Analytic gradients in variational quantum algorithms: Algebraic extensions of the parameter-shift rule to general unitary transformations

Artur F. Izmaylov [1,2,*] Robert A. Lang [1,2] and Tzu-Ching Yen [2]

[1]*Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada M1C 1A4*

[2]*Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 3H6*

(Received 26 July 2021; accepted 9 December 2021; published 29 December 2021)

Optimization of unitary transformations in variational quantum algorithms benefits highly from efficient evaluation of cost function gradients with respect to amplitudes of unitary generators. We propose several extensions of the parameter-shift rule to formulating these gradients as linear combinations of expectation values for generators with general eigenspectra (i.e., with more than two eigenvalues). Our approaches are exact and do not use any auxiliary qubits; instead they rely on a generator eigenspectrum analysis. Two main directions in the parameter-shift-rule extensions are (1) polynomial expansion of the exponential unitary operator based on a limited number of different eigenvalues in the generator and (2) decomposition of the generator as a linear combination of low-eigenvalue operators (e.g., operators with only two or three eigenvalues). These techniques have a range of scalings for the number of needed expectation values with the number of generator eigenvalues from quadratic (for polynomial expansion) to linear and even $\log_2$ (for generator decompositions). This allowed us to propose efficient differentiation schemes for commonly used two-qubit transformations (e.g., match gates, transmon gates, and fSim gates) and $\hat{S}^2$-conserving fermionic operators for the variational quantum eigensolver.

DOI: 10.1103/PhysRevA.104.062443

I. INTRODUCTION

Variational quantum algorithms (VQAs) currently provide the main route to employing noisy intermediate-scale quantum hardware without error correction to solve classically difficult optimization problems in quantum chemistry [1–3], information compression [4], machine learning [5–7], and number factorization [8]. The mathematical formulation of VQAs involves a cost function defined as follows:

$$E(\boldsymbol{\tau}) = \langle\bar{0}|\hat{U}^\dagger(\boldsymbol{\tau})\hat{H}\hat{U}(\boldsymbol{\tau})|\bar{0}\rangle, \tag{1}$$

where $\hat{H}$ is some Hermitian N-qubit operator (e.g., the quantum system Hamiltonian for quantum chemistry applications) and $\hat{U}(\boldsymbol{\tau})$ is a unitary transformation encoded on a quantum computer as a circuit operating on the initial state of N qubits $|\bar{0}\rangle \equiv |0\rangle^{\otimes N}$. To avoid deep circuits, $E(\boldsymbol{\tau})$ is optimized with respect to $\boldsymbol{\tau}$ components using a hybrid quantum-classical iterative process: (1) every set of $\boldsymbol{\tau}$ parameters is implemented on a quantum computer to measure the value of $E(\boldsymbol{\tau})$ and (2) results of quantum measurements are passed to a classical computer to suggest a next set of $\boldsymbol{\tau}$ parameters.

Naturally, this hybrid scheme becomes more efficient if a quantum computer can provide gradients of $E(\boldsymbol{\tau})$ with respect to $\boldsymbol{\tau}$ components. In molecular problems, analytical gradients with respect to circuit parameters are not only useful for variational energy optimization, but also in the calculation of analytical nuclear energy gradients and nonadiabatic couplings in variational quantum eigensolver extensions for excited states [9,10]. Usual parametrizations of unitary transformations are organized as products of exponential functions of some Hermitian generators $\{\hat{G}_k\}$:

$$\hat{U}(\boldsymbol{\tau}) = \prod_k \exp(i\tau_k\hat{G}_k). \tag{2}$$

The choice of efficient generators is generally a challenging problem the solution of which often relies on heuristics of a concrete field (e.g., in quantum chemistry there is a large variety of techniques developed recently [11–20]). Due to general noncommutativity of generators, τ_k gradients can be written as

$$\begin{aligned}\frac{\partial E}{\partial \tau_k} &= \frac{\partial}{\partial \tau_k}\langle\bar{0}|\hat{U}_1^\dagger e^{-i\tau_k\hat{G}_k}\hat{U}_2^\dagger\hat{H}\hat{U}_2 e^{i\tau_k\hat{G}_k}\hat{U}_1|\bar{0}\rangle \\ &= i\langle\bar{0}|\hat{U}_1^\dagger e^{-i\tau_k\hat{G}_k}[\hat{U}_2^\dagger\hat{H}\hat{U}_2, \hat{G}_k]e^{i\tau_k\hat{G}_k}\hat{U}_1|\bar{0}\rangle,\end{aligned} \tag{3}$$

where $\hat{U}_{1,2}$ are $\hat{U}$ parts on the left-hand and right-hand sides of the $\hat{G}_k$ exponent. Evaluating the gradient as the expectation value in Eq. (3) requires extra efforts to accommodate for nonsymmetric distribution of unitary transformations around $\hat{H}$ (considering the simplest case when $\hat{G}_k$ is also unitary). This treatment requires introducing an auxiliary qubit and controlled unitaries in the circuit, which enhance depth of the circuit [21].

It was found that in cases when $\hat{G}_k$ has only two eigenvalues symmetrically distributed, $\{\pm\lambda\}$, the so-called parameter-shift rule (PSR) is applicable to Eq. (3) [21,22]:

$$\begin{aligned}\frac{\partial E}{\partial \tau_k} = \lambda\Big(&\langle\bar{0}|\hat{U}_1^\dagger e^{-i(\tau_k+s)\hat{G}_k}\hat{U}_2^\dagger\hat{H}\hat{U}_2 e^{i(\tau_k+s)\hat{G}_k}\hat{U}_1|\bar{0}\rangle \\ &-\langle\bar{0}|\hat{U}_1^\dagger e^{-i(\tau_k-s)\hat{G}_k}\hat{U}_2^\dagger\hat{H}\hat{U}_2 e^{i(\tau_k-s)\hat{G}_k}\hat{U}_1|\bar{0}\rangle\Big),\end{aligned} \tag{4}$$

*artur.izmaylov@utoronto.ca

where $s = \pi/(4\lambda)$. Clearly, this approach allows one to evaluate the expectation values in Eq. (4) using the same circuit as for $E(\boldsymbol{\tau})$ with only minor modifications of $\boldsymbol{\tau}$ parameters.

A natural question is how to extend the PSR to a general unitary transformation containing generators with more than two eigenvalues. The algebraic form of these extensions is set to be a linear combination of expectation values of Eq. (4). Such extensions are motivated by active developments in hardware (two-qubit gates [23,24] generally have four eigenvalues) and theory related to specific problems (new generators for solving quantum chemistry problems, e.g., spin-adapted fermionic rotations). Also, these extensions will allow one to reduce the number of optimized parameters if more complex generators can be considered.

Generators with more than two eigenvalues can naturally be decomposed to generators with two eigenvalues for which the PSR can be applied individually, as suggested by Crooks [25]. A few approaches to such decompositions were considered for some standard two-qubit gates [25]. But there was no attempt to systematically address the minimization of the number of terms in such decompositions or their extensions beyond two-qubit operators.

A naive application of the Ref. [25] decomposition scheme to generators of the unitary coupled cluster (UCC) approach can lead to exponential growth of the number of terms (Pauli products) with two eigenvalues [26]. Using a specifically tailored decomposition of UCC generators (fermionic-shift rule) Ref. [26] was able to address the exponential growth of expectation values needed for gradient evaluation in the Pauli product decomposition of UCC generators.

