

# Optimizing Parametrized Quantum Circuits with Free-Axis Single-Qubit Gates

HIROSHI C. WATANABE\*

Quantum Computing Center, Keio University  
JST PRESTO  
hcwatanabe@keio.jp

RUDY RAYMOND\*

IBM Quantum, IBM Japan  
Quantum Computing Center, Keio University  
Dept. of Computer Science, The Univ. of Tokyo  
rudyhar@jp.ibm.com

YU-YA OHNISHI

Materials Informatics Initiative, RD Technology  
& Digital Transformation Center, JSR  
Corporation, JSR BiRD  
yuuya\_oonishi@jsr.co.jp

ERIKO KAMINISHI

Quantum Computing Center, Keio University  
JST PRESTO  
kaminishi@keio.jp

MICHIHIKO SUGAWARA

Quantum Computing Center, Keio University  
sugawara.a6@keio.jp

## Abstract

Variational quantum algorithms, which utilize Parametrized Quantum Circuits (PQCs), are promising tools to achieve quantum advantage for optimization problems on near-term quantum devices. Their PQCs have been conventionally constructed from parametrized rotational angles of single-qubit gates around predetermined set of axes, and two-qubit entangling gates, such as CNOT gates. We propose a method to construct a PQC by continuous parametrization of both the angles and the axes of its single-qubit rotation gates. 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 half rotation or  $\pi$ . 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 Hamiltonians for quantum chemistry and combinatorial optimization. Because free-axis selection allows designing PQCs without specifying their single-qubit rotational axes, it may significantly improve the handiness of PQCs.

## 1 Introduction

Parametrized quantum circuit (PQC) is one of the most essential components of hybrid quantum-classical algorithms on near-term quantum devices [1–3]. 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. Variational quantum algorithms optimize these parameters so that their cost function is minimized. For example, in variational quantum eigensolver (VQE) [4], which treats fermionic or spin Hamiltonian of physics and chemistry, the cost function is the energy and the PQC represents its wavefunction. PQC that gives the minimum energy corresponds to the wavefunction of the ground state according to the variational principle in VQE. Quantum Approximate Optimization Algorithm (QAOA) [5] is another example using PQC.

---

\*Equal contribution

Combinatorial optimization problems are targets of QAOA and the PQC generates a certain quantum state, for which the measurement yields a probability of each bit string encoding the solutions of the problems.

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 [6]. Highly expressible PQC can also be created in hardware efficient (or heuristic) ansatz [7–12] 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, 13–18]. 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.

In recent years, various approaches are being explored to alleviate the problems in the design of PQCs. Adapt-VQE [19] employs operator pool to achieve chemical accuracy with shallower circuit starting based on unitary coupled-cluster method. It has been extended to hardware efficient ansatz [11]. Theoretical evaluations of expressibility of PQC is another direction to lead a proper design of PQC [20, 21]. The difficulty in optimizing the PQC with hardware efficient ansatz could be solved by sophisticating the optimization methods. There are two main types of optimization methods that employ gradient-based or gradient-free methods. For gradient-based optimization methods, analytical calculation of partial derivative in PQC [22, 23] and stochastic gradient decent for VQE [24, 25] are proposed and widely used. Although naive gradient-based methods often suffers from barren plateau [1, 13–18], new approaches to avoid it, e.g., layerwise learning of quantum neural network [26], are still eagerly explored. Gradient-free optimizer, such as NFT [27], Rotosolve/select [28], and Jacobi-Anderson [29], is another option in classical optimization. It is usually more robust against statistical error.

Nevertheless, while the design of PQCs has been optimized by using rotations of qubit states around predetermined axes [27, 28], 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 [27], but other rotational gates (such as,  $R_x$  and  $R_z$ ) may be necessary for different types of Hamiltonians [28]. Moreover, such additional rotations may facilitate shortcuts to quickly find the ground states as also argued in [28].

In this work<sup>1</sup>, 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 [27–29]. 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  $\theta$  on a single-qubit gate of a PQC. It turns out that an optimal axis of rotation  $\hat{\mathbf{n}} \in \mathbb{R}^3$  satisfies a linear system of equations  $\mathbf{R}(\theta)\hat{\mathbf{n}} = \mathbf{b}(\theta)$ , where the elements of the matrix  $\mathbf{R}(\theta)$  and the vector  $\mathbf{b}(\theta)$  can be determined by running the slightly-modified PQC. Furthermore, the optimal axis is an eigenvector of  $\mathbf{R}(\pi)$  when the angle  $\theta$  is limited to one-half of full rotation. Surprisingly, even under such limitation, we show that PQCs with Fraxis have better expressibility than those with other structural optimization, and can find ground states relatively faster for several important Hamiltonians.

The benefits of PQC with Fraxis will be undoubtedly optimal when the underlying quantum devices natively support free-axis single-qubit gates. However, PQC with Fraxis still brings some important benefits even without such native implementability. Firstly, it frees users from specifying which set of rotational gates to choose from and thus lowers the barrier in designing PQCs on near-term quantum devices. Secondly, even though Fraxis gates are eventually decomposed into fixed-axis single-qubit gates on available quantum devices (we will show the results of running PQCs with Fraxis gates on IBM Quantum devices), our results imply that the optimization of their parameters with Fraxis is more efficient than corresponding techniques on fixed-axis single-qubit gates. This is not only because Fraxis has more degrees of freedom, but also because it considers correlations between parameters in the optimization. The decomposition of Fraxis gates may result in additional single-qubit gates, but because of the high

---

<sup>1</sup>Preliminary results were presented in [30].

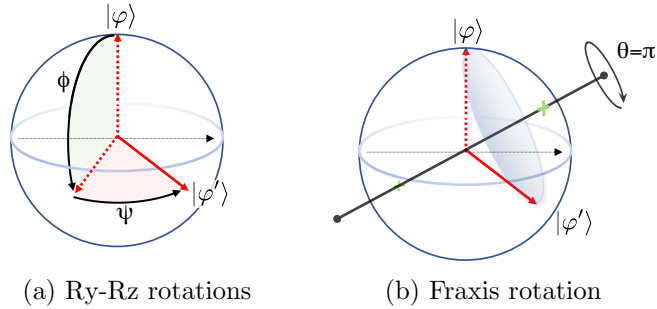


Figure 1: A comparison of obtaining a quantum state  $|\varphi'\rangle$  from (a) rotating  $|\varphi\rangle$  along the y-axis and then along the z-axis, i.e., applying a sequence of single-qubit gates  $R_y(\phi)$  and  $R_z(\psi)$ , and from (b) rotating  $|\varphi\rangle$  with a Fraxis gate.

quality of single-qubit gates on available quantum devices, this is not a problem. On the contrary, Fraxis improves the expressibility of PQCs without the need of adding layers of noisy two-qubit entangling gates. Thirdly, Fraxis gives a new formulation of sequential optimization for PQCs in the framework of matrix factorization that has opened paths for optimizing full degrees of freedom in single-qubit gates for quantum time simulation [31], and PQCs [32], albeit with more circuit evaluations.

The rest of the paper is arranged as follows. In Sec. 2, we describe theoretical foundation of our proposed methods by first giving motivating examples, and then deriving a system of linear equations that makes free-axis selection possible in an efficient manner. We then show the simplified version of Fraxis whose computational overhead is slightly better than known methods. In Sec. 3, we demonstrate numerical experiments showing the power of the proposed methods in comparison to known ones. Namely, Fraxis have better expressibility and they can be used to efficiently search for ground states of Hamiltonians for quantum chemistry and combinatorial optimization. Finally, the paper is concluded in Sec. 4 in which we also discuss possible future work extending the proposed methods.

## 2 Methods

We consider the problem of optimizing a PQC that consists of a sequence of parametrized unitary gates  $U = U_1 \dots U_D$  so that from a given initial state  $\rho$  the PQC generates a quantum state  $U\rho U^\dagger$  minimizing an objective function encoded in a Hamiltonian  $M$ . Namely, the value of  $\langle M \rangle \equiv \text{tr}(MU\rho U^\dagger)$  is minimized over all parameter 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 [28]. We then derive Fraxis in its general form, and present its simplified form that turns out to be computationally more efficient than the Rotoselect [28].

### 2.1 Motivating Examples

Here is an example why selecting axes of rotations can be better than fixing the axes and optimizing the continuous angles illustrated in Fig. 1. 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  $R_y R_z$  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  $\mathbf{r}_g = -\frac{1}{\sqrt{3}}(1, 1, 1)^T$ , and that of the initial state, which is  $\mathbf{r}_0 = (0, 0, 1)^T$ . 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  $\mathbf{r}_g$  from  $\mathbf{r}_0$  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 [28], need both  $R_y$  and  $R_z$  rotational gates, must order them correctly, and

tune the angles to obtain the ground state. See the left figure of Fig. 1.

