Variational quantum eigensolvers by variance minimization

Dan-Bo Zhang(张旦波)[1,2,†], Bin-Lin Chen(陈彬琳)[2], Zhan-Hao Yuan(原展豪)[3], and Tao Yin(殷涛)[4,‡]

[1] *Guangdong–Hong Kong Joint Laboratory of Quantum Matter, Frontier Research Institute for Physics, South China Normal University, Guangzhou* 510006, *China*

[2] *Guangdong Provincial Key Laboratory of Quantum Engineering and Quantum Materials, School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou* 510006, *China*

[3] *Guangzhou Educational Infrastructure and Equipment Center, Guangzhou* 510006, *China*

[4] *Yuntao Quantum Technologies, Shenzhen* 518000, *China*

(Received 23 May 2022; revised manuscript received 16 August 2022; accepted manuscript online 18 August 2022)

The original variational quantum eigensolver (VQE) typically minimizes energy with hybrid quantum-classical optimization that aims to find the ground state. Here, we propose a VQE based on minimizing energy variance and call it the variance-VQE, which treats the ground state and excited states on the same footing, since an arbitrary eigenstate for a Hamiltonian should have zero energy variance. We demonstrate the properties of the variance-VQE for solving a set of excited states in quantum chemistry problems. Remarkably, we show that optimization of a combination of energy and variance may be more efficient to find low-energy excited states than those of minimizing energy or variance alone. We further reveal that the optimization can be boosted with stochastic gradient descent by Hamiltonian sampling, which uses only a few terms of the Hamiltonian and thus significantly reduces the quantum resource for evaluating variance and its gradients.

Keywords: quantum computing, quantum algorithm, quantum chemistry

PACS: 03.67.Ac **DOI:** 10.1088/1674-1056/ac8a8d

1. Introduction

The variational principle is instrumental for understanding physical theories and also becomes a powerful tool for solving computational physics problems. In recent years, the power of quantum computing of near-term noisy quantum processors is expected to be exploited with variational methods,[1–24] which refers to hybrid quantum-classical algorithms for optimization. A representative example is variational quantum eigensolver (VQE),[1,3,7,15,19,22] which receives special interests due to its fundamental roles in quantum chemistry,[1,3,7,22,25] quantum many-body systems,[13,19] and many other applications.[9,26,27]

Interestingly, though called eigensolver, the original VQE is designed to solve the ground state for a Hamiltonian, since it uses the energy as a cost function to minimize, which promises to give an upper bound for the ground state energy. For being an eigensolver, many variants of VQE have been developed to solve excited states, which can be understood as minimizing energy in a subspace, constrained by symmetries,[28] enforced by orthogonality,[29] obtained by eliminating the space of lower energy states,[16] or using quantum subspace expansion.[5] As excited states are responsible for physical processes such as chemical reactions, advances in VQE for calculating excited states are arguably important.

While those variants of VQE can be successful to some extent, an alternative approach is to directly access whether a state is an eigenstate and use this criterion as a cost function to optimize. A simple answer for this is to use zero energy variance to judge an eigenstate, since an eigenstate should have zero energy variance. Variational methods based on the zero-energy variance principle that run on classical computers,[30–36] in fact, can be originated in the early age of quantum mechanics for solving the Helium atom.[30] Remarkably, zero variance puts a very strong constraint on the wavefunction structure, and it can be powerful for solving quantum many-body problems when a well-approximated wavefunction is required.[35] The zero-variance principle has been applied experimentally to self verify if a quantum simulator can prepare a ground state for a Hamiltonian.[13] It is, however, awaiting for exploiting the power of quantum computing to develop a variational quantum eigensolver that is based on minimizing energy variance, which can solve ground state and excited states on the same footing.

In this paper, we develop a variational quantum eigensolver based on minimizing energy variance, and we call it the variance-VQE. For solving excited states of a Hamiltonian, we give two approaches, which represent eigenstates in different ways. One uses a single wavefunction ansatz, where different parameters correspond to different eigenstates. The other uses a set of orthogonal wavefunction ansatzs, and incorporates energy variances into one cost function for optimization. We numerically solve the energy potential surface for

†Corresponding author. E-mail: dbzhang@m.scnu.edu.cn

‡Corresponding author. E-mail: tao.yin@artiste-qb.net

© 2022 Chinese Physical Society and IOP Publishing Ltd

molecules, which is a fundamental quantum chemistry problem, and demonstrate the properties of variance VQE as well as explore its potential advantages. We also show that optimizing a combination of energy and variance may be more efficient in finding a set of eigenstates with lowest energies, compared with one that optimizes energy or variance alone. Moreover, we investigate stochastic gradient descent for optimizing the variance-VQE with Hamiltonian sampling, which can be useful for optimization with fewer quantum resources. Our work demonstrates that VQE by minimizing energy variance can be useful for calculating excited states and also for working out an avenue to reach efficient optimization.

The paper is organized as follows. We first propose variance VQE in Section 2, and then apply it to solve excited states for quantum chemistry problems in Section 3. In Section 4, we investigate the stochastic gradient for optimizing the variance-VQE. Finally, we give a summary in Section 5.

2. Variance VQE

In this section, we first review the traditional approach of variational quantum eigensolver that minimizes the energy. Then, we formulate a new type of VQE that minimizes the energy variance and then discusses its optimization.

2.1. VQE by energy minimization

We consider a Hamiltonian H as a summation of local terms, $H=\sum_{i=1}^{N} c_i L_i$, where a local Hamiltonian L_i can be written as a tensor product of a few Pauli matrices. We denote $\boldsymbol{c}=(c_1,c_2,\ldots,c_N)^{\mathrm{T}}$ and $\boldsymbol{L}=(L_1,L_2,\ldots,L_N)^{\mathrm{T}}$. Thus we can write $H=\boldsymbol{c}^{\mathrm{T}}\boldsymbol{L}$.

