)^T$. The R_y and R_z gates rotate a quantum state along the y and z axes, respectively. Thus, it is not possible to obtain $\boldsymbol{r}_g$ from $\boldsymbol{r}_0$ by either R_y or R_z alone (or, even by applying R_z and R_y gate in that order). Thus, previously known approaches, such as, the Rotosolve and Rotoselect [16], need both R_y and R_z rotational gates, must order them correctly, and tune the angles to obtain the ground state.

On the other hand, it is well known that single-qubit unitaries are rotations of the Bloch sphere [19], and each rotation can be represented by a real unit vector $\hat{\boldsymbol{n}} = (n_x, n_y, n_z)^T \in \mathbb{R}^3$ and an angle $\theta \in [0, 2\pi]$ so that the unitary is

$$R_{\hat{\boldsymbol{n}}}(\theta) = \cos(\theta/2)I - i\sin(\theta/2)\left(n_xX + n_yY + n_zZ\right). \tag{1}$$

Notice that the $R_y(\theta)$ and $R_z(\theta)$ are instances of the above unitary. $R_{\hat{\boldsymbol{n}}}(\theta)$ rotates its input state by θ about the $\hat{\boldsymbol{n}}$ axis. With regards to the toy Hamiltonian H, by some algebra we can confirm that rotating the input state $|0\rangle$ by $\theta = \pi$ about the axis $\hat{\boldsymbol{n}} = \left(\frac{1}{\sqrt{2}}\sqrt{\frac{1}{2}+\frac{1}{2\sqrt{3}}}, \frac{1}{\sqrt{2}}\sqrt{\frac{1}{2}+\frac{1}{2\sqrt{3}}}, -\sqrt{\frac{1}{2}-\frac{1}{2\sqrt{3}}}\right)^T$ results in the ground state. Therefore, should we be able to optimize the axis of rotation of the corresponding unitary then we can obtain the ground state using only a single gate. We will show later that optimizing the axis of rotation indeed results in quantum circuits with better expressibility. In what follows, we first describe the steps to select axis of rotation in PQCs.

B. Optimizing PQCs

The PQCs are used to find an n-qubit quantum state $\rho \in \mathbb{C}^{2^n \times 2^n}$ to minimize an objective function encoded in a Hermitian matrix M. We follow a similar setting as the one in [16]. The quantum state ρ is generated by a PQC that consists of a sequence of unitary gates $U = U_1U_2 \dots U_D$ acting on n qubits so that each U_d is either a fixed two-qubit gate, or a parameterized single-qubit gate [29]. For ease of explanation, we consider only the single-qubit gates and consider the d-th unitary U_d as a rotational gate with parameter θ_d written as:

$$U_d = e^{-i\frac{\theta_d}{2}\hat{\boldsymbol{n}}_d \cdot \overrightarrow{\sigma}} = \cos(\theta_d/2)I - i\sin(\theta_d/2)\hat{\boldsymbol{n}}_d \cdot \overrightarrow{\sigma} = R_{\hat{\boldsymbol{n}}_d}(\theta_d), \tag{2}$$

where $\hat{\boldsymbol{n}}_d \cdot \overrightarrow{\sigma} = n_{d,x}X + n_{d,y}Y + n_{d,z}Z$, for $\hat{\boldsymbol{n}}_d \in \mathbb{R}^3$ so that $|\hat{\boldsymbol{n}}_d| = 1$.

We want to optimize the following expected value on the objective function over the choice of parameters of all single qubit gates, i.e., θ_d and $\hat{\boldsymbol{n}}_d$ for $d = 1, \dots, D$, that we can gather as $\boldsymbol{\theta} = (\theta_1, \dots, \theta_D)$ and $(\hat{\boldsymbol{n}}_1, \dots, \hat{\boldsymbol{n}}_D)$. To further simplify the notation, similar to [16] we write the expectation value as

$$\langle M \rangle = \mathrm{tr}\left(MU_D \dots U_{d+1}U_dU_{d-1} \dots U_1\rho U_1^\dagger \dots U_{d-1}^\dagger U_d^\dagger U_{d+1}^\dagger \dots U_D^\dagger\right), \tag{3}$$

which can be transformed as below because trace is invariant under cyclic permutations.

$$\langle M \rangle = \left\langle \hat{M} \right\rangle = \mathrm{tr}\left(\hat{M}U_d\hat{\rho}U_d^\dagger\right). \tag{4}$$

In the above, we utilize the following definitions

$$\hat{M} \equiv U_{d+1}^\dagger \dots U_D^\dagger M U_D \dots U_{d+1}, \tag{5}$$

$$\hat{\rho} \equiv U_{d-1} \dots U_1 \rho U_1^\dagger \dots U_{d-1}^\dagger. \tag{6}$$

In the hereafter, we simply write $\hat{M}$ and $\hat{\rho}$ as M and ρ, respectively, as they are clear from the context. Fixing all parameters of single-qubit gates excepting the d-th ones, the value $\langle M \rangle$ is a function of θ_d and $\hat{n}_d$ as below (see [16] for the proof).

$$\langle M \rangle_{\hat{\boldsymbol{n}}_d, \theta_d} = \cos^2\left(\frac{\theta_d}{2}\right) \mathrm{tr}\left(M\rho\right) + i \sin\left(\frac{\theta_d}{2}\right) \cos\left(\frac{\theta_d}{2}\right) \mathrm{tr}\left(M\left[\rho, H_d\right]\right) + \sin^2\left(\frac{\theta_d}{2}\right) \mathrm{tr}\left(M H_d \rho H_d\right), \tag{7}$$

where $H_d \equiv \hat{\boldsymbol{n}}_d \cdot \overrightarrow{\sigma}$, and $[A, B] \equiv AB - BA$.

Now, we remark an important result of [16, 17]: given a fixed $\hat{\boldsymbol{n}}_d$ the optimal θ_d^* can be computed efficiently because $\langle M \rangle_{\theta_d}$ is a sinusoidal form that can be fully characterized by evaluating the circuits on three different parameters of θ_d. The above fact is used to derive structural optimization of PQCs for finding the optimal angles $\boldsymbol{\theta}$ under fixed axes of rotations (e.g., Rotosolve in [16]), and for choosing $(\hat{\boldsymbol{n}}_1, \ldots, \hat{\boldsymbol{n}}_D) \in \{R_x, R_y, R_z\}^D$ where each of R_x, R_y and R_z corresponds to rotational axes $\hat{\boldsymbol{n}}^{(x)} = (1,0,0)^T$, $\hat{\boldsymbol{n}}^{(y)} = (0,1,0)^T$, and $\hat{\boldsymbol{n}}^{(z)} = (0,0,1)^T$, respectively. In the next section we show how we can further extend the choices of rotational axes to all unit vectors $\hat{\boldsymbol{n}}_d \in \mathbb{R}^3$.

