JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

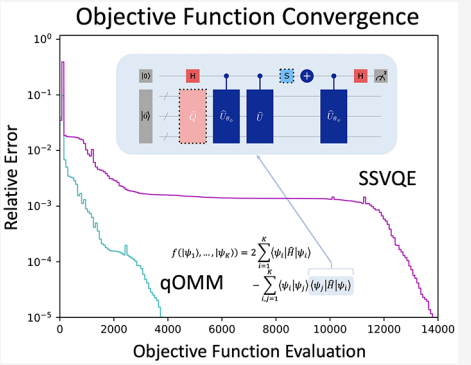
Quantum Orbital Minimization Method for Excited States Calculation on a Quantum Computer

Joel Bierman, Yingzhou Li,* and Jianfeng Lu*

Cite This: J. Chem. Theory Comput. 2022, 18, 4674−4689 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: We propose a quantum-classical hybrid variational algorithm, the quantum orbital minimization method (qOMM), for obtaining the ground state and low-lying excited states of a Hermitian operator. Given parametrized ansatz circuits representing eigenstates, qOMM implements quantum circuits to represent the objective function in the orbital minimization method and adopts a classical optimizer to minimize the objective function with respect to the parameters in ansatz circuits. The objective function has an orthogonality constraint implicitly embedded, which allows qOMM to apply a different ansatz circuit to each input reference state. We carry out numerical simulations that seek to find excited states of H_2, LiH, and a toy model consisting of four hydrogen atoms arranged in a square lattice in the STO-3G basis with UCCSD ansatz circuits. Comparing the numerical results with existing excited states methods, qOMM is less prone to getting stuck in local minima and can achieve convergence with more shallow ansatz circuits.



Downloaded via DUKE UNIV on September 1, 2022 at 12:55:13 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

1. INTRODUCTION

The development of quantum computing has boomed in recent years. The supremacy of quantum computing is demonstrated by several groups.[1,2] One of the most promising applications for noisy intermediate-scale quantum (NISQ) devices is the variational quantum eigensolver (VQE) under the full configuration interaction (FCI) framework.[3−8] FCI is a quantum chemistry method which discretizes the time-independent many-body Schrödinger equation numerically exactly. The ground state and low-lying excited states are calculated via solving a standard eigenvalue problem, while the problem dimension scales exponentially as the number of electrons in the system. The FCI framework fits naturally into quantum computing. Quantum algorithms for ground state eigensolvers under the FCI framework have been extensively developed in past decades.[9−13] Among them, VQE is a quantum-classical hybrid method and has the shortest circuit depth, which makes it the most widely applied ground state quantum eigensolver on NISQ devices. In short, VQE adopts a parametrized circuit as the variational ansatz $|\phi_\theta\rangle$ for the ground state. Then, the energy function $\langle\phi_\theta|\hat{H}|\phi_\theta\rangle$ is minimized with respect to the parameter θ, where $\hat{H}$ is the Hamiltonian operator.

In addition to the ground state, excited states play an important role in connecting computational results and experimental observations in quantum chemistry. In this paper, we propose a hybrid quantum-classical algorithm, named the quantum orbital minimization method (qOMM), to compute the excited state energies as well as the corresponding states under the FCI framework.

Existing approaches for excited states on quantum computers could be grouped into two categories: the excited state energy computation and the excited state vector computation.

Excited state energy computation on quantum computers evaluates the excited state energy without explicitly constructing the state vector. Two methods, quantum subspace expansion (QSE)[14−16] and quantum equation of motion (qEoM),[17] are representatives of this category. Both of them are based on perturbation theory starting from the ground state. They first construct the ground state $|\phi_{GS}\rangle$ via VQE. Then, they evaluate the expansion values on single excited states from the ground state $|\phi_{GS}\rangle$, i.e.,

$$A^{p_2q_2}_{p_1q_1} = \langle\phi_{GS}|\hat{a}^\dagger_{q_2}\hat{a}_{p_2}\hat{H}\hat{a}^\dagger_{p_1}\hat{a}_{q_1}|\phi_{GS}\rangle \quad (1)$$

and

$$B^{p_2q_2}_{p_1q_1} = \langle\phi_{GS}|\hat{a}^\dagger_{q_2}\hat{a}_{p_2}\hat{a}^\dagger_{p_1}\hat{a}_{q_1}|\phi_{GS}\rangle \quad (2)$$

for $\hat{a}^\dagger_p$ and $\hat{a}_p$ being creation and annihilation operators. Once these expansion values are available, QSE solves a generalized eigenvalue problem of matrix pencil (A, B) and obtains the refined ground state energy and low-lying excited state energies on the perturbed space of $|\phi_{GS}\rangle$. qEoM adopts these expansion values into the equation of motion expression and obtains the low-lying excited state energies. Double excitations and higher order excitations from the ground state could be further

Received: March 2, 2022
Published: July 25, 2022

ACS Publications © 2022 American Chemical Society

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689

evaluated and included to improve the accuracy of QSE and qEoM. However, including more excitations leads to a much higher computational cost.

Excited state vector computations include a wide spectrum of methods. The folded spectrum method[18] folds the spectrum of a Hamiltonian operator via a shift-and-square operation, i.e., $\hat{H} \rightarrow (\hat{H} - \lambda I)^2$, where λ is a chosen number close to the desired excited state energy. Then, $(\hat{H} - \lambda I)^2$ is used as the operator in the VQE method, and an excited state with the energy closest to λ is obtained via the optimization procedure in VQE. The witness-assisted variational eigenspectra solver[19,20] adopts an entropy term as the regularizer in the objective function to keep the minimized excited state close to its initial state, where the initial state is set to be an excitation from the ground state. The quantum deflation method[21,22] adopts the overlapping between the target state and the ground state as a penalty term in the objective function to keep the target state as orthogonal as possible to the ground state. The extension to more than one excited state in the quantum deflation method is straightforward. All of these methods above have their classical algorithm counterparts. On the other hand, the following two methods use a unique property of quantum computing; i.e., all quantum operators are unitary and orthogonal states after the actions of quantum operators are still orthogonal to each other. Subspace search VQE (SSVQE)[23] and multistate contracted VQE (MCVQE)[24] apply the same basic ansatz. They first prepare a sequence of non-parameterized mutually orthogonal states, $|\phi_1\rangle$, $|\phi_2\rangle$, ..., $|\phi_K\rangle$ for K excited states. A parametrized circuit $\hat{U}_\theta$ is then applied to these states. The unitarity of $\hat{U}_\theta$ (implemented on a quantum computer) ensures that $\hat{U}_\theta|\phi_1\rangle$, ..., $\hat{U}_\theta|\phi_K\rangle$ are mutually orthogonal. Hence, we could minimize the objective function

$$\sum_k \langle\phi_k|\hat{U}_\theta^\dagger \hat{H} \hat{U}_\theta|\phi_k\rangle \tag{3}$$

to obtain approximated ground states and excited states. The energies associated with $\hat{U}_\theta|\phi_1\rangle$, ..., $\hat{U}_\theta|\phi_K\rangle$ are usually not ordered. SSVQE further introduces a min−max problem to obtain a specific excited state. Also, in the SSVQE paper, a weighted objective function is employed to preserve the ordering, which will be the version we investigate numerically in our paper below. The MCVQE approach enriches the ansatz by introducing a unitary matrix applied across states in a classical post-processing step, i.e., $|\psi_k\rangle = \hat{U}_\theta \sum_l |\phi_l\rangle V_{lk}$ for V_{lk} being the (l, k)-th entry of a unitary matrix V. Hence, a standard eigenvalue problem is solved as the post-processing in MCVQE to obtain V.

In this paper, we propose to obtain the excited state vector using the objective function in the orbital minimization method (OMM)[25,26]

$$f(|\psi_1\rangle, ..., |\psi_K\rangle) = 2\sum_{i=1}^{K} \langle\psi_i|\hat{H}|\psi_i\rangle - \sum_{i,j=1}^{K} \langle\psi_i|\psi_j\rangle\langle\psi_j|\hat{H}|\psi_i\rangle \tag{4}$$

We refer to our method as the quantum orbital minimization method (qOMM) throughout this paper. An important property of eq 4 is that, even though the orthogonality conditions among $|\Psi_i\rangle$'s are not explicitly enforced, the minimizer of eq 4 consists of mutually orthogonal state vectors. For most other methods, the orthogonality condition is explicitly proposed as a constraint in the optimization problem. The implicit orthogonality property allows us to adopt a different ansatz (with regard to either the numerical parameter values, the structure of the ansatz, or both) for each excited state. In this paper, the variational ansatz class that we use is of the form

$$|\tilde{\psi}_k\rangle = \sum_{l=1}^{K} \hat{U}_{\theta_l}|\phi_l\rangle V_{lk} \tag{5}$$

where $|\phi_1\rangle$, ..., $|\phi_K\rangle$ are initial states, $\hat{U}_{\theta_1}$, ..., $\hat{U}_{\theta_K}$ are parametrized circuits with θ_1, ..., θ_K being the parameters, and V_{lk} is the (l, k)-th entry of an invertible matrix V that is applied during a classical post-processing step. Such an expression is more general than those used in other excited state vector methods, which typically either apply the same circuit $\hat{U}_\theta$ to all input states or do not apply the classical post-processing matrix V. Another feature of eq 4 is that the objective function does not include any hyper-parameter (unlike, e.g., the deflation approach), which makes the method easy to use in practice without parameter tuning. Furthermore, this approach shares the same advantages over QSE and qEoM as SSVQE and MCVQE. Namely, the excited states in qOMM are encoded into the objective function of an optimization problem and not as linear combinations of single and double excitations above the ground state in a post-processing step. This allows the method to (1) compute excited states for which triple and higher excitation terms have a non-negligible contribution without calculating higher order reduced density matrices and (2) treat the ground and excited states on the same footing.

Notice that the objective function (eq 4) includes the evaluation of the state inner product with respect to the Hamiltonian operator and the state inner product, i.e., $\langle\phi_i|\hat{H}|\psi_j\rangle$ and $\langle\psi_i|\psi_j\rangle$, respectively. In other existing excited state methods, neither of these two are evaluated directly. (In the deflation method,[21,22] the absolute value of the inner product, $|\langle\psi_i|\psi_j\rangle|$, is evaluated, while we need the actual value of the inner product without taking the absolute value.) Given our variational ansatz, we construct quantum circuits to evaluate both inner products, which are essentially inspired by the Hadamard test. When the evaluation of eq 4 is available on a quantum computer, we can adopt gradient-free optimizers to minimize the objective function with respect to parameters, θ_1, ..., θ_K.

Finally, we include a sequence of numerical results to demonstrate the efficiency and robustness of the proposed method. We test the following chemistry systems using the proposed method on quantum simulators: H_2 and LiH at their equilibrium configurations, H_2 at a stretched bond distance of twice the equilibrium distance, and a toy model consisting of four hydrogen atoms arranged in a square at near-equilibrium distances and a stretched bond distance configuration. We observe that qOMM has two major benefits compared to algorithms such as weighted-SSVQE (in this paper, when we refer to SSVQE, we implicitly mean the weighted version of SSVQE that uses one optimization step for finding multiple eigenvalues simultaneously) that enforce the orthogonality of the input states at every optimization step. (1) Enforcing the mutual orthogonality of the input states through the overlap terms allows for greater flexibility in the choice of ansatz applied to each input state, as well as greater flexibility in the choice of input states. The circuit applied to each input state must be identical in the SSVQE framework in order to enforce orthogonality, whereas qOMM allows us to apply a different ansatz circuit to each input state. (2) The tendency to get

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689

stuck in local minima, while not eliminated entirely, is diminished considerably. The first benefit is important because choosing a suitably expressive ansatz for excited states is often more difficult than the ground state VQE problem. In practice, this means that one is often able to use a more shallow and less expressive ansatz circuit, a feature that is crucial for applications on NISQ devices.

The rest of the paper is organized as follows. Section 2 illustrates the circuits we use to evaluate the inner products. The overall method is discussed in section 3. In section 4, we apply the method to various chemistry molecules to demonstrate the efficiency of the proposed method. Finally, section 5 concludes the paper and discusses future work.

