Full optimization of a single-qubit gate on the generalized sequential quantum optimizer

Kaito Wada,1 Rudy Raymond,2,3,4 Yuki Sato,5,3 and Hiroshi C. Watanabe5,6

1Department of Applied Physics and Physico-Informatics, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, 223-8522, Japan
2IBM Quantum, IBM Japan 19-21 Nihonbashi Hakozaki-cho, Chuou-ku, Tokyo, 103-8510, Japan
3Quantum Computing Center, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, 223-8522, Japan
4Department of Computer Science, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
5Toyota Central R&D Labs., Inc., Koraku Mori Building 10F, 1-4-14 Koraku, Bunkyo-ku, Tokyo 112-0004, Japan

We propose a new quantum algorithm based on analytically-maximum optimization of a single-qubit gate in parameterized quantum circuits (PQCs). Variational quantum algorithms on near-term quantum devices depend on the ability to optimize objective functions with hardware-efficient PQCs. Standard optimizations utilize single-parameter quantum gates in PQCs whose rotation angles along fixed axes are parameterized. The parameters are then updated by gradient or by sequential quantum optimization. A conventional sequential optimizer utilizes sinusoidal properties to directly find local optima regarding a single-qubit gate. A recent method proposed sequential optimization of multi-parameter single-qubit gates in PQCs whose parameters are the rotation axes of quantum gates with fixed half-rotation angles. Here, unifying and extending the existing methods, we derive the maximum optimization of a single-qubit gate from the framework of matrix factorization. We perform numerical experiments demonstrating the efficacy of the framework.

I. INTRODUCTION

Variational Quantum Algorithm (VQA) is a notable classical-quantum hybrid algorithm, which is feasible on near-term quantum devices. It is a versatile methodology applicable to quantum chemical calculation, combinatorial optimization problem, dynamics simulation, time evolving simulation, principle component analysis, linear and non-linear system solvers, and so on. The key component in VQAs is parameterized quantum circuit (PQC), also called ansatz, which is a sequence of quantum gates with classically controllable parameters.

The cost function of VQA is often formulated as the expectation value of Hamiltonian $H$ of a target system whose solution can be obtained from the eigensystem of $H$ computed with variational quantum eigensolver (VQE), e.g., for quantum chemical calculations using fermionic or spin Hamiltonian. Appropriate designs of PQCs are essential to express the quantum states of interest. PQCs are classified into physics-based ansatz and heuristic-based ansatz. A physics-based ansatz, including a unitary coupled cluster and Hamiltonian variational ansatz, can achieve efficient optimization to limit the Hilbert space spanned by the ansatz to the neighborhood of the target states. However, it is difficult to run on near-term devices because of its circuit depth. In contrast, a heuristic-based ansatz puts more weight on the feasibility on near-term devices, which hence generally results in shallower circuits with the uncertainty to express the target states.

A typical strategy to mitigate the uncertainty is by systematically increasing the circuit depth with adding layers of gates. However, increasing layers can cause the gradient of the cost function with regards to parameters of PQCs to exponentially vanish as the number of qubits grows, a phenomenon termed barren plateau. The barren plateau renders gradient-based approaches useless. Several remedies, such as layerwise-learning and parameter correlation, have been proposed but they can only cope with noiseless conditions. The only effective strategy for noise-induced barren plateau is to reduce circuit depth. Hence, there are two conflicting requirements of PQCs with heuristic ansatz; they should be deep enough to express target states but should be as shallow as possible to avoid the barren plateau.

A concept of circuit structure optimization implemented in Rotoselect draws attention to deal this dilemma, which is shared in Variable Ansatz (VAns) algorithm. Rotoselect is also a family of quantum sequential optimizers such as NFT (also termed Rosetosolve in [26]), where single-qubit gates are sequentially and analytically optimized without gradient. It allows to select the optimal (single-qubit) rotational gate in a PQC among $R_x(\theta)$, $R_y(\theta)$, and $R_z(\theta)$ by utilizing the periodicity of objective functions in $\theta$, as $R_x(\theta) = \cos(\theta/2)I - i\sin(\theta/2)X$ and similarly for $R_y$ and $R_z$. Since the PQCs with Rotoselect become a bit more flexible via choosing the rotation axes of the single-qubit gates through cost minimization, Rotoselect is regarded as structural optimization of fixed-depth PQCs. How-
ever, Rotoselect has two drawbacks: the rotation axis is selected from a finite and discrete gate set, and each parameter of PQCs is updated locally. As a result, Rotoselect is likely to be stuck in local optima. To deal with the drawbacks, PQCs consisting of single-qubit gates, whose rotational axes are arbitrary three-dimensional vectors, have been proposed \cite{29}. Like Rotoselect, free-axis selection (Fraxis) \cite{29} is a quantum sequential optimizer, but with more parameters (i.e., axes of rotations) to optimize. The optimal axis is obtained from a matrix diagonalization whose elements are computed from expectation values of \( H \) for quantum states generated from PQCs, where the gate of interest is replaced by a set of unitaries. Because the PQC with Fraxis gates optimizes the rotational axis of each single-qubit gate, it has higher degree of freedom and therefore, has better expressibility than a PQC with Rotoselect/select. Recently, PQCs with fully optimized single-qubit gates, i.e., both the rotation axis and angle of each single-qubit gate are simultaneously optimized, have been shown possible for time-evolving simulation \cite{11}.