The variational quantum eigensolver works as follows. One uses an ansatz $|\psi(\boldsymbol{\beta})\rangle = U(\boldsymbol{\beta})|R\rangle$ to represent a candidate ground state. Here $|R\rangle$ usually is taken as an approximation for the ground state of H, e.g., Hartree–Fock state for the quantum chemistry problem. $U(\boldsymbol{\beta})$ is a unitary operator parameterized with $\boldsymbol{\beta}$. The task then is to optimize $\boldsymbol{\beta}$ for some cost function. The traditional way is to minimize the energy defined as

$$\mathcal{E}(\boldsymbol{\beta}) = \langle\psi(\boldsymbol{\beta})|H|\psi(\boldsymbol{\beta})\rangle = \mathrm{Tr}[\psi(\boldsymbol{\beta})H]. \tag{1}$$

Here, we have denoted $\psi(\boldsymbol{\beta}) = |\psi(\boldsymbol{\beta})\rangle\langle\psi(\boldsymbol{\beta})|$. By designing, VQE based on minimizing energy is suitable for finding ground state of a Hamiltonian.

The optimization is completed with a hybrid quantum-classical algorithm. The quantum processor prepares $\psi(\boldsymbol{\beta})$ and performs measurements to evaluate $\mathcal{E}(\boldsymbol{\beta})$, which can be reduced into $\mathcal{E}(\boldsymbol{\beta}) = \boldsymbol{c}^{\mathrm{T}}\mathcal{L}(\boldsymbol{\beta})$, where $\mathcal{L}(\boldsymbol{\beta}) = \mathrm{Tr}(\psi(\boldsymbol{\beta})\boldsymbol{L})$. Here, quantum average of each component of $\boldsymbol{L}$ corresponds to a joint measurement on multiple qubits. The classical computer updates parameters $\boldsymbol{\beta}$ according to received data from the quantum processor, which forms a hybrid quantum-classical optimization loop.

2.2. VQE by variance minimization

As the name indicates, variational quantum eigensolver should aim at solving eigenstates. However, VQE, based on minimizing energy, is prone to find only an eigenstate with the lowest energy, named as the ground state. Other VQEs have been developed for solving excited states. Essentially, those VQEs are realized by minimizing energy in a subspace, which is enforced with symmetry, or by eliminating space of lower energy states with an orthogonality condition.

Can variational quantum eigensolver directly solve excited states? An answer for this is to design a cost function that can assign the same cost to all eigenstates and higher cost for other states. A natural choice is energy variance, which is zero only for eigenstates and positive for others. By minimizing energy variance, one can then find the eigenstates of a Hamiltonian. We call these two methods the minimized energy variance-VQE and energy-VQE, respectively.

Let us formulate variance-VQE. The energy variance of Hamiltonian H with a wavefunction $\psi(\boldsymbol{\beta})$ can be written as

$$\Delta(\boldsymbol{\beta}) \equiv \left\langle H^2\right\rangle_{\boldsymbol{\beta}} - \langle H\rangle_{\boldsymbol{\beta}}^2 \geq 0, \tag{2}$$

where $\langle *\rangle_{\boldsymbol{\beta}} = \langle\psi(\boldsymbol{\beta})|*|\psi(\boldsymbol{\beta})\rangle$. It is useful to introduce the quantum covariance matrix[37–39] defined as

$$\begin{aligned}\mathcal{G}_{ij}(\boldsymbol{\beta}) &= \left\langle L_i L_j\right\rangle_{\boldsymbol{\beta}} - \langle L_i\rangle_{\boldsymbol{\beta}}\left\langle L_j\right\rangle_{\boldsymbol{\beta}} \\ &= \mathrm{Tr}[\psi(\boldsymbol{\beta})L_iL_j] - \mathrm{Tr}[\psi(\boldsymbol{\beta})L_i]\,\mathrm{Tr}[\psi(\boldsymbol{\beta})L_j].\end{aligned} \tag{3}$$

Then, the energy variance can be expressed as

$$\Delta(\boldsymbol{\beta}) = \boldsymbol{c}^{\mathrm{T}}\mathcal{G}(\boldsymbol{\beta})\boldsymbol{c} \geq 0. \tag{4}$$

The non-negative energy variance consists of $\mathcal{G}(\boldsymbol{\beta})$, which is a semi-positive matrix. Note that each element of $\mathcal{G}(\boldsymbol{\beta})$ can be obtained on a quantum computer. Then, the energy variance can be evaluated based on the quantum covariance matrix.

As a variational method, one key ingredient for variance-VQE is whether the wavefunciton ansatz has enough expressive power to parameterize the eigenstates. Unlike energy-VQE which looks for a ground state, the chance that an ansatz can parameterize some eigenstates can be greater as there are 2^M eigenstates for a quantum system of M qubits, although it is in general hard for an ansatz to parameterize all eigenstates. Notably, with an intermediate-size variational quantum circuit, it is expected that lower-lying eigenstates may be parameterized, as those states typically will be characterized with smaller entanglement entropy.[40] This can be useful as most applications are interested in lower-lying eigenstates. In addition, the variance-VQE has the self-verifying property that it can verify whether it indeed obtains an eigenstate by checking if the energy variance is zero.

Another important aspect is optimization, namely minimizing the energy variance, which can be carried with gradient free or gradient descents. The gradient descent method requires to calculate the gradient of energy variance, which is

$$\frac{\partial \Delta(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \boldsymbol{c}^{\mathrm{T}} \frac{\partial \mathcal{G}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \boldsymbol{c}. \tag{5}$$

Each element $\partial \mathcal{G}_{ij}/\partial \boldsymbol{\beta}$ can be obtained using numeral differential or the shift rule.[41] The gradient descent updates $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta}^t = \boldsymbol{\beta}^{t-1} - \eta \frac{\partial \Delta(\boldsymbol{\beta}^{t-1})}{\partial \boldsymbol{\beta}}. \tag{6}$$

Since $\mathcal{G}(\boldsymbol{\beta})$ should be evaluated with $O(N^2)$ elements, calculating energy variance costs $O(N^2)$, which is larger than a cost of $O(N)$ for calculating energy (see Appendix for details as well as a demonstrative comparison using a 2-qubit Hamiltonian). As a result, direct gradient descent for variance-VQE can be more resource-costing than energy-VQE. Nevertheless, we can use Hamiltonian sampling to reduce this cost,[42] which estimates the energy variance and its gradient by sampling only a portion of the quantum covariance matrix. In Section 4, we will give a stochastic gradient descent algorithm for optimizing variance-VQE by Hamiltonian sampling,[42] which can be efficient under a minimal sampling rate and thus can reduce the overload significantly. Moreover, one may reduce the overload of evaluating the energy variance by using some smarter strategies that can simultaneously access a set of quantum covariance matrix elements,[43,44] or use the quantum power method introduced in Ref. [45].