2. INNER PRODUCT CIRCUITS

In the objective function of qOMM, the inner product of two states and the inner product of two states with respect to an operator are complex numbers. Both the real and imaginary parts are explicitly needed to construct the objective function. If two states are identical, then the inner products can be efficiently evaluated by the well-known Hadamard test. We propose quantum circuits for the evaluation of the inner products for the case in which they are different in this section.

To simplify the notations, we consider evaluating $\langle\psi|\phi\rangle$ and $\langle\psi|\hat{H}|\phi\rangle$, where $|\psi\rangle$ and $|\phi\rangle$ are given by a unitary circuit acting on $|0\rangle$, i.e., $|\psi\rangle = \hat{U}_{\theta_\psi}|0\rangle$ and $|\phi\rangle = \hat{U}_{\theta_\phi}|0\rangle$. Given the variational ansatz, the inner product $\langle\psi \mid \phi\rangle = \langle 0|\hat{U}_{\theta_\psi}^{\dagger}\hat{U}_{\theta_\phi}|0\rangle$ can be viewed as the inner product of $|0\rangle$ with respect to the operator $\hat{O} = \hat{U}_{\theta_\psi}^{\dagger}\hat{U}_{\theta_\phi}$. In principle, the transpose of the ansatz circuit can be constructed. After the ansatz circuit is compiled into basic gates, the transpose of the circuit is the composition of the transposed basic gates in the reverse ordering. Hence, $\langle\psi|\phi\rangle$ could then be evaluated using the Hadamard test, which requires the controlled ansatz circuit and the controlled transpose of the ansatz circuit. Instead, we adopt an idea inspired by the Hadamard test, which can be applied to evaluate $\langle\psi|\hat{H}|\phi\rangle$ without controlled transpose of the ansatz circuit. In the following, we discuss such two quantum circuits evaluating the inner products in detail.

2.1. $\langle\psi|\phi\rangle$ Evaluation. For complex-valued states $|\psi\rangle$ and $|\phi\rangle$, the inner product $\langle\psi|\phi\rangle$ is also a complex number. The evaluation of $\langle\psi|\phi\rangle$ is divided into the evaluations of the real and imaginary parts. We will introduce the circuit for evaluation of the real part of $\langle\psi|\phi\rangle$ in detail, and the circuit for the imaginary part could be constructed analogously.

Figure 1 illustrates the precise quantum circuit used to evaluate the real part of $\langle\psi|\phi\rangle$. Here we list expressions at all five stages, as shown in the figure

$$(H \otimes I)|0\rangle|\vec{0}\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|\vec{0}\rangle \tag{S_1}$$

$$C\text{-}\hat{U}_{\theta_\phi}\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|\vec{0}\rangle = \frac{1}{\sqrt{2}}(|0\rangle|\vec{0}\rangle + |1\rangle|\phi\rangle) \tag{S_2}$$

$$(X \otimes I)\frac{1}{\sqrt{2}}(|0\rangle|\vec{0}\rangle + |1\rangle|\phi\rangle) = \frac{1}{\sqrt{2}}(|1\rangle|\vec{0}\rangle + |0\rangle|\phi\rangle) \tag{S_3}$$

$$C\text{-}\hat{U}_{\theta_\psi}\frac{1}{\sqrt{2}}(|1\rangle|\vec{0}\rangle + |0\rangle|\phi\rangle) = \frac{1}{\sqrt{2}}(|1\rangle|\psi\rangle + |0\rangle|\phi\rangle) \tag{S_4}$$

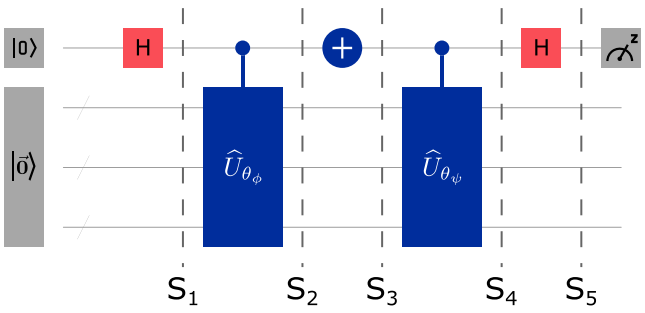


Figure 1. Inner product circuit.

$$(H \otimes I)\frac{1}{\sqrt{2}}(|1\rangle|\psi\rangle + |0\rangle|\phi\rangle) = \frac{1}{2}(|0\rangle(|\phi\rangle + |\psi\rangle) + |1\rangle(|\phi\rangle - |\psi\rangle)) \tag{S_5}$$

where H is the Hadamard gate, X is the Pauli-X gate, I is the identity mapping, and $C\text{-}\hat{U}$ is the controlled $\hat{U}$ gate. Conducting the measurement on the ancilla qubit, we have the probability measuring zero being

$$\begin{aligned}(\text{measurement} = 0) &= \frac{1}{4}(\langle\phi| + \langle\psi|)(|\phi\rangle + |\psi\rangle) \\ &= \frac{1}{2} + \frac{1}{2}\text{Re}\langle\psi \mid \phi\rangle\end{aligned} \tag{6}$$

where $\text{Re}(\cdot)$ denotes the real part of the complex number. Hence, if we execute the quantum circuit as in Figure 1 for M shots and count the number of zeros, the $\text{Re}\langle\psi \mid \phi\rangle$ could be well-approximated given M (which will scale as $O\left(\frac{1}{\epsilon^2}\right)$ for some target accuracy ϵ due to statistical sampling noise) is reasonably large.

The imaginary part of $\langle\psi \mid \phi\rangle$ could be estimated in a similar way. Notice that the real part in eq 6 comes from $\langle\psi \mid \phi\rangle + \langle\phi \mid \psi\rangle$. If either $|\psi\rangle$ or $|\phi\rangle$ is replaced by $\imath|\psi\rangle$ or $\imath|\phi\rangle$, then the real part in eq 6 corresponds to the imaginary part of the inner product instead. One way to implement the replacement is to add a phase gate on the ancilla qubit after stage eq 2. The resulting quantum circuit is as that in Figure 2 where the light blue dashed box is enabled. We then need to execute the circuit for another M shots to obtain the approximation of $\text{Im}\langle\psi \mid \phi\rangle$.

In addition to variational ansatz $|\psi\rangle = \hat{U}_{\theta_\psi}|0\rangle$ and $|\phi\rangle = \hat{U}_{\theta_\phi}|0\rangle$, we could extend the quantum circuit to variational ansatz $|\psi\rangle = \hat{U}_{\theta_\psi}\hat{Q}|0\rangle$ and $|\phi\rangle = \hat{U}_{\theta_\phi}\hat{Q}|0\rangle$, where $\hat{Q}$ is a common quantum circuit preparing the initial state. If such ansatzes are adopted, we could enable the light pink dashed box in Figure 2 such that the circuit $\hat{Q}$ is applied on $|0\rangle$ as a common circuit. The

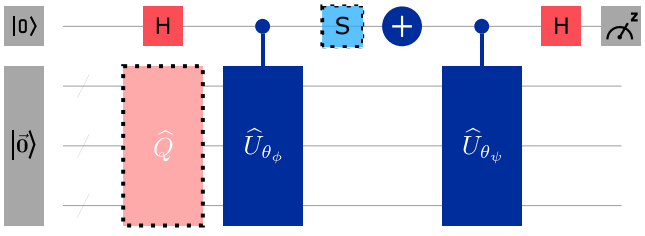


Figure 2. Extended inner product circuit.

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689

real and imaginary parts of the inner product are evaluated similarly as before.

2.2. $\langle\psi|\hat{H}|\phi\rangle$ Evaluation. The Hamiltonian operator considered in this paper is under the FCI framework and admits the form

$$\hat{H} = \sum_{p,q=1}^{N_{orb}} h_{pq}\hat{a}_p^\dagger\hat{a}_q + \sum_{p,q,r,s=1}^{N_{orb}} v_{pqrs}\hat{a}_p^\dagger\hat{a}_q^\dagger\hat{a}_s\hat{a}_r \tag{7}$$

where h_{pq} and v_{pqrs} are one-body and two-body integrals, respectively. The Hamiltonian operator $\hat{H}$, in general, is not a unitary operator. Hence, it cannot be represented by a quantum circuit directly. On the other hand, the excitation operators, $\hat{a}_p^\dagger\hat{a}_q$ and $\hat{a}_p^\dagger\hat{a}_q^\dagger\hat{a}_s\hat{a}_r$, are unitary and can be represented by quantum circuits. The detailed circuits of these excitation operators depend on the encoding scheme. All methods proposed in this paper do not rely on the encoding scheme. The value $\langle\psi|\hat{H}|\phi\rangle$ is then calculated as a weighted summation of circuit results

$$\langle\psi|\hat{H}|\phi\rangle = \sum_{p,q=1}^{N_{orb}} h_{pq}\langle\psi|\hat{a}_p^\dagger\hat{a}_q|\phi\rangle + \sum_{p,q,r,s=1}^{N_{orb}} v_{pqrs}\langle\psi|\hat{a}_p^\dagger\hat{a}_q^\dagger\hat{a}_s\hat{a}_r|\phi\rangle \tag{8}$$

Thus, it suffices to consider circuits for the evaluation of $\langle\psi|\hat{U}|\phi\rangle$, where $\hat{U}$ denotes a general unitary operator.

Figure 3 illustrates the circuit for $\langle\psi|\hat{U}|\phi\rangle$. This circuit can be understood in the same way as the circuit of $\langle\psi|\phi\rangle$. The value $\langle\psi|\hat{U}|\phi\rangle$ can be explicitly written as

$$\langle\psi|\hat{U}|\phi\rangle = (\langle 0|\hat{U}_{\theta_\psi}^\dagger)(\hat{U}\hat{U}_{\theta_\phi}|0\rangle) = \langle\psi\,|\,\tilde{\phi}\rangle \tag{9}$$

where the construction circuit for $|\tilde{\phi}\rangle$ is $\hat{U}\hat{U}_{\theta_\phi}$. Hence, Figure 3 is essentially the circuit in Figure 1 and Figure 2 with a modified constructing circuit for $|\phi\rangle$.

3. QUANTUM ORBITAL MINIMIZATION METHOD

The orbital minimization method originates in the field of density functional theory (DFT). The minimization problem admits the following linear algebra form

$$\min_{X\in^{N\times K}} \mathrm{tr}((2I - X^*X)X^*\hat{H}X) \tag{10}$$

where X is a matrix with each column denoting a state vector, $\hat{H}$ is the Hamiltonian matrix, I is the identity matrix of size $K \times K$, and $\mathrm{tr}(\cdot)$ denotes the trace operator. When $\hat{H}$ is Hermitian negative definite, all local minima of eq 10 are of the form[27]

$$X = QV \tag{11}$$

where $Q \in^{N\times K}$ are K eigenvectors of $\hat{H}$ associated with the smallest K eigenvalues and $V \in^{K\times K}$ is an arbitrary unitary matrix. Substituting eq 11 back into eq 10, we find that all of these local minima are of the same objective value and they are global minima with minimizers consisting of mutually orthogonal vectors. This property has been proven analytically.[27] Unlike many other optimization problems for solving the low-lying eigenvalue problem, OMM as in eq 10 does not have any explicit orthogonal constraint and the orthogonality in eq 11 could be achieved without any orthogonalization step. This property is the key reason for us to investigate the application of eq 10 in quantum computing, where explicit orthogonalization is difficult to implement in the presence of hardware and environmental noise. Furthermore, despite the fact that minimizing eq 10 is a non-convex problem, its property that all local minimizers are also global minimizers which encode mutually orthogonal vectors in the case where we are minimizing over all possible vectors[27] (absent the constraint of vectors expressable as quantum circuits) makes it a compelling objective function to investigate in the quantum setting. The numerical simulations in this work serve to provide some insight into the extent to which these favorable properties in the classical setting carry over to the quantum setting where we are constrained by vectors expressable as quantum circuit ansatzes.