Here we demonstrate how to do the generator decomposition systematically and how to avoid cases of exponential increase of terms in such decompositions. We provide three generalizations of the PSR based on somewhat different algebraic ideas the main unifying theme of which is consideration of the generator eigenspectrum. In the first approach, we use the fact that the exponential function of the generator with K eigenvalues can be presented as a $K-1$ degree polynomial. This allows us to extend the PSR by using a larger number of expectation values in a linear combination to cancel all higher powers of the generator. In the second approach, we decompose the generator into a sum of commuting operators with a smaller number of unique eigenvalues. The third approach uses a decomposition over noncommutative operators with a low number of eigenvalues. While the first approach can be seen as a generalization of Ref. [26] to generators beyond those found in UCC, the second and third approaches have no apparent connections with previous works on efficient evaluation of gradients.

Note that even though there are multiple generalizations of the PSR to higher derivatives for amplitudes of generators with two eigenvalues [27,28] and stochastic techniques for gradients of an arbitrary generator [29], they will not be considered here, since our focus is on deterministic expression for gradients involving expectation values of Hermitian operators.

One problem that is related to using analytical gradients in variational algorithms is the problem of barren plateaus (exponentially vanishing gradients) [30–32]. The theoretical developments discussed in this paper do not consider this problem, since it is related to specifics of the variational problem and a choice of generators.

II. THEORY

A. Polynomial expansion

Energy partial derivatives [Eq. (3)] can be rewritten as

$$\frac{\partial E}{\partial \tau} = i\langle e^{-i\tau\hat{G}}\hat{H}_2\hat{G}e^{i\tau\hat{G}}\rangle - i\langle \hat{G}e^{-i\tau\hat{G}}\hat{H}_2e^{i\tau\hat{G}}\rangle, \tag{5}$$

where we removed the k subscript for simplicity and introduced short notation: $\langle \ldots \rangle = \langle \bar{0}|\hat{U}_1^\dagger \ldots \hat{U}_1|\bar{0}\rangle$, $\hat{H}_2 = \hat{U}_2^\dagger \hat{H}\hat{U}_2$. Asymmetry of operators' placement around $\hat{H}$ and potential nonunitarity of $\hat{G}$ make the obtained expectation values more challenging to measure. However, if $\hat{G}$ has a finite number of different eigenvalues, there is a way to rewrite this difference as a linear combination of terms measurable on a quantum computer without any modifications of the $E(\boldsymbol{\tau})$ measurement scheme:

$$\frac{\partial E}{\partial \tau} = \sum_n C_n \langle e^{-i(\tau+\theta_n)\hat{G}}\hat{H}_2 e^{i(\tau+\theta_n)\hat{G}}\rangle, \tag{6}$$

where θ_n and C_n are coefficients to be defined. Details on obtaining this generalization of the PSR are given in the Appendix. The key quantity that defines θ_n and C_n is the number of different eigenvalues in $\hat{G}$, which will be denoted as L. L defines a finite polynomial expression for the exponential operator

$$e^{i\theta\hat{G}} = \sum_{n=0}^{L-1} a_n(\theta)(i\hat{G})^n, \tag{7}$$

where $a_n(\theta)$'s are constants obtained by solving a linear system of equations, and C_n's are evaluated from another system of linear equations using $a_n(\theta)$'s with fixed θ_n's (see the Appendix for further details). To illustrate the process in the simplest case of $L = 2$, where $\hat{G}$'s eigenvalues are ± 1, hence $\hat{G}^2 = \hat{1}$ and

$$e^{i\theta\hat{G}} = a_0(\theta)\hat{1} + a_1(\theta)(i\hat{G}), \tag{8}$$

where $a_0(\theta) = \cos(\theta)$ and $a_1(\theta) = \sin(\theta)$. To obtain the energy derivative we need only two terms in Eq. (6):

$$\frac{\partial E}{\partial \tau} = \frac{1}{\sin(2\theta)}\Big[\langle e^{-i(\tau+\theta)\hat{G}}\hat{H}_2 e^{i(\tau+\theta)\hat{G}}\rangle - \langle e^{-i(\tau-\theta)\hat{G}}\hat{H}_2 e^{i(\tau-\theta)\hat{G}}\rangle\Big]. \tag{9}$$

For example, $\hat{G}$ satisfying the described conditions can be any tensor product of Pauli operators for different qubits.

For $L = 3$ with symmetric spectrum $\{0, \pm 1\}$, the Appendix shows that four expectation values are enough to obtain the analytic gradient. It was shown recently that all fermionic operators $\hat{G} = \hat{a}_p^\dagger \ldots \hat{a}_q^\dagger \hat{a}_r \ldots \hat{a}_s - \hat{a}_s^\dagger \ldots \hat{a}_r^\dagger \hat{a}_q \ldots \hat{a}_p$ used in the UCC method have this spectrum [26]. Techniques developed in Ref. [26] also found the gradient expressions requiring four expectation values for such operators, and were able to reduce it to only two expectation values for real unitary transformations acting on real wave functions. The polynomial expansion

can be seen as a generalization of Ref. [26] to generators beyond the fermionic rotations used in the UCC method.

The polynomial expansion for general $\hat{G}$ with L eigenvalues will produce the gradient expression with the number of expectation values that scales as $\approx L^2$. If there are some relations between different eigenvalues, they can be used to reduce the number of expectation values by exploiting freedom in the choice of θ_n and C_n parameters (see the Appendix for more details).

B. Generator decompositions

To address the $\approx L^2$ scaling of the number of expectation values in the polynomial expansion approach, instead of Eq. (6) we will use the following alternative:

$$\frac{\partial E}{\partial \tau} = \sum_n C_n \langle e^{-i\theta_n \hat{O}_n} e^{-i\tau \hat{G}} \hat{H}_2 e^{i\tau \hat{G}} e^{i\theta_n \hat{O}_n} \rangle, \tag{10}$$

where we introduced new operators $\{\hat{O}_n\}$. $\hat{O}_n$'s are required to have only a few eigenvalues (two or three) and to sum to $\hat{G}$:

$$\hat{G} = \sum_{n=1}^{K} d_n \hat{O}_n, \tag{11}$$

where d_n are real coefficients.

1. Involutory example

To illustrate how $\{\hat{O}_n\}$ decomposition can be used in the gradient evaluation, let us assume that $\hat{O}_n$'s have only two eigenvalues ± 1. To define C_n and θ_n let us consider the following pairs:

$$\begin{aligned} &\langle e^{-i\theta_n \hat{O}_n} e^{-i\tau \hat{G}} \hat{H}_2 e^{i\tau \hat{G}} e^{i\theta_n \hat{O}_n} \rangle - \langle e^{i\theta_n \hat{O}_n} e^{-i\tau \hat{G}} \hat{H}_2 e^{i\tau \hat{G}} e^{-i\theta_n \hat{O}_n} \rangle \\ &= i \sin(2\theta_n) \Big[\langle e^{-i\tau \hat{G}} \hat{H}_2 e^{i\tau \hat{G}} \hat{O}_n \rangle - \langle \hat{O}_n e^{-i\tau \hat{G}} \hat{H}_2 e^{i\tau \hat{G}} \rangle \Big]. \end{aligned} \tag{12}$$

Here, we used the involutory property of $\{\hat{O}_n\}$ to convert their exponents according to Eq. (8). This consideration shows that to obtain the energy derivative via expansion in Eq. (10) we should select $\pm\theta_n$ pairs with coefficient $C_{\pm n} = d_n / \sin(\pm 2\theta_n)$. The number of the expectation values in Eq. (10) is $2K$. Unfortunately, K depends not only on the number of $\hat{G}$ eigenvalues but also on their distribution and degeneracies (or multiplicities). However, it is easy to formulate the best case scenario where $K = \log_2(L)$, where adding $K\hat{O}_n$ operators produces $\hat{G}$ the spectrum of which has 2^K eigenvalues $\{\lambda_j\}$:

$$\lambda_j = \sum_{n=1}^{K} d_n b_{nj}, \; b_{nj} = \{\pm 1\}. \tag{13}$$

Starting with some $\hat{G}$, it is not necessary that its eigenvalues will be encoded so efficiently with involutory operators $\hat{O}_n$'s, yet this best case scenario shows great potential for the decomposition approach.