3. The variance-VQE for excited states

As all eigenstates always have zero variance, optimizing the variance-VQE by minimizing energy variance may lead to any eigenstate. Thus, the variance-VQE can be used for obtaining excited states. In this section, we propose different approaches for solving excited states of molecule Hamiltonians with the variance VQE. We first show that the cost function variance-VQE can have many global minimums of zero variance, corresponding to different excited states. However, this approach can be inefficient if a specified eigenstate is required. To overcome this, we propose the variance-VQE that solves a set of excited states jointly, using a set of orthogonal wavefunction ansatzs. We further demonstrate that optimization of a linear combination of energy and variance can be more efficient than optimizing energy or variance alone.

For demonstration, we use the hydrogen molecule H_2 as an example. We would consider a more complicated system, H_4, in Section 4. All Hamiltonians are calculated with *OpenFermion*,[46] and the numeral simulation for the variance VQE is conducted with the HiQ simulator framework.[47] Under STO-3G basis, we use four qubits to describe the Hamiltonian of H_2. A UCC ansatz is taken with trotter step $k = 1$ (see Appendix B), and there are 9 parameters (6 for single excitations and 3 for double excitations). The reference state is chosen as $|0011\rangle$, or others in the subspace of two electrons.

Though demonstrated with UCC ansatz for quantum chemistry problems, it should be stressed that the framework of variance-VQE can be applied for solving general quantum systems with proper chosen ansatz. The choosing of ansatz is often domain specific and can be important for efficiently applying VQE. For instance, to solve quantum lattice models one may adopt the Hamiltonian variational ansatz.[48]

3.1. Single ansatz with multiple global minimums

For the variance VQE, the criterion for being eigenstates is that energy variance $\Delta(\boldsymbol{\beta}) = 0$. If a wavefunction ansatz can represent all eigenstates, then we can expect that there are many solutions of $\boldsymbol{\beta}$ corresponding to $\Delta(\boldsymbol{\beta}) = 0$. In practice, as high energy eigenstates usually have higher complexity, an ansatz may only capture some eigenstates. In the optimization, one can select those minimums that are close to zero. This is indeed one advantage of the variance-VQE: one can verify if an eigenstate is obtained by checking whether the variance is zero.[13] A direct way to calculate excited states with the variance VQE is as follows. First, chose an ansatz $|\psi(\boldsymbol{\beta})\rangle$ that can express considered excited states. For quantum chemistry, we can use some modified UCC ansatzs with enough single-particle and double-particle excitations (see Appendix A). Second, minimizing $\Delta(\boldsymbol{\beta})$ with random choice of initial parameters $\boldsymbol{\beta}_0 \in [0, 2\pi]^K$. Third, select minimums of $\Delta(\boldsymbol{\beta})$ that are close to zero. Lastly, with optimized parameters, we can calculate their corresponding energies.

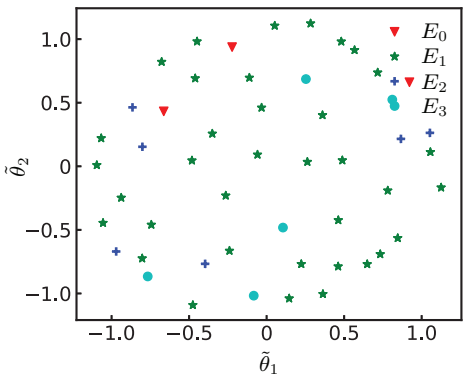


Fig. 1. Distribution of global minimums in the parameter space $\boldsymbol{\beta}$, where the high dimension $\boldsymbol{\beta}$ is visualized on a plane by dimensional reduction, using multi-dimensional scaling method. E_0, E_1, E_2, and E_3 correspond to energy levels in ascending order.

As for demonstration, we calculate the spectrum of H_2. With uniformly sampled initial $\boldsymbol{\beta}$, we optimize the energy variance with default optimizer in *Scipy*. All obtained minimums are almost zero (less than 10^{-8}), and their corresponding energies are eigen-energies successfully. However, we find the chance to be a given eigenstate varies, e.g., solutions to ground states are far less than excited states. To illustrate this, we

visualize the distribution of solutions in the parameter space, as seen in Fig. 1, by projecting the 9-dimensional parameters onto a plane, using a multidimensional scaling method that can preserve well information of distance. Note that we use a distance metric $d(\boldsymbol{\beta},\boldsymbol{\beta}') = \sum_i \cos(\theta_i - \theta_i')$ as θ is an angle such that $\theta_1 = \theta_2 \mod 2\pi$. This indicates that the ground state is relatively hard to find in this method. It is also observed that multiple points of parameters may correspond to excited states with the same energy. For non-degenerate eigenstates, corresponding to E_0, E_2, E_3 respectively, it can be explained that different $U(\boldsymbol{\beta})$ can transform $|R\rangle$ to the same target state. As excited states for E_1 are three-fold degenerate (will be revealed later), there are many solutions of $U(\boldsymbol{\beta})$ that transforms $|R\rangle$ to this subspace. We remark that the above phenomena hold for other reference states, e.g., $|0101\rangle$ and $|0110\rangle$.

3.2. A set of orthogonal ansatzs

While variance-VQE can be optimized to obtain all eigenstates for a single ansatz with different optimized parameters, it may be hard to get the ground state or other specified states as it requires choosing a proper initial parameter. Here we develop another method, using a set of orthogonal ansatz wavefunctions.