In the quantum computing setting, each column of X is represented by the ansatz $|\psi_i\rangle = \hat{U}_{\theta_i}|\phi_i\rangle$. By the nature of quantum computing, if quantum error is not taken into account, $|\psi_i\rangle$ is always of unit length for all i. Hence, the $2I - X^*X$ term in eq 10 has one on its diagonal. Substituting the variational ansatz into eq 10, we obtain the optimization problem for our quantum excited state problem

$$\min_{\theta_1,\ldots,\theta_K} g(\theta_1,\ldots,\theta_K) \tag{12}$$

for

$$g(\theta_1,\ldots,\theta_K) = \sum_{i=1}^{K}\langle\phi_i|\hat{U}_{\theta_i}^\dagger\hat{H}\hat{U}_{\theta_i}|\phi_i\rangle - \sum_{\substack{i,j=1\\ i\neq j}}^{K}\langle\phi_i|\hat{U}_{\theta_i}^\dagger\hat{U}_{\theta_j}|\phi_j\rangle\langle\phi_j\,|\,\hat{U}_{\theta_j}^\dagger\hat{H}\hat{U}_{\theta_i}|\phi_i\rangle \tag{13}$$

If the variational space is rich enough such that vectors in QV as in eq 11 can be well-approximated by $\{\hat{U}_{\theta_i}|\phi_i\rangle\}_{i=1}^K$, then qOMM can achieve a good approximation to the global minimum of its linear algebra counterpart (eq 10). In practice, if we obtain the global minimum value of eq 10 for qOMM (though the value is not known in advance), then we are confident that the ansatz is rich enough and the optimizer works

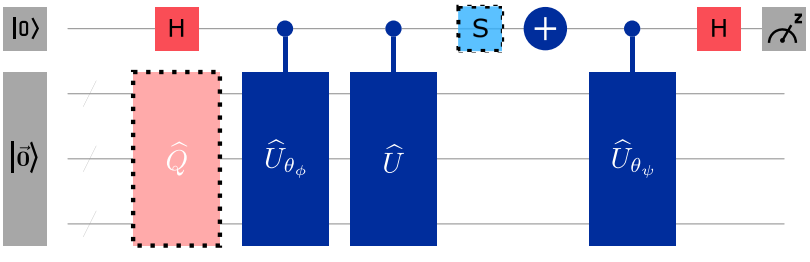


Figure 3. Extended inner product circuit with a unitary gate.

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689

perfectly in finding the global minimum for parameters $\theta_1, ..., \theta_K$. However, we lack the theoretical understanding of the energy landscape of eq 12. In part, the dependence of g on $\theta_1, ..., \theta_K$, like that in the deep neural network, is highly non-linear and non-convex. It could be analyzed for some simple ansatz circuit. However, for most complicated circuits, the analysis remains difficult. On the other hand, from the linear algebra point of view, when the variational space is rich enough, we could view the problem as optimizing the vectors X directly in eq 10. The nature of a quantum computer poses unit length constraints on each column of X. Adding extra unit length constraints would dramatically change the energy landscape of the original problem (eq 10). We defer the theoretical understanding of eq 12 to future work. In this paper, we focus on the numerical performance of eq 12.

For numerical optimization of the objective function g, note that its gradient with respect to a parameter $\theta_{i\alpha}$ is given by

$$\begin{aligned}\frac{dg}{d\theta_{i\alpha}} &= 2\mathrm{Re}\langle\phi_i|\frac{d\hat{U}_{\theta_i}^{\dagger}}{d\theta_{i\alpha}}\hat{H}\hat{U}_{\theta_j}|\phi_j\rangle \\ &\quad - 2\mathrm{Re}\sum_{\substack{j=1\\ j\neq i}}^{K}\langle\phi_i|\frac{d\hat{U}_{\theta_i}^{\dagger}}{d\theta_{i\alpha}}\hat{U}_{\theta_j}|\phi_j\rangle\langle\phi_j|\hat{U}_{\theta_j}^{\dagger}\hat{H}\hat{U}_{\theta_i}|\phi_i\rangle \\ &\quad - 2\mathrm{Re}\sum_{\substack{j=1\\ j\neq i}}^{K}\langle\phi_i|\hat{U}_{\theta_i}^{\dagger}\hat{U}_{\theta_j}|\phi j\rangle\langle\phi_j|\hat{U}_{\theta_j}^{\dagger}\hat{H}\frac{d\hat{U}_{\theta_i}}{d\theta_{i\alpha}}|\phi_i\rangle\end{aligned} \tag{14}$$

where $\theta_{i\alpha}$ is the α-th parameter in θ_i. From eq 14, we notice that the gradient evaluation requires an efficient quantum circuit for $\frac{d\hat{U}_{\theta_i}}{d\theta_{i\alpha}}$. The existence of such quantum circuits is ansatz dependent. Assume that we can evaluate eq 14 for all parameters. Then, the update on parameters follows a gradient descent method as

$$\theta^{(t+1)} = \theta^{(t)} - \tau\nabla_\theta g(\theta^{(t)}) \tag{15}$$

for $\theta^{(t+1)}$ and $\theta^{(t)}$ being all parameters at $(t+1)$-th and (t)-th iteration, respectively, and τ is the step size. Since all evaluations of quantum circuits involve randomness, in practice, this amounts to using the stochastic gradient descent method. Other advanced first order methods recently developed in deep neural network literature, e.g., ADAM, AdaGrad, coordinate descent method, etc., could be used to accelerate. These methods could achieve better performance than the vanilla stochastic gradient descent method.

When the gradient circuit is not available, we can adopt zeroth-order optimizers to solve eq 12. In this work, we use L-BFGS-B[28] as the optimizer, which uses a two-point finite difference gradient approximation and thus does not require explicit implementation of the circuit for $\frac{d\hat{U}_{\theta_i}}{d\theta_{i\alpha}}$. We find this optimizer to work well in the noiseless simulations we consider here, but we note that other promising candidates for noisy simulations and experimental calculations exist, which could also be used for our proposed objective function. In particular, using the simultaneous perturbation stochastic approximation (SPSA) for small molecule VQE calculations has been demonstrated experimentally on superconducting hardware[29] and a quantum natural gradient version of SPSA also exists.[30]

When the optimization problem is solved, we will obtain a set of parameters $\theta_1, ..., \theta_K$ such that $\hat{U}_{\theta_1}|\phi_1\rangle, ..., \hat{U}_{\theta_K}|\phi_K\rangle$ approximate vectors in QV as in eq 11. Numerically, we find that V is very close to an identity matrix, while a post-processing procedure still improves the accuracy of the approximation. The post-processing procedure is slightly different from that in MCVQE. We construct two matrices with their (i,j)-th elements being

$$A_{ij} = \langle\phi_i|\hat{U}_{\theta_i}^{\dagger}\hat{H}\hat{U}_{\theta_j}|\phi_j\rangle \quad \text{and} \quad B_{ij} = \langle\phi_i|\hat{U}_{\theta_i}^{\dagger}\hat{U}_{\theta_j}|\phi_j\rangle \tag{16}$$

Then, a generalized eigenvalue problem with matrix pencil (A, B) is solved as

$$AR = BR\Lambda \tag{17}$$

where R denotes the eigenvectors and Λ is a diagonal matrix with eigenvalues of (A, B) being its diagonal entries. The desired eigenvalues of $\hat{H}$ are in Λ, and the corresponding eigenvectors admit $\sum_{j=1}^{K}\hat{U}_{\theta_j}|\phi_j\rangle R_{ji}$ for $i = 1, ..., K$. The eigenvector, which is the excited state of $\hat{H}$, can be explicitly constructed as a linear combination of unitary operators[31] on a quantum computer.

qOMM has several unique features compared to existing excited state vector methods. The targets of the folded spectrum method[18] and witnessing eigenspectra solver[19,20] are different from that of qOMM. The quantum deflation method[21,22] addresses the excited states one by one, and the optimization problems therein are of different difficulty than that in qOMM. The most related excited state vector methods to qOMM are SSVQE and MCVQE. Since SSVQE and MCVQE are very similar to each other, we focus on the difference among qOMM and SSVQE in this paper. Both the qOMM and weighted SSVQE algorithms[23] take a set of input states and seek to simultaneously optimize the parametrized circuit and obtain the low-lying eigenvalues and eigenvectors of the Hermitian operator. Both methods achieve their goal via solving a single optimization problem. There are two major differences between them. The first one is how the mutual orthogonality of the eigenvectors is enforced. The construction of the weighted SSVQE objective function ensures that the set of states being evaluated is mutually orthogonal throughout optimization iteration. The proposed qOMM implicitly embeds this constraint into the objective function such that the states being evaluated are only guaranteed to be mutually orthogonal at the global minimum. The first difference leads to the second major difference. SSVQE adopts the ansatz circuits for different states that must admit a mutual orthogonal property. Hence, they restrict their ansatz circuits to be a set of identical parametrized circuits acting on mutually orthogonal initial states. qOMM does not constrain the ansatz circuits in this way. qOMM could adopt as its set of parametrized circuits any combination of ansatz circuits that have been tested in the literature, e.g., ansatz circuits in MCVQE[24] and SSVQE,[23] ansatz circuits in the deflation method,[21,22] etc. In addition to these ansatz circuits, qOMM could also adopt other ansatz circuits that have not been tested by excited state methods. In this work, we adopt UCCSD[32] as our main ansatz circuit so that the particle number is preserved throughout.

4. NUMERICAL RESULTS

We present the numerical results of our proposed method obtained from a classically simulated quantum computer. We apply it to the problem of finding low-lying eigenvalues of the electronic structure Hamiltonian for near-equilibrium configurations of two different molecules as well as the toy model

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689

consisting of four hydrogen atoms arranged in a square lattice. The numerical tests in this section serve to explore what advantages or disadvantages may arise for each of the qOMM and weighted SSVQE approaches (using weight vectors of the form $[n, n-1, ..., 1]$ for finding n states) in different situations. Section 4.1 presents the results for H_2 and the hydrogen square model obtained from noiseless simulations. Section 4.2 presents the results obtained from simulating H_2 with a depolarizing noise model. Section 4.3 presents the results for LiH obtained from noiseless simulations.

All codes for these simulations are implemented using the Qiskit[33] library. The electronic structure Hamiltonians were generated using molecular data from PySCF[34] in the STO-3G basis and mapped to qubit Hamiltonians using the Jordan−Wigner mapping for the noiseless simulations and the Parity mapping for the noisy simulations. The relative accuracies of the results for both methods are calculated by comparing them to their numerically exact counterparts obtained by numerically exact diagonalization of the Hamiltonians. Circuits involved in noiseless qOMM simulations were conducted using Qiskit's *StateVector* simulator. Circuits involved in noiseless SSVQE simulations were conducted using Qiskit's *QasmSimulator* in conjunction with the *AerPauliExpectation* method for computing expectation values in order to improve runtime performance. Both of these methods produce ideal, noiseless results. Simulations for finding the energy levels of H_2 were performed at an interatomic distance of 0.735 Å, those of LiH were performed at an interatomic distance of 1.595 Å, and those of the hydrogen square lattice model were performed at an interatomic distance of 1.23 Å. Simulation results for the hydrogen square model and H_2 at stretched bond distances can also be found in Appendix A. For the LiH simulations, we freeze the two core electrons to reduce the problem size from 12 qubits to 10 qubits.