2. Efficient generator decompositions

$\hat{O}_n$ operators optimal for the generator decomposition depend on the spectrum of $\hat{G}$. We assume that $\hat{G}$ can be written in terms of a few qubit or fermionic operators. The number of involved qubits or fermionic spin orbitals should not exceed the limit when the dimensionality of a faithful representation for involved operators becomes too large to do matrix algebra on a classical computer.

The mathematical basis for representing $\hat{G}$ as matrix $\mathbf{G}$ is that qubit or fermionic operators expressing $\hat{G}$ can be considered as basis elements of a Lie algebra. Using a faithful representation of this Lie algebra one can work with corresponding matrices instead of operators. $\mathbf{G}$ can be diagonalized $\mathbf{G} = \mathbf{V}^\dagger \mathbf{D} \mathbf{V}$ to obtain the guidance on choice of optimal $\hat{O}_n$'s. To minimize the number of $\hat{O}_n$ operators, one would build them from decomposition $\mathbf{D} = \sum_n \mathbf{D}_n$, where $\mathbf{D}_n$ are diagonal matrices with a few (two to three) different eigenvalues. Then, $\hat{O}_n$ is obtained via the inverse representation map of $\mathbf{O}_n = \mathbf{V}^\dagger \mathbf{D}_n \mathbf{V}$. The caveat is that even the decomposition of the diagonal matrix $\mathbf{D}$ can be done in various ways differing in the number of necessary $\mathbf{O}_n$'s. A simple example illustrating various possibilities is

$$\begin{aligned} &\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \\ &= 3(\mathbf{P}_1 - \mathbf{P}_2) + (\mathbf{P}_4 - \mathbf{P}_3), \end{aligned} \tag{14}$$

where $\mathbf{P}_j$ is a 4×4 matrix with 1 on the (j, j)th element and zeros everywhere else. In this example, the most optimal choice is the second expansion, two operators with two symmetric eigenvalues each. It also shows that even though the eigensubspace projector expansion [last in Eq. (14)] is the most straightforward, it is not necessarily the most optimal.

Using the form of the $\mathbf{D}_n$ matrix one can optimize the number and the form of $\hat{O}_n$ operators for a particular generator. The result of this optimization is not explicitly representable in some closed form for an arbitrary generator. Instead, here we provide several heuristics that can generate shorter expansions than those from the eigensubspace projector expansion; the latter can always be used as a conservative option.

3. Commutative Cartan subalgebra decomposition

The basis of our algebraic heuristics is a Cartan subalgebra (CSA) decomposition for $\hat{G}$ [33]. This decomposition can be done for an element of any compact Lie algebra. Here we will use it for $\hat{G}$ realized as an element of the N-qubit operator algebra, $\mathfrak{su}(2^N)$:

$$\hat{G} = \sum_n C_n \hat{P}_n, \; \hat{P}_n = \prod_{j=1}^{N} \hat{\sigma}_j, \tag{15}$$

where C_n are coefficients, and $\hat{\sigma}_j = \{\hat{x}_j, \hat{y}_j, \hat{z}_j, \hat{\mathbb{1}}_j\}$ is one of the Pauli operators or identity for the jth qubit. $\mathfrak{su}(2^N)$

contains $4^N - 1$ generators $\hat{P}_n$ due to the exclusion of the tensor product of N identity operators. The largest Abelian (or Cartan) subalgebra in $\mathfrak{su}(2^N)$ that we will involve in the decomposition is a set of $\hat{P}_n$'s that contain only $\hat{z}_j$ operators, denoted as $\hat{Z}_n$'s. $\hat{Z}_n$'s have only two distinct eigenvalues, ± 1, which is convenient for our decomposition. The CSA decomposition of $\hat{G}$ is

$$\hat{G} = \hat{V}^\dagger \left(\sum_{n=1}^{K} c_n \hat{Z}_n \right) \hat{V}, \tag{16}$$

where c_n are coefficients and $\hat{V}$ is a unitary transformation:

$$\hat{V} = \prod_k e^{i\tau_k \hat{P}_k}, \tag{17}$$

where τ_k are real amplitudes and $\hat{P}_k$'s are all Pauli products that are not in the CSA. Clearly, each term in the sum of Eq. (16) has eigenvalues $\pm c_n$, therefore we can choose each $\hat{O}_n = c_n \hat{V}^\dagger \hat{Z}_n \hat{V}$.

The CSA decomposition in Eq. (16) can be done by expanding the left- and right-hand sides of Eq. (16) in a basis of $\mathfrak{su}(2^N)$ Lie algebra of $\hat{P}_n$'s to find coefficients c_n and amplitudes τ_k for $\hat{V}$. This decomposition is unique in terms of the number of $\hat{Z}_n$ terms, which suits our purpose to obtain the number of two-eigenvalue operators in the $\hat{G}$ decomposition.

Since all $\hat{O}_n$ operators commute, one can rewrite the gradient expression as an application of the PSR to each $\hat{O}_n$ operator in $\hat{G}$:

$$\frac{\partial E}{\partial \tau} = \sum_n \frac{1}{\sin(2\theta_n)} \Big[\langle e^{-i(\tau\hat{G}+\theta_n\hat{O}_n)} \hat{H}_2 e^{i(\tau\hat{G}+\theta_n\hat{O}_n)} \rangle - \langle e^{-i(\tau\hat{G}-\theta_n\hat{O}_n)} \hat{H}_2 e^{i(\tau\hat{G}-\theta_n\hat{O}_n)} \rangle \Big]. \tag{18}$$

The involved unitary transformations can be rewritten as

$$e^{\pm i(\tau\hat{G}\pm\theta_n\hat{O}_n)} = \hat{V}^\dagger \prod_{m=1}^{K} e^{\pm i c_m(\tau \pm \delta_{nm}\theta_n)\hat{Z}_m} \hat{V}. \tag{19}$$

This form is convenient for implementation of these operators as a circuit.