On the other hand, the effect of two rotations is the same as that of a single rotation due to the consequence of Euler's rotation theorem. It is well known that single-qubit unitaries are rotations of the Bloch sphere [33], and each rotation can be represented by a real unit vector  $\hat{\mathbf{n}} = (n_x, n_y, n_z)^T \in \mathbb{R}^3$  and an angle  $\theta \in [0, 2\pi]$  so that the unitary is

$$R_{\hat{\mathbf{n}}}(\theta) = \cos(\theta/2)I - i \sin(\theta/2) (n_x X + n_y Y + n_z Z). \quad (1)$$

Notice that the  $R_y(\theta)$  and  $R_z(\theta)$  are instances of the above unitary.  $R_{\hat{\mathbf{n}}}(\theta)$  rotates its input state by  $\theta$  about the  $\hat{\mathbf{n}}$  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  $\hat{\mathbf{n}} = \left(\frac{1}{\sqrt{2}}\sqrt{\frac{1}{2} + \frac{1}{2\sqrt{3}}}, \frac{1}{\sqrt{2}}\sqrt{\frac{1}{2} + \frac{1}{2\sqrt{3}}}, -\sqrt{\frac{1}{2} - \frac{1}{2\sqrt{3}}}\right)^T$  results in the ground state. See the right figure of Fig. 1. Therefore, if we can optimize the axis of rotation of the corresponding unitary then we can obtain the ground state using only a single gate. Even if the quantum devices do not natively support free-axis gates, we can still obtain benefits in the formulation of circuit optimization with Fraxis. We will show and argue later that optimizing the axis of rotation indeed results in quantum circuits with better expressibility and convergence. In what follows, we first describe the steps to select axis of rotation in PQCs.

## 2.2 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 [28]. The quantum state  $\rho$  is generated by a PQC that consists of a sequence of unitary gates  $U = U_1 U_2 \dots U_D$  acting on  $n$  qubits so that each  $U_d$  is either a fixed two-qubit gate, or a parametrized single-qubit gate. 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  $2 \times 2$  identity matrix acting on other qubits. 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  $\theta_d$  written as:

$$\begin{aligned} U_d &\equiv e^{-i\frac{\theta_d}{2}\hat{\mathbf{n}}_d \cdot \vec{\sigma}} = \cos(\theta_d/2)I - i \sin(\theta_d/2)\hat{\mathbf{n}}_d \cdot \vec{\sigma} \\ &= R_{\hat{\mathbf{n}}_d}(\theta_d), \end{aligned}$$

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

We want to optimize the following expected value on the objective function over the choice of parameters of all single qubit gates, i.e.,  $\theta_d$  and  $\hat{\mathbf{n}}_d$  for  $d = 1, \dots, D$ , that we can gather as  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_D)$  and  $(\hat{\mathbf{n}}_1, \dots, \hat{\mathbf{n}}_D)$ . To further simplify the notation, similar to [28] we write the expectation value as

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

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

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

In the above, we utilize the following definitions

$$\begin{aligned} \hat{M} &\equiv U_{d+1}^\dagger \dots U_D^\dagger M U_D \dots U_{d+1}, \\ \hat{\rho} &\equiv U_{d-1} \dots U_1 \rho U_1^\dagger \dots U_{d-1}^\dagger. \end{aligned}$$

In the hereafter, we simply write  $\hat{M}$  and  $\hat{\rho}$  as  $M$  and  $\rho$ , 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  $\theta_d$  and  $\hat{\mathbf{n}}_d$  as below (see [28] for the proof).

$$\langle M \rangle_{\hat{\mathbf{n}}_d, \theta_d} = \cos^2(\theta_d/2) \text{tr}(M\rho) + i \sin(\theta_d/2) \cos(\theta_d/2) \text{tr}(M[\rho, H_d]) + \sin^2(\theta_d/2) \text{tr}(M H_d \rho H_d), \quad (2)$$

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

Now, we remark an important result of [27, 28]: given a fixed  $\hat{\mathbf{n}}_d$  the optimal  $\theta_d^*$  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  $\theta_d$ . 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 [28]), and for choosing  $(\hat{\mathbf{n}}_1, \dots, \hat{\mathbf{n}}_D) \in \{R_x, R_y, R_z\}^D$  where each of  $R_x, R_y$  and  $R_z$  corresponds to rotational axes  $\hat{\mathbf{n}}^{(x)} = (1, 0, 0)^T$ ,  $\hat{\mathbf{n}}^{(y)} = (0, 1, 0)^T$ , and  $\hat{\mathbf{n}}^{(z)} = (0, 0, 1)^T$ , respectively. In the next section we show how we can further extend the choices of rotational axes from all unit vectors  $\hat{\mathbf{n}}_d \in \mathbb{R}^3$ .

### 2.3 The Proposed Method: Free-Axis Selection

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

By the Lagrange multipliers, the objective function to optimize is

$$(\hat{\mathbf{n}}^*, \lambda^*) = \min_{\hat{\mathbf{n}}, \lambda} f(\hat{\mathbf{n}}, \lambda) \equiv \langle M \rangle_{\hat{\mathbf{n}}} - \lambda (n_x^2 + n_y^2 + n_z^2 - 1), \quad (3)$$

where  $\lambda$  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

$$\begin{aligned} \frac{\partial f}{\partial n_x} &= \alpha_\theta \text{tr}(M[\rho, X]) \\ &+ \sin^2\left(\frac{\theta}{2}\right) (2n_x \text{tr}(MX\rho X) + n_y \text{tr}(MX\rho Y + MY\rho X) + n_z \text{tr}(MX\rho Z + MZ\rho X)) - 2\lambda n_x, \end{aligned} \quad (4)$$

$$\begin{aligned} \frac{\partial f}{\partial n_y} &= \alpha_\theta \text{tr}(M[\rho, Y]) \\ &+ \sin^2\left(\frac{\theta}{2}\right) (2n_y \text{tr}(MY\rho Y) + n_x \text{tr}(MX\rho Y + MY\rho X) + n_z \text{tr}(MY\rho Z + MZ\rho Y)) - 2\lambda n_y, \end{aligned} \quad (5)$$

$$\begin{aligned} \frac{\partial f}{\partial n_z} &= \alpha_\theta \text{tr}(M[\rho, Z]) \\ &+ \sin^2\left(\frac{\theta}{2}\right) (2n_z \text{tr}(MZ\rho Z) + n_x \text{tr}(MX\rho Z + MZ\rho X) + n_y \text{tr}(MY\rho Z + MZ\rho Y)) - 2\lambda n_z. \end{aligned} \quad (6)$$

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

$$\frac{\partial f}{\partial n_x} = \frac{\partial f}{\partial n_y} = \frac{\partial f}{\partial n_z} = 0. \quad (7)$$

Defining  $r_x, r_y, r_z, r_{(x+y)}, r_{(x+z)}$  and  $r_{(y+z)}$  as in (9)–(14), and noticing the following identities:

$$\begin{aligned} \text{tr}(MX\rho Y + MY\rho X) &= 2r_{(x+y)} - r_x - r_y, \\ \text{tr}(MX\rho Z + MZ\rho X) &= 2r_{(x+z)} - r_x - r_z, \\ \text{tr}(MY\rho Z + MZ\rho Y) &= 2r_{(y+z)} - r_y - r_z, \end{aligned}$$

we can arrange (4)–(6) and (7) to obtain the following equation.

$$\left[ \sin^2\left(\frac{\theta}{2}\right) \begin{pmatrix} 2r_x & 2r_{(x+y)} - r_x - r_y & 2r_{(x+z)} - r_x - r_z \\ 2r_{(x+y)} - r_x - r_y & 2r_y & 2r_{(y+z)} - r_y - r_z \\ 2r_{(x+z)} - r_x - r_z & 2r_{(y+z)} - r_y - r_z & 2r_z \end{pmatrix} - 2\lambda^* \mathbf{I} \right] \begin{pmatrix} n_x^* \\ n_y^* \\ n_z^* \end{pmatrix} = -\alpha_\theta \begin{pmatrix} \text{tr}(M[\rho, X]) \\ \text{tr}(M[\rho, Y]) \\ \text{tr}(M[\rho, Z]) \end{pmatrix} \quad (8)$$

In (8), we use the following definitions which are the expectation values obtained 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.

$$r_x \equiv \text{tr}(MX\rho X), \quad (9)$$

$$r_y \equiv \text{tr}(MY\rho Y), \quad (10)$$

$$r_z \equiv \text{tr}(MZ\rho Z), \quad (11)$$

$$r_{(x+y)} \equiv \text{tr}\left(M\left(\frac{X+Y}{\sqrt{2}}\right)\rho\left(\frac{X+Y}{\sqrt{2}}\right)\right), \quad (12)$$

$$r_{(x+z)} \equiv \text{tr}\left(M\left(\frac{X+Z}{\sqrt{2}}\right)\rho\left(\frac{X+Z}{\sqrt{2}}\right)\right), \quad (13)$$

$$r_{(y+z)} \equiv \text{tr}\left(M\left(\frac{Y+Z}{\sqrt{2}}\right)\rho\left(\frac{Y+Z}{\sqrt{2}}\right)\right). \quad (14)$$

Equation (8) can be simply written as:

$$(\sin^2(\theta/2)\mathbf{R} - 2\lambda^*\mathbf{I})\hat{\mathbf{n}}^* = -\alpha_\theta\mathbf{b}, \quad (15)$$

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

$$\mathbf{R} \equiv 2 \begin{pmatrix} r_x & r_{(x+y)} - \frac{r_x+r_y}{2} & r_{(x+z)} - \frac{r_x+r_z}{2} \\ r_{(x+y)} - \frac{r_x+r_y}{2} & r_y & r_{(y+z)} - \frac{r_y+r_z}{2} \\ r_{(x+z)} - \frac{r_x+r_z}{2} & r_{(y+z)} - \frac{r_y+r_z}{2} & r_z \end{pmatrix} \quad (16)$$

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 (15) has a unique solution. In fact, the value of  $\lambda^*$  is the solution to the following identity:

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

Equation (15) gives a method to determine an optimal axis of rotation against a fixed  $\theta$  by estimating the matrix  $\mathbf{R}$  and vector  $\mathbf{b}$ . We notice that we can further simplify the equation when  $\theta = \pi$  that corresponds to replacing the single-qubit gate with a  $\pi$ -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.