The orthogonality of ansatzs for different eigenstates can be easily enforced by using $|\psi_n(\boldsymbol{\beta})\rangle = U(\boldsymbol{\beta})|R_n\rangle$, where $\{|R_n\rangle\}$ are orthogonal to each other. Note that all ansatz uses the same $U(\boldsymbol{\beta})$ and the orthogonality of $\{|\psi_n(\boldsymbol{\beta})\rangle\}$ only holds under the same $\boldsymbol{\beta}$. To optimize $\boldsymbol{\beta}$, we use a cost function that is equal-weighting summation of all energy variances, which can be written as

$$\mathcal{C}_{\text{var}}(\boldsymbol{\beta}) = \sum_{n=0}^{k-1} w_n \Delta_n(\boldsymbol{\beta}), \tag{7}$$

where $\Delta_n(\boldsymbol{\beta})$ is the energy variance of H for the state $|\psi_n(\boldsymbol{\beta})\rangle$, and $w_n = 1/k$. The formula is similar to that of subspace-search VQE, which requires a specified weighting $\{w_n\}$ that enforces an ordering of eigenstates. We note that Eq. (7) has been used in Ref. [35] to calculate all eigenstates of quantum many-body systems with unitary matrix product state ansatz.

While other methods usually calculate excited states one by one, this method incorporates all into a single cost function to optimize. In spirit, it can be viewed as a multi-task learning.[49] As those tasks are closely related, solving them in a package may benefit each other.

Solving excited states then can be carried out by optimizing $\mathcal{C}_{\text{var}}(\boldsymbol{\beta})$ with gradient descent,

$$\boldsymbol{\beta}^t = \boldsymbol{\beta}^{t-1} - \eta \frac{\partial \mathcal{C}_{\text{var}}(\boldsymbol{\beta}^{t-1})}{\partial \boldsymbol{\beta}}. \tag{8}$$

Other gradient-free methods are also applicable. We present the numeral results for H_2. Reference states are computational basis on a subspace of two occupied electrons, which has 6 states. As the UCC ansatz is particle-conserving, the wavefunction ansatz is in the subspace of 2 particles. The energy potential curve fits perfectly to that by exact diagonalization, as seen in Fig. 2. It is interesting to note that there are only four energy levels, while the subspace is 6-dimensional. This can be explained that the first excited state is three-fold degenerate.

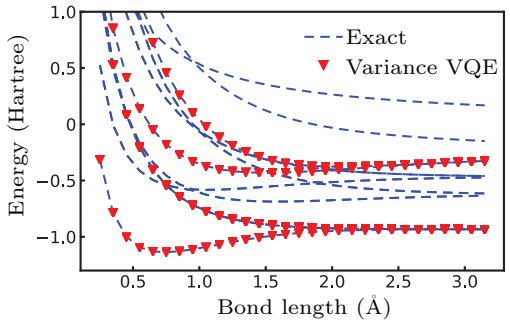


Fig. 2. The spectrum of H_2 with different bond lengths. Blue dash lines are results from exact diagonalization in the whole Hilbert space. Red markers are obtained from the variance-VQE in a subspace of two electrons that H_2 is electronic neutral.

Moreover, evaluation of the cost function Eq. (7) can be more efficient with ancillary qubits, which is given in Appendix C. This is useful when lots of eigenstates are needed, where a direct summation can be impractical. For instance, there is an exponentially large number of terms when all eigenstates are needed.

It is interesting to compare the result of the variance-VQE with orthogonal ansatz to that of subspace-search VQE.[29] The subspace-search VQE focuses on a subspace of eigenstates function and optimizes a number of k low-energy eigenstates simultaneously. The cost function is a weighted combination of all energies,

$$\mathcal{C}(\boldsymbol{\beta}) = \sum_{n=0}^{k-1} w_n \mathcal{E}_n(\boldsymbol{\beta}) = \sum_{n=0}^{k-1} w_n \text{Tr}\left[\psi_n(\boldsymbol{\beta}) H\right], \tag{9}$$

where $|\psi_n(\boldsymbol{\beta})\rangle = U(\boldsymbol{\beta})|n\rangle$. The weightings satisfy $w_0 > w_1 > w_2 > \cdots > w_{k-1}$, such that energies at optimized $\boldsymbol{\beta}^*$ will satisfy $\mathcal{E}_0(\boldsymbol{\beta}^*) \leq \mathcal{E}_1(\boldsymbol{\beta}^*) \leq \cdots \leq \mathcal{E}_{k-1}(\boldsymbol{\beta}^*)$ (see the proof in Ref. [29]). It is noted that the weightings can be learned, as in Ref. [20], which self adjusts the weightings when variational preparing Gibbs state for a quantum system.

One requirement for the subspace-search VQE is that the order of eigenstates needs to be given. It is possible that the optimization may go a long way to make all eigenstates rest in a given order, as revealed in Fig. 3. This can increase the complexity of the optimization. To illuminate this, we compare the evolution of energies in the optimization process for both subspace-search VQE and variance-VQE. The ordering of eigenstates in the subspace-search VQE is fixed by reference states $|0011\rangle, |0101\rangle, |1001\rangle, |0110\rangle, |1010\rangle, |1100\rangle$,

which give to ascending order of energies after optimization. As seen in Fig. 3, initialized with the same parameters and thus the same energies, evolutions of energies for subspace-VQE will have complicated trajectories toward convergence. In contrast, all orthogonal ansatz wavefunctions flow directly to the nearest eigenstates in the variance-VQE. We also check other ordering of eigenstates for the subspace-search VQE, and find the same result.

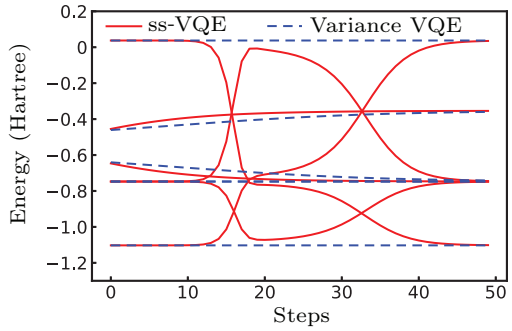


Fig. 3. Evolution of energy levels in the process of optimization. Two methods are compared: the subspace-search VQE and the variance-VQE. The bond length for H_2 is $\lambda = 1$.