The set of initial states for both algorithms and all molecular Hamiltonians were chosen to be the Hartree−Fock state and low-lying single-particle excitations above it. For example, if we wish to find three energy levels for four-qubit H_2, in the Jordan−Wigner mapping, this would be the set $\{|01\rangle_\alpha|01\rangle_\beta, |01\rangle_\alpha|10\rangle_\beta, |10\rangle_\alpha|01\rangle_\beta\}$. For LiH, this would be the set $\{|00001\rangle_\alpha|00001\rangle_\beta, |00001\rangle_\alpha|00010\rangle_\beta, |00010\rangle_\alpha|00001\rangle_\beta\}$, where the subscripts α and β denote the spin group [in the Qiskit implementation of n-qubit basis states in the Jordan−Wigner representation, the first $\frac{n}{2}$ qubits encode spin-up (α) and the second $\frac{n}{2}$ qubits encode spin-down (β); thus, we have chosen these three basis states because they are elements of the two-particle, spin magnetization $S_z = 0$ subspace of the full 2^n-dimensional Fock space]. Given that the UCCSD[32] ansatz preserves the numbers of spin-up and spin-down, the above choice of initial states constrains these variational quantum algorithms to search only the desired subspace, lessening the computational difficulty and producing physically meaningful results at the same time. For this reason, UCCSD is our ansatz of choice for these simulations. If we wish to find only two energy levels, we would omit the highest-energy state from the set. This choice of the set of states allows us to study two different types of algorithm initializations: one in which all of the ansatz parameters are randomly sampled according to a uniform distribution on $[-2\pi, 2\pi)$ and another one in which they are all set to zero. With the random parameter initialization, we can study the robustness of the algorithms with respect to their starting point in the parameter space. With the "zero vector" initialization, the UCCSD ansatz circuit is initialized to the identity and the states after the initial ansatz circuit application remain the Hartree−Fock state and low-lying single-particle excitations above the Hartree−Fock states. For most molecules, these states from Hartree−Fock calculation are good approximations of the ground state and low-lying excited states under FCI. The optimization from the "zero vector" initialization could be viewed as a local optimization and improves the performance of optimizers upon the random initialization. In Qiskit's implementation of ansatz circuits such as UCCSD, one can increase the expressiveness of the ansatz by repeating the corresponding quantum circuit block pattern r times, which comes at the cost of both the circuit depth and the number of variational parameters increasing by a factor of r. In this paper, we refer to an ansatz circuit that consists of the UCCSD circuit block pattern repeated r times as r-repetition UCCSD (or simply r-UCCSD). We run both algorithms for several different numbers of repetitions in order to account for the fact that the necessary ansatz expressiveness for convergence is not known *a priori*. Such a study allows us to explore the dependence of convergence success on the circuit depth for each algorithm.

When using a random initialization, we run each algorithm 10 times for each setting. In sections 4.1 and 4.3, for each setting, we plot one run roughly representative of the average convergence. In the noisy results in section 4.2, we plot the run which obtained the lowest objective function convergence. The complete set of all 10 runs for all of these simulations is plotted in the appendix, and their convergence success rates are summarized in Table 1 for completeness. When using the "zero vector" initialization, we run each algorithm only once. Running the algorithm multiple times for this initialization is neither necessary nor useful in the absence of noise because the outcome is deterministic for a given initial point.

4.1. H_2 and Hydrogen Square. We begin by presenting the results for the four- and eight-qubit systems we consider here: H_2 and the hydrogen square model. Figure 4 illustrates the convergence of the objective function for each algorithm when attempting to calculate the first three energy levels of the H_2 molecule at the equilibrium bond distance.

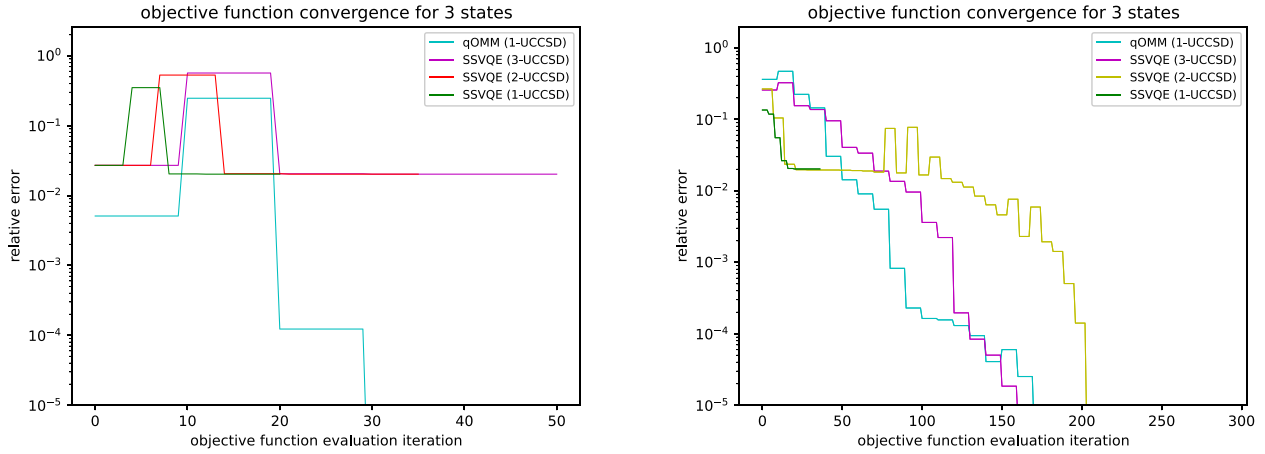
Figure 4b demonstrates how qOMM and SSVQE both converge to the global minimum, but notably, SSVQE cannot do so with just one-repetition UCCSD, requiring two repetitions in order to converge to within a relative accuracy of 10^{-5}. The success rate is further reported in Table 1. Figure 4a illustrates an attempt to solve the same problem, except the ansatz parameters are all initialized to zero. In this instance, we see that qOMM converges more quickly by a factor of roughly 5 compared to its random initialization counterpart, while SSVQE gets stuck in a local minimum and does not converge regardless of the ansatz circuit depth used. In both figures and all later convergence figures in this paper, we depict the convergence curve as the relative error (defined as $\frac{|f_{\text{exact}} - f_{\text{measured}}|}{|f_{\text{exact}}|}$) against the number of the objective function evaluations, which is considered the most expensive operation in VQE-type algorithms. The L-BFGS-B implementation used in this paper uses a two-point finite difference method such that each parameter is perturbed slightly from a given reference point common to all of them. Thus, for an n-parameter objective function, $n + 1$ function evaluations are required for each parameter update step in order to estimate the gradient. Hence, in all convergence figures, we observe stair-like curves of width $n + 1$.

Figure 5 illustrates the convergence results for the hydrogen square lattice at an interatomic distance of 1.23 Å. We can see

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689

Table 1. Success Rate of Given Problem Instances Converging to within a Relative Accuracy of 10^{-5} for a Given Number of UCCSD Repetitions Using a Randomized Ansatz Parameter Initialization

		UCCSD repetitions						
molecule	algorithm	1-rep	2-rep	3-rep	4-rep	5-rep	6-rep	7-rep
H_2 (0.735 Å)	qOMM (3 states)	100%						
	SSVQE (3 states)	0%	70%	100%				
H_2 (1.47 Å)	qOMM (3 states)	100%						
	SSVQE (3 states)	0%	90%	100%				
hydrogen square (1.23 Å)	qOMM (2 states)	0%	100%					
	qOMM (3 states)	0%	100%					
	SSVQE (2 states)		0%	100%				
	SSVQE (3 states)			0%	100%			
hydrogen square (2.46 Å)	qOMM (2 states)	0%	90%					
	qOMM (3 states)	0%	100%					
	SSVQE (2 states)		0%	100%				
	SSVQE (3 states)			0%	100%			
LiH (1.595 Å)	qOMM (2 states)	0%	100%					
	qOMM (3 states)	0%	100%					
	qOMM (4 states)		100%					
	qOMM (5 states)		100%					
	qOMM (6 states)		100%					
	qOMM (7 states)		100%					
	SSVQE (2 states)	0%	80%					
	SSVQE (3 states)	0%	0%	70%				
	SSVQE (4 states)			0%	100%			
	SSVQE (5 states)				10%	100%		
	SSVQE (6 states)					10%	100%	
	SSVQE (7 states)						60%	100%



(a) Zero vector initialization (b) Random initialization

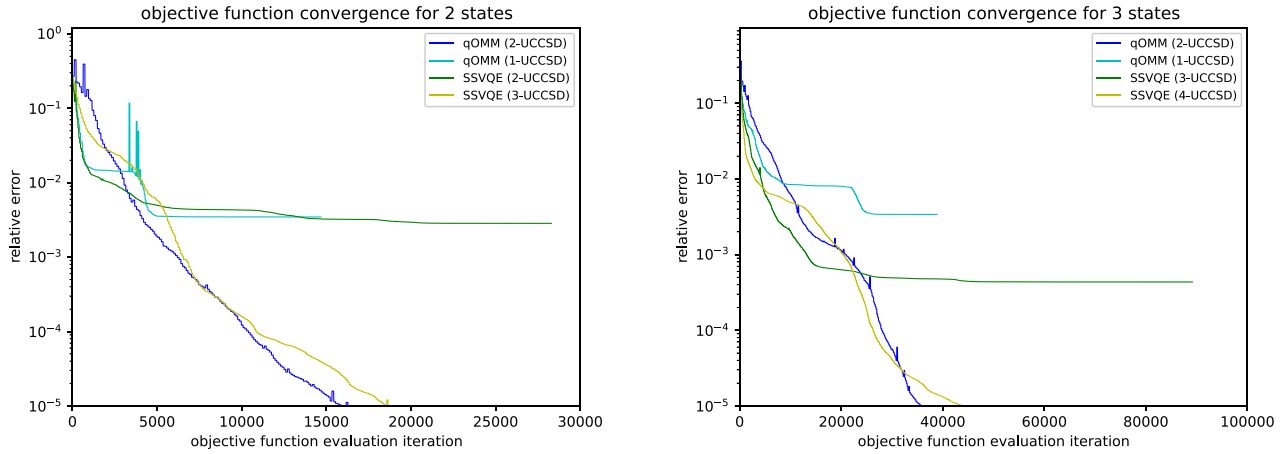
Figure 4. Convergence (noise-free) of the relative error $\frac{|f_i - f_{exact}|}{|f_{exact}|}$ of qOMM and SSVQE for three H_2 states at an interatomic distance of 0.735 Å.

from these figures that, similarly to H_2, the difference between SSVQE and qOMM is primarily in the number of UCCSD repetitions needed to converge. When calculating two states, qOMM requires 2-UCCSD, whereas SSVQE requires 3-UCCSD. When calculating three states, qOMM still requires only 2-UCCSD, whereas SSVQE requires 4-UCCSD. These observations are further tabulated in Table 1.

4.2. Noisy H_2 Simulations. We now run H_2 simulations in the presence of a classically simulated noise model on Qiskit's *AerSimulator*. Each one-qubit gate in the circuit is modeled as being accompanied by a local depolarizing channel with some probability of error p_{error}. Two-qubit gates are accompanied by a tensor product of two local depolarizing channels acting on each qubit. No error mitigation strategies are employed. We use the Parity mapping in order to use symmetry considerations to reduce the H_2 Hamiltonian to a two-qubit representation.[35] We construct ansatzes for each algorithm using the circuit block pattern shown in Figure 6.

This pattern can be repeated an arbitrary number of times to construct increasingly expressive ansatzes in the same way that the number of UCCSD repetitions was varied in the noiseless simulations. The Qiskit compiler is used to compile all circuits to the set of basis gets consisting of R_z, R_y, CNOT, $\sqrt{X}$, and the identity. For all problem instances, we use the minimum number

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689



(a) Two states (b) Three states

Figure 5. Convergence (noise-free) of the relative error $\frac{|f_i - f_{\text{exact}}|}{|f_{\text{exact}}|}$ of qOMM and SSVQE for the hydrogen square model at an interatomic distance of 1.23 Å using a random parameter initialization.