## 2.4 Simplified Free-Axis Selection

We have derived a method to select a rotational axis minimizing the objective function by fixing  $\theta$ . We refer to it as  $\theta$ -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  $\hat{\mathbf{n}}^*$  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  $\pi$ -Fraxis in Algorithm 1. The stopping criterion can be either reaching the maximum of the number of repetitions or the lack of progress in reaching lower energy values. In the hereafter, for ease of notation we simply call it Fraxis and omit the prefix  $\pi$ -.

We note an important and nice property of the Fraxis: its computational resources turn out to be less than those of Rotoselect [28] 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  $\mathbf{R}_d$ . The landscape of  $\langle M \rangle_{\hat{\mathbf{n}}_d, \pi}$  as in ((2)) is completely determined by the eigenvalues and eigenvectors of  $\mathbf{R}_d$ . We record this property in the following lemma.

---

**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.,  $\hat{\mathbf{n}}_d \in \mathbb{R}^3$  such that  $|\hat{\mathbf{n}}_d| = 1$  for  $d = 1, 2, \dots, D$ .

**repeat**

**for**  $d = 1, 2, \dots, D$  **do**

    Fix all axes  $\hat{\mathbf{n}}_j$  for  $j \neq d$

    Compute elements of  $\mathbf{R}_d$  as in (9)–(14)

    Compute the eigenvectors  $\mathbf{r}_d^{(1)}, \mathbf{r}_d^{(2)}, \mathbf{r}_d^{(3)}$  of  $\mathbf{R}_d$

$\hat{\mathbf{n}}_d \leftarrow \arg \min_{\mathbf{r} \in \{\mathbf{r}_d^{(1)}, \mathbf{r}_d^{(2)}, \mathbf{r}_d^{(3)}\}} \langle M \rangle_{\mathbf{r}, \pi}$

**until** satisfying stopping criterion

**Output** Axes of rotation  $(\hat{\mathbf{n}}_1, \dots, \hat{\mathbf{n}}_D)$

---

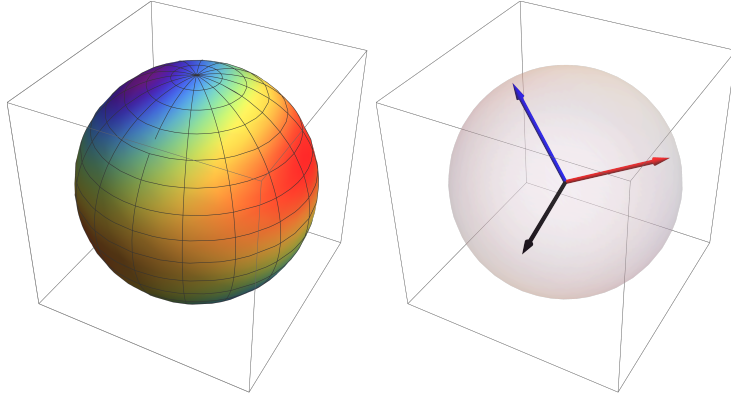


Figure 2: (left) An illustration of the energy landscape of Fraxis determined by the matrix  $\mathbf{R}_d$ . Red, green and blue areas correspond to high, medium, low energy values. (right) Three eigenvectors corresponding to minimum (blue), maximum (red) and saddle (black) points.

**Lemma 1.** Let  $\lambda_d^{(i)}$  and  $\mathbf{r}_d^{(i)}$  be, respectively, the  $i$ -th eigenvalue and eigenvector of  $\mathbf{R}_d$  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  $\mathbf{r} \in \mathbb{R}^3$  it holds that  $\lambda_d^{(1)}/2 \leq \langle M \rangle_{\mathbf{r}, \pi} \leq \lambda_d^{(3)}/2$ , and therefore  $\hat{\mathbf{n}}_d = \mathbf{r}_d^{(1)}$ .

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

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

We remark that the cost of energy estimations can be significantly reduced for special Hamiltonians (e.g.,  $k$ -local Hamiltonians [34–36]) if we use locally-biased classical shadows [37]. 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 will see later those drawbacks are not significant.

We also remark that conventional stochastic and gradient approaches may still be compelling for the

global optimization in certain cases, such as, the parallelization of variational quantum eigensolvers [3]. Fraxis can easily be combined with those approaches. For instance, we can analytically evaluate the gradient of the cost functions using (4) – (6). The obtained gradient is then passed to the conventional gradient-based optimizers. For this reason, unlike Rotosolve that must update the gates sequentially, Fraxis can be used in the parallelization of variational quantum algorithms, and therefore is potential for avoiding the drawbacks of Rotosolve as mentioned in [3]. 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.

Two  $\pi$ -Fraxis gates can realize an arbitrary rotation gate shown in (1) as in the lemma below. This is tight since the identity gate requires two  $\pi$ -Fraxis gates.

**Lemma 2.** *Let an arbitrary single-qubit gate as in (1) with parameters  $(\hat{\mathbf{n}}, \theta)$ . Then,  $R_{\hat{\mathbf{n}}}(\theta) = R_{\hat{\mathbf{n}}_1}(\pi)R_{\hat{\mathbf{n}}_2}(\pi)$  provided  $\hat{\mathbf{n}}_1$  and  $\hat{\mathbf{n}}_2$  satisfy  $\hat{\mathbf{n}} = \hat{\mathbf{n}}_1 \times \hat{\mathbf{n}}_2$  and  $\theta = 2 \arccos(-\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2)$ , where  $\times$  and  $\cdot$  denote, respectively, the outer and inner products.*

*Proof.* By (1), we can easily see that

$$R_{\hat{\mathbf{n}}_i}(\pi) = -i(n_{i,x}X + n_{i,y}Y + n_{i,z}Z) \equiv -i\hat{\mathbf{n}}_i \cdot \hat{\sigma},$$

where  $\hat{\sigma} \equiv (X, Y, Z)$  is a vector of Pauli matrices  $X, Y$  and  $Z$ . Utilizing the above identity, we can confirm that

$$\begin{aligned} R_{\hat{\mathbf{n}}_1}(\pi)R_{\hat{\mathbf{n}}_2}(\pi) &= -(\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2)I - i((\hat{\mathbf{n}}_1 \times \hat{\mathbf{n}}_2) \cdot \hat{\sigma}) \\ &= R_{\hat{\mathbf{n}}_1 \times \hat{\mathbf{n}}_2}(2 \arccos(-\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2)) \end{aligned}$$

□

We end this section with an important note on the implementability of  $\pi$ -Fraxis gates on IBM Quantum devices that do not natively support Fraxis gates. The  $\pi$ -Fraxis gate is decomposed into native single-qubit gates as follows.

$$R_{\hat{\mathbf{n}}}(\pi) = R_z(\pi + \phi) \cdot \sqrt{X} \cdot R_z(\pi + \theta) \cdot \sqrt{X} \cdot R_z(\pi - \phi), \quad (17)$$

where  $\theta = 2 \arctan\left(\sqrt{n_x^2 + n_y^2}/n_z\right)$  and  $\phi = \arctan(n_y/n_x)$ . Because the  $R_z$  gates on IBM Quantum devices do not require *pulse* operation [38], the  $\pi$ -Fraxis gate can be realized with the actual cost of two  $\sqrt{X}$  gates, which is essentially the same as the  $R_yR_z$  gate. We can also notice from the decomposition that the  $\pi$ -Fraxis gate has two degrees of freedom, which is also the same as the  $R_yR_z$  gate. Optimizing the  $R_yR_z$  gate with Rotosolve [27, 28] proceeds with sequentially optimizing the parameters of  $R_y$  and  $R_z$  one at a time, while the  $\pi$ -Fraxis gate optimizes the two parameters simultaneously. There is a discussion on how to optimize correlated parameters with Rotosolve but, unlike the  $\pi$ -Fraxis, there is no known closed-form solution [27].

### 3 Numerical Results

In this section, we first apply  $\theta$ -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  $\pi$ -Fraxis. Next, using Qiskit [39] we demonstrate the high expressibility of PQC with the simplified  $\pi$ -Fraxis (or, Fraxis) in contrast to PQC with its traditional counterparts. For studying expressibility, we show the one-qubit PQC with Fraxis can generate quantum states covering the Bloch sphere more uniformly than its counterparts. We then show evidences of better expressibility of multi-qubit PQC with Fraxis when measured in KL divergences under different number of entangled layers. Finally, we present numerical experiments on simulators and near-term quantum devices showing Fraxis can further improve the efficacy of variational quantum circuits for the Heisenberg model, molecular, and combinatorial optimization Hamiltonians.



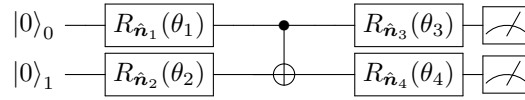


Figure 3: A 2-qubit PQC with  $\theta$ -Fraxis.