C. The Proposed Method: Free-Axis Selection

Hereafter we describe how to optimize over the choice of a unit vector $\hat{\boldsymbol{n}}_d \in \mathbb{R}^3$ that defines the rotation axis with regards to a fixed rotation angle θ_d. For simplicity, we drop the subscript d. We want to determine the values of n_x, n_y and n_z to minimize $\langle M \rangle_{\hat{\boldsymbol{n}}}$.

By the Lagrange multipliers, the objective function to optimize is

$$(\hat{\boldsymbol{n}}^*, \lambda^*) = \min_{\hat{\boldsymbol{n}}, \lambda} f(\hat{\boldsymbol{n}}, \lambda) \equiv \langle M \rangle_{\hat{\boldsymbol{n}}} - \lambda\left(n_x^2 + n_y^2 + n_z^2 - 1\right), \tag{8}$$

where λ is a scalar whose value can be determined later.

For ease of notation, let us define $\alpha_\theta \equiv i \sin\left(\frac{\theta}{2}\right) \cos\left(\frac{\theta}{2}\right)$, and assume that $\alpha_\theta \neq 0$ (it is zero if $\theta = 0$, where there is no need to change axis of rotation, or if $\theta = \pi$ which we will deal later). Differentiating f with respect to n_x, n_y, and n_z we obtain that the following system of linear equations has to be satisfied by an optimal $(\hat{\boldsymbol{n}}^*, \lambda^*)$. The derivation of the system of linear equations is given in Appendix (VI A).

$$\left[\sin^2\left(\frac{\theta}{2}\right) \begin{pmatrix} 2r_x & 2r_{(x+y)} - r_x - r_y & 2r_{(x+z)} - r_x - r_z \\ 2r_{(x+y)} - r_x - r_y & 2r_y & 2r_{(y+z)} - r_y - r_z \\ 2r_{(x+z)} - r_x - r_z & 2r_{(y+z)} - r_y - r_z & 2r_z \end{pmatrix} - 2\lambda^* \mathbf{I}\right] \begin{pmatrix} n_x^* \\ n_y^* \\ n_z^* \end{pmatrix} = -\alpha_\theta \begin{pmatrix} \mathrm{tr}\left(M\left[\rho, X\right]\right) \\ \mathrm{tr}\left(M\left[\rho, Y\right]\right) \\ \mathrm{tr}\left(M\left[\rho, Z\right]\right) \end{pmatrix} \tag{9}$$

In the above equation, we use the following definitions which are the expectation values obtain from running the circuits each of which its d-th single-qubit gate is replaced with $X, Y, Z, (X+Y)/\sqrt{2}, (X+Z)/\sqrt{2}$ and $(Y+Z)/\sqrt{2}$ gate, respectively.

$$r_x \equiv \mathrm{tr}\left(MX\rho X\right), \tag{10}$$
$$r_y \equiv \mathrm{tr}\left(MY\rho Y\right), \tag{11}$$
$$r_z \equiv \mathrm{tr}\left(MZ\rho Z\right), \tag{12}$$
$$r_{(x+y)} \equiv \mathrm{tr}\left(M\left(\frac{X+Y}{\sqrt{2}}\right)\rho\left(\frac{X+Y}{\sqrt{2}}\right)\right), \tag{13}$$
$$r_{(x+z)} \equiv \mathrm{tr}\left(M\left(\frac{X+Z}{\sqrt{2}}\right)\rho\left(\frac{X+Z}{\sqrt{2}}\right)\right), \tag{14}$$
$$r_{(y+z)} \equiv \mathrm{tr}\left(M\left(\frac{Y+Z}{\sqrt{2}}\right)\rho\left(\frac{Y+Z}{\sqrt{2}}\right)\right). \tag{15}$$

Eq. (9) can be simply written as:

$$\left(\sin^2\left(\theta/2\right)\mathbf{R} - 2\lambda^*\mathbf{I}\right)\hat{\boldsymbol{n}}^* = -\alpha_\theta \mathbf{b}, \tag{16}$$

where $\mathbf{b} \equiv \left(\mathrm{tr}\left(M\left[\rho, X\right]\right), \mathrm{tr}\left(M\left[\rho, Y\right]\right), \mathrm{tr}\left(M\left[\rho, Z\right]\right)\right)^T$, and

$$\mathbf{R} \equiv \begin{pmatrix} 2r_x & 2r_{(x+y)} - r_x - r_y & 2r_{(x+z)} - r_x - r_z \\ 2r_{(x+y)} - r_x - r_y & 2r_y & 2r_{(y+z)} - r_y - r_z \\ 2r_{(x+z)} - r_x - r_z & 2r_{(y+z)} - r_y - r_z & 2r_z \end{pmatrix}. \tag{17}$$

Algorithm 1 Free-axis selection (Fraxis)

Input A Hermitian matrix M of the objective function, a PQC with single-qubit gates $U_1 \dots U_D$, and a stopping criterion
Initialize: Choose random axes, i.e., $\hat{\boldsymbol{n}}_d \in \mathbb{R}^3$ such that $|\hat{\boldsymbol{n}}_d| = 1$ for $d = 1, 2, \dots, D$.
repeat
  for d = 1, 2, ... D **do**
    Fix all axes $\hat{\boldsymbol{n}}_j$ for $j \neq d$
    Compute elements of the $\mathbf{R}_d$ matrix as in Eq. (10)–(15)
    Compute the eigenvectors $\boldsymbol{r}_d^{(1)}, \boldsymbol{r}_d^{(2)}, \boldsymbol{r}_d^{(3)}$ of $\mathbf{R}_d$
    $\hat{\boldsymbol{n}}_d \leftarrow \arg\min_{\boldsymbol{r} \in \{\boldsymbol{r}_d^{(1)}, \boldsymbol{r}_d^{(2)}, \boldsymbol{r}_d^{(3)}\}} \langle M \rangle_{\boldsymbol{r},\pi}$, where $\langle M \rangle_{\boldsymbol{r},\theta}$ as in Eqs. (7)
until satisfying stopping criterion
Output Axes of rotation $(\hat{\boldsymbol{n}}_1, \dots, \hat{\boldsymbol{n}}_D)$

We can clearly see that $\mathbf{R}$ is a symmetric matrix and therefore its eigenvalues are real numbers. Moreover, provided that $\sin(\frac{\theta}{2})\cos(\frac{\theta}{2}) \neq 0$ there exists $\lambda \in \mathbb{R}$ such that Eq. (16) has a unique solution. In fact, the value of λ^* is the solution to the following identity:

$$1 = (\hat{\boldsymbol{n}}^*)^T \hat{\boldsymbol{n}}^* = (\alpha_\theta)^2 \mathbf{b}^T \left(\left(\sin^2(\theta/2)\mathbf{R} - 2\lambda \mathbf{I} \right)^{-1} \right)^2 \mathbf{b}.$$

Eq. (16) gives a method to determine an optimal way to select axis of rotation against a fixed θ by estimating the matrix $\boldsymbol{R}$ and vector $\boldsymbol{b}$. We notice that we can further simplify the equation when $\theta = \pi$ that corresponds to replacing the single-qubit gate with a π-rotation gate around a particular axis. Although this type of gates is less general than an arbitrary rotational gate, we will see in later sections that not only its expressibility is higher, but also its computational cost to optimize the parameter is better than other ansatz.