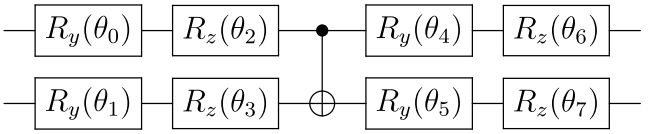
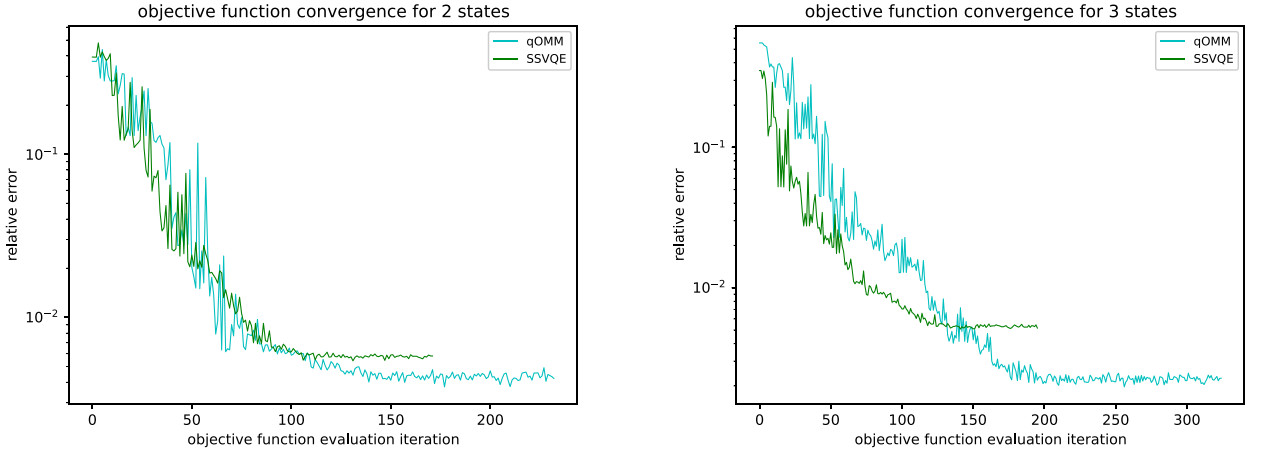


Figure 6. Base circuit block pattern used for noisy H_2 simulations.

of repetitions necessary to converge in the absence of noise in order to ensure that we are not simply measuring the ability of the ansatz to represent the global minimum. This corresponds to one repetition for qOMM and two repetitions for SSVQE. We use COBYLA as the classical optimization subroutine. This optimizer is more suitable for noisy simulations than L-BFGS-B due to its lack of a need to calculate gradient information. 10^6 circuit samples are used to evaluate all inner product terms and expectation values. p_{error} is set to 10^{-3}. Both SSVQE and qOMM are run 10 times with randomly initialized parameters. The convergence of the runs which achieved the lowest objective function values is given in Figure 7. The results for all 10 runs are given in Appendix A. We can see from these figures that both SSVQE and qOMM demonstrate a robustness to noise, although they do not achieve the same accuracy as the noiseless results.

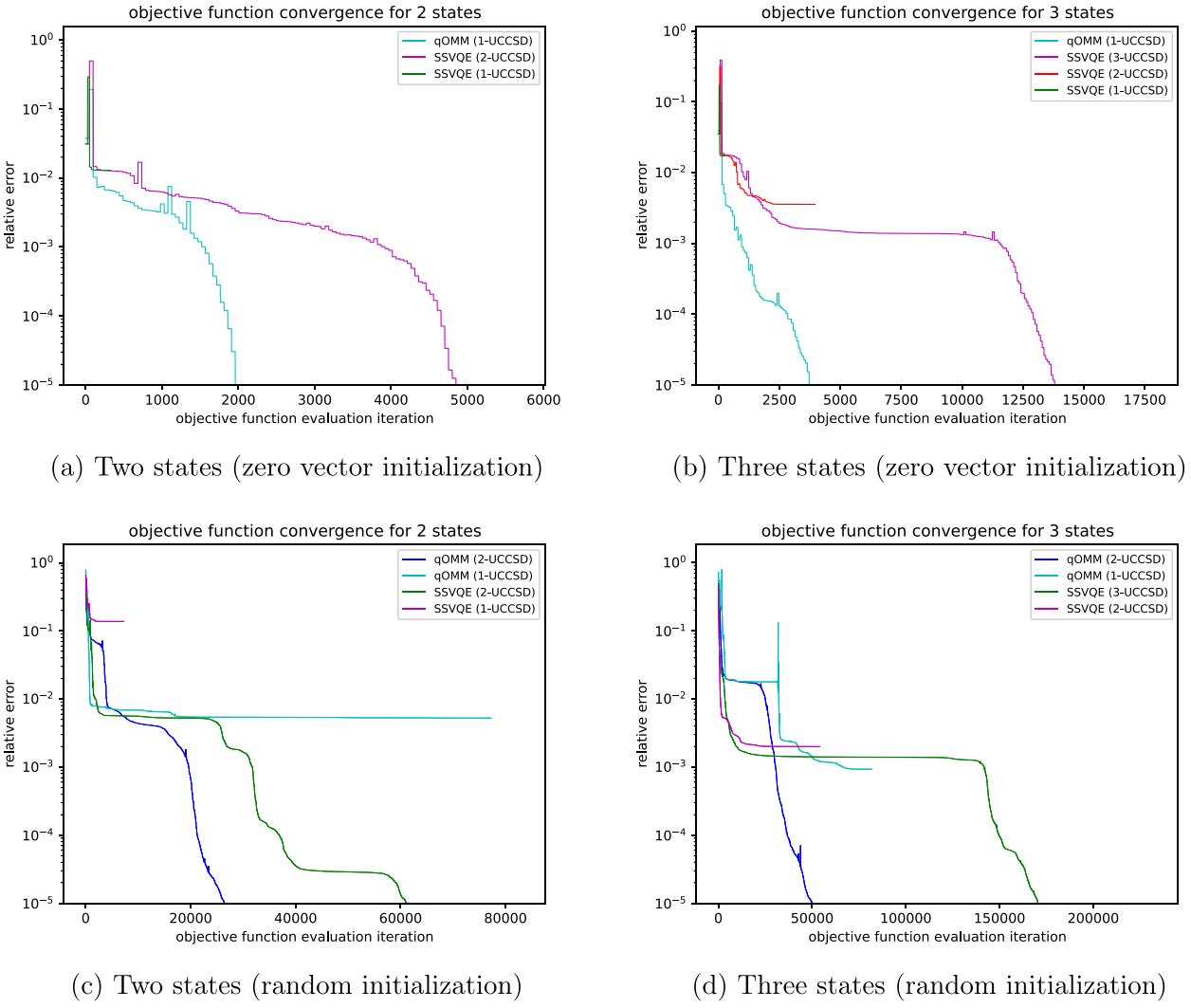
4.3. LiH. We now present the results for the 10-qubit system we consider here: LiH. Figure 8 illustrates the convergence of the objective function for both algorithms with various UCCSD ansatzes when calculating up to the first three energy levels. We can see from Figure 8c that, when the ansatz parameters are randomly initialized, both algorithms demonstrate the ability to converge within an accuracy of 10^{-5} but require two-repetition UCCSD to do so. From Figure 8a, we can see that, when all of the parameters are initialized to zero, both algorithms converge to within a relative accuracy of 10^{-5} much more quickly than their randomly initialized counterparts. Notably, qOMM requires only one-repetition UCCSD to achieve this convergence, while SSVQE requires two repetitions. When three states are calculated, qOMM can achieve a convergence below 10^{-5} with two-repetition UCCSD, whereas SSVQE requires



(a) Two states (b) Three states

Figure 7. Convergence of the relative error $\frac{|f_i - f_{\text{exact}}|}{|f_{\text{exact}}|}$ of qOMM and SSVQE for H_2 at an interatomic distance of 0.735 Å, where each circuit gate is modeled as having a probability of local depolarizing error $p_{\text{error}} = 0.001$.

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689



(a) Two states (zero vector initialization)

(b) Three states (zero vector initialization)

(c) Two states (random initialization)

(d) Three states (random initialization)

Figure 8. Convergence of the relative error $\frac{|f_i - f_{\text{exact}}|}{|f_{\text{exact}}|}$ of qOMM and SSVQE for LiH.

three repetitions. From Figure 8b we can see that when the ansatz parameters are all initialized to zero, both algorithms quickly converge to the global minimum. Notably, qOMM requires only 1-UCCSD repetition with this initialization. When attempting to find larger numbers of states, we can see from Table 1 that qOMM can converge with 2-UCCSD for all numbers of states considered, whereas SSVQE requires an increasing number of ansatz repetitions to find increasing numbers of states. In the following, we discuss the numerical results on LiH mainly from three perspectives: convergence plateau, initialization, and ansatz circuit depth.

Convergence Plateau. In many of the LiH convergence plots we show here, both qOMM and SSVQE initially converge at a rapid rate but then hit a plateau and stall the convergence until they escape from the "bad" region of the energy landscape. As long as the algorithm escapes from the "bad" region, the convergence rate resumes being rapid. These "bad" regions of the energy landscape (regardless of what their nature may be) seem to be the main obstacle for the convergence of both algorithms, which is more severe to SSVQE. Studying this apparent feature of the energy landscape may provide some insight as to how both algorithms may be improved and what limitations are inherent to the construction of the cost functions. From a numerical linear algebra point of view, such a convergence behavior is not very surprising. The objective function in SSVQE is convex, while with the consideration of the orthogonality constraint and the parametrization of the ansatz circuits, the energy landscape of SSVQE becomes non-convex. On the other hand, the objective function of qOMM, by itself, is non-convex but has no spurious local minima.[27,36,37] When the parametrization of the ansatz circuits is taken into consideration, the energy landscape of qOMM is non-convex and could have spurious local minima. For non-convex energy landscapes, usual optimizers, including L-BFGS-B, are efficient in a neighborhood of local minima whereas, around strict saddle points, without second-order information, the optimizers could stall there for a long time. From a practical point of view, we can further investigate the convergence of each eigenvalue. Since the issue is more severe for SSVQE, we focus on the SSVQE rather than qOMM. The construction of the SSVQE objective function allows us to obtain the estimated values for each eigenvalue at every function evaluation (in addition to the objective function as a whole) at a negligible additional computational cost. This allows us to gain some intuition as to why the SSVQE objective function convergence is observed to plateau in many of the LiH runs before steeply converging below a relative accuracy of 10^{-5}.

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689

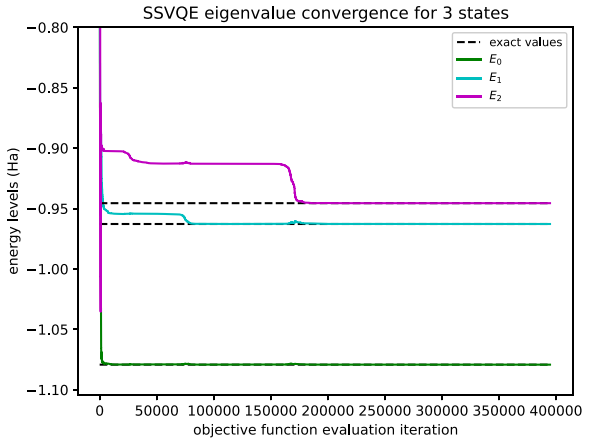


Figure 9. Convergence of the first three energy levels of LiH for SSVQE using a random parameter initialization.

In Figure 9 and Figure 10, we plot the eigenvalues and their relative accuracies, respectively, for one of the ten randomly initialized SSVQE runs. We can see from Figure 9 that while E_0 converges rapidly, the convergence of E_1 and E_2 plateau and do not converge for many more function evaluations. From Figure 10 we can see that E_1 does not escape from its plateau until the accuracy of E_0 is sufficiently decreased. This is followed by the accuracies of E_0 and E_1 increasing together. E_2 does not escape from its plateau until the accuracies of E_0 and E_1 are both sufficiently decreased, after which the accuracies of all three energy levels collectively increase, allowing the objective function to converge to its global minimum. While we only illustrate this for one of the ten runs, we observe that this qualitative behavior is typical for the other nine runs. Besides the non-convexity reason from objective functions, other reasons for this behavior could be related to two aspects of the weighted sum nature of the SSVQE algorithm. First, the same ansatz circuit is applied to all input states, meaning any change in the parameters has a direct impact on all the expectation value terms simultaneously. Second, the energy levels contribute unequally to the overall cost function, with the lower energy levels contributing more than the higher energy levels, an effect that becomes more pronounced due to the presence of the weight vector. Because qOMM enforces orthogonality (at the global minimum) through its overlap terms, the energy expectation value terms are associated with different independent parameter values. This means that changes in parameters that directly affect one of the energy expectation value terms in eq 4 do not directly change the others. They only change the overlap terms. This affords qOMM a flexibility that allows it to spend much fewer iterations escaping from the "bad" landscape regions. From Figure 8b and Figure 8a we see that starting with a good initial guess can reduce this effect, resulting in significantly improved convergence.