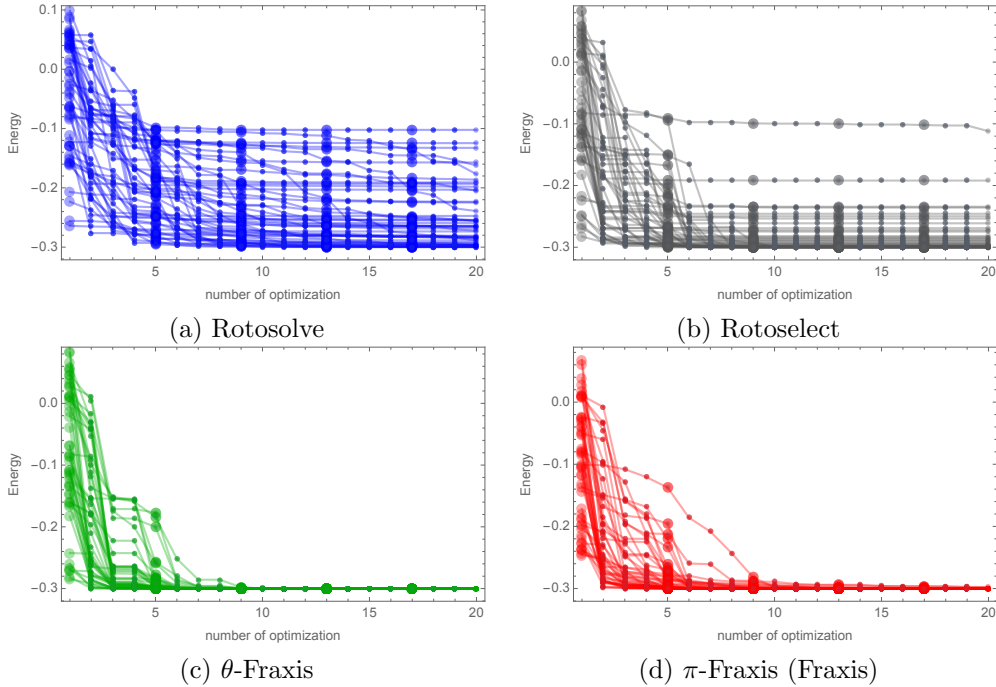


Figure 4: Comparison of energy convergence as a function of number of iterations for optimization. (a) Rotosolve (blue), (b) Rotoselect (gray), (c)  $\theta$ -Fraxis (green), and (d) Fraxis (red). Common initial angular parameter set is used for Rotosolve, Rotoselect and  $\theta$ -Fraxis, while  $(\theta_1, \theta_2, \theta_3, \theta_4)$  are overridden to  $(\pi, \pi, \pi, \pi)$  in Fraxis trials.

### 3.1 Comparison with Other Structural Optimization

For comparison, we tested Rotosolve, Rotoselect,  $\theta$ -Fraxis, and Fraxis methods to a 2-qubit PQC system in Fig. 3 for finding the ground state of the Hamiltonian

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

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,  $(\hat{n}_1, \hat{n}_2, \hat{n}_3, \hat{n}_4)$ . We applied the methods to start from the same 50 random initial parameter sets.

Figure 4 shows the optimized energy as a function of iterative number of optimization. One can see that  $\theta$ -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,  $\theta$ -Fraxis makes it possible to optimize rotation axis continuously with respect to given  $\theta$ , which results in fine convergence nature as shown in Fig. 4. We also applied the simplified Fraxis with  $\theta = \pi$  (or,  $\pi$ -Fraxis) to the same problem. Note that the initial parameters for axes are the same in Rotosolve, Rotoselect and  $\theta$ -Fraxis cases, while rotation angles  $(\theta_1, \theta_2, \theta_3, \theta_4)$  are set to

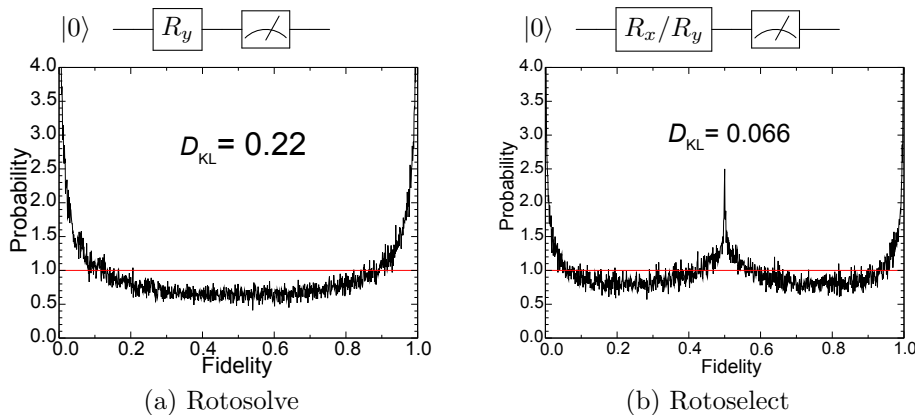


Figure 5: Fidelity distribution obtained by single-qubit circuits with (a) Rotosolve( $R_y$ ), and (b) Rotoselect( $R_x/R_y$ ). Horizontal lines in red represent Haar-uniform distribution. The evaluated KL divergence values are displayed over the respective plots

$(\pi, \pi, \pi, \pi)$ . The result in Fig. 4 shows that the convergence behaviour of  $\pi$ -Fraxis is still better than those of Rotosolve/select. We also remark that  $\pi$ -Fraxis has comparable performance to  $\theta$ -Fraxis. Nevertheless, estimation cost for matrix-elements in  $\pi$ -Fraxis is reduced from 9 to 6, which suggests  $\pi$ -Fraxis is more practical than  $\theta$ -Fraxis. Recall that we abbreviate  $\pi$ -Fraxis as Fraxis and will further investigate expressibility of Fraxis ansatz/circuit to ensure the usability of this simplified method in the next subsection.

### 3.2 Circuit Expressibility

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

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

In the case of Fraxis, the rotation angle is fixed at  $\theta = \pi$  in (1), which is equivalent the constraint condition  $\phi + \lambda = \pi$  in (19). Hence, a single gate controlled by Fraxis is not universal any more. 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  $\pi$  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. Moreover, it can also be shown that a universal 1-qubit gate in (19) can be realized by at most 2 Fraxis gates, and therefore a hard instance of general quantum circuit can be turned into a hard instance of Fraxis ansatz whose size is in the same order of the general quantum circuit.

To quantitatively evaluate the expressibility of the Fraxis circuit in comparison with Rotosolve and Rotoselect, we evaluate Kullback-Leibler (KL) divergence as

$$\begin{aligned} I\mathcal{C}\mathcal{E}(C) &= D_{\text{KL}}(P(C, F) || P_{\text{Haar}}(F)) \\ &= \int_0^1 P(C, F) \log \frac{P(C, F)}{P_{\text{Haar}}(F)} dF, \end{aligned} \quad (20)$$

where  $F$  is the fidelity between two parametrized 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  $C$ .  $P_{\text{Haar}}(F)$  is a probability distribution sampled uniformly according to the Haar measure, in which  $|\psi\rangle$  is uniformly distributed in Hilbert space. According to [40] the Haar-measure distribution is derived as,  $P_{\text{Haar}}(F) = (N-1)(1-F)^{N-2}$ , where  $N$  is the dimension of Hilbert space. Here, a product state of  $\otimes^n |0\rangle$  is employed as the initial state.

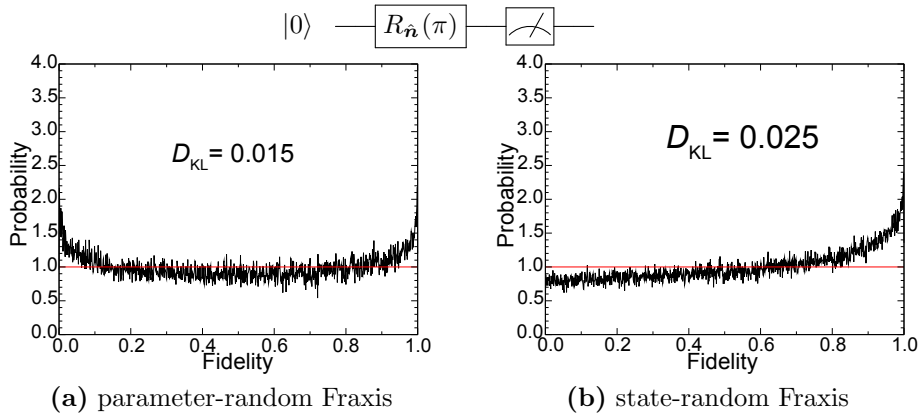


Figure 6: Fidelity distribution obtained by single-qubit circuits with (a) parameter-random Fraxis, and (b) state-random Fraxis. Horizontal lines in red represent Haar-uniform distribution. The evaluated KL divergence values are displayed over the respective plots.

First, we confirmed the expressibility of single-qubit circuits as shown in Figs. 5 and 6, 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  $R_y(\theta)$  gate. Then, the rotation angles  $\theta$  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  $\theta$  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

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

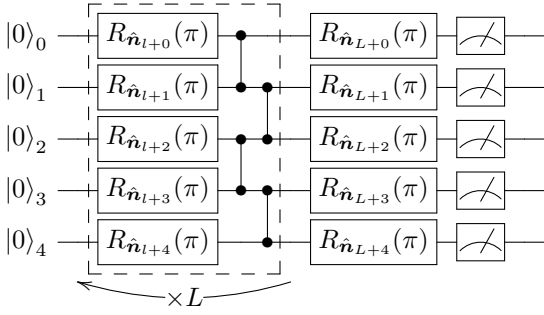
Substituting (21) into (20) gives theoretical value of  $D_{KL} = \log(4/\pi) = 0.24$  for Rotosolve with  $R_y$ -ansatz. As mentioned in the previous study [20], we confirmed the fact that the KL divergence is largely sensitive to the bin width of the pdf histogram, and thus the proper width is not 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 5), 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.