4. Noncommutative Cartan subalgebra decomposition

An alternative representation of $\hat{G}$ is a sum of noncommuting two-eigenvalue operators:

$$\hat{G} = \sum_{n=1}^{K'} c_n \hat{V}_n^\dagger \hat{Z}_n \hat{V}_n, \tag{20}$$

where $\hat{V}_n$'s are defined in the same way as $\hat{V}$. This decomposition defines $\hat{O}_n = c_n \hat{V}_n^\dagger \hat{Z}_n \hat{V}_n$, and due to differences in $\hat{V}_n$'s, different $\hat{O}_n$'s do not necessarily commute. The main advantage of the noncommutative decomposition is that it uses not only coefficients c_n for reproducing the spectrum of $\hat{G}$ but also some parameters in $\hat{V}_n$'s, $\lambda_j = \lambda_j(\{V_n\}_{n=1}^{K'}, \{c_n\}_{n=1}^{K'})$. This dependence provides an opportunity for the noncommutative decomposition to represent $\hat{G}$ with a lower number of terms $K' < K$ [cf. Eqs. (16) and (20)].

To construct the noncommutative decomposition we fix the number of terms K' to values lower than K in Eq. (16) and minimize the difference between the left- and right-hand sides of Eq. (20) using c_n and $\tau_k^{(n)}$ (amplitudes of $\hat{V}_n$). The choice of $\hat{Z}_n$ in Eq. (20) is insignificant because $\hat{V}_n$ can always transform one CSA operator into another.

Noncommutativity of $\hat{O}_n$ operators does not preclude use of the shift rule to each $\hat{O}_n$ operator to obtain components of the derivative for the $\hat{G}$ amplitude:

$$\frac{\partial E}{\partial \tau} = \sum_n \frac{1}{\sin(2\theta_n)} \Big[\langle e^{-i\theta_n\hat{O}_n} e^{-i\tau\hat{G}} \hat{H}_2 e^{i\tau\hat{G}} e^{i\theta_n\hat{O}_n} \rangle - \langle e^{i\theta_n\hat{O}_n} e^{-i\tau\hat{G}} \hat{H}_2 e^{i\tau\hat{G}} e^{-i\theta_n\hat{O}_n} \rangle \Big]. \tag{21}$$

To measure such expectation values there is overhead related to noncompatibility of eigenstates for individual $\hat{O}_n$ and $\hat{G}$. Thus, one needs to explore for each class of $\hat{G}$ operators, whether the potential reduction in the number of terms in Eq. (20) is not diminished by a possible higher circuit depth.

III. APPLICATIONS

We will consider application of the generator decompositions for gradient evaluations of several classes of challenging operators: (1) two-qubit generators, (2) three-qubit generators, and (3) generators of $\hat{S}^2$-conserving fermionic rotations. Our choice was motivated not only by inapplicability of the PSR for these generators due to the multitude of eigenvalues but also because advantages of all three decomposition techniques can be illustrated on them. To compare results of the proposed decompositions with a previous general scheme from Ref. [21], we start this section with reviewing the latter.

A. Gradients via linear combination of unitaries

Denoting $\hat{V} = e^{i\tau\hat{G}}$ and $\partial_\tau \hat{V} = i\hat{G}e^{i\tau\hat{G}}$, one can rewrite Eq. (5) as

$$\frac{\partial E}{\partial \tau} = \langle \hat{V}^\dagger \hat{H}_2 \partial_\tau \hat{V} \rangle + \langle \partial_\tau \hat{V}^\dagger \hat{H}_2 \hat{V} \rangle. \tag{22}$$

To use a measurement scheme introduced in Ref. [21], $\partial_\tau \hat{V}$ needs to be decomposed as a linear combination of unitaries (LCU). Since $e^{i\tau\hat{G}}$ is already a unitary operation, the decomposition is only needed for the $i\hat{G}$ part:

$$\partial_\tau \hat{V} = \sum_{k=1}^{K_u} c_k \hat{W}_k e^{i\tau\hat{G}}, \tag{23}$$

where $i\hat{G}$ is decomposed in linear combinations of unitaries $\hat{W}_k$ with coefficients c_k. A typical choice of $\hat{W}_k$'s is a set of Pauli products $i\hat{P}_k$ comprising $\hat{G}$. Also, one can use the commutative CSA decomposition Eq. (16) to obtain a potentially more compact set of $\hat{W}_k$'s. Note though that the decomposition in Eq. (23) is less flexible than the one in Eq. (11) because the latter does not require unitarity of $\hat{O}_n$ operators.

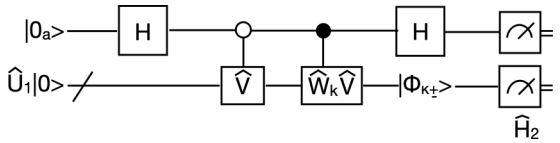

FIG. 1. Measurement scheme for terms in Eq. (25). States $|\Phi_{k\pm}\rangle = (\hat{V} \pm \hat{W}_k\hat{V})\hat{U}_1|0\rangle/(2\sqrt{p_\pm})$ are obtained after measurement of the ancilla qubit, where $p_\pm = \langle 0|\hat{U}_1^\dagger(\hat{V} \pm \hat{W}_k\hat{V})^\dagger(\hat{V} \pm \hat{W}_k\hat{V})\hat{U}_1|0\rangle/4$ is the probability for the ancilla qubit results to be ± 1, respectively. H is the Hadamard gate.

The LCU decomposition for $\partial_\tau \hat{V}$ allows one to rewrite

$$\frac{\partial E}{\partial \tau} = \sum_{k=1}^{K_u} c_k[\langle \hat{V}^\dagger \hat{H}_2 \hat{W}_k \hat{V}\rangle + \text{c.c.}] \tag{24}$$

$$= \frac{1}{2}\sum_{k=1}^{K_u} c_k[\langle \hat{V}^\dagger(1+\hat{W}_k)^\dagger \hat{H}_2(1+\hat{W}_k)\hat{V}\rangle - \langle \hat{V}^\dagger(1-\hat{W}_k)^\dagger \hat{H}_2(1-\hat{W}_k)\hat{V}\rangle]. \tag{25}$$

Each term in square brackets of Eq. (25) can be obtained via quantum measurement using the circuit depicted in Fig. 1. Extra features required for the circuit on Fig. 1 are one ancilla qubit and controlled versions of unitaries. These features are not needed for a regular circuit measuring expectation values in Eqs. (6) and (10). In what follows we will only compare the number of expectation values needed to be measured in the LCU scheme and the proposed approaches. If this number is lower or the same in our schemes, absence of the ancilla qubit and controlled unitaries in our schemes makes these approaches more advantageous than the LCU scheme.

B. Two-qubit generators

Any two-qubit generator has not more than four different eigenvalues, and thus the eigenvalue decomposition scheme will need eight expectation values for a gradient evaluation. The CSA decomposition [Eq. (16)] of any two-qubit generator results in at most three $\hat{Z}_n$'s ($\hat{z}_1, \hat{z}_2, \hat{z}_1\hat{z}_2$), which leads to not more than six expectation values for each gradient. In all considered two-qubit gates, commuting and noncommuting CSA decompositions provided the same number of $\hat{O}_n$'s. Since all $\hat{O}_n$'s provided by CSA decompositions are unitary, they can be used in the LCU scheme. Therefore, the number of terms required for measurements in the LCU and our schemes is the same, but the LCU scheme will require an extra qubit and controlled versions of unitary operations.