We also briefly note that although we have chosen to measure convergence speed in terms of the number of calls to the objective function, another valid measure of convergence speed would be the total number of optimization iterations if it can be assumed that one can parallelize multiple calls to the objective function. This is because one optimization iteration can require multiple calls to the objective function. In the noiseless results we have presented here, each optimization iteration of the L-BFGS-B optimizer requires n+1 calls to the objective function for n total parameters. For problem instances we have presented here where qOMM has more parameters than SSVQE, the speedup of qOMM over SSVQE would be even greater by this measure. For instance, in Figure 8d, qOMM with 2-UCCSD has 144 total parameters and SSVQE with 3-UCCSD has 72.

Initialization. Initialization plays an important role in the convergence for both qOMM and SSVQE. Even for the corresponding VQE ground state problem with one-repetition UCCSD, as in Figure 11, random parameter initialization on average converges in about 100 times more objective function evaluations than the zero vector parameter initialization. Both initializations in Figure 11 converge to a relative error 10^{-5}. In excited state computing, the two different initialization strategies for both qOMM and SSVQE not only differ in the number of iterations but also differ in the convergence results. Comparing Figures 8a and 8c, for SSVQE with two-repetition UCCSD, random parameter initialization converges more than ten times slower than that of zero vector parameter initialization. A similar result can be observed in Figures 8b and 8d for SSVQE with three-repetition UCCSD. Besides the difference in the number of iterations, optimization starting from random parameter initializations sometimes cannot find the global minima, as can be seen from Table 1. Overall, in all results included in this paper, zero vector parameter initializations outperform their

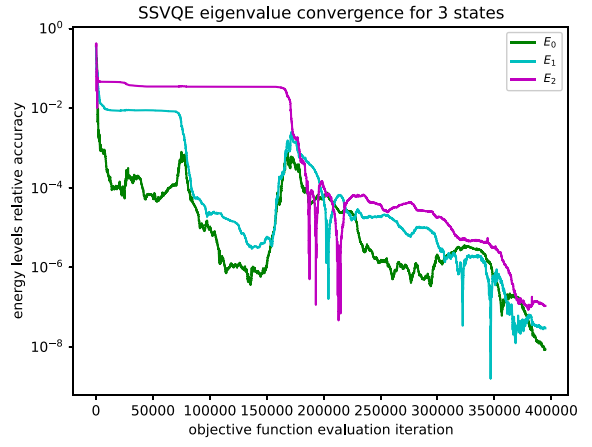


Figure 10. Relative accuracy $\frac{|E_{n,i} - E_{n,\text{exact}}|}{|E_{n,\text{exact}}|}$ of the convergence of the first three energy levels of LiH for SSVQE using a random parameter initialization.

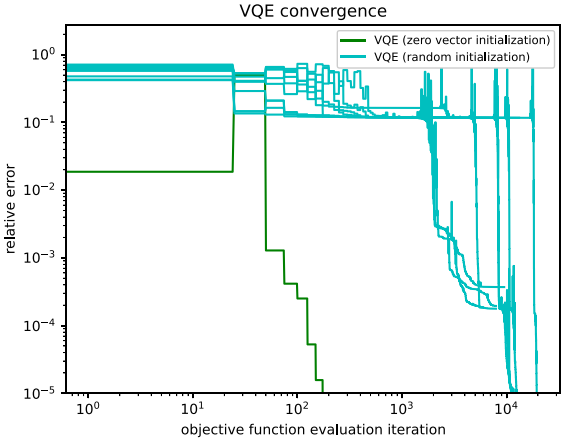


Figure 11. Convergence of the relative error $\frac{|E_{0,i} - E_{0,\text{exact}}|}{|E_{0,\text{exact}}|}$ of VQE for the ground state problem of LiH using a random parameter initialization.

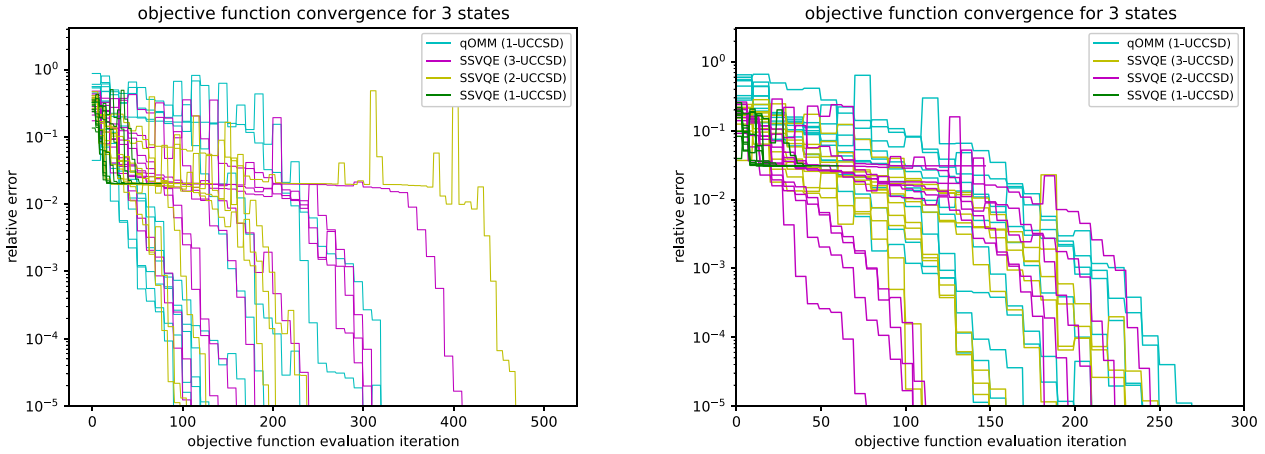
https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689

random parameter initialization counterparts for both algorithms. However, we still include numerical results for random parameter initializations to demonstrate the different performance between qOMM and SSVQE. This is because the zero vector initializations initialize the states to be excitations from Hartree−Fock states and only work for a limited number of low-lying states. If more excited states are needed, the zero vector initialization could even be a bad initialization, which is due to the fact that the excitation energies from Hartree−Fock may not be in the consistent order with that of FCI excitation energies. This "zero vector" initialization is demonstrated primarily not to advocate for its use as a scalable initialization strategy, but to illustrate what the improved convergence could look like if it can be assumed that one has a reasonable initialization strategy. The flexibility of variational ansatzes in qOMM would have another advantage when many excited states are needed. In qOMM, we could use zero vector parameter initialization for a few states and random parameter initialization for the remaining states without known good initializations. Such a mixed initialization strategy could avoid being trapped in bad local minima and benefit from the fast convergence of those states with zero vector parameter initialization. SSVQE, unfortunately, cannot use such a mixed initialization strategy due to the implicit orthogonality among initialized states. This motivates further investigation into developing and benchmarking initialization strategies more sophisticated than the ones we consider in this work. For example, the MCVQE paper[24] proposes an efficient quantum circuit implementation of CIS states for excited state initializations. This initialization would likely improve the convergence results of the simulations for both qOMM and SSVQE presented in this paper. Furthermore, because the optimization stage of MCVQE can be seen as a special case of SSVQE that uses a CIS initialization and an equal weighting of the states, performing analogous simulations using a CIS initialization would offer insight into how the convergence of qOMM compares to that of MCVQE. The approach of using chemically motivated initializations inspired from classical computational chemistry is likely to further improve the convergence of algorithms such as qOMM and SSVQE. The main challenge in this approach is that high level chemically inspired initializations require efficient low-level circuit implementation, which is highly nontrivial.

Ansatz Circuit Depth. Based on our numerical results, increasing the ansatz circuit depth, i.e., increasing the repetition in UCCSD, has two major impacts. First, the expressiveness increases as the ansatz circuit depth increases. In the ground state VQE, as in Figure 11, one-repetition UCCSD is sufficient for the ansatz to approximate the ground state to chemical accuracy. When two states of LiH are needed, one-repetition UCCSD, as in Figure 8, is not able to simultaneously approximate the ground state and the first excited states, whereas two-repetition UCCSD is able to. From the qOMM results in Figure 8a, we know that one-repetition UCCSD is able to approximate the ground state and the first excited state. When three states are needed, as in Figure 8b, we need three-repetition UCCSD to simultaneously approximate three states. Again, one-repetition UCCSD is still able to approximate any of these three states. More generally, 2-UCCSD is sufficient for all LiH cases studied when using qOMM, whereas SSVQE requires n-UCCSD in order to reliably find n states. The results for the hydrogen square model in Figure 5 demonstrates a similar pattern, except the situation is more severe for SSVQE, which requires more than n-UCCSD for finding n states. A more detailed analytical description of the ability of the UCCSD ansatz to simultaneously represent an increasing number of excited states as the number of repetitions is increased is not yet known. qOMM could adopt many UCCSDs, UCCSD with many repetitions, or a mixture of them as its variational ansatz. SSVQE, however, could only benefit from increasing the repetitions in UCCSD. The second impact of ansatz circuit depth relates to optimization. Another field full of non-convex optimization and parametrized ansatz is the deep neural network. An interesting theoretical result[38] therein shows that enlarging the parameter space, i.e., increasing the number of parameters, would make the optimization problem easier, i.e., the energy landscape would be full of the approximated global minimum. Although we have not established similar results for quantum ansatz circuits, we observe the behavior from our numerical results. In Figure 8b, one-repetition UCCSD is able to characterize the three states for LiH using qOMM. However, all ten random parameter initializations got stuck as can be seen in Table 1. As we increase the circuit depth from one-repetition to two-repetition UCCSD, qOMM rapidly converges to a global minimum for all 10 random parameter initializations.

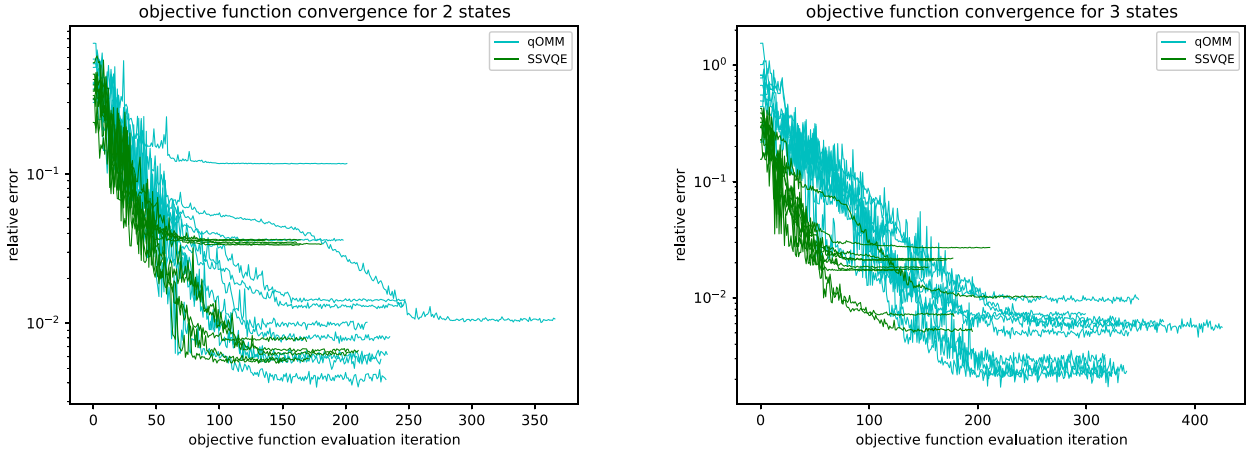
We conclude this section by discussing the circuit depth involved in calculating the inner product terms in the qOMM objective function. Currently, the only method for computing these terms that we are aware of involves running the circuit in Figure 1. At a minimum, the inner product circuit will incur a circuit depth twice that of the original ansatz. There will be additional circuit depth due to the need to compile controlled versions of the ansatz to the finite basis gate set of the machine. The extent of this additional depth compared to the original ansatz will depend on a number of factors including the basis gate set and qubit connectivity of the hardware, the choice of ansatz, and the efficiency of the compilation algorithm used. We can give some intuition for the circuit depths involved for the particular systems we studied here for a particular basis gate set. Using the Qiskit compiler, the gate depth for the 1-UCCSD ansatz used in our LiH simulations is approximately 1400 gates when compiled to the basis gate set consisting of R_z, $\sqrt{X}$, X, CNOT, and the identity. The corresponding circuit depth for computing the inner product terms is approximately 19,600 gates. These counts would be scaled roughly by a factor of n for n-UCCSD. This underscores the importance of future work that addresses how to reduce these gate counts. In particular, finding (if they exist) more efficient implementations of the inner product computations will be important for the applicability of qOMM in the NISQ era. We emphasize that, while there are several known subroutines for computing the overlap of two states for other applications such as quantum machine learning,[39] these generally compute real quantities of the form $|\langle\psi_1 | \psi_2\rangle|$, whereas for qOMM we need the full complex quantity $\langle\psi_1|\psi_2\rangle$. The observation that SSVQE appears to require an increasingly expressive ansatz for increasingly larger systems and an increasing number of states is also a concern. Future work that addresses how to mitigate this effect will be important for the applicability of SSVQE in the NISQ era. In particular, it has been shown that certain conditions can induce cost function landscapes with gradient magnitudes that vanish exponentially (i.e., barren plateaus) with the number of qubits, rendering convergence infeasible. This has been shown to occur in instances where one uses a hardware-efficient ansatz in combination with a non-local cost function[40] and in the presence of

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689



(a) 0.735 Å bond distance (b) 1.47 Å bond distance

Figure 12. Convergence of the relative error $\frac{|f_i - f_{\text{exact}}|}{|f_{\text{exact}}|}$ of qOMM and SSVQE for H_2 using a random parameter initialization.