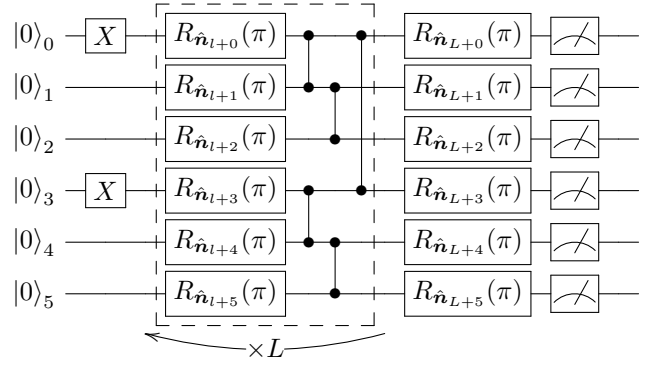
In the Fraxis circuits, while the rotation angles are fixed  $\theta = \pi$ , the rotation axis  $\hat{n} = (n_x, n_y, n_z)^T$  are generated randomly. Since the Fraxis ansatz optimizes the target gate in a deterministic way without systematic search in parameter space, it is not straightforward to compare the expressibility with conventional parametrized 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  $\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^T$  and  $\theta$  and  $\phi$  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



(a) 5-qubit circuit A



(b) 6-qubit circuit B

Figure 7: PQC with Fraxis-ansatz

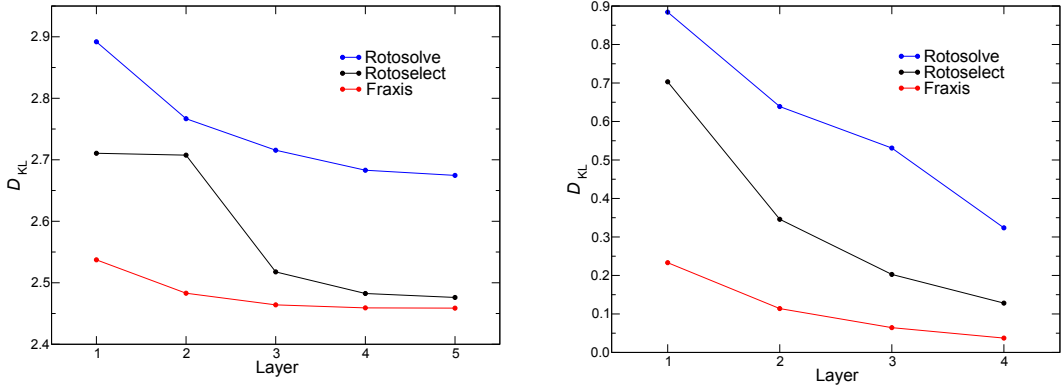


Figure 8: Relation between KL divergences (lower is better) and number of layers for circuits A (left) and B (right)

sphere of the Bloch sphere without loss of generality because of the symmetry of  $\pi$  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. 6 In the parameter random, the rotation axis  $\hat{n}$  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 6. 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. 7 consisting of 5 and 6 qubits respectively. Fig. 8 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.

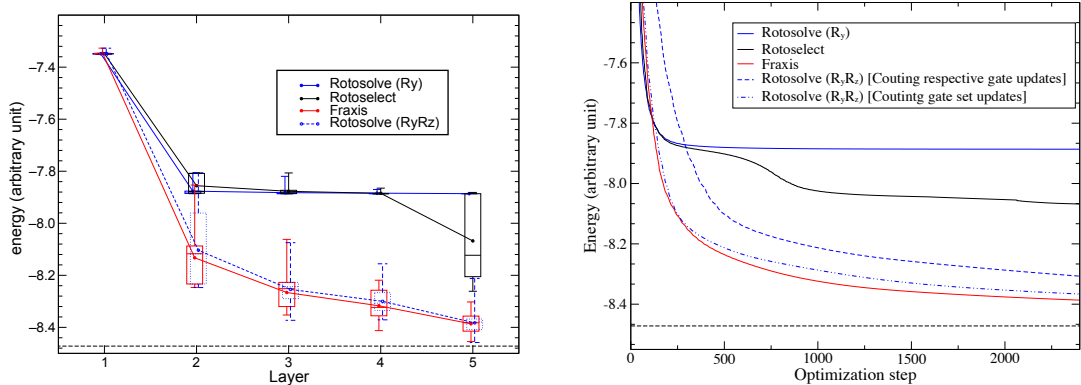


Figure 9: (Left) Relation between number of layers for circuit A and the optimized energy after 100 iterations for Rotosolve, Rotoselect, and Fraxis. The averaged energy is represented by filled circles and lines. The boxplots represent respective quantiles. The exact ground state energy is displayed with dash line. (Right) The achieved energy against the number of optimization steps for Rotosolve, Rotoselect, and Fraxis. At both subfigures, the Rotosolve is also applied to  $R_y R_z$  gates whose degrees of freedom are the same as Fraxis.

### 3.3 Heisenberg Model

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

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

where  $J = h = 1$ . We employed circuit A in Fig. 7, in Rotosolve, Rotoselect and Fraxis ansatz, respectively. In Rotosolve, all gates were represented by  $R_y$ . 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 9 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. 8, the obtained energy are consistent among all ansatze for  $L = 1$ . In the multi-layer calculation with the Rotosolve ansatz using  $R_y$  gates, the energy were not improved except for at  $L = 2$ . It is probably because the ansatz consisting of only  $R_y$  gates cannot cover the ground state of this model. For this reason, we also run the Rotosolve ansatz using  $R_y R_z$  gates and confirm that the obtained energies were much improved but still slightly inferior to Fraxis as shown in both subfigures. The figures present empirical evidence that Fraxis is still better than the Rotosolve having the same degree of freedom with  $R_y R_z$  gates.

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. 8. 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.

### 3.4 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 con-

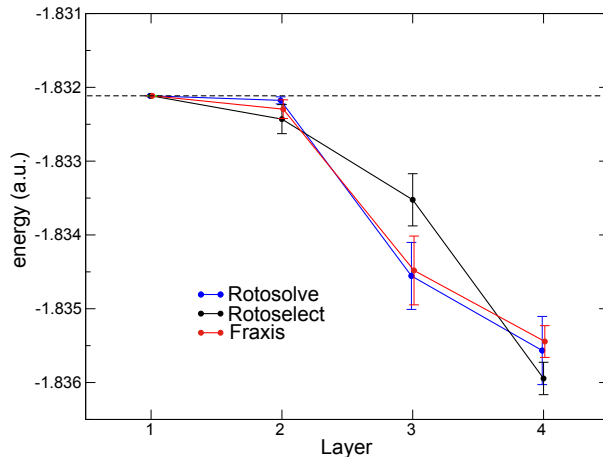


Figure 10: The average of optimized energy after 100 iterations with standard error. The Hartree-Fock state energy is displayed with dash line.

struction 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. 3.2.

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. 10. 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 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 temporarily 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 expressibility 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 help narrowing the search space.

### 3.5 Hamiltonian for MaxCut

We want to show the applicability of the Fraxis gates when the optimal quantum state is not spanned by quantum states with real amplitudes, and when its approximation is sufficient (i.e., its exact quantum state is not required). A quantum relaxation for combinatorial optimization in [41] is one of the methods that rely on finding such quantum states to solve MaxCut problems. Given a graph  $G = (V, E)$  with  $|V|$  vertices and  $|E|$  edges, the MaxCut problem, which is NP-hard, asks for assigning each vertex in the

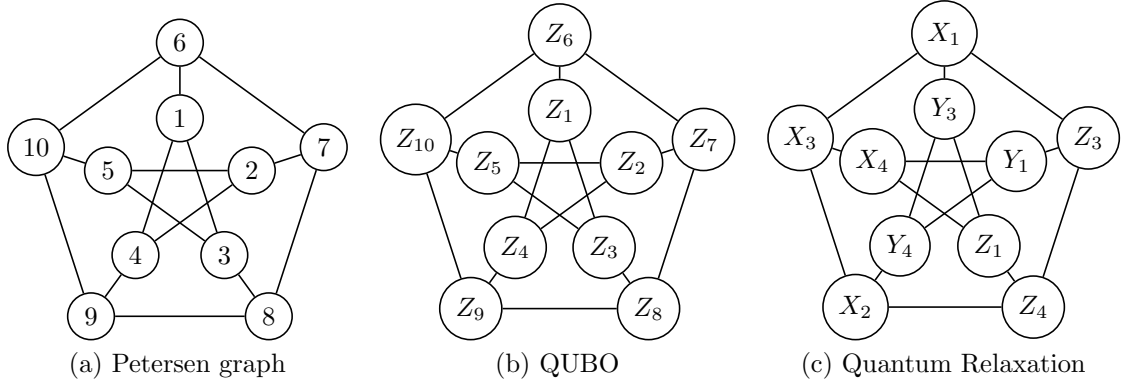


Figure 11: The Petersen graph (Left), the labelling of its vertices to construct the Hamiltonian for QUBO (Center), and the corresponding labelling to construct the Hamiltonian for Quantum Relaxation (Right) whose optimal quantum state can be highly entangled with non real-valued amplitudes.