D. Simplified Free-Axis Selection

We have derived a method to select a rotational axis minimizing the objective function by fixing θ. We refer to it as θ-Fraxis. The method requires us to run the PQC nine times to estimate the elements of $\mathbf{R}$ and $\mathbf{b}$, which is more than previous methods. However, we can reduce the number of PQC runs by fixing $\theta = \pi$ as we do not need $\mathbf{b}$ in this case. Moreover, the unit vector $\hat{\boldsymbol{n}}^*$ becomes an eigenvector of $\mathbf{R}$. The axis that minimizes the objective function is exactly the eigenvector whose eigenvalue is the smallest. We summarize the π-Fraxis in Algorithm 1. In the hereafter, for ease of notation we simply call it Fraxis and omit the prefix π-.

Finally, we note an important and nice property of the Fraxis: its computational resources turn out to be less than those of Rotoselect [16] despite the fact that it selects axis of rotation in continuous manner. While Rotoselect requires evaluating seven energy estimations with the PQC, at each for-loop of Algorithm 1 Fraxis needs exactly six energy estimations to compute all elements of $\boldsymbol{R}_d$. The landscape of $\langle M \rangle_{\hat{\boldsymbol{n}}_d,\pi}$ as in Eq. (7) is completely determined by the eigenvalues and eigenvectors of $\boldsymbol{R}_d$. We record this property in the following lemma.

Lemma 1. *Let $\lambda_d^{(i)}$ and $\boldsymbol{r}_d^{(i)}$ be, respectively, the i-th eigenvalue and eigenvector of $\boldsymbol{R}_d$ at each for-loop of Algorithm 1 for $i = 1, 2, 3$ so that $\lambda_d^{(1)} \leq \lambda_d^{(2)} \leq \lambda_d^{(3)}$. Then, for all unit vector $\boldsymbol{r} \in \mathbb{R}^3$ it holds that $\lambda_d^{(1)}/2 \leq \langle M \rangle_{\boldsymbol{r},\pi} \leq \lambda_d^{(3)}/2$, and therefore $\hat{\boldsymbol{n}}_d = \boldsymbol{r}_d^{(1)}$.*

Proof. It follows directly from transforming Eq. (7) to obtain $\langle M \rangle_{\boldsymbol{r},\pi} = (1/2)\ \boldsymbol{r}^T \cdot \boldsymbol{R}_d \cdot \boldsymbol{r}$, and therefore the minimum (maximum) value is achieved when $\boldsymbol{r}$ is the eigenvector whose eigenvalue is minimum (maximum). □

Figure 1 shows the relation between landscape of $\langle M \rangle_{\hat{\boldsymbol{n}}_d,\pi}$ and three eigenvectors $\boldsymbol{r}_d^{(1)}, \boldsymbol{r}_d^{(2)}, \boldsymbol{r}_d^{(3)}$ drawn for simple model example. Eigenvectors denoted by blue and red arrows corresponds to the rotation axes for minimum and maximum energy values, respectively.

Finally, we remark that the cost of energy estimations can be significantly reduced for special Hamiltonians (e.g., k-local Hamiltonians [20–22]) if we use locally-biased classical shadows [23]. Also, fixing the rotation angle of local gates could be advantageous feature for device control, such as simplification of microwave control pulse calibration for superconducting qubits. Some potential drawbacks of the Fraxis method are its reduced expressibility (such as, lacking ability to represent identity gates), and its overfitting against the R_y-ansatz circuits that are popular for molecular Hamiltonians (because such ansatz are sufficient for searching quantum states with real-valued probability amplitudes). We show in practice those drawbacks are not significant in the next section.

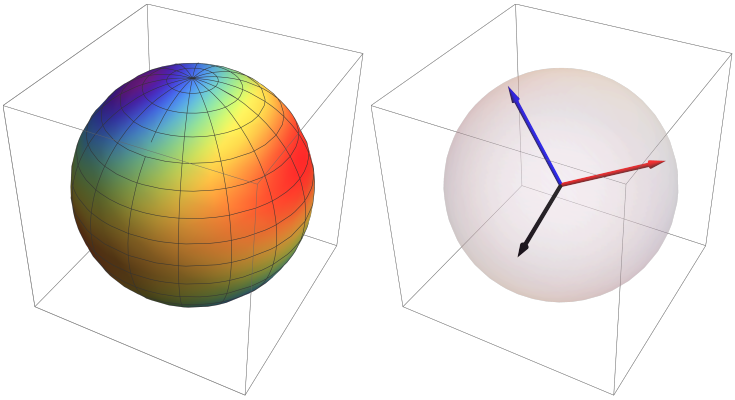


FIG. 1: (left) An illustration of the energy landscape of Fraxis determined by the matrix $\boldsymbol{R}_d$. Red, green and blue areas correspond to high, medium, low energy values. (right) Three eigenvectors corresponding to minimum (blue), maximum (red) and saddle (black) points.

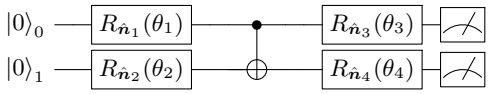


FIG. 2: A 2-qubit PQC with θ-Fraxis.

III. RESULTS

In this section, we first apply θ-Fraxis to simple 2-qubit system and compare it to Rotosolve and Rotoselect, and show its advantage in energy convergence, together with the potential of the simplified π-Fraxis. Next, using Qiskit [24] we demonstrate the high expressibility of PQCs with the simplified π-Fraxis (or, Fraxis) in contrast to PQCs with its traditional counterparts. For studying expressibility, we show the one-qubit PQCs with Fraxis can generate quantum states covering the Bloch sphere more uniformly that its counterparts. We then show evidences of better expressibility of multi-qubit PQCs with Fraxis when measured in KL divergences under different number of entangled layers. Finally, we present numerical experiments showing Fraxis can further improve the efficacy of variational quantum circuits for the Heisenberg model and molecular Hamiltonians.