3.3. Minimizing a combination of energy and variance

For the orthogonal ansatz, if we only choose a subset of the Hilbert space or subspace with a fixed number, then the variance-VQE can not promise that the solutions are lowest energy states. Just as the previous approach uses one ansatz, the final result depends on the initial parameter and the optimization process. On the other hand, for a cost function with an equal-weighting energies for orthogonal ansatz, it can get the lowest total energy, but can not promise each optimized state to be an eigenstate. While subspace-search VQE can solve this problem, the optimization can not be efficient by assigning an order for ansatzs as eigenstates. Here, we demonstrate that a simple combination of energies and energy variances can make the best of both to find low energy excited states.[33] The cost function as a combination of energies and energy variances can be written as

$$\mathcal{C}_{\rm mix}(\boldsymbol{\beta}) = \sum_{n=0}^{k-1} \mathcal{E}_n(\boldsymbol{\beta}) + \eta_v \sum_{n=0}^{k-1} \Delta_n(\boldsymbol{\beta}). \tag{10}$$

Here, all weights are the same, and η_v is a mixer factor. By minimizing $\mathcal{C}_{\rm mix}(\boldsymbol{\beta})$, the first term of Eq. (10) flows to low energy subspace while the second term makes sure that optimized states are eigenstates. A gradient descent process for optimizing $\mathcal{C}_{\rm mix}(\boldsymbol{\beta})$ is

$$\boldsymbol{\theta}^t = \boldsymbol{\theta}^{t-1} - \eta_{\rm A} \sum_n \frac{\partial \mathcal{E}_n(\boldsymbol{\theta}^{t-1})}{\partial \boldsymbol{\theta}} - \eta_{\rm A}\eta_v \sum_n \frac{\partial \Delta_n(\boldsymbol{\theta}^{t-1})}{\partial \boldsymbol{\theta}}. \tag{11}$$

In this regard, the optimization includes gradient information of both energy and energy variance, and η_v plays a role in adjusting the step size between them. In Fig. 4, we show the result of mixed VQE with different η_v, and it can be seen that nonzero η_v is critical to get the right eigenstates, and increasing η_v can raise the efficiency for optimization. Moreover, if only energy variance is used in the cost function, the solution may not be the lowest energy eigenstates. Thus, the result demonstrates that a combination of energy and energy variance can be useful and efficient for VQE to solve low-energy excited states.

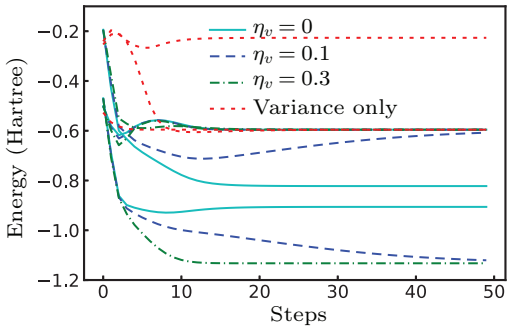


Fig. 4. Optimization processes for mixed cost function with different mixing factor η_v of energy variances. Three orthogonal ansatzs are used. The bond length for H_2 is $\lambda = 0.8$.

4. Hamiltonian sampling

So far, we have demonstrated that the variance-VQE can efficiently solve excited states for a Hamiltonian. However, the overload of calculating variance and its gradient descent is massive since it scales with the number of Hamiltonian terms as $O(N^2)$. In this section, we propose stochastic gradient descent for the variance-VQE by Hamiltonian sampling, which can significantly reduce the overload.

The Hamiltonian sampling randomly choose some components of $\boldsymbol{c}$ (set $c_i = 0$ if c_i is not chosen), which is denoted as $\tilde{\boldsymbol{c}}$. Then, the estimated energy variance is

$$\tilde{\Delta}(\boldsymbol{\beta}) = \frac{|\boldsymbol{c}|^2}{|\tilde{\boldsymbol{c}}|^2} \tilde{\boldsymbol{c}}^{\rm T} \mathcal{G}(\boldsymbol{\beta}) \tilde{\boldsymbol{c}}. \tag{12}$$

A prefactor $|\boldsymbol{c}|^2/\tilde{\boldsymbol{c}}^2$ is added to account for the fact that coefficients for different terms of a Hamiltonian vary largely, so there should be a term for normalization. Whenever $\tilde{c}_i = 0$, elements $\mathcal{G}(\boldsymbol{\beta})_{ij}$ do not need to be evaluated. For a sampling rate s (defined as a ratio that c_i is sampled), the number of elements in $\mathcal{G}(\boldsymbol{\beta})_{ij}$ required to be calculated can be reduced to a factor s^2.

With a Hamiltonian sampling, the gradient can be estimated as $\partial\tilde{\Delta}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$, which shall have a distribution due to sampling. With this gradient, a stochastic gradient descent algorithm can be applied for optimizing the variance-VQE, which updates parameters as

$$\boldsymbol{\beta}^t = \boldsymbol{\beta}^{t-1} - \eta \frac{\partial \tilde{\Delta}(\boldsymbol{\beta}^{t-1})}{\partial \boldsymbol{\beta}}. \tag{13}$$

We apply the stochastic gradient descent for solving eigenstates of H_2 and H_4 with the variance-VQE. The molecule H_4 is investigated in a trapezoidal structure. Under

sto-6g basis, we use a Hamiltonian of 6 qubits to describe H_4. The UCC ansatz is chosen as $k = 1$ with 21 parameters, including 15 for single-particle excitations and 6 for double-particle excitations (we only consider double-particle excitations from occupied orbitals to empty orbitals). Note that there are more than one hundred terms for the Hamiltonian, which means that Hamiltonian sampling is badly demanded to evaluate the variance.

Numeral simulation results are displayed in Fig. 5 for solving two eigenstates with orthogonal ansatzs. For H_2, it shows that Hamiltonian sampling with a small sampling rate $s = 0.1, 0.2$ can have comparable convergent behavior with the case without sampling. However, the variance ceases to converge at small values, due to the fluctuation of energy for a subsystem even when the whole system has zero energy variance. To solve this issue, we can turn to no sampling (sampling rate $s = 1$) in the later stage of optimization. For H_4, similar phenomena can be observed. It is noted that only one ansatz reaches the true eigenstate, while the other fail (a local minimum or the ansatz lacks the capacity to express this eigenstate), which can be seen from the nonzero energy variance. This again verifies that the variance-VQE can self-verify whether an eigenstate is obtained.