(a) Two states (b) Three states

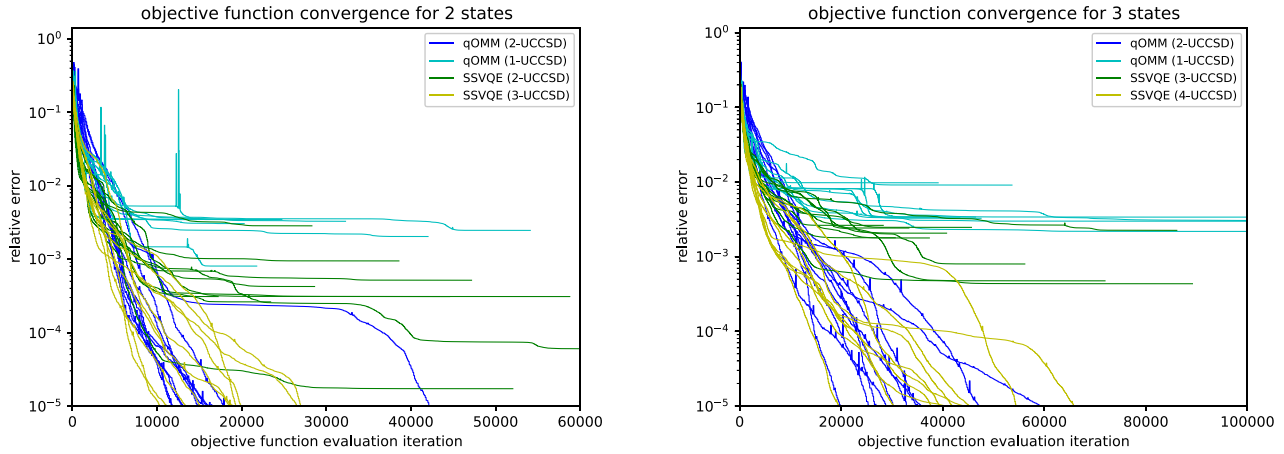
Figure 13. Convergence of the relative error $\frac{|f_i - f_{\text{exact}}|}{|f_{\text{exact}}|}$ of qOMM and SSVQE for H_2 at an interatomic distance of 0.735 Å, where each circuit gate is modeled as having a probability of local depolarizing error $p_{\text{error}} = 0.001$.

noise when one uses an ansatz with circuit depth that scales linearly or faster with the number of qubits.[41] Because both of these types of barren plateaus worsen with increasing circuit depth, the extent to which the two methods studied in this paper are susceptible to this effect will likely depend on how the circuit depth required for each algorithm to converge scales with increasing problem size. Such an investigation would require the simulation of larger chemical systems and would be an interesting direction of future research. We also note that there are various strategies that could be employed to facilitate the study of larger systems by representing the active space with as few qubits as possible, effectively delaying the onset of barren plateaus. A simple, well-known example of this idea was used in this work when we froze the two core orbitals in LiH to reduce the problem size from 12 qubits to 10. Another strategy would be to optimize the basis set under a fixed qubit budget, as is done in OptOrbFCI in the context of classical computational chemistry.[8] Generalizing this work to quantum variational algorithms such as VQE, SSVQE, and qOMM will be the topic of future work.

5. CONCLUSION

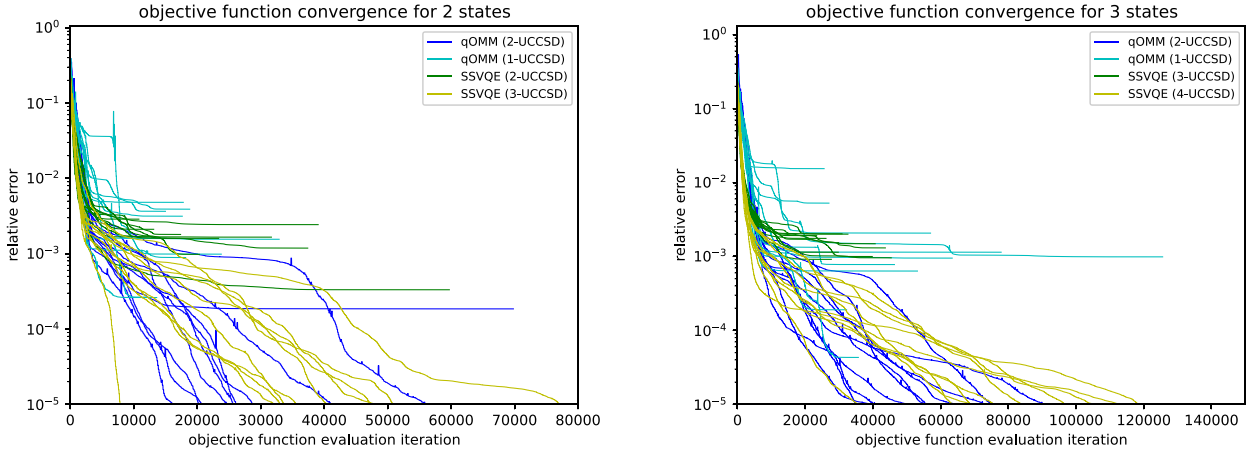
In this work, we have proposed a method which can calculate the low-lying eigenvalue/eigenstate pairs of a Hermitian operator on quantum hardware. We have classically simulated the algorithm for three different electronic structure Hamiltonians: H_2, LiH, and the hydrogen square lattice. We have compared it to SSVQE, another algorithm with the same goal, but one which enforces the orthogonality of the input states explicitly at every optimization step. We showed that a small problem such as H_2 is not difficult enough to meaningfully distinguish any differences in performance for these two approaches when using a random initialization strategy, but when one compares them to more moderately sized problems such as 10-qubit LiH, noticeable differences in performance begin to emerge. In this scenerio, qOMM converges with fewer objective function evaluations and with a less expressive ansatz in all but one of the simulations considered here. Each of these methods incurs additional circuit depth beyond the ground state problem that should be addressed by future works in order to increase their applicability in the NISQ era. This

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689



(a) Two states (b) Three states

Figure 14. Convergence of the relative error $\frac{|f_i - f_{\text{exact}}|}{|f_{\text{exact}}|}$ of qOMM and SSVQE for the hydrogen square model at an interatomic distance of 1.23 Å using a random parameter initialization.



(a) Two states (b) Three states

Figure 15. Convergence of the relative error $\frac{|f_i - f_{\text{exact}}|}{|f_{\text{exact}}|}$ of qOMM and SSVQE for the hydrogen square model at an interatomic distance of 2.46 Å using a random parameter initialization.

corresponds to finding more efficient methods for computing the inner product terms in qOMM and reducing the ansatz circuit depth needed for SSVQE to converge efficiently. One potential direction for attempting to improve SSVQE would be to investigate the extent to which incorporating adaptive ansatz strategies[42−44] would alleviate this apparent feature of SSVQE. These methods were developed in order to find depth-efficient ansatz circuits for problem instances for which doing so by manual heuristic guesswork is difficult. This is precisely the problem from which we observed SSVQE to suffer in this work. qOMM could benefit from these methods as well. Additionally, one could investigate whether or not modifying the SSVQE weight vector to adaptively change over the course of the optimization could mitigate the objective function plateau effect we observed in our simulations.

The LiH simulations demonstrate that, in the context of these types of excited state methods, even a crude initialization strategy can greatly reduce not only the number of optimization iterations needed to converge but also the circuit depth needed to do so. Another potentially interesting topic would be to develop and benchmark more powerful and versatile initialization strategies that are tailored specifically to these excited state methods. This would also improve the applicability of these methods in the NISQ era.

A. ADDITIONAL CONVERGENCE PLOTS

Here we provide all 10 runs for each of the randomly initialized tests for which only one run was shown in section 4. We also provide some additional tests such as H_2 and the hydrogen square model at stretched bond distances and up to seven states of LiH. The success rates for the noise-free simulations are summarized in Table 1. Figure 12 depicts all 10 randomly initialized runs for the H_2 results discussed in section 4.1. Figure 13 depicts all of the 10 randomly initialized runs for the noisy H_2 results discussed in section 4.2. Figure 14 depicts all of the 10 randomly initialized runs for the hydrogen square toy model at equilibrium bond distance discussed in section 4.1. Figure 15 further depicts 10 randomly initialized

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689

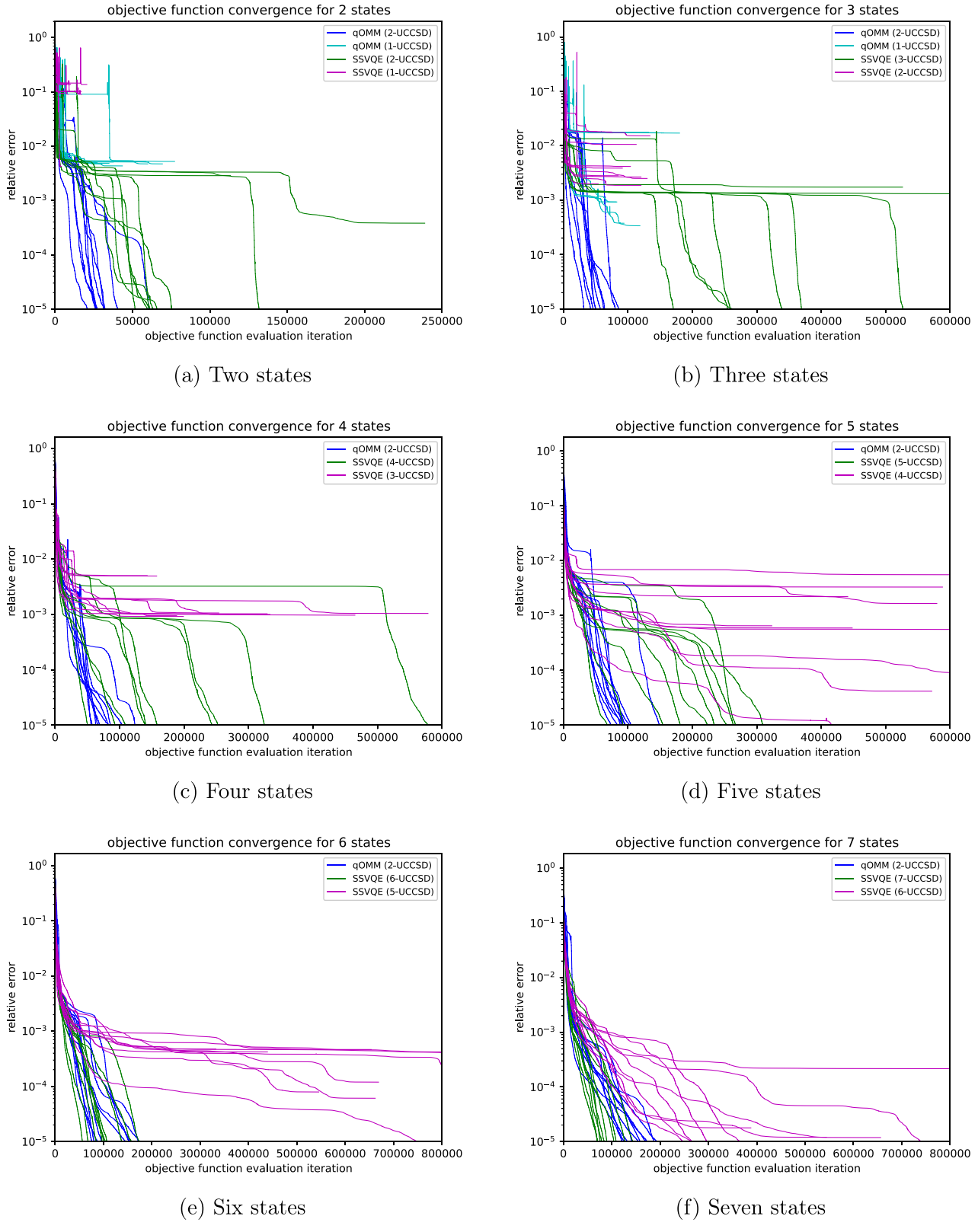


Figure 16. Convergence of the relative error $\frac{|f_i - f_{\text{exact}}|}{|f_{\text{exact}}|}$ of qOMM and SSVQE for LiH using randomly initialized parameters.

runs for this same system at a stretched bond distance. Figure 16 is an extension of the LiH results discussed in section 4.3, where all 10 randomly initialized runs for the results given in that section are depicted here as well as randomly initialized results for up to seven states.