A. Comparison with other structural optimization

For comparison, we tested Rotosolve, Rotoselect, θ-Fraxis, and Fraxis methods to a 2-qubit PQC system in Fig. 2 for finding the ground state of the Hamiltonian

$$H = 0.1(XX + YY + ZZ) + 0.01(IZ + ZI). \tag{18}$$

The ground state energy is $E_0 = -0.3$. The variational parameters to be optimized are four rotation angles $(\theta_1, \theta_2, \theta_3, \theta_4)$ and their corresponding rotational axes, $(\hat{\boldsymbol{n}}_1, \hat{\boldsymbol{n}}_2, \hat{\boldsymbol{n}}_3, \hat{\boldsymbol{n}}_4)$. We applied the methods to start from the same 50 random initial parameter sets.

Fig. 3 shows the optimized energy as a function of iterative number of optimization. One can see that θ-Fraxis shows better convergence compared to Rotosolve or Rotoselect, while those methods seem to suffer from being trapped in local minima. This is because Rotosolve and Rotoselect mainly optimize rotation angles $(\theta_1, \theta_2, \theta_3, \theta_4)$ and leave rotation axes unchanged (Rotosolve) or selected from the limited choice, R_x, R_y or R_z (Rotoselect). Although Rotoselect result exhibits better convergence nature, i.e., less likely to be trapped in local minima when compared to Rotosolve, some parameter sets could not lead to the exact ground state energy E_0.

On the other hand, θ-Fraxis makes it possible to optimize rotation axis continuously with respect to given θ, which results in fine convergence nature as shown in Fig. 3. We also applied the simplified Fraxis with $\theta = \pi$ (or, π-Fraxis) to the same problem. Note that the initial parameters for axes are same as in Rotosolve, Rotoselect and θ-Fraxis cases, while rotation angles $(\theta_1, \theta_2, \theta_3, \theta_4)$ are set to (π, π, π, π). The result in Fig. 3 shows that the convergence behaviour of π-Fraxis is still better than those of Rotosolve/select. We also remark that π-Fraxis has comparable performance to

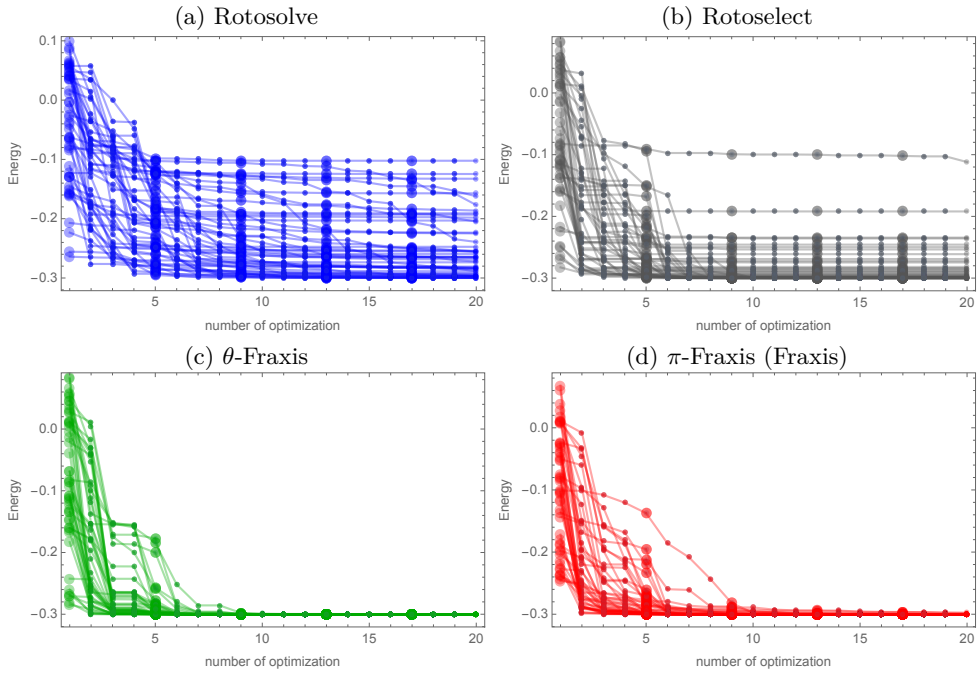


FIG. 3: Comparison of energy convergence as a function of number of iterations for optimization. (upper left : Blue lines) Rotosolve, (upper right : Gray lines) Rotoselect, (lower left : Green lines) θ-Fraxis, (lower right : Red lines) Fraxis. Common initial angular parameter set is used for Rotosolve, Rotoselect and θ-Fraxis, while $(\theta_1, \theta_2, \theta_3, \theta_4)$ are overridden to (π, π, π, π) in Fraxis trials.

θ-Fraxis. Nevertheless, estimation cost for matrix-elements in π-Fraxis is reduced from 9 to 6, which suggests π-Fraxis is more practical than θ-Fraxis. Recall that we abbreviate π-Fraxis as Fraxis and will further investigate expressibility of Fraxis ansatz/circuit to ensure the ability of this simplified method in the next subsection.

B. Circuit Expressibility

In general, a universal gate for a single qubit is represented as

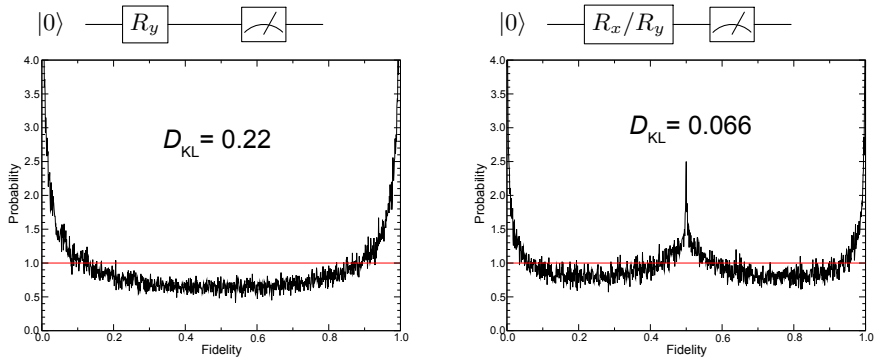
$$U = \begin{bmatrix} \cos\frac{\psi}{2} & -e^{i\lambda}\sin\frac{\psi}{2} \\ e^{i\phi}\sin\frac{\psi}{2} & e^{i(\phi+\lambda)}\cos\frac{\psi}{2} \end{bmatrix}. \tag{19}$$