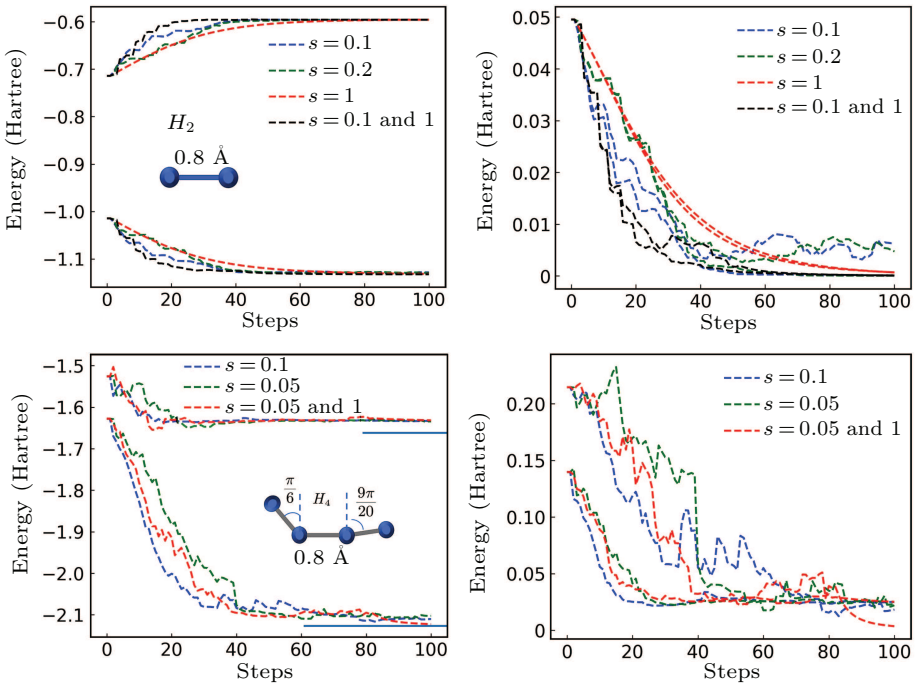


Fig. 5. Optimization of the variance-VQE for H_2 (upper row) and H_4 (down row) with different Hamiltonian sampling rates s. Sampling rates can be adjustable in an optimization process, for instance, $s = 0.1$ and 1 means that s changes from 0.1 to 1 in the later stage of optimization. The learning rate is 0.05 for all cases.

We can explain why Hamiltonian sampling fails to work around the optimized point of parameters, where gradients by Hamiltonian sampling can give meaningless direction for optimization. This is because energy variance is not additive: if $H = H_A + H_B$ and the wavefunction is $|\psi\rangle$, then the variance of H is not a summation of the variances of H_A and H_B. For example, if $|\psi\rangle$ is an eigenstate of H, $|\psi\rangle$ will not be an eigenstate of H_A or H_B, if A and B are entangled. As a result, even at the optimized parameter of zero variance, gradients by Hamiltonian sampling can be finite. Such a property is in sharp contrast to the case of machine learning. Machine learning usually uses a set of independent and identically distributed samples, and the cost function is a summation of independent contributions from all samples. Thus, gradients can be estimated with a batch of samples, giving stochastic gradient descents with well convergence behavior.[50]

From the above discussion, a practical strategy is to use Hamiltonian sampling with a small sampling rate to optimize the variance to a small value, then turn to a large sampling rate or no sampling. Also, one may then use energy-VQE. We note that stochastic gradient descent with few shots of measurements can further reduce the overload,[42,51] which leaves for further investigation.

5. Summary

In summary, we have proposed the variance-VQE that optimizes the energy variance to obtain eigenstates for a Hamiltonian. Compared with the VQE that minimizes the energy which is prone to find ground state, the variance-VQE naturally is suitable for solving and self-verifying arbitrary eigenstates. We have adopted different strategies to solve excited states with the variance-VQE, using molecules H_2 and H_4 as examples. Remarkably, it has been shown that optimizing a set of orthogonal ansatzs with the same parameterized circuit

can be very efficient for calculating a set of eigenstates. Also, we have demonstrated that minimizing a linear combination of energy and variance can outperform the case of optimizing either energy or variance alone. Moreover, we have proposed stochastic gradient descent for minimizing the variance by Hamiltonian sampling, using only a few terms of the Hamiltonian to estimate the variance and its gradient. It has been shown that the variance can decrease quickly under a small sampling rate. While numeral simulations suggested that the variance fails to converge, Hamiltonian sampling can be useful in the early stage to locate the zone of optimized parameters. Our work has demonstrated that the variance-VQE can be useful and practical for variational solving excited states in a quantum system on a quantum computer.

Appendix A: Comparison between measurements of energy and variance

For a Hamiltonian $H=\sum_{i=1}^{N}c_iL_i$, where L_i is a Pauli string, it is known that measurement of the energy with regard to a wavefunction $|\psi\rangle$ can be carried out by measuring every L_i separately and then making a linear combination of results of N Pauli measurements, namely,

$$\langle\psi|H|\psi\rangle=\sum_{i=1}^{N}c_i\langle\psi|L_i|\psi\rangle.$$

The capacity of decomposing the energy measurement to simple Pauli measurements is key for implementing variational quantum eigensolver.

The situation is similar for measuring the energy variance on the quantum computer. One can make a decomposition for the first term in Eq. (2),

$$\langle\psi|H^2|\psi\rangle=\sum_{i=1}^{N}\sum_{j=1}^{N}c_ic_j\langle\psi|L_iL_j|\psi\rangle.$$

As L_iL_j is a string of Pauli operators, measuring the energy variance requires $O(N^2)$ Pauli measurements.