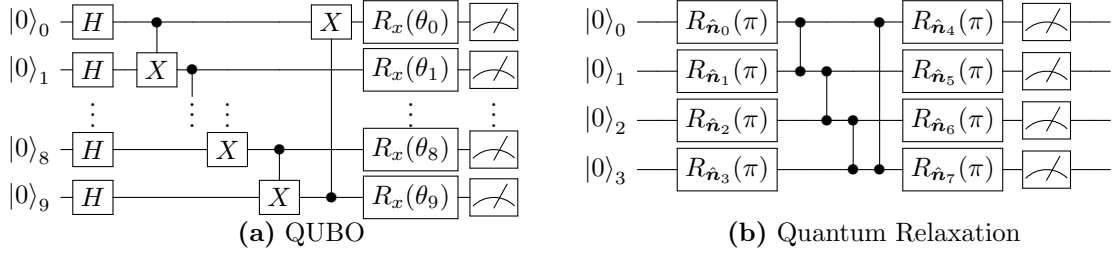


Figure 12: PQC for QUBO and Quantum Relaxation of the Petersen graph

graph to either 1 or  $-1$  so that the sum of edges connecting vertices with different values is maximized. Namely, it finds a configuration  $m \in \{-1, 1\}^{|V|}$  that maximizes a cost function  $\text{cut}(m)$ , where

$$\text{cut}(m) \equiv \sum_{(i,j) \in E} \frac{1}{2}(1 - m_i m_j). \quad (23)$$

Figure 11 shows a particular example of MaxCut instances considered in [41]. The left subfigure shows the Petersen graph that consists of 10 nodes and 15 edges. The center subfigure illustrates the Quadratic Unconstrained Binary Optimization (QUBO) formulation of the MaxCut problem of the Petersen graph, where the Hamiltonian of the QUBO formulation is defined as

$$H_{\text{QUBO}} \equiv \sum_{(i,j) \in E} \frac{1}{2}(1 - Z_i Z_j), \quad (24)$$

where  $Z_i Z_j$  denotes the tensor product of Pauli matrices  $Z$  at the  $i$ -th and  $j$ -th qubit, and  $I$  at the rest of the qubits. Popular quantum optimization methods, such as, QAOA [5], utilize the QUBO formulation to encode a binary variable into a qubit. Namely, the QUBO formulates a 10-qubit Hamiltonian to solve the MaxCut problem of the Petersen graph to obtain an optimal 10-qubit quantum state. On the other hand, the Quantum Relaxation utilizes an encoding that can map three binary variables into a qubit. We refer the readers to [41] for more details on the mapping and the resulting Hamiltonian. For the Petersen graph, the Hamiltonian of the Quantum Relaxation is given as

$$H_{\text{Relax}} \equiv \sum_{(i,j) \in E} \frac{1}{2}(1 - 3 W_i W_j), \quad (25)$$

where  $W_i \in \{X_1, Y_1, Z_1, \dots, Z_4\}$  is the Pauli matrix that can be inferred from the label of the vertex as shown in the center subfigure of Fig. 11. Here, we can observe that the  $H_{\text{QUBO}}$  is a diagonal matrix

(because it consists of the weighted sum of tensor products of Pauli  $I$  and  $Z$ ), while the  $H_{\text{Relax}}$  is not (because it consists of the weighted sum of Pauli  $I, X, Y$  and  $Z$ ). Consequently, it is easy to see that the optimal quantum state of the former is a computational basis, while that of the latter is not necessarily so. The striking difference is in the size of the Hamiltonian:  $2^{10} \times 2^{10}$  for the former and  $2^4 \times 2^4$  for the latter. Because of its smaller Hamiltonian, the Quantum Relaxation requires smaller PQCs to find an optimal quantum state of the underlying problem.

One of major obstacles is how to recover the binary solution from the optimal quantum state  $\rho^*$ , which is unlike that of QUBO, can be highly entangled in non computational basis. A simple *magic* rounding method that maps  $\rho^*$  into a random quantum state which is a product of a one-qubit state in the form of  $\frac{1}{2} \left( I + \frac{1}{\sqrt{3}} (\pm X \pm Y \pm Z) \right)$  (such states are called *magic states* and hence the method name) is shown to guarantee the solution of MaxCut whose cut value is 5/9 of the optimal one. Moreover, such guarantee holds not only for the optimal  $\rho^*$ , but also for any quantum state  $\rho_{\text{relax}}$  satisfying  $\text{Tr}(H_{\text{Relax}}\rho_{\text{relax}}) \geq \text{cut}(m)$ . Namely, it suffices to find a quantum state whose relaxed Hamiltonian value exceeds that of the optimal binary solution. The quality of the rounding from such quantum state is determined by the Hamiltonian value: the higher it is the better. Notice that although the approximation guarantee 5/9 is lower than the best of classical one [42], in practice the MaxCut solution of Quantum Relaxation can be much better. See [41] for more details.

We provide evidences that PQCs with Fraxis gates can obtain better quantum states for Hamiltonians of QUBO and Quantum Relaxation. We followed the construction of Hamiltonians as in [41] for the Petersen graph, as shown in Fig. 11 and test PQCs with single-qubit parametrized gates, as shown in Fig. 12, to obtain  $\rho_{\text{relax}}$ . The PQCs on the left of Fig. 12 were used for QUBO Hamiltonian with the Hadamard gates in the first layer of single-qubit gates so that the initial states is the superposition of all possible cuts as suggested in [5]. The entangling gates are fixed to cyclic CX gates with single-qubit gates varied from fixed axes to Fraxis (the figure depicts the fixed axes ones). The PQCs on the right were for the Quantum Relaxation Hamiltonian with cyclic CZ gates as entanglers and the single-qubit gates varied similarly.

First, we carried out 20 independent optimizations from  $H^{\otimes 10} |0\rangle$  for the QUBO Hamiltonian using PQCs with fixed-axis single-qubit gates optimized with Rotosolve and Rotoselect against those with Fraxis. The top figure in Fig. 13 shows the plots of the expected cut values against the number of parameter updates of the single-qubit gates of the PQCs. We observe that fixing the single-qubit gates to  $R_y$  or  $R_x$  and updating their parameters with Rotosolve fails as the cut values do not improve from the initial value. This is surprising as PQCs with  $R_y$  or  $R_x$  gates are often suggested, such as in [43], for the Hamiltonian whose ground states are of real-valued amplitudes. On the other hand, Rotoselect and Fraxis successfully find the optimal cuts after two sweeps of parameter updates despite their larger search space of complex-valued amplitudes.

Next, we carried out 20 independent optimizations from different initial conditions using the ansatz with cyclic entangler to find the ground state of the Quantum Relaxation Hamiltonian using PQCs with Rotosolve, Rotoselect and Fraxis. Unlike the QUBO and the quantum chemistry Hamiltonians, the eigenstate of the Quantum Relaxation Hamiltonian cannot be generated by the  $R_y$  ansatz alone, and thus,  $R_y$  ansatz may be unfairly disadvantageous. Hence, we independently carried out two types of Rotosolve optimizations with both  $R_y$  and  $R_y R_z$  ansätze. We confirm that Rotosolve on PQCs with  $R_y R_z$  ansatz are better than that on PQCs with only  $R_y$  ansatz, or even with fixed-axis single-qubit gates optimized with Rotoselect (namely, parametrized gates chosen from  $R_x, R_y$ , and  $R_z$  rotations). Here, we can see that PQCs with Fraxis obtain better cut values when compared to the rest of PQCs. Hence, freeing the design of PQCs on the choices of single-qubit gates is not only convenient but also can result in better optimization values.

### 3.6 Experimental Evaluation on Near-Term Quantum Devices

We compared the performance of Rotosolve and Fraxis using two types of 2-qubit Hamiltonian on the 5-qubit `ibmq_manila` quantum device of the IBM Quantum Systems. The experimental evaluations were performed on the device from May 22 to May 24, 2021 by *fairshare* run mode. The employed PQCs were homologous of that in Fig. 3 but all  $\theta$ -Fraxis gates were replaced by  $\pi$ -Fraxis gates.