Here, we show such fully optimization of single-qubit gates, termed as Free Quaternion Selection (FQS) after the quaternion representation of a single-qubit gate, is also applicable to the general form of the VQA cost function. Furthermore, we show that the FQS generalizes other existing quantum sequential optimizers, such as NFT, Rotoselect/select, and Fraxis; all known quantum sequential optimizers can be formulated as matrix factorization. Because the single-qubit gates of PQCs with FQS have the highest degree of freedom, FQS can optimize more effectively the circuit structure to express target quantum states with limited depth when compared to Fraxis and Rotoselect/select as illustrated in Fig. 1.

The remainder of this paper is organized as follows. In Sec. II we describe the theoretical aspects of our proposed method. We first give the quaternion representation for a single-qubit gate, and then construct the FQS for VQA. We also present the FQS as the generalized form of NFT, Rotoselect/select, and Fraxis. In Sec. III we provide numerical experiments to demonstrate the effectiveness of our proposed method. Finally, we conclude this study in Sec. IV.

II. THEORY

A. Quaternion representation for single-qubit gate

A general single-qubit gate is conventionally represented as

\[
R_n(\psi) := \cos \left( \frac{\psi}{2} \right) I - i \sin \left( \frac{\psi}{2} \right) n \cdot \vec{\sigma}, \quad (1)
\]

where \( I \) and \( \vec{\sigma} = (X, Y, Z) \) denote the 1-qubit identity operator and the Pauli matrices. The parameters \( n \) and \( \psi \) correspond to a rotational axis and angle in the Bloch sphere, respectively. Here, we show another way to parameterize the general single-qubit gate. Since the rotational axis \( n \) is a three-dimensional real unit vector, we can write it in the polar coordinate system with the zenith angle \( \theta \) and the azimuth angle \( \phi \) as

\[
n = n(\theta, \phi) = (\cos \theta, \sin \theta \cos \phi, \sin \theta \sin \phi). \quad (2)
\]

Substituting Eq. (2) into Eq. (1), we obtain the quaternion representation of a single-qubit gate as

\[
R_n(\psi, \theta, \phi) = q(\psi, \theta, \phi) \cdot \vec{\varsigma}, \quad (3)
\]

where an unit quaternion \( q = (q_0, q_1, q_2, q_3) \) (i.e., \( q \in \mathbb{R}^4, |q| = 1 \)) is parameterized with \((\psi, \theta, \phi)\) as

\[
q_0 = \cos \left( \frac{\psi}{2} \right),
\]

\[
q_1 = \sin \left( \frac{\psi}{2} \right) \cos \theta,
\]

\[
q_2 = \sin \left( \frac{\psi}{2} \right) \sin \theta \cos \phi,
\]

\[
q_3 = \sin \left( \frac{\psi}{2} \right) \sin \theta \sin \phi. \quad (4)
\]

Here, \( \vec{\varsigma} := (s_0, s_1, s_2, s_3) \) is an extension of the Pauli matrices defined as

\[
s_0 = I, \quad s_1 = -iX, \quad s_2 = -iY, \quad s_3 = -iZ. \quad (5)
\]

Thus, we can identify a point on the three-dimensional spherical surface with a single-qubit gate. Based on this identification, we write \( R_n(\psi) \) as \( R(q) \) for simplicity in the rest of this paper. Note that if we focus on the conventional single-qubit rotation gate with one-parameter, such as \( R_x(\theta) \), this gate can be identified with a point on the one-dimensional spherical surface, i.e., the unit circle.
FIG. 2. A quantum circuit to evaluate the FQS matrix elements. An optimal single-qubit gate $R(q_d)$ of interest is computed from evaluating the circuit by replacing the gate with ten different gates as described in Sec. 4C.

B. Our algorithm: Free Quaternion Selection for Variational Quantum Algorithm

Let us consider an $n$-qubit parameterized quantum circuit $U$ consisting of $D$ parameterized single-qubit gates $\{R(q_d)\}_{d=1}^D$ and parameter-free gates such as CNOT gate. Tuning the parameters, we aim to solve an optimization task with the following objective function

$$\sum_{k=1}^K \text{tr} \left[ \rho_k U \left( \{q_d\}_{d=1}^D \right) \right],$$

where $\rho_k$ is an $n$-qubit initial state from a training set, and $H_k$ is a some observable. In the following, without loss of generality we focus on a single expected value in the objective function (i.e., $K = 1$), which can simply be considered as the minimization of energy for the Hamiltonian $H := H_1$ and the input state $\rho_{in} := \rho_1$. The extension of subsequent discussion to the whole objective function is trivial due to the linearity. For the energy minimization, we focus on a sequential optimization regarding $R(q_d)$ where all parameters are fixed except for the $d$th single-qubit gate as shown in Fig. 2.