1. Transmon gates

These gates are generated by [23]

$$\hat{G} = \hat{x}_1 - b\hat{z}_1\hat{x}_2 + c\hat{x}_2. \tag{26}$$

Applying $\hat{W}(\tau) = \exp(i\tau\hat{y}_1\hat{x}_2)$ to each term of $\hat{G}$,

$$\hat{W}^\dagger(\tau)\hat{x}_1\hat{W}(\tau) = \cos(2\tau)\hat{x}_1 - \sin(2\tau)\hat{z}_1\hat{x}_2, \tag{27}$$

$$\hat{W}^\dagger(\tau)\hat{z}_1\hat{x}_2\hat{W}(\tau) = \cos(2\tau)\hat{z}_1\hat{x}_2 + \sin(2\tau)\hat{x}_1, \tag{28}$$

$$\hat{W}^\dagger(\tau)\hat{x}_2\hat{W}(\tau) = \hat{x}_2, \tag{29}$$

one can choose τ_0 so that $\cos(2\tau_0) = 1/\sqrt{1+b^2}$ and $\sin(2\tau_0) = b/\sqrt{1+b^2}$, then $\hat{G}$ can be represented as

$$\hat{G} = \hat{W}^\dagger(\tau_0)[\sqrt{1+b^2}\hat{x}_1 + c\hat{x}_2]\hat{W}(\tau_0). \tag{30}$$

To arrive at the form of Eq. (16), $\hat{V}$ needs to be defined as

$$\hat{V} = e^{i\frac{\pi}{4}(\hat{y}_1+\hat{y}_2)}\hat{W}(\tau_0) \tag{31}$$

and then $\hat{O}_1 = \sqrt{1+b^2}\hat{V}^\dagger\hat{z}_1\hat{V}$ and $\hat{O}_2 = c\hat{V}^\dagger\hat{z}_2\hat{V}$. This decomposition allows one to evaluate the gradient using only four expectation values.

2. Match gates

Generators of these gates are linear combinations of the following operators [18,34]:

$$\{\hat{x}_1\hat{x}_2,\ \hat{y}_1\hat{y}_2,\ \hat{x}_1\hat{y}_2,\ \hat{y}_1\hat{x}_2,\ \hat{z}_1,\ \hat{z}_2\}. \tag{32}$$

This set forms a subalgebra of $\mathfrak{su}(4)$ that is a direct sum of two $\mathfrak{su}(2)$ algebras:

$$\mathcal{A}_1 = \left\{\frac{\hat{z}_1+\hat{z}_2}{2}, \frac{\hat{x}_1\hat{y}_2+\hat{y}_1\hat{x}_2}{2}, \frac{\hat{x}_1\hat{x}_2-\hat{y}_1\hat{y}_2}{2}\right\}, \tag{33}$$

$$\mathcal{A}_2 = \left\{\frac{\hat{z}_1-\hat{z}_2}{2}, \frac{\hat{y}_1\hat{x}_2-\hat{x}_1\hat{y}_2}{2}, \frac{\hat{x}_1\hat{x}_2+\hat{y}_1\hat{y}_2}{2}\right\}. \tag{34}$$

Each $\mathfrak{su}(2)$ has only one Cartan element. The CSA decomposition of any match-gate generator provides two $\hat{O}_n$'s, which are results of conjugation of two CSA elements, $\hat{z}_1$ and $\hat{z}_2$, with unitaries ($\hat{V}$'s) from the two $SU(2)$ groups corresponding to the $\mathfrak{su}(2)$ algebras.

3. fSim gates

The fSim gate generator [24,35] is

$$\hat{G}_{\text{fSim}} = \frac{\theta}{2}(\hat{x}_1\hat{x}_2 + \hat{y}_1\hat{y}_2) + \frac{\phi}{4}(1-\hat{z}_1)(1-\hat{z}_2). \tag{35}$$

Its CSA decomposition results in three $\hat{Z}_n$'s, therefore to do gradients with respect to the overall amplitude τ in $\exp(i\tau\hat{G}_{\text{fSim}})$ will require six expectation values. $\hat{G}_{\text{fSim}}$ can be split into

$$\hat{O}_1 = \frac{\theta}{2}(\hat{x}_1\hat{x}_2 + \hat{y}_1\hat{y}_2), \tag{36}$$

$$\hat{O}_2 = \frac{\phi}{4}(1-\hat{z}_1)(1-\hat{z}_2), \tag{37}$$

which have three and two eigenvalues, respectively, thus the θ and ϕ gradients of $\exp[i\hat{G}_{\text{fSim}}(\theta,\phi)]$ will require four and two expectation values.

C. Three-qubit generators

Considering the two-qubit generators, we were not able to find a case where the noncommutative CSA decomposition had an advantage over the commutative CSA scheme. Here, we give an example of a three-qubit transformation where this advantage is clear. Consider a generator

$$\hat{G} = \hat{U}^\dagger\hat{z}_1\hat{U} + \hat{z}_2 \tag{38}$$

that requires only two $\hat{O}_n$'s using the noncommuting scheme. We choose a three-qubit unitary $\hat{U} = \exp(\hat{A})$, where $\hat{A}$ is an anti-Hermitian operator with the following matrix representation: $A_{i,i} = 0$, $A_{i,j<i} = 1$, and $A_{i,j>i} = -1$. The CSA

decomposition of $\hat{G}$

$$\hat{G} \approx \hat{V}^\dagger(1.250\hat{z}_1\hat{z}_2 + 0.045\hat{z}_1\hat{z}_2\hat{z}_3 + 0.014\hat{z}_1\hat{z}_3 + 0.658\hat{z}_2 - 0.045\hat{z}_2\hat{z}_3 + 0.014\hat{z}_3)\hat{V} \quad (39)$$

indicates that there are at least six $\hat{O}_n$'s (12 expectation values) for the commutative decomposition scheme.

D. $\hat{S}^2$-conserving fermionic generators

One of the approaches to construct a pool of generators for application of VQAs to solving the electronic structure problems is adding symmetry conserving conditions [36]. Usual UCC single and double operators

$$\hat{\kappa}_i^a = a_a^\dagger a_i - a_i^\dagger a_a, \quad (40)$$

$$\hat{\kappa}_{ji}^{ab} = a_a^\dagger a_b^\dagger a_i a_j - a_j^\dagger a_i^\dagger a_b a_a \quad (41)$$

conserve the number of electrons but generally not the electron spin. Unitary generators that commute with the electron-spin operators, $\hat{S}_z$ and $\hat{S}^2$, can be obtained by anti-Hermitization of singlet spherical tensor operators [37]. A general spherical tensor operator, $\hat{T}^{S,M}$, is defined as

$$[\hat{S}_\pm, \hat{T}^{S,M}] = \sqrt{S(S+1) - M(M\pm 1)}\hat{T}^{S,M\pm 1}, \quad (42)$$

$$[\hat{S}_z, \hat{T}^{S,M}] = M\hat{T}^{S,M}, \quad (43)$$

where S and M are the electron spin and its projection to the z axis, respectively. Equation $\hat{S}^2 = \hat{S}_-\hat{S}_+ + \hat{S}_z(\hat{S}_z + 1)$ can be used to show that any singlet spherical tensor operator $\hat{T}^{0,0}$ will commute with $\hat{S}_z$ and $\hat{S}^2$.