Let us use a simple 2-qubit Hamiltonian $H=\sigma_1^x+\sigma_2^x+\sigma_1^y+\sigma_2^y$ for demonstration. To measure the energy, one needs to implement Pauli measurements of $\{\sigma_1^x,\sigma_2^x,\sigma_1^y,\sigma_2^y\}$. The set of Pauli measurement turns to be $\{\sigma_1^z,\sigma_2^z,\sigma_1^x\sigma_2^x,\sigma_1^x\sigma_2^y,\sigma_1^y\sigma_2^y,\sigma_1^y\sigma_2^x\}$.

Appendix B: UCC ansatz

The UCC ansatz is widely used in the field of quantum chemistry, as it can represent a parameterized wavefunction of electronic structure of a molecule efficiently. The parameterized wavefunction can be written as

$$|\psi(\boldsymbol{\theta})\rangle=\mathrm{e}^{T-T^\dagger}|R\rangle. \tag{B1}$$

Here, $T=T_1+T_2$ usually includes single-particle excitations T_1 and double-particle excitations T_2. We adopt a notation that does not distinguish occupied and empty states in the reference state $|R\rangle$ (which typically is a Hartree–Fock state). Then T_1 and T_2 can be expressed as

$$\begin{aligned}T_1&=\sum_{pq}\theta_q^pc_p^\dagger c_q,\\ T_2&=\sum_{pqrs}\theta_{rs}^{pq}c_p^\dagger c_q^\dagger c_rc_s.\end{aligned} \tag{B2}$$

To implement Eq. (B1), a trotterization of $\mathrm{e}^{T-T^\dagger}$ is required. However, $\{t_q^p\}$ and $\{t_{rs}^{pq}\}$ are parameters that should be optimized first. This makes a formula as Eq. (B1) can not be directly implemented on a quantum computer. Due to the variational flexibility, one can use a few Trotter steps (even a single Trotter step may work well). The case of a single Trotter step can be written as[22]

$$|\psi(\boldsymbol{\theta})\rangle=\prod_{pq}\mathrm{e}^{\hat{t}_{pq}}\prod_{pqrs}\mathrm{e}^{\hat{t}_{pqrs}}|R\rangle, \tag{B3}$$

where $\hat{t}_{pq}=\theta_q^p(c_p^\dagger c_q-c_q^\dagger c_p)$ and $\hat{t}_{pqrs}=\theta_{rs}^{pq}(c_p^\dagger c_q^\dagger c_rc_s-c_s^\dagger c_r^\dagger c_pc_q)$. For more Trotter steps, one can use[22]

$$|\psi(\boldsymbol{\theta})\rangle=\prod_{k=1}^{n}\prod_{pq}\mathrm{e}^{\hat{t}_{pq}^k}\prod_{pqrs}\mathrm{e}^{\hat{t}_{pqrs}^k}|R\rangle. \tag{B4}$$

Appendix C: Efficient evaluation of cost function

While the goal is to solve a number of k excited states, evaluation of the cost function (the total energy variance) is time-consuming for large k, since it should calculate energy variance one by one. For instance, for a system of M qubits, the total number of eigenstates is 2^M. Here, we discuss how to efficiently evaluate $\mathcal{C}_{\mathrm{var}}(\boldsymbol{\beta})$, similar to Ref. [35]. Let us first consider $k=2^M$. Using

$$I=\sum_{n=0}^{2^M-1}|\psi_n(\boldsymbol{\beta})\rangle\langle\psi_n(\boldsymbol{\beta})|, \tag{C1}$$

Eq. (7) can be rewritten as

$$\begin{aligned}\mathcal{C}_{\mathrm{var}}(\boldsymbol{\beta})&=\mathrm{Tr}[H^2\sum_{n=0}^{2^M-1}|\psi_n(\boldsymbol{\beta})\rangle\langle\psi_n(\boldsymbol{\beta})|]\\ &\quad-\sum_{n=0}^{2^M-1}\langle\psi_n(\boldsymbol{\beta})|H|\psi_n(\boldsymbol{\beta})\rangle^2\\ &=\mathrm{Tr}[H^2]-\sum_{n=0}^{2^M-1}\langle\psi_n(\boldsymbol{\beta})|H|\psi_n(\boldsymbol{\beta})\rangle^2.\end{aligned} \tag{C2}$$

The first term is equal to $|\boldsymbol{c}|^2$. The second still involves a summation of exponential number terms. However, we can sum up them in a quantum parallel way. By preparing a state,

$$\begin{aligned}|\phi(\boldsymbol{\beta})\rangle&=U(\boldsymbol{\beta})\otimes U(\boldsymbol{\beta})\otimes I\frac{1}{\sqrt{2^M}}\sum_{n=0}^{2^M-1}|n\rangle|n\rangle|n\rangle\\ &=\sum_{n=0}^{2^M-1}|\psi_n(\boldsymbol{\beta})\rangle|\psi_n(\boldsymbol{\beta})\rangle|n\rangle,\end{aligned} \tag{C3}$$

then it can be verified that

$$\langle\phi(\beta)|H\otimes H\otimes I|\phi(\beta)\rangle=\frac{1}{2^M}\sum_{n=0}^{2^M-1}\langle\psi_n(\beta)|H|\psi_n(\beta)\rangle^2.$$

This indicates that the cost function can be efficiently calculated, at the price of increasing the number of qubits from M to $3M$, which makes a good deal since it can sum up a number of 2^M terms.

For general $k<2^W$, we can adopt a similar idea. We introduce W (requiring $2^W\geq k>2^{W+1}$) ancilla qubits to prepare

$$|\varphi(\beta)\rangle=U(\beta)\otimes I\frac{1}{\sqrt{k}}\sum_{n=0}^{k-1}|n\rangle|n\rangle=\sum_{n=0}^{k-1}|\psi_n(\beta)\rangle|n\rangle,$$

then the first part of Eq. (7) is evaluated as

$$k\langle\varphi(\beta)|H^2|\varphi(\beta)\rangle. \tag{C4}$$

To evaluate the second part of Eq. (7), the summation in Eq. (C4) should be replaced by $k-1$. The total number of qubits is $M+M+W=2M+W$.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 12005065) and the Guangdong Basic and Applied Basic Research Fund (Grant No. 2021A1515010317).