The energy expectation of the variational state is written as

$$\langle H \rangle(q_d) := \text{tr} \left[ \rho_{in} U_1^\dagger R(q_d) U_2^\dagger H_1 U_2 R(q_d) U_1 \right],$$

where $U_1$ and $U_2$ are the quantum circuits before and after $R(q_d)$, respectively. $H'$ and $\rho_{in}'$ are defined as

$$H' := U_1^\dagger H U_2, \quad \rho_{in}' := U_1 \rho_{in} U_1^\dagger.$$ (8)

Here we omit the subscript $d$ for simplicity. Substituting Eq. (3) into Eq. (7), we can obtain the following quadratic form

$$\langle H \rangle(q) = q^\top S q,$$ (9)

where superscript $\top$ denotes a transpose operation, and $S = (S_{\mu\nu})$ is a $4 \times 4$ real-symmetric matrix whose elements are defined as

$$S_{\mu\nu} := \frac{1}{2} \text{tr} \left[ \rho_{in}' \left( \sigma_{\mu}' H_{\nu}' + \sigma_{\nu}' H_{\mu}' \right) \right]$$

$$= \begin{pmatrix}
\frac{1}{2} \text{tr}(H_{\mu}' \rho_{in}'), & \frac{1}{2} \text{tr}(H_{\nu}' \rho_{in}'), & \frac{1}{2} \text{tr}(H_{\nu}' \rho_{in}'), & \frac{1}{2} \text{tr}(H_{\nu}' \rho_{in}'), \\
\text{tr}(H_{\nu}' \rho_{in}') & \text{tr}(H_{\nu}' \rho_{in}'), & \text{tr}(H_{\nu}' \rho_{in}'), & \text{tr}(H_{\nu}' \rho_{in}'), \\
\frac{1}{2} \text{tr}(H_{\mu}' \rho_{in}'), & \frac{1}{2} \text{tr}(H_{\mu}' \rho_{in}'), & \text{tr}(H_{\mu}' \rho_{in}'), & \text{tr}(H_{\mu}' \rho_{in}'), \\
\text{tr}(H_{\nu}' \rho_{in}') & \text{tr}(H_{\nu}' \rho_{in}'), & \text{tr}(H_{\nu}' \rho_{in}'), & \text{tr}(H_{\nu}' \rho_{in}').
\end{pmatrix}.$$ (10)

See Appendix A for derivation of the quadratic form.

The matrix $S$ can be obtained by running and measuring ten quantum circuits, each corresponding to the element in the upper diagonal of $S$, as detailed in the next subsection. Note that the minimization of the quadratic form is exactly achieved by calculating the eigenvector corresponding to the lowest eigenvalue of $S$. In addition, this optimization over the whole $SU(2)$ is a generalization of other sequential optimizers [20, 28, 29], which optimize only a part of $SU(2)$. To clarify this point, we show they can be derived from our general framework at the end of this section. Notice that a special FQS, which applies to the objective functions in a special form, was proposed for time-evolving simulation [11]. Our formulation is also regarded as the extension of that special FQS. Since FQS


can select the optimal gate from the whole $SU(2)$ for minimizing the energy expectation in Eq. (4), it can incorporate the multi-parameter correlation in $SU(2)$ and thus achieves better performance for optimizing PQCs than other sequential optimizers. Rotoselect, Rotosolve, and their variants [20, 23, 30, 31] can be used to optimize a general single-qubit gate by first decomposing the multi-parameter gate using $ZYZ$-decomposition [32] to obtain three single-parameter gates. In that case, each parameter of the decomposed gate is optimized locally in contrast to FQS. Similarly, a general single-qubit gate is decomposed into two Fraxis gates defined as $R_{\alpha}(\pi)$. Then, Fraxis [29] separately updates these gates but simultaneously optimizes two parameters within the gate. Recent works [33, 34] extend Rotosolve to deal with non-Pauli
single-parameter gates which are different from multi-parameter Fraxis and FQS that have closed-form solutions.

C. Evaluation of the S elements

To determine the optimal single-qubit gate for the cost minimization, we evaluate the matrix $S$ constructing the quadratic form Eq. (9). All the elements of $S$ in Eq. (10) are in fact calculated from ten expected values classified into three types as below.