First, we calculated the 2-qubit Heisenberg model which is given in (18). Figure 14 shows four



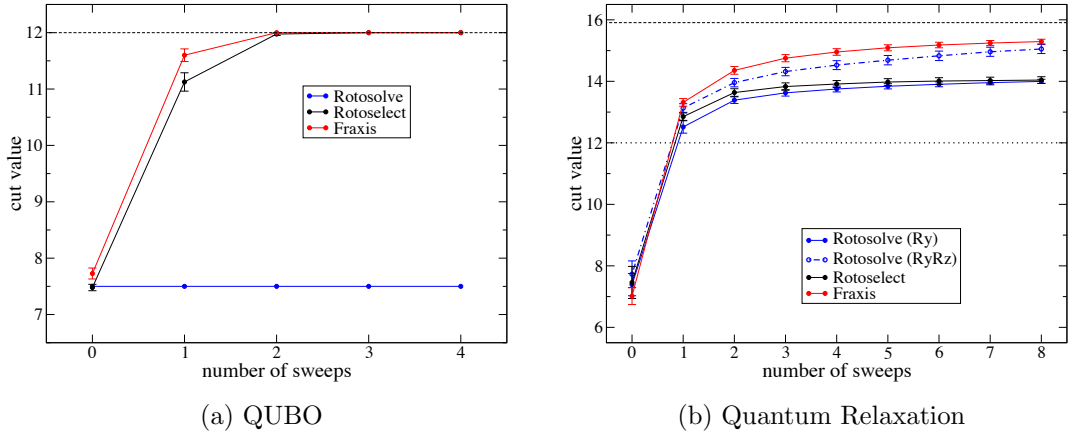


Figure 13: The cut values (higher is better) of quantum states against the number of sweeps updating parameters of PQC with different single-qubit gates for QUBO (Top) and Quantum Relaxation (Bottom) averaged over 20 independent trajectories for the Maxcut on the Petersen graph. The dash and dotted lines represent the optimal quantum relaxed value and the optimal binary solution, resp. For both QUBO and Quantum Relaxation, PQC with Fraxis gates find better quantum states.

energy trajectories for Rotosolve and Fraxis, respectively, where the initial parameter sets were randomly generated. Here, the energy expectation values were evaluated from 8192 measurements (or, shots) with readout error mitigation. Although, according to the simulator result in Fig. 4, Rotosolve can be trapped in local minima, it seemed to converge to the ground state in all four trials. However, its convergence seemed to largely vary among the trials. In contrast, Fraxis showed better convergence (e.g., less number of optimization steps to reach the ground state in all four trials) compared to Rotosolve in consistency with the simulator results in Sec. 3.1.

Next, we also benchmarked the  $H_2$  molecular Hamiltonian with STO-3G basis set as shown in Fig. 15. In an ideal case, the energy should monotonically decrease in both Rotosolve and Fraxis in course of optimization step as shown with the simulator in Sec. 3.1, although it was not the case. Such unwarranted behavior did not vary between Rotosolve and Fraxis. However, it seemed to be more distinct with  $H_2$  Hamiltonian rather than Heisenberg model, which may reflect difference in number of Pauli terms and corresponding accumulated errors. As a result, it is difficult to obtain the converged value of  $H_2$  under Hartree-Fock energy. Thus, for effective utilization of NISQ in chemistry, it may be necessary to utilize other error correction such as state tomography as demonstrated in literature [44]. Since PQC consisting of  $R_y$  gates are sufficient to express the ground state, as discussed before, Rotosolve may be more advantageous in converging efficiency because of its confined search space. Nevertheless we did not find any distinct difference between Rotosolve and Fraxis. In fact, one of the four trials of Rotosolve seemed to be trapped in the Hartree-Fock energy up to 20 optimization steps, while all four trials of Fraxis could find energies lower than the Hartree-Fock energy within the same steps. Although the search space of Fraxis is not confined within the real-valued probability amplitudes, Fraxis can still find the ground state effectively.

Overall, we can observe from Figs. 14 and 15 that on 2-qubit systems the performance of Fraxis (and Rotosolve) with real quantum devices is similar to that with noiseless simulators. This hints the potential usefulness of Fraxis-ansatz for a variety of Hamiltonians with near-term quantum devices.

## 4 Conclusion

In this paper, we proposed a method of free-axis selection optimizing rotational axes of single-qubit gates in PQC. The Fraxis ansatz can achieve high expressibility even with shallow PQC. We confirmed the expressibilities estimated with KL divergence for both single- and multi-qubit PQC with Fraxis are significantly higher when compared to optimizing the rotation angles. Nevertheless, only six energy esti-

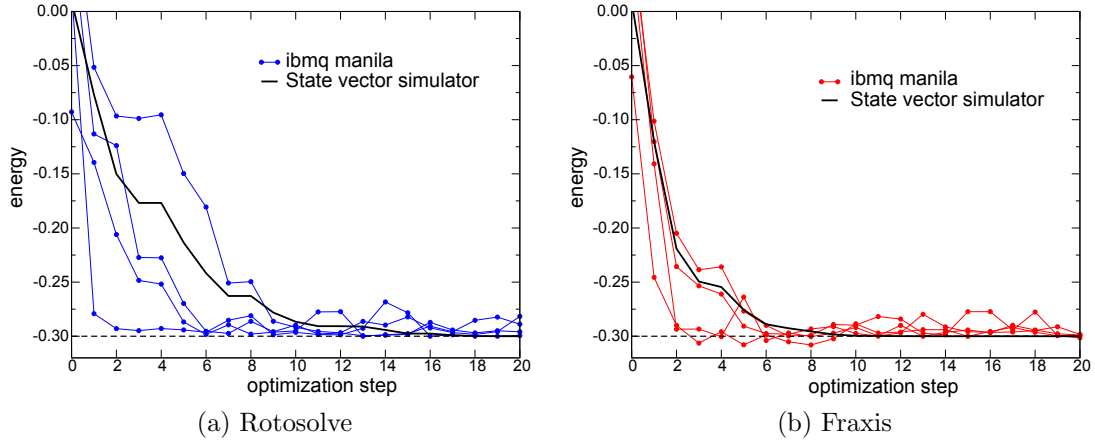


Figure 14: Comparison of energy convergence as a number of optimization steps for Heisenberg model on the `ibmq_manila` device with readout error mitigation. (a) Rotosolve (blue), (b) Fraxis (red). Both method are applied to start from the 4 random initial parameter sets.

mations are required for each gate update to find the optimal axis, which is more efficient than Rotoselect that requires seven of them. Fraxis ansatz seems to be more effective in dealing with general quantum systems, such as, the Heisenberg model whose ground states may not be real-valued amplitudes like those of molecular Hamiltonian models. Meanwhile, [45] recently showed that a restricted Fraxis ansatz can improve QAOA for MaxCut despite the ground states are real-valued amplitudes. We confirmed numerically that Fraxis ansatz could be better at similar cases to find ground states of Hamiltonian for MaxCut.

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 and/or by gradient-based approaches to avoid local minima, which we plan to pursue as future work. Another promising approach is by simultaneously updating  $k$  axes, which is similarly shown in [27, 29] 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 [27, 29] and for this reason we may have to combine Fraxis with other techniques, such as, matrix completion [46, 47]. Fraxis’ higher number of energy estimations for classical optimization of parameters may be unavoidable due to its higher expressibility. Thanks to the matrix factorization framework of Fraxis, sequential arbitrary single-qubit optimization [32] and sequential time evolution of quantum systems in [31] have been successfully developed. The codes of Fraxis and its extension are publicly available.

## Acknowledgment

H.C.W. was supported by JSPS under Grant Numbers 20K03885 and 20H05518, and JST PRESTO under Grant Number JPMJPR17GC. E.K. was supported by JSPS under Grant Number 20K14388 and JST PRESTO under Grant Number JPMJPR2011. In addition, H.C.W., E.K., and M.S. were supported by the MEXT Quantum Leap Flagship Program under Grant Number JPMXS0118067285 and JPMXS0120319794. R.R. would like to thank colleagues at IBM Quantum, Yutaka Shikano, and Harumichi Nishimura for insightful technical discussion. The authors would like to thank Naoki Yamamoto, Michael Lubasch, and Christiane Koch for their valuable comments.

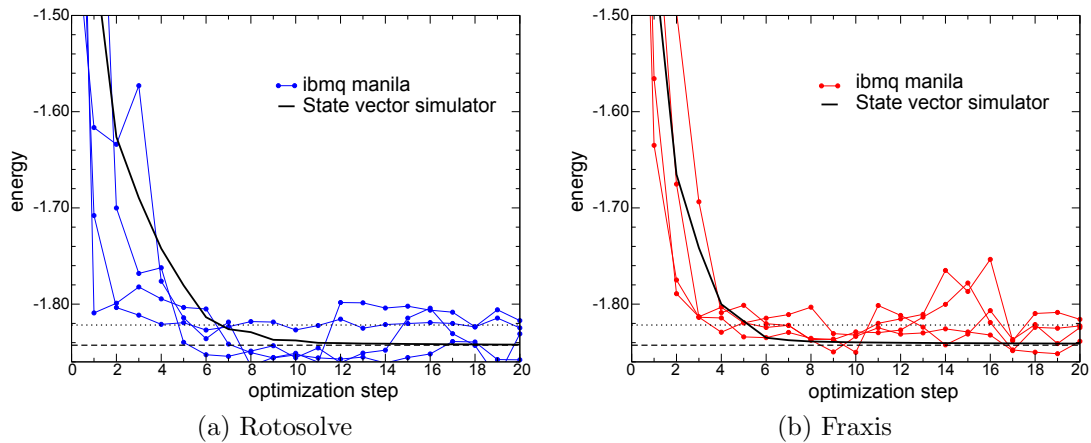


Figure 15: Comparison of energy convergence as a number of optimization steps for  $H_2/STO-3G$  on the `ibmq_manila` device with readout error mitigation. (a) Rotosolve (blue), (b) Fraxis (red). Both method are applied to start from the 4 random initial parameter sets. The exact ground state energy is displayed with dashed line. The Hartree-Fock state is displayed with dotted line.

## Code availability

The Fraxis code used to create and the data in this article is executable on Python using Qiskit version 0.36 [39], and can be found on GitHub, <https://github.com/KQCC-Chemistry/SeQpt>.