In the case of Fraxis, the rotation angle is fixed at $\theta = \pi$ in Eq. 1, which is equivalent the constraint condition $\phi + \lambda = \pi$ in Eq. 19. Hence, a single gate controlled by Fraxis is not universal any more. For instance, the identity gate is not included with the Fraxis expression. However, it should be noted that any input state in a single qubit system can be transformed to an arbitrary state. To intuitively understand this, let us suppose one qubit state in $|0\rangle$. This $|0\rangle$ state can be transformed to any state on the Bloch sphere by π rotation around the axis that passes through the dividing point of $|0\rangle$ and the target state. Therefore, the Fraxis ansatz seems to retain high expressibility. To quantitatively evaluate the expressibility of the Fraxis circuit in comparison with Rotosolve and Rotoselect, we evaluate Kullback-Leibler (KL) divergence as

$$\mathcal{E}(\mathcal{C}) \;=\; D_{\mathrm{KL}}(P(C,F)||P_{\mathrm{Haar}}(F)) = \int_0^1 P(C,F)\log\frac{P(C,F)}{P_{\mathrm{Haar}}(F)}dF, \tag{20}$$

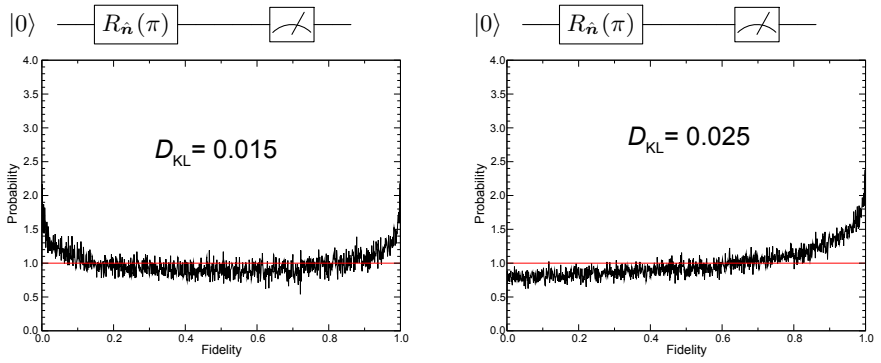
where F is the fidelity between two parameterized random states $|\psi(\theta)\rangle, |\psi(\theta')\rangle$ which is for two pure states defined as $F = |\langle\psi|\psi'\rangle|^2$. Here, $P(C,F)$ is a *probability distribution function* (pdf) of fidelity F of a circuit ansatz C. $P_{\mathrm{Haar}}(F)$ is a probability distribution sampled uniformly according to the Haar measure, in which $|\psi\rangle$ is uniformly distributed in Hilbert space. According to [25] the Haar-measure distribution is derived as, $P_{\mathrm{Haar}}(F) = (N-1)(1-F)^{N-2}$, where N is the dimension of Hilbert space. Here, a product state of $\otimes^n |0\rangle$ is employed as the initial state.

First, we confirmed the expressibility of single-qubit circuits as shown in Figs. 4 and 5, where $|0\rangle$ is employed as the initial state. Since the rotation axis is fixed in Rotosolve ansatz, the expressibility is equivalent to that of a single



(a) Rotosolve **(b)** Rotoselect

FIG. 4: Fidelity distribution obtained by single-qubit circuits with (a) Rotosolve(R_y), and (b) Rotoselect(R_x/R_y). Horizontal lines in red represent Haar-uniform distribution. The evaluated KL divergence values are displayed over the respective plots



(a) parameter-random Fraxis **(b)** state-random Fraxis

FIG. 5: Fidelity distribution obtained by single-qubit circuits with (a) parameter-random Fraxis, and (b) state-random Fraxis. Horizontal lines in red represent Haar-uniform distribution. The evaluated KL divergence values are displayed over the respective plots.

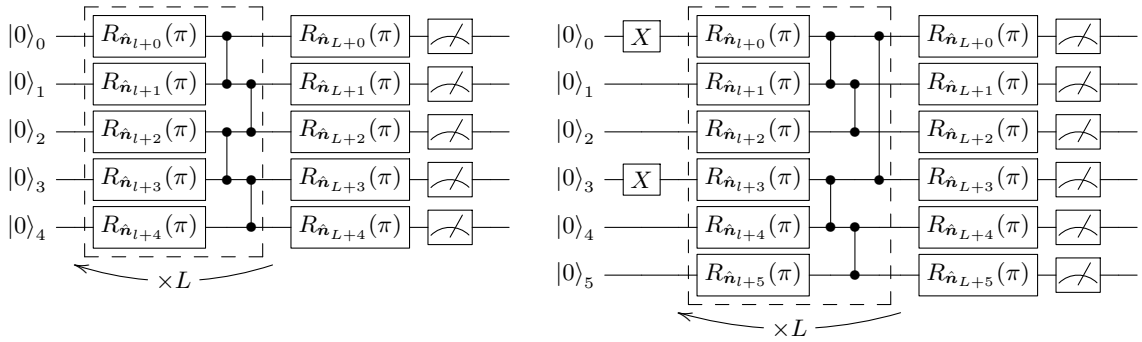
$R_y(\theta)$ gate. Then, the rotation angles θ were randomly generated in the range of $(-\pi, \pi]$. Due to symmetry, the fidelity is analytically given as $F = f(\theta) = |\langle 0|\psi(\theta)\rangle|^2 = \cos^2(\theta/2)$ where θ is uniformly distributed. Then, the pdf of the fidelity $P(F)$ can be defined as $P(F) \propto |f^{-1}(F)/dF|$. Because $f^{-1}(F) = -2\arccos(F)$, we obtain

$$P(F) = \frac{1}{\pi\sqrt{F(1-F)}}. \tag{21}$$

Substituting Eq.(21) into Eq.(20) gives theoretical value of $D_{KL} = \log(4/\pi) = 0.24$ for Rotosolve with R_y-ansatz. As mentioned in the previous study [26], we confirmed the fact that the KL divergence is largely sensitive to the bin width of the pdf histogram, and thus the proper width is not trivial. Here, we evaluated the histogram with bin width of 0.001 and 100,000 fidelity sampling and obtained numerically the $D_{KL} = 0.22$ for the Rotosolve (as in Fig 4), in close agreement with the analytical value. Notice from the figure that the high value of the KL divergence is due to discrepancies of probability distributions around the edges, i.e., $F = 0, 1$ and a wide range centered at $F = 0.5$. For fair comparison, all expressibility evaluations are based on bin width of 0.001 and 100,000 sampling throughout this paper.