Type-A:

$$\text{tr} [\rho_n' H']$$

Type-B: for $k = 1, 2, 3$,

$$\text{tr} \left[ \rho_n' \left( \frac{I \pm i \sigma_k}{\sqrt{2}} \right)^\dagger H' \left( \frac{I \pm i \sigma_k}{\sqrt{2}} \right) \right]$$

Type-C: for $(k, m) = (1, 2), (1, 3), (2, 3)$

$$\text{tr} \left[ \rho_n' \left( \frac{\sigma_k + \sigma_m}{\sqrt{2}} \right)^\dagger H' \left( \frac{\sigma_k + \sigma_m}{\sqrt{2}} \right) \right].$$

Here, Type-A is the $(0, 0)$-element of $S$ and corresponds to the expected value of $H$ on the PQC when the single-qubit gate of interest, as in Fig. 2, is replaced with identity. The other diagonal elements are produced by the type-A and type-B values with the following identity:

$$\text{tr}(H' \sigma_k \rho_n' \sigma_k) + \text{tr}(H' \rho_n')$$

$$= \text{tr} \left[ \rho_n' \left( \frac{I - i \sigma_k}{\sqrt{2}} \right)^\dagger H' \left( \frac{I - i \sigma_k}{\sqrt{2}} \right) \right]$$

$$+ \text{tr} \left[ \rho_n' \left( \frac{I + i \sigma_k}{\sqrt{2}} \right)^\dagger H' \left( \frac{I + i \sigma_k}{\sqrt{2}} \right) \right].$$

(11)

Note that the Type-B corresponds to the expected value of $H$ on the PQC when the single-qubit gate of interest, as in Fig. 2, is replaced with, respectively, $R_x(-\pi/2), R_y(-\pi/2), R_y(\pi/2), R_z(-\pi/2),$ and $R_z(\pi/2)$. In contrast, subtracting the type-B values with different sign yields the other elements in the first row directly. The remaining off-diagonal elements are produced by the type-C expected values and the already obtained diagonal elements with the following identity:

$$\frac{1}{2} [\text{tr}(H' \sigma_i \rho_n' \sigma_i) + \text{tr}(H' \sigma_j \rho_n' \sigma_j)]$$

$$= \text{tr} \left[ \rho_n' \left( \frac{\sigma_i + \sigma_j}{\sqrt{2}} \right)^\dagger H' \left( \frac{\sigma_i + \sigma_j}{\sqrt{2}} \right) \right]$$

$$- \frac{1}{2} [\text{tr}(H' \sigma_i \rho_n' \sigma_i) + \text{tr}(H' \sigma_j \rho_n' \sigma_j)].$$

(12)

Note that the Type-C values correspond to the expected values of $H$ on the PQC when the single-qubit gate to be optimized, as in Fig. 2, is replaced with, respectively, Fraxis gate $R_m(\pi)$ for $m \in \{(1/\sqrt{2}, 1/\sqrt{2}, 0), (1/\sqrt{2}, 0, 1/\sqrt{2}), (0, 1/\sqrt{2}, 1/\sqrt{2})\}$. All expected values of Type-A, B, and C can be evaluated with direct measurements without any control operation such as the Hadamard test. Since the degree of freedom for a $4 \times 4$ real-symmetric matrix is ten, the number of required direct measurements should be optimal.

D. Unification of sequential quantum optimizers

FQS generalizes all other sequential optimizers, such as, Rotosolve (NFT), Rotoselect, Fraxis, and $\theta$-Fraxis; those methods can be regarded as special cases of FQS. For NFT, a single-qubit gate is restricted to a fixed axis $m$ such as $U_{\text{NFT}} := R_m(\psi)$. Then the corresponding objective function in quadratic form is

$$\text{tr} \left[ \rho_n' U_{\text{NFT}}^\dagger H' U_{\text{NFT}} \right] = c^\top \left( S_0 \tilde{S}_0 \cdot m \cdot m^\top \tilde{S} m \right) c,$$  

(13)

where $c := (\cos \psi/2, \sin \psi/2)^\top$ and $\tilde{S}_0 := (S_{01}, S_{02}, S_{03})$. $\tilde{S}$ denotes the lower right $3 \times 3$ part of the $S$ matrix. The derivation of the quadratic form is detailed in Appendix A. The real symmetric matrix in Eq. (13) can be regarded as a contraction of the FQS matrix $S$ with respect to the rotational axis $m$ reducing its degree of freedom to three.

In Rotoselect, three contracted matrices in Eq. (13) are constructed for $m \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, and the lowest eigenvalue is selected after three separate diagonalization procedures. Since the three matrices share $S_{00}$, the total number of circuit evaluations can be reduced to seven. As for Fraxis, the target gate $U_{\text{Fraxis}} := R_m(\pi)$ is simply expressed by the quaternion $q = (0, n)$. Thus, substituting $q = (0, n)$ into the objective function, we can reproduce the previous results derived in [29] as follows

$$\text{tr} \left[ \rho_n' U_{\text{Fraxis}}^\dagger H' U_{\text{Fraxis}} \right] = n^\top \tilde{S} n.$$  

(14)

Note that $\theta$-Fraxis in which the rotation angle is fixed to arbitrary values (i.e., fixed $q_0$) is regarded as minimizing Eq. (10) for $q^\prime := (q_1, q_2, q_3) \in \mathbb{R}^3$ under the constraint $|q|^2 = 1 - q_0^2$, which results in solving simultaneous equations rather than diagonalization.