There are standard approaches for producing spherical tensor operators [37]; they involve very similar techniques to those used for generating spin-adapted configuration state functions [38,39]. Individual single excitations are not $\hat{T}^{0,0}$ operators; therefore, one needs to group more than one excitation to obtain singlet operators:

$$\hat{T}_{ia}^{0,0} = \hat{\kappa}_{i_\alpha}^{a_\alpha} + \hat{\kappa}_{i_\beta}^{a_\beta}, \quad (44)$$

where $a_\alpha(a_\beta)$ and $i_\alpha(i_\beta)$ are the spin orbitals arising from the $\alpha(\beta)$ spin parts of the ath and ith spatial orbitals. For double and higher excitations or deexcitations, the seniority number Ω (the number of unpaired electrons created by the operator) correlates well with the number of individual excitation or deexcitation pairs in construction of singlet operators:

$$\Omega = 0: \quad \hat{T}_{iiaa}^{0,0} = \hat{\kappa}_{i_\alpha i_\beta}^{a_\alpha a_\beta}, \quad (45)$$

$$\Omega = 2: \quad \hat{T}_{iiab}^{0,0} = \hat{\kappa}_{i_\alpha i_\beta}^{a_\alpha b_\beta} + \hat{\kappa}_{i_\alpha i_\beta}^{a_\beta b_\alpha}, \quad (46)$$

$$\Omega = 2: \quad \hat{T}_{ijaa}^{0,0} = \hat{\kappa}_{i_\alpha j_\beta}^{a_\alpha a_\beta} + \hat{\kappa}_{i_\beta j_\alpha}^{a_\alpha a_\beta}, \quad (47)$$

$$\Omega = 4: \quad \hat{T}_{ijab}^{0,0} = \sum_{s,\bar{s}\in\{\alpha,\beta\}} \hat{\kappa}_{i_s j_{\bar{s}}}^{a_{\bar{s}} b_s}. \quad (48)$$

Generators in Eqs. (44)–(48) are required for spin-conserving UCC single and double *Ansätze*. Note that these generators can be used to add electronic correlation to an initial state of any electron-spin symmetry (not necessarily closed-shell singlet) without altering the spin state.

The spectra of the spin-conserving generators are reported in Table I. It is important to note that the zero eigenvalue has much higher multiplicity than the nonzero eigenvalues for the single and double spherical tensor operators. Due to large differences between multiplicities of different eigenvalues in the singlet operators' spectra, their decomposition following Eq. (16) was found to be inefficient in K. We found that it usually takes a lot of $\{\pm 1\}$-eigenvalued operators to create large variations in eigenvalues' multiplicities. Furthermore, due to parity symmetry of the spectra, it is natural to introduce alternative $\hat{O}_n$'s in Eq. (10) which have three eigenvalues $\{0, \pm\lambda_n\}$. Explicit forms of the $\hat{O}_n$ operators for each singlet spherical operator are as follows:

TABLE I. The eigenvalues and number of Pauli products for the singlet single and double fermionic operators. Multiplicities are provided as subscripts for eigenvalues.

Operators	Eigenvalues	Number of $\hat{P}_k$'s
$\hat{T}_{ia}^{0,0}$	$\{0_6, \pm i_4, \pm i2_1\}$	4
$\hat{T}_{iiaa}^{0,0}$	$\{0_{14}, \pm i_1\}$	8
$\hat{T}_{ijaa}^{0,0}$, $\hat{T}_{iiab}^{0,0}$	$\{0_{52}, \pm i_4, \pm i\sqrt{2}_2\}$	16
$\hat{T}_{ijab}^{0,0}$	$\{0_{186}, \pm i_{16}, \pm i\sqrt{2}_{16}, \pm i2_2, \pm i2\sqrt{2}_1\}$	32

$$\hat{T}_{ia}^{0,0} = \hat{O}_1 + \hat{O}_2:$$

$$\lambda \in \{0, \pm i\}: \hat{O}_1 = \hat{\kappa}_{i_\alpha}^{a_\alpha}\left(\hat{n}_{i_\beta} - \hat{n}_{a_\beta}\right)^2 + \hat{\kappa}_{i_\beta}^{a_\beta}(\hat{n}_{i_\alpha} - \hat{n}_{a_\alpha})^2,$$

$$\lambda \in \{0, \pm i2\}: \hat{O}_2 = \hat{T}_{ia}^{0,0} - \hat{O}_1,$$

$$\hat{T}_{iiab}^{0,0} = \hat{O}_1 + \hat{O}_2:$$

$$\lambda \in \{0, \pm i\}: \hat{O}_1 = \hat{\kappa}_{i_\alpha i_\beta}^{a_\beta b_\alpha}\left(\hat{n}_{a_\alpha} - \hat{n}_{b_\beta}\right)^2 + \hat{\kappa}_{i_\alpha i_\beta}^{a_\alpha b_\beta}\left(\hat{n}_{a_\beta} - \hat{n}_{b_\alpha}\right)^2,$$

$$\lambda \in \{0, \pm i\sqrt{2}\}: \hat{O}_2 = \hat{T}_{iiab}^{0,0} - \hat{O}_1,$$

$$\hat{T}_{ijaa}^{0,0} = \hat{O}_1 + \hat{O}_2:$$

$$\lambda \in \{0, \pm i\}: \hat{O}_1 = \hat{\kappa}_{i_\beta j_\alpha}^{a_\alpha a_\beta}\left(\hat{n}_{i_\alpha} - \hat{n}_{j_\beta}\right)^2 + \hat{\kappa}_{i_\alpha j_\beta}^{a_\alpha a_\beta}\left(\hat{n}_{i_\beta} - \hat{n}_{j_\alpha}\right)^2,$$

$$\lambda \in \{0, \pm i\sqrt{2}\}: \hat{O}_2 = \hat{T}_{ijaa}^{0,0} - \hat{O}_1,$$

$$\hat{T}_{ijab}^{0,0} = \sum_{i=1}^{4}\hat{O}_i:$$

$$\lambda \in \{0, \pm i2\}: \hat{O}_1 = \sum_{s,\bar{s}\in\{\alpha,\beta\}} \hat{\kappa}_{i_s j_{\bar{s}}}^{a_{\bar{s}} b_s}\Big[\hat{n}_{a_s}\hat{n}_{i_{\bar{s}}}(1 - \hat{n}_{j_s}) + \hat{n}_{b_{\bar{s}}}\hat{n}_{j_s}(1 - \hat{n}_{i_{\bar{s}}}) - \hat{n}_{a_s}\hat{n}_{b_{\bar{s}}}(\hat{n}_{i_{\bar{s}}} - \hat{n}_{j_s})^2\Big],$$

$$\lambda \in \{0, \pm i2\sqrt{2}\}: \hat{O}_2 = \sum_{s,\bar{s}\in\{\alpha,\beta\}} \hat{\kappa}_{i_s j_{\bar{s}}}^{a_{\bar{s}} b_s}\Big[\hat{n}_{i_{\bar{s}}}\hat{n}_{j_s}(1 - \hat{n}_{a_s}) + \hat{n}_{a_s}\hat{n}_{b_{\bar{s}}}(1 - \hat{n}_{i_{\bar{s}}}) - \hat{n}_{j_s}\hat{n}_{b_{\bar{s}}}(\hat{n}_{i_{\bar{s}}} - \hat{n}_{a_s})^2\Big],$$

$$\lambda \in \{0, \pm i\sqrt{2}\}: \hat{O}_3 = \sum_{s,\bar{s}\in\{\alpha,\beta\}} \hat{\kappa}_{i_s j_{\bar{s}}}^{a_{\bar{s}} b_s}\Big[(\hat{n}_{i_{\bar{s}}} - \hat{n}_{b_{\bar{s}}})^2 + (\hat{n}_{j_s} - \hat{n}_{a_s})^2\Big] - 2\left(\hat{O}_1 + \hat{O}_2\right),$$

$$\lambda \in \{0, \pm i\}: \hat{O}_4 = \hat{T}_{ijab}^{0,0} - \sum_{i=1}^{3}\hat{O}_i,$$

where $\hat{n}_p = \hat{a}_p^\dagger \hat{a}_p$. For $\Omega = 0$, $\hat{T}^{0,0}_{iiaa}$ has only one nonzero eigenvalue and thus does not require a decomposition.