In the Rotoselect circuit, the rotation gates are usually selected from $R_x(\theta)$, $R_y(\theta)$ and $R_z(\theta)$. Since the $R_z(\theta)$ rotation does not change the initial state $|0\rangle$, we here randomly selected either $R_x(\theta)$ or $R_y(\theta)$ with equal probabilities as well as the rotation angle that uniformly distributed in the range of $(-\pi, \pi]$. In other words, the resulting pdf of the fidelity is a mixture of those of $|\langle 0| R_y(\theta')R_y(\theta)|0\rangle|^2$ and $|\langle 0| R_x(\theta')R_y(\theta)|0\rangle|^2$. While the former term contributes in the same way as Rotosolve, the cross evaluation between R_x and R_y contributes in the peak at $F = 0.5$ that in turn lifts the pdf closer to the Haar-random. This gives a better expressibility of Rotoselect ansatz.

In the Fraxis circuits, while the rotation angles are fixed $\theta = \pi$, the rotation axis $\hat{\boldsymbol{n}} = (n_x, n_y, n_z)^T$ are generated randomly. Since the Fraxis ansatz optimizes the target gate in a deterministic way without systematic search in



(a) 5-qubit circuit A **(b)** 6-qubit circuit B

FIG. 6: PQCs with Fraxis-ansatz

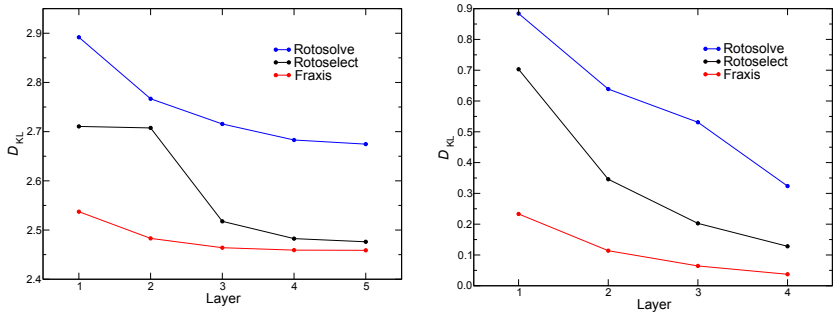


FIG. 7: Relation between KL divergences and number of layers for circuits A (left) and B (right)

parameter space, it is not straightforward to compare the expressibility with conventional parameterized circuits. Indeed, the description of randomness of Fraxis ansatz seems not to be unique and we have to arbitrarily choose to describe randomness in either parameter space (parameter random) or Hilbert space (state random).

In parameter random, the rotation axis is represented as $\hat{\boldsymbol{n}} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^T$ and θ and ϕ are randomly sampled with uniform probability in ranges of $[0, \pi)$ and $[-\pi, \pi)$, respectively. In state random, each element of rotation axis (i.e., n_x, n_y, n_z) was randomly generated according to the normal distribution, and then normalized by the sum of their squares.

For qualitative understanding of the difference between parameter and state randoms, let us consider the surface area on Bloch sphere. Here, We can consider only the rotation axis pointing to the north sphere of the Bloch sphere without loss of generality because of the symmetry of π rotation. A set of rotation axis pointing south from of latitude 45 degree transforms $|0\rangle$ to a state in the south sphere on Bloch sphere, while the rest of the axis group transforms $|0\rangle$ to a state in the north sphere.

Since the surface area south from latitude 45 degree is larger than the rest, the uniform distribution of rotation axis on Bloch sphere results in larger population of transformed state on south sphere. This bias will cause asymmetry of the pdf of fidelity as shown in Fig. 5 In the parameter random, the rotation axis $\hat{\boldsymbol{n}}$ is likely to be more populated around north/south pole on the Bloch sphere. Accordingly, the rotated states are also populated around $|0\rangle$, which increases the probability around $F = 0$ and 1 as shown in Fig 5. Although the KL divergences vary depending on how the random states are generated in Fraxis, it is notable that both pdfs are almost uniform close to Haar measure even with the fixed rotation angle $\theta = \pi$ for 1-qubit systems.

Next, we evaluated expressibility of two multi-layer PQCs as shown in Fig. 6 consisting of 5 and 6 qubits respectively. Fig. 7 shows the relation between the KL divergence and number of layers. It is notable that Fraxis always obtained smaller KL divergences than Rotosolve and Rotoselect.

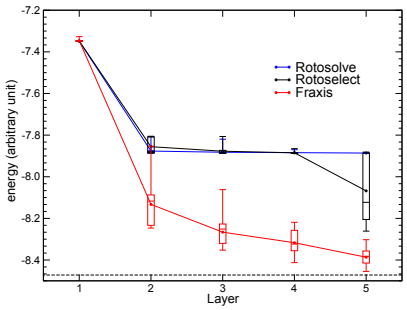


FIG. 8: Relation between number of layer for circuit A and the optimized energy after 100 iterations. The averaged energy is represented by filled circles and lines. The boxplots represent respective quantiles. The exact ground state energy is displayed with dash line.

C. Heisenberg Model

We performed energy optimization using the Hamiltonian that is one dimensional Heisenberg model under the periodic boundary condition with 5-qubit as

$$H = J \sum_{(i,j)\in\mathcal{E}} (X_i X_j + Y_i Y_j + Z_i Z_j) + h \sum_{i\in\mathcal{V}} Z_i, \tag{22}$$

where $J = h = 1$. We employed circuit A in Fig. 6, in Rotosolve, Rotoselect and Fraxis ansatze, respectively. Although the gate update order is arbitrary and can affect the final results, we here conducted gate update as: starting from the first layer, (i) update the single-qubit gate at the 0-th qubit in the layer, (ii) update each single-qubit gate in the ascending order of qubits in the layer, (iii) move to the next layer in ascending order, and (iv) go back to (i) from the first layer until maximum number of iterations is reached.

Figure 8 shows the optimized energy after 100 iteration and thus the optimization is not necessarily converged. Although KL divergence shows large difference among the ansatze in Fig 7, the obtained energy are consistent among all ansatze for $L = 1$. In the multi-layer calculation with the Rotosolve ansatz, the energy were not improved except for at $L = 2$. It is probably because the ansatz consisting of only Ry gates cannot cover the ground state of this model. It is also notable that although the expressibility of Rotoselect was not improved by increasing the number of layers from 1 to 2, the obtained energy was obviously improved in Rotosolve. Fraxis showed significant improvement even at $L = 3$, 4, and 5, although the improvement of the expressibility was not as large as those of Rotosolve and Rotoselect in Fig. 7. Altogether, although Fraxis showed the best performance, it seems that its performance cannot be explained solely by the expressibility measured with KL divergence. And thus, it is also important to discuss the circuit performance with the entanglement capability. However, it should be stressed that the Fraxis ansatz always showed the best performance with the given circuit and the number of layers.