## References

- [1] M. Cerezo *et al*, “Variational Quantum Algorithms,” *arXiv:2012.09265*, 2020.
- [2] K. Bharti *et al*, “Noisy intermediate-scale quantum (NISQ) algorithms,” *arXiv:2101.08448*, 2021.
- [3] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, “The variational quantum eigensolver: a review of methods and best practices,” 2021. [Online]. Available: <https://arxiv.org/abs/2111.05176>
- [4] A. Peruzzo *et al*, “A variational eigenvalue solver on a photonic quantum processor,” *Nat. Commun.*, vol. 5, no. 1, p. 4213, 2014. [Online]. Available: <https://doi.org/10.1038/ncomms5213>
- [5] E. Farhi, J. Goldstone, and S. Gutmann, “A quantum approximate optimization algorithm,” 2014. [Online]. Available: <https://arxiv.org/abs/1411.4028>
- [6] J. Preskill, “Quantum Computing in the NISQ era and beyond,” *Quantum*, vol. 2, p. 79, Aug. 2018. [Online]. Available: <https://doi.org/10.22331/q-2018-08-06-79>
- [7] A. Kandala *et al*, “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,” *Nature*, vol. 549, pp. 242–246, 2017.
- [8] P. K. Barkoutsos *et al*, “Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions,” *Phys. Rev. A*, vol. 98, p. 022322, Aug 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.98.022322>
- [9] M. Ganzhorn *et al*, “Gate-Efficient Simulation of Molecular Eigenstates on a Quantum Computer,” *Phys. Rev. Applied*, vol. 11, p. 044092, Apr 2019. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevApplied.11.044092>

- [10] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, “Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm,” *npj Quantum Inf.*, vol. 6, no. 1, p. 10, 2020. [Online]. Available: <https://doi.org/10.1038/s41534-019-0240-1>
- [11] H. L. Tang *et al.*, “Qubit-ADAPT-VQE: An Adaptive Algorithm for Constructing Hardware-Efficient Ansätze on a Quantum Processor,” *PRX Quantum*, vol. 2, p. 020310, Apr. 2021. [Online]. Available: <https://link.aps.org/doi/10.1103/PRXQuantum.2.020310>
- [12] N. V. Tkachenko *et al.*, “Correlation-Informed Permutation of Qubits for Reducing Ansatz Depth in VQE,” *arXiv:2009.04996*, 2020.
- [13] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, “Barren plateaus in quantum neural network training landscapes,” *Nat. Commun.*, vol. 9, 2018.
- [14] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, “Cost function dependent barren plateaus in shallow parametrized quantum circuits,” *Nat. Commun.*, vol. 12, 2021.
- [15] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, “Absence of barren plateaus in quantum convolutional neural networks,” *arXiv:2011.02966*, 2020.
- [16] Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Albrecht, and A. T. Sornborger, “Barren plateaus preclude learning scramblers,” *Phys. Rev. Lett.*, vol. 126, p. 190501, May 2021. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.126.190501>
- [17] C. Zhao and X.-S. Gao, “Analyzing the barren plateau phenomenon in training quantum neural network with the ZX-calculus,” *arXiv:2102.01828*, 2021.
- [18] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, “Effect of barren plateaus on gradient-free optimization,” *arXiv:2011.12245*, 2020.
- [19] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, “An adaptive variational algorithm for exact molecular simulations on a quantum computer,” *Nat. Commun.*, vol. 10, no. 1, pp. 1–9, 2019.
- [20] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, “Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms,” *Adv. Quantum Technol.*, vol. 2, no. 12, p. 1900070, Oct. 2019. [Online]. Available: <http://dx.doi.org/10.1002/qute.201900070>
- [21] L. Funcke, T. Hartung, K. Jansen, S. Kühn, and P. Stornati, “Dimensional Expressivity Analysis of Parametric Quantum Circuits,” *Quantum*, vol. 5, p. 422, Mar. 2021. [Online]. Available: <https://doi.org/10.22331/q-2021-03-29-422>
- [22] J. Li, X. Yang, X. Peng, and C.-P. Sun, “Hybrid Quantum-Classical Approach to Quantum Optimal Control,” *Phys. Rev. Lett.*, vol. 118, p. 150503, Apr. 2017. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.118.150503>
- [23] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, “Quantum circuit learning,” *Phys. Rev. A*, vol. 98, p. 032309, Sep. 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.98.032309>
- [24] R. Sweke *et al.*, “Stochastic gradient descent for hybrid quantum-classical optimization,” *Quantum*, vol. 4, p. 314, 2020.
- [25] A. W. Harrow and J. C. Napp, “Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms,” *Phys. Rev. Lett.*, vol. 126, p. 140502, Apr. 2021. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.126.140502>
- [26] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, “Layerwise learning for quantum neural networks,” *Quantum Mach. Intell.*, vol. 3, no. 1, pp. 1–11, 2021.

- [27] K. M. Nakanishi, K. Fujii, and S. Todo, “Sequential minimal optimization for quantum-classical hybrid algorithms,” *Phys. Rev. Research*, vol. 2, p. 043158, Oct. 2020. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevResearch.2.043158>
- [28] M. Ostaszewski, E. Grant, and M. Benedetti, “Structure optimization for parameterized quantum circuits,” *Quantum*, vol. 5, p. 391, Jan. 2021. [Online]. Available: <https://doi.org/10.22331/q-2021-01-28-391>
- [29] R. M. Parrish, J. T. Iosue, A. Ozaeta, and P. L. McMahon, “A Jacobi Diagonalization and Anderson Acceleration Algorithm for Variational Quantum Algorithm Parameter Optimization,” *arXiv:1904.03206*, 2019.
- [30] H. C. Watanabe, R. Raymond, Y.-Y. Ohnishi, E. Kaminishi, and M. Sugawara, “Optimizing parameterized quantum circuits with free-axis selection,” in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, 2021, pp. 100–111.
- [31] K. Wada, R. Raymond, Y.-y. Ohnishi, E. Kaminishi, M. Sugawara, N. Yamamoto, and H. C. Watanabe, “Simulating time evolution with fully optimized single-qubit gates on parametrized quantum circuits,” *Phys. Rev. A*, vol. 105, p. 062421, Jun 2022. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.105.062421>
- [32] K. Wada, R. Raymond, Y. Sato, and H. C. Watanabe, “Full optimization of a single-qubit gate on the generalized sequential quantum optimizer,” 2022. [Online]. Available: <https://arxiv.org/abs/2209.08535>
- [33] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.
- [34] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation*. Boston, MA, USA: American Mathematical Society, 2002.
- [35] J. Kempe, A. Kitaev, and O. Regev, “The Complexity of the Local Hamiltonian Problem,” *SIAM J. Comput.*, vol. 35, no. 5, pp. 1070–1097, May 2006. [Online]. Available: <http://dx.doi.org/10.1137/S0097539704445226>
- [36] R. Oliveira and B. M. Terhal, “The complexity of quantum spin systems on a two-dimensional square lattice,” *Quantum Info. Comput.*, vol. 8, no. 10, pp. 900–924, Nov. 2008. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2016985.2016987>
- [37] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, “Measurements of quantum hamiltonians with locally-biased classical shadows,” *Communications in Mathematical Physics*, vol. 391, no. 3, pp. 951–967, May 2022. [Online]. Available: <https://doi.org/10.1007/s00220-022-04343-8>
- [38] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, “Efficient  $z$  gates for quantum computing,” *Phys. Rev. A*, vol. 96, p. 022330, Aug 2017. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.96.022330>
- [39] H. Abraham *et al*, “Qiskit: An Open-source Framework for Quantum Computing,” 2019.
- [40] K. Życzkowski and H.-J. Sommers, “Average fidelity between random quantum states,” *Phys. Rev. A*, vol. 71, p. 032313, Mar. 2005. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.71.032313>
- [41] B. Fuller, C. Hadfield, J. R. Glick, T. Imamichi, T. Itoko, R. J. Thompson, Y. Jiao, M. M. Kagele, A. W. Blom-Schieber, R. Raymond *et al.*, “Approximate solutions of combinatorial problems via quantum relaxations,” *arXiv preprint arXiv:2111.03167*, 2021.
- [42] M. X. Goemans and D. P. Williamson, “Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming,” *J. ACM*, vol. 42, no. 6, p. 1115–1145, nov 1995. [Online]. Available: <https://doi.org/10.1145/227683.227684>

- [43] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Müller, W. Riess, G. Salis, J. Smolin, I. Tavernelli, and K. Temme, “Quantum optimization using variational algorithms on near-term quantum devices,” *Quantum Science and Technology*, vol. 3, no. 3, p. 030503, jun 2018.
- [44] Q. Gao *et al*, “Applications of quantum computing for investigations of electronic transitions in phenylsulfonyl-carbazole TADF emitters,” *npj Comput. Mater.*, vol. 7, no. 1, pp. 1–9, 2021.
- [45] L. C. G. Govia, C. Poole, M. Saffman, and H. K. Krovi, “Freedom of mixer rotation-axis improves performance in the quantum approximate optimization algorithm,” 2021.
- [46] E. J. Candès and B. Recht, “Exact Matrix Completion via Convex Optimization,” *Found. Comput. Math.*, vol. 9, no. 6, p. 717, Apr 2009. [Online]. Available: <https://doi.org/10.1007/s10208-009-9045-5>
- [47] E. J. Candès and T. Tao, “The Power of Convex Relaxation: Near-Optimal Matrix Completion,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.