It is worth noting that the required number of circuit evaluations for each sequential quantum optimizer coincides with the degrees of freedom for the real-symmetric matrix of respective methods, i.e., $3=1+2+3$ circuit evaluations for NFT/Rotosolve, $6=1+2+3$ circuit evaluations for Fraxis, and $10=1+2+3+4$ circuit evaluations for FQS.
FIG. 3. Grouped-layer PQC with ladder entangler

FIG. 3. Grouped-layer PQC with ladder entangler

FIG. 4. Cascading-block PQC

FIG. 4. Cascading-block PQC

FIG. 5. Averaged trajectories of the VQE resulting energy for 5-qubit 1-dimensional Heisenberg model on the three-layer PQC with the cascade structure. The horizontal dashed line represents the exact ground energy. Note that the 24 gate updates corresponds to one sweep.

FIG. 5. Averaged trajectories of the VQE resulting energy for 5-qubit 1-dimensional Heisenberg model on the three-layer PQC with the cascade structure. The horizontal dashed line represents the exact ground energy. Note that the 24 gate updates corresponds to one sweep.

III. NUMERICAL EXPERIMENTS

We have seen that FQS generalizes known sequential optimizers such as NFT, Rotosolve/select and Fraxis. Here, we will show how FQS are better for VQE in Heisenberg model and fidelity maximization. We compare FQS with Fraxis because Fraxis has been shown to be comparable or better than NFT and Rotosolve/select [29].

A. Heisenberg model

To benchmark the performance of FQS in comparison with Fraxis, we carried out VQE optimization for 1-dimensional Heisenberg model with five qubits whose Hamiltonian is written as

$$H = J \sum_{i=1}^{5} \sum_{n=x,y,z} \sigma_{i}^{(n)} \sigma_{i+1}^{(n)} + h \sum_{i=1}^{5} \sigma_{i}^{(z)}$$  (15)

where the periodic boundary condition is employed as $\sigma_{6}^{(n)} = \sigma_{1}^{(n)}$ and $J = h = 1$. We employed two types of layerwise PQC as in Figs. 3 and 4. In both ansätze, one layer consists of the consistent number of single-qubit gates and thus comparable number of single-qubit and control gates in total. In the Fraxis optimization, we prepared 20 initial conditions in state-random manner, where the rotation axes $n$ in Eq. (1) were randomly selected based on the uniform probability distribution on the Bloch sphere. For fair comparison, the same initial states were used for the FQS optimizations.

Figure 6 shows averaged trajectories of independent 20
VQE simulations based on the cascading-block PQC. It is obvious that the FQS trajectory is lower than that of Fraxis for all optimization steps, which evinces the optimization efficiency. This efficiency can be attributed to two characters of FQS. Firstly, a single-qubit gate is generally decomposed into three fixed-axis rotation gates such as $R_z(\phi)R_y(\theta)R_z(\lambda)$. The fact that FQS can fully optimize single-qubit gates indicates these three parameters are simultaneously optimized taking the parameter correlation into account. Considering that Fraxis outperforms the other sequential quantum optimizers such as Rotosolve/NFT and Rotoselect as in [29], the present result is a clear demonstration that the optimization efficiency reflects the incorporated parameter correlation. Secondly, FQS can deal with single-qubit gates that have higher expressibility than Fraxis, which is advantageous in making shortcut linking to the ground state.

Figure 6 shows the optimized energy after 100 sweeps based on the ansaetze in Figs. 3 and 4. In a single sweep, all gates were updated once in ascending order of the gate set index as labeled in Figs. 3 and 4. When compared the two ansaetze in the same number of layers, the cascading-block PQC achieved lower energy in Fig. 6. In particular, the $L$-layer cascading-block PQC was comparable to the grouped-layer PQC with more layer than $L$. Note that the mean and the minimum value of the resulting energies by FQS are equal or lower than those of Fraxis. We believe the two advantages for efficient optimization described above also bring benefits to find better solution.

It is also notable that the difference between FQS and Fraxis becomes distinct as the number of layers increases in both ansaetze. We attribute this behavior to the difference in entanglement tuning between Fraxis and FQS. To understand this, it is useful to consider the difference between FQS and Fraxis ansaetze, which is related to the degree of freedom of the rotational angle in single-qubit gate, i.e., the first term in Eq. (1). Suppose PQCs consists of parameterized single-qubit gates and following fixed entangling gates, the PQC can be expanded as

$$
\prod_{d=1}^{D} U_{\text{ent}}^{(d)} R(q^{(d)}_{\mu}) = \prod_{d=1}^{D} \sum_{\mu} q_{\mu}^{(d)} U_{\text{ent}}^{(d)} y^{(d)}_{\mu}
$$

$$
= \prod_{d=1}^{D} \left( q_0^{(d)} U_{\text{ent}}^{(d)} y^{(d)}_0 + \sum_{i} q_i^{(d)} U_{\text{ent}}^{(d)} y^{(d)}_i \right)
$$

$$
= q_0^{(D)} \cdots q_0^{(2)} q_0^{(1)} U_{\text{ent}}^{(D)} \cdots U_{\text{ent}}^{(2)} U_{\text{ent}}^{(1)}
+ \sum_{d} \sum_{i} q_0^{(d)} \cdots q_0^{(d+1)} q_i^{(d+1)} q_0^{(d)} \cdots q_0^{(d-1)} q_0^{(d-1)} \cdots q_0^{(1)}
$$