D. Molecular Hamiltonian Model: H_2/6-31G

Here, we evaluated the electronic energy of H_2 molecules with bond length of 0.75 Å. First we performed the Hartree-Fock calculation with 6-31G basis set, which was followed by the fermionic Hamiltonian construction with active space [2,3]. Then, the fermionic Hamiltonian was converted to the qubit Hamiltonian by Jordan-Wigner transformation. The optimization continued until either reaching to convergence or 100 iterations, where respective gates were sequentially updated as in the previous section. Since the optimization strongly depends on the initial conditions, we evaluated statistical average by randomly generating the initial parameters. Paying attention to the statistical error, the independent optimizations were performed 20 to 300 times with different initial conditions. In the case of Fraxis, the initial states were generated in the manner of parameter random as described in Sec. III B.

In agreement with the result of the Heisenberg model, the obtained energy did not vary among ansatze at $L = 1, 2$ regardless of the diverse expressibility and all optimizations were stuck in local minimum at the Hartree-Fock energy level in Fig. 9. Note that at $L = 3$ the averaged optimized energy by Rotosolve is lower than that by Rotoselect. For qualitative understanding of this non-intuitive results, it should be noted that in quantum chemistry, the coefficients of the CI expansion are in usual real values, which indicates that the eigenstates can be expressed by only R_y gates. We suppose the high expressibility can bring two conflicting consequences, (1) inclusion of the crucial piece of Hilbert

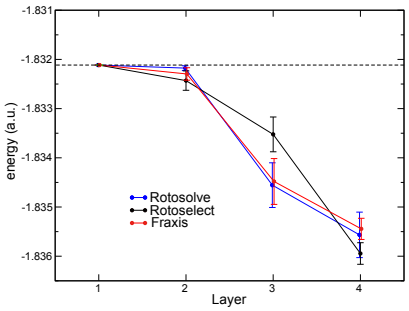


FIG. 9: The average of optimized energy after 100 iterations with standard error. The Hartree-Fock state energy is displayed with dash line.

space spanned by the ansatz, which may lead to successfully the ground state. This also includes the ability to express the ground state itself, (2) inessential extension of search space irrelevant to the efficient optimization. We suppose that the latter effects emerged in the case of Rotoselect at $L = 3$ because the Rotoselect ansatz greedily selects a gate that yields the lowest energy at each optimization state. Indeed, we confirmed that Rotoselect optimization did not necessarily select R_y gate, which may not be the best choice to find the global minimum, although it can temporally gain the largest energy decrease. In contrast, Fraxis resulted in the energy at the same level as Rotosolve. Although we found the selected axes in Fraxis were not necessarily on the X-Z plain of the Bloch sphere expression, it may be possible to express effective path to the ground state by utilizing complex space if the expresssibility is sufficiently high. As supporting evidences, the optimized efficiency levels of Rotoselect and Fraxis are reversed at $L = 4$, which is probably because as the number of layer increases, the confined search space in Rotosolve and Rotoselect may outweight the expressibility.

IV. CONCLUSION

In this paper, we proposed a new method for optimizing rotational axes of single-qubit gates in a PQC. Since it makes full use of the expressibility of a single qubit gate, the Fraxis ansatz can achieve high expressibility even with shallow PQCs. Indeed, we confirmed the expressibilities estimated with KL divergence for both single- and multi-qubit PQCs with Fraxis are significantly higher when compared to optimizing the rotation angles. Nevertheless, only six energy estimations are required for each gate update to find the optimal axis, which is more efficient than Rotoselect that requires seven of them.

Although Fraxis enables to find the optimal rotation axis for single qubit gate, the optimal order of gate update still remains untouched. Indeed, as seen in the application to a H_2 molecules, many optimization runs were trapped in Hartree-Fock states, which also happened with Rotosolve and Rotoselect. This result implies the optimal choice for a single qubit rotation is not necessarily the best strategy for the entire system to avoid local minima and Barren plateaus. This may be avoided by randomizing the order of updates. Another possible remedy is by simultaneously updating k axes, which is similarly shown in [17, 18] to update k angles simultaneously. However, updating k axes by Fraxis requires 6^k energy estimations in contrast to 3^k energy estimation of structural optimization in [17, 18] and for this reason we may have to combine Fraxis with other techniques, such as, matrix completion [27, 28].

On the other hand, we also remark that conventional stochastic and gradient approaches may still be compelling for the global optimization in certain cases. In that sense, Fraxis can be combined with those approaches. For instance, we can analytically evaluate the gradient of the cost functions using Eqs. (23) – (25). The obtained gradient is then passed to the conventional gradient-based optimizers. In addition, Fraxis has high affinity with stochastic optimizations, because it gives the exact energy landscape of the single qubit rotation accessible. This can be used to sample the rotation axis at random according to the estimated energy instead of choosing the (local) minimum one. We believe that Fraxis effectively elicits the potential of PQCs for the progress of quantum computation.

V. ACKNOWLEDGEMENTS

H.C.W. was supported by JSPS Grant Numbers 20K03885 and 20H05518, and JST PRESTO Grant number JPMJPR17GC. E.K was supported by JSPS Grant Number 20K14388 and JST PRESTO Grant number JPMJPR2011.

In addition, H.C.W., E.K., and M.S. were supported by the MEXT Quantum Leap Flagship Program Grant Number JPMXS0118067285 and JPMXS0120319794. R.R. would like to thank colleagues at IBM Quantum and Yutaka Shikano for technical discussion. The authors would like to thank Naoki Yamamoto for insightful advises.