In electronic structure calculations, owing to time-reversal symmetry of the electronic Hamiltonians, unitary transformations generating real-valued wave functions are considered. Therefore, the technique developed in Ref. [26] to reduce the number of expectation values for real fermionic wave functions from four to two is applicable for the singlet spherical tensor operators as well. This leads to not more than eight expectation values needed for evaluating gradients in the most complicated case of $\Omega = 4$.

Note that in this case $\hat{O}_n$'s are not unitary operators, and thus they cannot be used by the LCU scheme. To estimate the number of fragments for measurements within the LCU scheme we can use the number of Pauli products within each fermionic rotation (see Table I). It is clear that the number of Pauli products within singlet fermionic rotations can be up to eight times larger than the corresponding number of $\hat{O}_n$ operators, which makes our scheme much more preferable than the LCU approach.

IV. CONCLUSIONS

We considered two approaches to generalization of the parameter-shift rule based on the polynomial expansion of exponentially parametrized unitary transformations and the generator decompositions. As in the original parameter-shift-rule application, these approaches provide gradient expressions as linear combinations of expectation values, where the main criterion for efficiency is the number of different expectation values.

Both of the considered approaches depend on the eigenspectrum of the generator for the differentiated unitary transformation, but in different ways. The performance of the polynomial expansion depends only on the number of different eigenvalues, while that of the generator decompositions depends also on the generator eigensubspaces and how well their structures can be reproduced by decomposing operators. The polynomial expansion approach scales quadratically with the number of generator eigenvalues and provides efficient expression only for two- and three-eigenvalue generators [40]. For generators with a larger number of eigenvalues it is more beneficial to employ the generator decomposition technique. This technique provides efficient schemes (in terms of the number of needed expectation values) for all considered generators. Also, compared to the LCU decomposition technique used for an arbitrary generator before [21], our approach does not require ancilla qubits.

The generator decomposition approach has several variations differing in low-eigenvalue operators used for the decomposition. The most conservative approach is to use projectors on individual eigensubspaces; the number of expectation values scales linearly with the number of the generator eigenvalues. It was found to be superior to other decompositions if one of the generator eigenvalues has much higher multiplicity than the other eigenvalues, as in the case of S^2-conserving fermionic operators.

Another alternative for decomposing generators is using the CSA. For some generators the different eigenvalues of which can be related via linear combinations with binary coefficients and have similar degeneracies, the CSA decomposition can reduce the generator expansion to scale as $\log_2$ of the number of eigenvalues. Results of the CSA decomposition can be further improved if one will allow generation of noncommutative terms. The CSA based approaches showed that any two-qubit transmon and match gates require only four expectation values for their gradients.

Note added. Recently, two more proposals generalizing the PSR via methods identical to our polynomial expansion were submitted [41,42].

ACKNOWLEDGMENTS

A.F.I. is grateful to Jakob Kottmann, Abhinav Anand, Alireza M. Khah, Lisa Jeffrey, and Ilya G. Ryabinkin for stimulating discussions and acknowledges financial support from the Google Quantum Research Program, Early Researcher Award, and the Natural Sciences and Engineering Research Council of Canada.

APPENDIX

Here we derive Eqs. (6) and (7) for generator $\hat{G}$ that has L eigenvalues. First, to find $a_i(\theta)$'s in Eq. (7) we use an eigenspace projector decomposition of $\hat{G}$:

$$\hat{G} = \sum_{n=1}^{L} \hat{P}_n \lambda_n, \tag{A1}$$

where λ_n are different eigenvalues of $\hat{G}$ and $\hat{P}_n$ are projectors on the corresponding eigensubspaces. Convenient properties of these projectors are their orthogonality and idempotency ($\hat{P}_n \hat{P}_m = \delta_{nm} \hat{P}_n$). These properties allows us to connect the exponential function

$$e^{i\theta \hat{G}} = \sum_{n=1}^{L} \hat{P}_n e^{i\theta \lambda_n} \tag{A2}$$

with its polynomial expansion

$$\sum_{k=0}^{L-1} a_k(\theta)(i\hat{G})^k = \sum_{n=1}^{L} \hat{P}_n \sum_{k=0}^{L-1} a_k(\theta)(i\lambda_n)^k. \tag{A3}$$

Due to linear independence of projector operators this results in a system of linear equations with $\{i^k a_k(\theta)\}$ as variables:

$$\sum_{k=0}^{L-1} \lambda_n^k [i^k a_k(\theta)] = e^{i\theta\lambda_n}, \quad n = 1, \ldots, L. \tag{A4}$$

The matrix involved in this system of equations is the Vandermonde matrix ($\lambda_n^k = W_{nk}$), the determinant of which is nonzero as long as the eigenvalues are different. Inverting the Vandermonde matrix provides $a_k(\theta)$ solutions:

$$a_k(\theta) = i^{-k} \sum_n W^{-1}_{kn} e^{i\theta\lambda_n}. \tag{A5}$$

Since λ_n's are real, it is easy to show the following relations:

$$a_{2k}(\theta) = a^*_{2k}(-\theta), \tag{A6}$$

$$a_{2k+1}(\theta) = -a^*_{2k+1}(-\theta). \tag{A7}$$

Second, C_n coefficients in Eq. (6) can be found as solutions of a linear system of equations. This system can be formulated by rewriting Eq. (6) as

$$\sum_n C_n \langle e^{-i\theta_n \hat{G}} \tilde{H} e^{i\theta_n \hat{G}} \rangle = \sum_n C_n \sum_{k,k'=0}^{L-1} \langle \hat{G}^{k'} \tilde{H} \hat{G}^k \rangle A_{kk'}(\theta_n) = i\big[\langle \tilde{H}\hat{G} \rangle - \langle \hat{G}\tilde{H} \rangle\big], \tag{A8}$$

where

$$\tilde{H} = e^{-i\tau \hat{G}} \hat{U}_2^\dagger \hat{H} \hat{U}_2 e^{i\tau \hat{G}}, \tag{A9}$$

$$A_{kk'}(\theta_n) = a_k(\theta_n) a_{k'}^*(\theta_n) i^{k+k'} (-1)^k. \tag{A10}$$

Accounting for linear independence of $\langle \hat{G}^{k'} \tilde{H} \hat{G}^k \rangle$ terms one can obtain C_n from equations

$$\sum_n A_{kk'}(\theta_n) C_n = B_{kk'}, \tag{A11}$$

where $B_{10} = -B_{01} = i$ and $B_{kk'} = 0$ for all other kk'. Depending on the choice of θ_n, the number of needed C_n's to satisfy L^2 equations can vary, but it cannot exceed L^2.