AUTHOR INFORMATION

Corresponding Authors

Yingzhou Li − *School of Mathematical Sciences, Fudan University, Shanghai 200433, China*; Email: yingzhouli@fudan.edu.cn

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689

Jianfeng Lu − *Department of Physics, Duke University, Durham, North Carolina 27708-0187, United States; Department of Mathematics and Department of Chemistry, Duke University, Durham, North Carolina 27708-0187, United States*; Email: jianfeng@math.duke.edu

Author

Joel Bierman − *Department of Physics, Duke University, Durham, North Carolina 27708-0187, United States;* orcid.org/0000-0002-4311-4548

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00218

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported in part by the US National Science Foundation under award CHE-2037263 and by the US Department of Energy via grant DE-SC0019449.

■ REFERENCES

(1) Arute, F.; Arya, K.; Babbush, R.; Bacon, D.; Bardin, J. C.; Barends, R.; Biswas, R.; Boixo, S.; Brandao, F. G. S. L.; Buell, D. A.; et al. Quantum supremacy using a programmable superconducting processor. *Nature* **2019**, *574*, 505−510.

(2) Zhong, H.-S.; Wang, H.; Deng, Y.-H.; Chen, M.-C.; Peng, L.-C.; Luo, Y.-H.; Qin, J.; Wu, D.; Ding, X.; Hu, Y.; et al. Quantum computational advantage using photons. *Science* **2020**, *370*, 1460−1463.

(3) Blunt, N. S.; Smart, S. D.; Booth, G. H.; Alavi, A. An excited-state approach within full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2015**, *143*, 134117.

(4) Holmes, A. A.; Umrigar, C. J.; Sharma, S. Excited states using semistochastic heat-bath configuration interaction. *J. Chem. Phys.* **2017**, *147*, 164111.

(5) Schriber, J. B.; Evangelista, F. A. Adaptive configuration interaction for computing challenging electronic excited states with tunable accuracy. *J. Chem. Theory Comput.* **2017**, *13*, 5354−5366.

(6) Li, Y.; Lu, J.; Wang, Z. CoordinateWise descent methods for leading eigenvalue problem. *SIAM J. Sci. Comput.* **2019**, *41*, A2681−A2716.

(7) Wang, Z.; Li, Y.; Lu, J. Coordinate descent full configuration interaction. *J. Chem. Theory Comput.* **2019**, *15*, 3558−3569.

(8) Li, Y.; Lu, J. Optimal orbital selection for full configuration interaction (OptOrbFCI): Pursuing basis set limit under budget. *J. Chem. Theory Comput.* **2020**, *16*, 6207−6221.

(9) Abrams, D. S.; Lloyd, S. Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors. *Phys. Rev. Lett.* **1999**, *83*, 5162−5165.

(10) Aspuru-Guzik, A.; Dutoi, A. D.; Love, P. J.; Head-Gordon, M. Simulated quantum computation of molecular energies. *Science* **2005**, *309*, 1704−1707.

(11) Peruzzo, A.; McClean, J.; Shadbolt, P.; Yung, M. H.; Zhou, X. Q.; Love, P. J.; Aspuru-Guzik, A.; O'Brien, J. L. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **2014**, *5*, 4213.

(12) Bauer, B.; Bravyi, S.; Motta, M.; Chan, G. K.-L. Quantum algorithms for quantum chemistry and quantum materials science. *Chem. Rev.* **2020**, *120*, 12685−12717.

(13) McArdle, S.; Endo, S.; Aspuru-Guzik, A.; Benjamin, S. C.; Yuan, X. Quantum computational chemistry. *Rev. Mod. Phys.* **2020**, *92*, 015003.

(14) McClean, J. R.; Kimchi-Schwartz, M. E.; Carter, J.; De Jong, W. A. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A* **2017**, *95*, 042308.

(15) Colless, J. I.; Ramasesh, V. V.; Dahlen, D.; Blok, M. S.; Kimchi-Schwartz, M. E.; McClean, J. R.; Carter, J.; De Jong, W. A.; Siddiqi, I. Computation of molecular spectra on a quantum processor with an error-resilient algorithm. *Physical Review X* **2018**, *8*, 011021.

(16) Ganzhorn, M.; Egger, D. J.; Barkoutsos, P.; Ollitrault, P.; Salis, G.; Moll, N.; Roth, M.; Fuhrer, A.; Mueller, P.; Woerner, S.; et al. Gate-efficient simulation of molecular eigenstates on a quantum computer. *Physical Review Applied* **2019**, *11*, 044092.

(17) Ollitrault, P. J.; Kandala, A.; Chen, C.-F.; Barkoutsos, P. K.; Mezzacapo, A.; Pistoia, M.; Sheldon, S.; Woerner, S.; Gambetta, J. M.; Tavernelli, I. Quantum equation of motion for computing molecular excitation energies on a noisy quantum processor. *Phys. Rev. Res.* **2020**, *2*, 043140.

(18) McClean, J. R.; Romero, J.; Babbush, R.; Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* **2016**, *18*, 023023.

(19) Santagati, R.; Wang, J.; Gentile, A. A.; Paesani, S.; Wiebe, N.; McClean, J.; Bonneau, D.; Silverstone, J. W.; Morley-Short, S.; Shadbolt, P. J.; et al. Finding excited states of physical Hamiltonians on a silicon quantum photonic device. *Front. Opt.* **2017**, FM4E.2.

(20) Santagati, R.; Wang, J.; Gentile, A. A.; Paesani, S.; Wiebe, N.; McClean, J. R.; Morley-Short, S.; Shadbolt, P. J.; Bonneau, D.; Silverstone, J. W.; et al. Witnessing eigenstates for quantum simulation of Hamiltonian spectra. *Sci. Adv.* **2018**, *4*, No. eaap9646.

(21) Higgott, O.; Wang, D.; Brierley, S. Variational quantum computation of excited states. *Quantum* **2019**, *3*, 156.

(22) Jones, T.; Endo, S.; McArdle, S.; Yuan, X.; Benjamin, S. C. Variational quantum algorithms for discovering Hamiltonian spectra. *Phys. Rev. A* **2019**, *99*, 62304.

(23) Nakanishi, K. M.; Mitarai, K.; Fujii, K. Subspace-search variational quantum eigensolver for excited states. *Physical Review Research* **2019**, *1*, 033062.

(24) Parrish, R. M.; Hohenstein, E. G.; McMahon, P. L.; Martínez, T. J. Quantum computation of electronic transitions using a variational quantum eigensolver. *Phys. Rev. Lett.* **2019**, *122*, 230401.

(25) Ordejon, P.; Drabold, D. A.; Grumbach, M. P.; Martin, R. M. Unconstrained minimization approach for electronic computations that scales linearly with system size. *Phys. Rev. B* **1993**, *48*, 14646.

(26) Mauri, F.; Galli, G.; Car, R. Orbital formulation for electronic-structure calculations with linear system-size scaling. *Phys. Rev. B* **1993**, *47*, 9973.

(27) Lu, J.; Thicke, K. Orbital minimization method with l^1 regularization. *J. Comput. Phys.* **2017**, *336*, 87−103.

(28) Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing* **1995**, *16*, 1190−1208.

(29) Kandala, A.; Mezzacapo, A.; Temme, K.; Takita, M.; Brink, M.; Chow, J. M.; Gambetta, J. M. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **2017**, *549*, 242−246.

(30) Gacon, J.; Zoufal, C.; Carleo, G.; Woerner, S. Simultaneous perturbation stochastic approximation of the quantum Fisher information. *Quantum* **2021**, *5*, 567.

(31) Childs, A. M.; Wiebe, N. Hamiltonian simulation using linear combinations of unitary operations. *Quantum Info. Comput.* **2012**, *12*, 901−924.

(32) Romero, J.; Babbush, R.; McClean, J. R.; Hempel, C.; Love, P. J.; Aspuru-Guzik, A. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology* **2019**, *4*, 014008.

(33) Anis, M. S.; Abraham, H.; AduOffei; Agarwal, R.; Agliardi, G.; Aharoni, M.; Akhalwaya, I. Y.; Aleksandrowicz, G.; Alexander, T.; Amy, M. *Qiskit: An Open-source Framework for Quantum Computing* **2021**, DOI: 10.5281/zenodo.2562111.

(34) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; et al. PySCF: the python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, No. e1340.

(35) Bravyi, S.; Gambetta, J. M.; Mezzacapo, A.; Temme, K. Tapering off qubits to simulate fermionic Hamiltonians. *arXiv*, January 27, 2017. DOI: 10.48550/arXiv.1701.08213.
(36) Gao, W.; Li, Y.; Lu, B. Triangularized Orthogonalization-free Method for Solving Extreme Eigenvalue Problems. *arXiv*, May 25, 2020. DOI: 10.48550/arXiv.2005.12161.
(37) Gao, W.; Li, Y.; Lu, B. Global Convergence of Triangularized Orthogonalization-free Method. *arXiv*, October 12, 2021. DOI: 10.48550/arXiv.2110.06212.
(38) Allen-Zhu, Z.; Li, Y.; Liang, Y. Learning and generalization in overparameterized neural networks, going beyond two layers. *Advances in Neural Information Processing Systems*. 2019.
(39) Cincio, L.; Subaşı, Y.; Sornborger, A. T.; Coles, P. J. Learning the quantum algorithm for state overlap. *New J. Phys.* **2018**, *20*, 113022.
(40) Cerezo, M.; Sone, A.; Volkoff, T.; Cincio, L.; Coles, P. J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat. Commun.* **2021**, *12*, 1791.
(41) Wang, S.; Fontana, E.; Cerezo, M.; Sharma, K.; Sone, A.; Cincio, L.; Coles, P. J. Noise-induced barren plateaus in variational quantum algorithms. *Nat. Commun.* **2021**, *12*, 6961.
(42) Grimsley, H. R.; Economou, S. E.; Barnes, E.; Mayhall, N. J. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nat. Commun.* **2019**, *10*, 3007.
(43) Tang, H. L.; Shkolnikov, V.; Barron, G. S.; Grimsley, H. R.; Mayhall, N. J.; Barnes, E.; Economou, S. E. Qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor. *PRX Quantum* **2021**, *2*, 020310.
(44) Yordanov, Y. S.; Armaos, V.; Barnes, C. H.; Arvidsson-Shukur, D. R. Qubit-excitation-based adaptive variational quantum eigensolver. *Communications Physics* **2021**, *4*, 228.

https://doi.org/10.1021/acs.jctc.2c00218
J. Chem. Theory Comput. 2022, 18, 4674−4689