$$
\times U_{\text{ent}}^{(D)} \cdots U_{\text{ent}}^{(d+1)} y^{(d)}_{\text{ent}} U_{\text{ent}}^{(d+1)} \cdots U_{\text{ent}}^{(1)}
+ \cdots
$$

$$
+ \sum_{i,j,\cdots,k} q_k^{(D)} \cdots q_j^{(2)} q_i^{(1)} U_{\text{ent}}^{(D)} y^{(1)}_{\text{ent}} k U_{\text{ent}}^{(1)} y^{(1)}_{\text{ent}} j U_{\text{ent}}^{(2)} y^{(2)}_{\text{ent}} j U_{\text{ent}}^{(2)} y^{(1)}_{\text{ent}} i
$$

where $U_{\text{ent}}^{(d)}$ denotes the fixed entangling gate associated with the $d$th single-qubit gate $R(q^{(d)}_{\mu})$. Note that the unitary operators except for the last term in the expansion consist of $D$ entangling gates and the various number of single-qubit gates. Considering the input state $|0\rangle^\otimes n$, the unitary from the first term returns a product state. For another example, the unitaries whose single-qubit gates are located in only the last part of the circuit such as $U_{\text{ent}}^{(D)} U_{\text{ent}}^{(D)} U_{\text{ent}}^{(1)} \cdots U_{\text{ent}}^{(1)}$ return weakly-entangled states. In contrast, the last term, which is the only term shared by both FQS and Fraxis, can cause higher entanglement by modulating the single-qubit gates between respective entangling gate. Altogether, FQS may possess the ability of tuning entanglement by combining the states in various entanglement level, and this ability will be harnessed as the layer increases.

B. Fidelity maximization

As another example to elicit FQS performance, we selected the fidelity (the higher the better) as the cost function of the VQE optimization, where the reference states were independently prepared with Haar random generator in Qiskit [35] in each optimization. Figure 8 shows the results after 100-sweep optimization where we employed the grouped-layer PQC with cyclic entangler in Fig. 7 and the cascading-block PQC in Fig. 4. As expected, FQS showed better performance than Fraxis in both ansaetze, where the advantage was more distinct as the number of layers $L$ increased. However, the results of Fraxis were comparable to those of FQS in the shallow PQCs. In particular, Fraxis resulted in better fidelity when the cascading-layer PQC of $L=1$ was employed, which seems to be contradictory to the fact that FQS is upwardly compatible with Fraxis. There are two possible causes: larger Hilbert space due to extra degree of freedom is irrelevant to the optimization, and/or the optimization with FQS were stuck at local minima due to the imbalance of parameter correlation.

To check if the former cause holds, we evaluated the entanglement entropy of 1000 quantum states independently generated by PQC with the randomized parameters. Note that the entanglement entropy is a measure for a bipartite system and there are several possible partitions of the 5-qubit system. Here, we classified the partitions into two groups; one to four qubits and two to
three qubits. Then we assessed averaged entanglement entropy for the respective groups. Figure 9 shows that the quantum states on PQCs consisting of FQS gates show higher entanglement entropy over all number of circuit layers, which implies FQS can be advantageous in expression of highly entangled states. Here, it should be noted that the randomly-generated target states are likely to be highly entangled states because the majority of Hilbert space corresponds to highly-entangled states [36, 37]. Therefore, it is likely that the Hilbert space additionally spanned by the FQS ansatz may involve the target states in fidelity maximization with the cascading-block PQC of \( L = 1 \).

With regards to the second cause that the result with the \( L = 1 \) cascading-block PQC may arise from the imbalance of the parameter correlation in FQS, we notice that FQS could not sufficiently incorporate the inter-gate correlation, while the intra-gate correlation was perfectly taken into account. Indeed, 20% of the FQS runs on cascading-block PQC of \( L = 1 \) were stuck at local minimum, while it is not the case for Fraxis. These results imply that FQS may not necessarily guide the optimization to better trajectories, although the FQS application to a single gate is trivially superior to Fraxis. Note that these shortcomings can be mitigated by increasing circuit layers to provide more degrees of freedom as shown in [38].

**IV. CONCLUSION**

We proposed a generalized framework of sequential quantum optimizers for VQAs based on the quaternion
expression of a single-qubit gate. Accordingly, we developed a new method, FQS, that allows for the full optimization of a parametrized single-qubit gate. The FQS optimization corresponds to the factorization of the $4 \times 4$ real-symmetric matrix consisting of the expectation values measured on quantum states generated by the slightly modified PQCs. In addition, we showed that other sequential quantum optimizers can be derived from the generalized framework. We benchmarked the FQS applications to VQE in the Heisenberg model and fidelity maximization. Compared to other sequential quantum optimizers, FQS achieved higher efficiency in optimization by taking into account the intra-gate parameter correlation. Since PQCs with FQS possess higher expressibility, we believe FQS is instrumental in designing shallower circuits on near-term quantum devices.