Minimization of the number of C_n coefficients and thus the number of expectation values depends on the $\hat{G}$ spectrum. For example, if every λ_n has its negative counterpart, $-\lambda_n$, then the even (odd) degree functions $a_{2k}(\theta)$ $[a_{2k+1}(\theta)]$ are real even (odd) θ functions. This condition allows one to reduce the number of needed C_n and θ_n parameters to $\approx L^2/2$, where θ_n's are chosen in pairs $\{\pm\theta_k\}_{k=1}^{N_p/2}$. Thus, in the case of $L = 2$, the number of C_n's is only 2 because $\theta_n = \pm\theta$ creates some dependencies in $A_{kk'}(\theta_n)$ elements.

For $L = 3$ and $\lambda_n \in \{0, \pm 1\}$ one can derive the following polynomial expansion of the exponential $\hat{G}$ operator:

$$e^{i\theta\hat{G}} = 1 + i\sin(\theta)\hat{G} + [\cos(\theta) - 1]\hat{G}^2. \tag{A12}$$

Taking $\theta_{1,2} = \pm\theta$ does not eliminate terms $\langle \hat{G}\tilde{H}\hat{G}^2 \rangle$ and $\langle \hat{G}^2\tilde{H}\hat{G} \rangle$ in the PSR expression, therefore another pair of θ's $\theta_{3,4} = \pm 2\theta$ is needed to eliminate these terms and to obtain the gradient of energy in this case. Here, we present the final expression

$$i\langle [\tilde{H}, \hat{G}] \rangle = (\alpha\Delta_1 - \Delta_2)\beta, \tag{A13}$$

where

$$\alpha = \frac{\sin(2\theta)(\cos(2\theta) - 1)}{\sin(\theta)(\cos(\theta) - 1)}, \tag{A14}$$

$$\beta = \frac{1}{2\sin(2\theta)} \left[\frac{1 - \cos(2\theta)}{1 - \cos(\theta)} - 1 \right]^{-1}, \tag{A15}$$

$$\Delta_k = \langle e^{-ik\theta\hat{G}} \tilde{H} e^{ik\theta\hat{G}} \rangle - \langle e^{ik\theta\hat{G}} \tilde{H} e^{-ik\theta\hat{G}} \rangle. \tag{A16}$$

This results in four expectation values required to obtain the gradient with respect to the amplitude of the $L = 3$ $\hat{G}$ with the symmetric eigenvalue spectrum $\lambda_n \in \{0, \pm 1\}$.

[1] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, Nat. Commun. **5**, 4213 (2014).

[2] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, Chem. Rev. **119**, 10856 (2019).

[3] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Rev. Mod. Phys. **92**, 015003 (2020).

[4] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum Sci. Technol. **2**, 045001 (2017).

[5] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Phys. Rev. A **101**, 032308 (2020).

[6] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Phys. Rev. A **98**, 032309 (2018).

[7] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Quantum Sci. Technol. **4**, 043001 (2019).

[8] E. R. Anschuetz, J. P. Olson, A. Aspuru-Guzik, and Y. Cao, arXiv:1808.08927.

[9] K. Arimitsu, Y. O. Nakagawa, S. Koh, W. Mizukami, Q. Gao, and T. Kobayashi, arXiv:2107.12705.

[10] S. Yalouz, E. Koridon, B. Senjean, B. Lasorne, F. Buda, and L. Visscher, arXiv:2109.04576.

[11] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Quantum Sci. and Technol. **4**, 014008 (2018).

[12] I. G. Ryabinkin, T.-C. Yen, S. N. Genin, and A. F. Izmaylov, J. Chem. Theory Comput. **14**, 6317 (2018).

[13] I. G. Ryabinkin, R. A. Lang, S. N. Genin, and A. F. Izmaylov, J. Chem. Theory Comput. **16**, 1055 (2020).

[14] H. L. Tang, E. Barnes, H. R. Grimsley, N. J. Mayhall, and S. E. Economou, PRX Quantum **2**, 020310 (2021).

[15] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, npj Quantum Inf **6**, 10 (2020).

[16] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, J. Chem. Theory Comput. **15**, 311 (2019).

[17] I. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, and I. Tavernelli, J. Chem. Phys. **152**, 124107 (2020).

[18] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, Quantum Sci. Technol. **4**, 045005 (2019).

[19] N. V. Tkachenko, J. Sud, Y. Zhang, S. Tretiak, P. M. Anisimov, A. T. Arrasmith, P. J. Coles, L. Cincio, and P. A. Dub, PRX Quantum **2**, 020337 (2021).

[20] R. A. Lang, I. G. Ryabinkin, and A. F. Izmaylov, J. Chem. Theory Comput. **17**, 66 (2021).

[21] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Phys. Rev. A **99**, 032331 (2019).

[22] J. Li, X. Yang, X. Peng, and C.-P. Sun, Phys. Rev. Lett. **118**, 150503 (2017).

[23] J. M. Chow, A. D. Córcoles, J. M. Gambetta, C. Rigetti, B. R. Johnson, J. A. Smolin, J. R. Rozen, G. A. Keefe, M. B. Rothwell, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **107**, 080502 (2011).

[24] B. Foxen, C. Neill, A. Dunsworth, P. Roushan, B. Chiaro, A. Megrant, J. Kelly, Z. Chen, K. Satzinger, R. Barends, F. Arute,

K. Arya, R. Babbush, D. Bacon, J. C. Bardin, S. Boixo, D. Buell, B. Burkett, Y. Chen, R. Collins *et al.*, Phys. Rev. Lett. **125**, 120504 (2020).
[25] G. E. Crooks, arXiv:1905.13311.
[26] J. S. Kottmann, A. Anand, and A. Aspuru-Guzik, arXiv:2011.05938.
[27] T. Hubregtsen, F. Wilde, S. Qasim, and J. Eisert, Single-component gradient rules for variational quantum algorithms, arXiv:2106.01388.
[28] A. Mari, T. R. Bromley, and N. Killoran, Phys. Rev. A **103**, 012405 (2021).
[29] L. Banchi and G. E. Crooks, Quantum **5**, 386 (2021).
[30] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nat. Commun. **9**, 4812 (2018).
[31] M. Cerezo and P. J. Coles, Quantum Sci. Technol. **6**, 035006 (2021).
[32] A. W. Harrow and J. C. Napp, Phys. Rev. Lett. **126**, 140502 (2021).
[33] T.-C. Yen and A. F. Izmaylov, PRX Quantum **2**, 040320 (2021).
[34] R. Jozsa and A. Miyake, Proc. R. Soc. A **464**, 3089 (2008).
[35] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, Nature (London) **574**, 505 (2019).
[36] A. F. Izmaylov, M. Díaz-Tinoco, and R. A. Lang, Phys. Chem. Chem. Phys. **22**, 12980 (2020).
[37] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, New York, 2000).
[38] X. Li and J. Paldus, Theor. Chem. Acc. **133**, 1467 (2014).
[39] I. Shavitt, Int. J. Quantum Chem. **14**, 5 (2009).
[40] In the three-eigenvalue case the symmetry of the eigenspectrum is essential.
[41] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, General parameter-shift rules for quantum gradients, arXiv:2107.12390.
[42] O. Kyriienko and V. E. Elfving, Generalized quantum circuit differentiation rules, Phys. Rev. A **104**, 052417 (2021).