[1] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, et al., arXiv:2012.09265 (2020).
[2] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, et al., arXiv preprint arXiv:2101.08448 (2021).
[3] J. Preskill, Quantum **2**, 79 (2018), ISSN 2521-327X, URL `https://doi.org/10.22331/q-2018-08-06-79`.
[4] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature **549**, 242 (2017), ISSN 0028-0836.
[5] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, et al., Phys Rev A **98**, 022322 (2018), URL `https://link.aps.org/doi/10.1103/PhysRevA.98.022322`.
[6] M. Ganzhorn, D. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, et al., Phys. Rev. Applied **11**, 044092 (2019), URL `https://link.aps.org/doi/10.1103/PhysRevApplied.11.044092`.
[7] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, npj Quantum Information **6**, 10 (2020), ISSN 2056-6387, URL `https://doi.org/10.1038/s41534-019-0240-1`.
[8] H. L. Tang, V. O. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, arXiv:1911.10205 (2019).
[9] N. V. Tkachenko, J. Sud, Y. Zhang, S. Tretiak, P. M. Anisimov, A. T. Arrasmith, P. J. Coles, L. Cincio, and P. A. Dub, arXiv:2009.04996 (2020).
[10] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nat Commun **9** (2018), ISSN 2041-1723.
[11] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Nat Commun **12** (2021), ISSN 2041-1723.
[12] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, arXiv preprint arXiv:2011.02966 (2020).
[13] Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Albrecht, and A. T. Sornborger, arXiv preprint arXiv:2009.14808 (2020).
[14] C. Zhao and X.-S. Gao, arXiv preprint arXiv:2102.01828 (2021).
[15] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, arXiv preprint arXiv:2011.12245 (2020).
[16] M. Ostaszewski, E. Grant, and M. Benedetti, Quantum **5**, 391 (2021), ISSN 2521-327X, URL `https://doi.org/10.22331/q-2021-01-28-391`.
[17] K. M. Nakanishi, K. Fujii, and S. Todo, Phys. Rev. Research **2**, 043158 (2020), URL `https://link.aps.org/doi/10.1103/PhysRevResearch.2.043158`.
[18] R. M. Parrish, J. T. Iosue, A. Ozaeta, and P. L. McMahon, arXiv:1904.03206 (2019).
[19] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
[20] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Boston, MA, USA, 2002), ISBN 0821832298.
[21] J. Kempe, A. Kitaev, and O. Regev, SIAM J. Comput. **35**, 1070 (2006), ISSN 0097-5397, URL `http://dx.doi.org/10.1137/S0097539704445226`.
[22] R. Oliveira and B. M. Terhal, Quantum Info. Comput. **8**, 900 (2008), ISSN 1533-7146, URL `http://dl.acm.org/citation.cfm?id=2016985.2016987`.
[23] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, arXiv e-prints arXiv:2006.15788 (2020), 2006.15788.
[24] H. Abraham, AduOffei, R. Agarwal, I. Y. Akhalwaya, G. Aleksandrowicz, T. Alexander, M. Amy, E. Arbel, Arijit02, A. Asfaw, et al., *Qiskit: An open-source framework for quantum computing* (2019).
[25] K. Życzkowski and H.-J. Sommers, Phys. Rev. A **71**, 032313 (2005), URL `https://link.aps.org/doi/10.1103/PhysRevA.71.032313`.
[26] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Advanced Quantum Technologies **2**, 1900070 (2019), ISSN 2511-9044, URL `http://dx.doi.org/10.1002/qute.201900070`.
[27] E. J. Candès and B. Recht, Foundations of Computational Mathematics **9**, 717 (2009), ISSN 1615-3383, URL `https://doi.org/10.1007/s10208-009-9045-5`.
[28] E. J. Candès and T. Tao, IEEE Transactions on Information Theory **56**, 2053 (2010).
[29] Here we make a simplification for ease of reading as U_d is in fact a $2^n \times 2^n$ matrix which is a tensor product of two-qubit gate or a single-qubit gate acting on a particular qubit with 2×2 identity matrix acting on other qubits.

VI. APPENDIX

A. Deriving Equations for Selecting Axis of Rotation with Fixed Angles

In this section we explain the derivation of Eq. (9), i.e., how to optimize over the choice of a unit vector $\hat{\boldsymbol{n}} \in \mathbb{R}^3$ that defines the rotation axis with regards to a fixed rotation angle θ. We want to determine the values of n_x, n_y and n_z to minimize $\langle M \rangle_{\hat{\boldsymbol{n}}}$ constrained with a unit $\hat{\boldsymbol{n}}$.

By the Lagrange multipliers, the objective function to optimize is

$$(\hat{\boldsymbol{n}}^*, \lambda^*) = \min_{\hat{\boldsymbol{n}}, \lambda} f(\hat{\boldsymbol{n}}, \lambda) \equiv \langle M \rangle_{\hat{\boldsymbol{n}}} - \lambda \left(n_x^2 + n_y^2 + n_z^2 - 1 \right).$$

Differentiating f with respect to n_x, n_y, and n_z we obtain

$$\begin{aligned}\frac{\partial f}{\partial n_x} = {} & \alpha_\theta \mathrm{tr}\left(M\left[\rho, X\right]\right) \\ & + \sin^2\left(\frac{\theta}{2}\right)\left(2n_x \mathrm{tr}\left(MX\rho X\right) + n_y \mathrm{tr}\left(MX\rho Y + MY\rho X\right) + n_z \mathrm{tr}\left(MX\rho Z + MZ\rho X\right)\right) \\ & - 2\lambda n_x, \end{aligned} \tag{23}$$

$$\begin{aligned}\frac{\partial f}{\partial n_y} = {} & \alpha_\theta \mathrm{tr}\left(M\left[\rho, Y\right]\right) \\ & + \sin^2\left(\frac{\theta}{2}\right)\left(2n_y \mathrm{tr}\left(MY\rho Y\right) + n_x \mathrm{tr}\left(MX\rho Y + MY\rho X\right) + n_z \mathrm{tr}\left(MY\rho Z + MZ\rho Y\right)\right) \\ & - 2\lambda n_y, \end{aligned} \tag{24}$$

$$\begin{aligned}\frac{\partial f}{\partial n_z} = {} & \alpha_\theta \mathrm{tr}\left(M\left[\rho, Z\right]\right) \\ & + \sin^2\left(\frac{\theta}{2}\right)\left(2n_z \mathrm{tr}\left(MZ\rho Z\right) + n_x \mathrm{tr}\left(MX\rho Z + MZ\rho X\right) + n_y \mathrm{tr}\left(MY\rho Z + MZ\rho Y\right)\right) \\ & - 2\lambda n_z. \end{aligned} \tag{25}$$

At optimality, the value $\hat{\boldsymbol{n}}^* = (n_x^*, n_y^*, n_z^*)^T$ satisfies:

$$\frac{\partial f}{\partial n_x} = \frac{\partial f}{\partial n_y} = \frac{\partial f}{\partial n_z} = 0. \tag{26}$$

Defining $r_x, r_y, r_z, r_{(x+y)}, r_{(x+z)}$ and $r_{(y+z)}$ as in Eqs. (10)–(15), and noticing the following identities:

$$\mathrm{tr}\left(MX\rho Y + MY\rho X\right) = 2r_{(x+y)} - r_x - r_y, \tag{27}$$

$$\mathrm{tr}\left(MX\rho Z + MZ\rho X\right) = 2r_{(x+z)} - r_x - r_z, \tag{28}$$

$$\mathrm{tr}\left(MY\rho Z + MZ\rho Y\right) = 2r_{(y+z)} - r_y - r_z, \tag{29}$$

we can transform Eq. (23)–(25) and Eq. (26) to obtain Eq. (9).