V. ACKNOWLEDGEMENT

H.C.W. was supported by JSPS Grant Numbers 20K03885 and the MEXT Quantum Leap Flagship Program Grant Number JPMSX0118067285 and JP-MX50120319794. We would like to thank Dr. Michihiko Sugawara, Dr. Yu-ya Ohnishi, Dr. Eriko Kaminishi, and Dr. Naoki Yamamoto for technical discussion.

[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., Nature Reviews Physics 3, 625 (2021).
[2] A. Peruzzo et al., Nat. Commun. 5, 4213 (2014).
[3] A. Kandala et al., Nature 549, 242 (2017).
[4] Q. Gao et al., npj Comput. Mater. 7, 1 (2021).
[5] B. Fuller, C. Hadfield, J. R. Glick, T. Iadecola, and P. P. Orth, PRX Quantum 2, 043017 (2022).
[6] Y.-X. Yao, N. Gomes, F. Zhang, C.-Z. Wang, K.-M. Ho, T. Iadecola, and P. P. Orth, PRX Quantum 2, 030307 (2021).
[7] M. Benedetti, M. Fiorentini, and M. Lubasch, Physical Review Research 3, 101103/physrevresearch.3.033083 (2021).
[8] K. Wada, R. Raymond, Y.-y. Ohnishi, K. Hirano, and M. Fiorentini, EPJ Quantum Technology 9, 5 (2022).
[9] C. Zoufal, R. V. Mishmash, N. Sharma, N. Kumar, A. Sheshadri, A. Dˇeshmuˇnik, N. Ibrahim, J. Gacon, and S. Woerner, arXiv preprint arXiv:2205.03045 (2022).
[10] Y. Li and S. C. Benjamin, Physical Review X 7, 021050 (2017).
[11] Y.-X. Yao, N. Gomes, F. Zhang, C.-Z. Wang, K.-M. Ho, T. Iadecola, and P. P. Orth, PRX Quantum 2, 030307 (2021).
[12] M. Benedetti, M. Fiorentini, and M. Lubasch, Physical Review Research 3, 101103/physrevresearch.3.033083 (2021).
[13] K. Wada, R. Raymond, Y.-y. Ohnishi, E. Kaminishi, M. Sugawara, N. Yamamoto, and H. C. Watanabe, Physical Review A 105, 062422 (2021).
[14] R. LaRose, A. Tikku, E. O’Neel-Judy, L. Cincio, and P. J. Coles, npj Quantum Information 5, 1 (2019).
[15] M. Cerezo, K. Sharma, A. Arrasmith, and P. J. Coles, arXiv preprint arXiv:2004.01372 (2020).
[16] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles, arXiv preprint arXiv:1909.05820 (2019).
[17] X. Xu, J. Sun, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Science Bulletin 66, 2181 (2021).
[18] M. Lubasch, J. Joo, P. Moinier, M. Kiffner, and D. Jaksch, Physical Review A 101, 010301 (2020).
[19] R. Demirджan, D. Gunlycke, C. A. Reynolds, J. D. Doyle, and S. Tafur, arXiv preprint arXiv:2208.11780 (2022).
[20] J. Tilly, H. Chen, S. Cao, D. Piccozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Runger, G. H. Booth, et al., arXiv preprint arXiv:2111.05176 (2021).
[21] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, Physical Review A 95, 020501 (2017).
[22] D. Wecker, M. B. Hastings, and M. Troyer, Physical Review A 92, 042303 (2015).
[23] R. Wiersema, C. Zhou, Y. de Sereville, J. F. Carrasquilla, Y. B. Kim, and H. Yuen, PRX Quantum 1, 020319 (2020).
[24] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Nat. Commun. 12, 10.1038/s41467-021-21728-w (2021).
[25] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, Quantum Mach. Intell. 3, 1 (2021).
[26] T. Volkoff and P. J. Coles, Quantum Science and Technology 6, 025008 (2021).
[27] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Nature communications 12, 1 (2021).
[28] M. Ostaszewski, E. Grant, and M. Benedetti, Quantum 5, 391 (2021).
[29] M. Biknis, M. Cerezo, G. Verdon, P. J. Coles, and L. Cincio, arXiv preprint arXiv:2103.06712 (2021).
[30] K. M. Nakanishi, K. Fujii, and S. Todo, Physical Review Research 2, 043158 (2020).
[31] H. C. Watanabe, R. Raymond, Y.-Y. Ohnishi, E. Kaminishi, and M. Sugawara, in 2021 IEEE International Conference on Quantum Computing and Engineering (QCE) (IEEE, 2021) pp. 100–111.
[32] J. G. Vidal and D. O. Theis, Calculus on parameterized quantum circuits (2018).
[33] E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and K. Kim, Physical Review A 107, 020501 (2018).
[34] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, arXiv preprint arXiv:2109.11676 (2021).
As for the NFT \([28]\) and Rotosolve/Rotoselect \([26]\), they can optimize the axis-fixed single-qubit gate
\(\rho'_{in} R(q) \rho'_{in} \) in the last line is defined by symmetrization of \(A := (A_{\mu\nu})\) as follows
\[
S_{\mu\nu} := \frac{A_{\mu\nu} + A_{\nu\mu}}{2} = \frac{1}{2} \tr \left[ \rho_{in} \left( H' \sigma^\dagger \sigma^\mu + \sigma^\mu H' \sigma^\nu \right) \right].
\]
where the matrix \(S := (S_{\mu\nu})\) in the last line is defined by symmetrization of \(A := (A_{\mu\nu})\) as follows
\[
S_{\mu\nu} := \frac{A_{\mu\nu} + A_{\nu\mu}}{2} = \frac{1}{2} \tr \left[ \rho_{in} \left( H' \sigma^\dagger \sigma^\mu + \sigma^\mu H' \sigma^\nu \right) \right].
\]
Here, the \(S\) is obviously a real symmetric matrix since \(\sigma^\mu H' \sigma^\nu + \sigma^\nu H' \sigma^\mu\) is a Hermitian operator.

We write \(\tilde{S}\) as the lower right \(3 \times 3\) part in \(S\) matrix. Then, the quadratic form for Fraxis (more precisely, \(\pi\)-Fraxis algorithm) \([29]\) is represented with \(\tilde{S}\). Actually, Fraxis algorithm can deal with the single-qubit gate \(U_{\text{Fraxis}} := R_n(\pi)\) and the corresponding quaternion is \(q = (0, n)^T\), which is simply located on the two-dimensional spherical surface. Thus, from Eq. (A.1) we directly obtain the quadratic form for Fraxis as
\[
\tr \left[ \rho_{in} U_{\text{Fraxis}}^\dagger H' U_{\text{Fraxis}} \right] = n^\top \tilde{S} n = n^\top \begin{pmatrix}
\tr(H'\sigma_1 \rho_{in}^\dagger \sigma_1) & \frac{1}{2} [\tr(H' \sigma_1 \rho_{in}^\dagger \sigma_2) + \tr(H' \sigma_2 \rho_{in}^\dagger \sigma_1)] & \frac{1}{2} [\tr(H' \sigma_1 \rho_{in}^\dagger \sigma_3) + \tr(H' \sigma_3 \rho_{in}^\dagger \sigma_1)] \\
\frac{1}{2} [\tr(H' \sigma_2 \rho_{in}^\dagger \sigma_1) + \tr(H' \sigma_1 \rho_{in}^\dagger \sigma_2)] & \tr(H' \sigma_2 \rho_{in}^\dagger \sigma_2) & \frac{1}{2} [\tr(H' \sigma_2 \rho_{in}^\dagger \sigma_3) + \tr(H' \sigma_3 \rho_{in}^\dagger \sigma_2)] \\
\frac{1}{2} [\tr(H' \sigma_3 \rho_{in}^\dagger \sigma_1) + \tr(H' \sigma_1 \rho_{in}^\dagger \sigma_3)] & \frac{1}{2} [\tr(H' \sigma_3 \rho_{in}^\dagger \sigma_2) + \tr(H' \sigma_2 \rho_{in}^\dagger \sigma_3)] & \tr(H' \sigma_3 \rho_{in}^\dagger \sigma_3)
\end{pmatrix} n.
\]
As for the NFT \([28]\) and Rotosolve/Rotoselect \([26]\), they can optimize the axis-fixed single-qubit gate \(U_{\text{NFT}} := R_m(\psi)\), where \(m\) is a fixed rotational axis. Substituting the quaternion \(q = (\cos \psi/2, m \sin \psi/2)^T\) corresponding to this gate into the quadratic form of FQS, we obtain
\[
\tr \left[ \rho_{in} U_{\text{NFT}}^\dagger H' U_{\text{NFT}} \right] = S_{00} \cos^2 \frac{\psi}{2} + \sum_{i=1,2,3} 2 S_{0i} m_i \sin \frac{\psi}{2} \cos \frac{\psi}{2} \\
+ \sum_{i,j=1,2,3} \sin^2 \frac{\psi}{2} S_{ij} m_i m_j \\
= \left( \cos \frac{\psi}{2} \sin \frac{\psi}{2} \right) \begin{pmatrix}
S_{00} & \tilde{S}_0 \cdot m & m^\top \tilde{S} m \\
\tilde{S}_0 \cdot m & \tilde{S}_0 \cdot m & \tilde{S} m \\
m^\top \tilde{S} m & m^\top \tilde{S} m & \tilde{S}_0 \cdot m
\end{pmatrix} \left( \cos \frac{\psi}{2} \sin \frac{\psi}{2} \right) 
\]
(A.4)
where $\vec{S}_0 := (S_{01}, S_{02}, S_{03})$. Here, Rotoselect simply solve the minimization problems of the quadratic form for different axis $m \in \Lambda$ and select the optimal axis for minimizing the energy, where $\Lambda$ is a predefined subset of rotational axis such as $\Lambda = \{(1,0,0), (0,1,0), (0,0,1)\}$. 