References

[1] Yung M H, Casanova J, Mezzacapo A, McClean J, Lamata L, Aspuru-Guzik A and Solano E 2014 *Sci. Rep.* **4** 3589
[2] Farhi E, Goldstone J and Gutmann S 2014 arXiv:1411.4028
[3] McClean J R, Romero J, Babbush R and Aspuru-Guzik A 2016 *New J. Phys.* **18** 023023
[4] Li Y and Benjamin S C 2017 *Phys. Rev. X* **7** 021050
[5] McClean J R, Kimchi-Schwartz M E, Carter J and de Jong W A 2017 *Phys. Rev. A* **95** 042308
[6] Shen Y, Zhang X, Zhang S, Zhang J N, Yung M H and Kim K 2017 *Phys. Rev. A* **95** 020501
[7] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 *Nature* **549** 242
[8] Hempel C, Maier C, Romero J, McClean J, Monz T, Shen H, Jurcevic P, Lanyon B P, Love P, Babbush R, Aspuru- Guzik A, Blatt R and Roos C F 2018 *Phys. Rev. X* **8** 031022
[9] Anschuetz E R, Olson J P, Aspuru-Guzik A and Cao Y 2018 arXiv:1808.08927
[10] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 *Phys. Rev. A* **98** 032309
[11] Moll N, Barkoutsos P, Bishop L S, Chow J M, Cross A, Egger D J, Filipp S, Fuhrer A, Gambetta J M, Ganzhorn M, Kandala A, Mezzacapo A, Müller P, Riess W, Salis G, Smolin J, Tavernelli I and Temme K 2018 *Quantum Sci. Technol.* **3** 030503
[12] Endo S, Sun J, Li Y, Benjamin S C and Yuan X 2020 *Phys. Rev. Lett.* **125** 010501
[13] Kokail C, Maier C, van Bijnen R, Brydges T, Joshi M K, Jurcevic P, Muschik C A, Silvi P, Blatt R, Roos C F and Zoller P 2019 *Nature* **569** 355
[14] Takeshita T, Rubin N C, Jiang Z, Lee E, Babbush R and McClean J R 2020 *Phys. Rev. X* **10** 011004
[15] McArdle S, Jones T, Endo S, Li Y, Benjamin S C and Yuan X 2019 *npj Quantum Inf.* **5** 75
[16] Higgott O, Wang D and Brierley S 2019 *Quantum* **3** 156
[17] Wu J and Hsieh T H 2019 *Phys. Rev. Lett.* **123** 220502
[18] Verdon G, Broughton M, McClean J R, Sung K J, Babbush R, Jiang Z, Neven H and Mohseni M 2019 arXiv:1907.05415
[19] Liu J G, Zhang Y H, Wan Y and Wang L 2019 *Phys. Rev. Res.* **1** 023025
[20] Liu J G, Mao L, Zhang P and Wang L 2021 *Mach. Learn.: Sci. Technol.* **2** 025011
[21] Benedetti M, Lloyd E, Sack S and Fiorentini M 2019 *Quantum Sci. Technol.* **4** 043001
[22] Grimsley H R, Economou S E, Barnes E and Mayhall N J 2019 *Nat. Commun.* **10** 3007
[23] Zhang D B and Yin T 2020 *Phys. Rev. A* **101** 032311
[24] Yuan Z H, Yin T and Zhang D B 2021 *Phys. Rev. A* **103** 012413
[25] Arute F 2020 rXiv:2004.04174
[26] Xu X, Sun J, Endo S, Li Y, Benjamin S C and Yuan X 2021 *Sci. Bull.* **66** 2181
[27] Lubasch M, Joo J, Moinier P, Kiffner M and Jaksch D 2020 *Phys. Rev. A* **101** 010301
[28] Greene-Diniz G and Ramo D M 2019 rXiv:1910.05168
[29] Nakanishi K M, Mitarai K and Fujii K 2019 *Phys. Rev. Res.* **1** 033062
[30] Bartlett J H, Gibbons J J and Dunn C G 1935 *Phys. Rev.* **47** 679
[31] Siringo F and Marotta L 2005 *Eur. Phys. J. C* **44** 293
[32] Umrigar C J, Wilson K G and Wilkins J W 1988 *Phys. Rev. Lett.* **60** 1719
[33] Umrigar C J and Filippi C 2005 *Phys. Rev. Lett.* **94** 150201
[34] Khemani V, Pollmann F and Sondhi S L 2016 *Phys. Rev. Lett.* **116** 247204
[35] Pollmann F, Khemani V, Cirac J I and Sondhi S L 2016 *Phys. Rev. B* **94** 041116
[36] Vicentini F, Biella A, Regnault N and Ciuti C 2019 *Phys. Rev. Lett.* **122** 250503
[37] Bairey E, Arad I and Lindner N H 2019 *Phys. Rev. Lett.* **122** 020504
[38] Chertkov E and Clark B K 2018 *Phys. Rev. X* **8** 031029
[39] Qi X L and Ranard D 2019 *Quantum* **3** 159
[40] Alba V, Fagotti M and Calabrese P 2009 *J. Stat. Mech.: Theory and Experiment* **2009** P10020
[41] Schuld M, Bergholm V, Gogolin C, Izaac J and Killoran N 2019 *Phys. Rev. A* **99** 032331
[42] Sweke R, Wilde F, Meyer J, Schuld M, Fährmann P K, Meynard-Piganeau B and Eisert J 2019 arXiv:1910.01155v1
[43] Huggins W J, McClean J, Rubin N, Jiang Z, Wiebe N, Whaley K B and Babbush R 2019 arXiv:1907.13117
[44] Cotler J and Wilczek F 2020 *Phys. Rev. Lett.* **124** 100401
[45] Seki K and Yunoki S 2021 *PRX Quantum* **2** 010333
[46] McClean J R *et al.* 2020 *Quantum Sci. Technol.* **5** 034014
[47] Huawei HiQ team *A High-performance Quantum Computing Simulator and Programming Framework*
[48] Wiersema R, Zhou C, de Sereville Y, Carrasquilla J F, Kim Y B and Yuen H 2020 *PRX Quantum* **1** 020319
[49] Evgeniou T and Pontil M 2004 *Regularized multi-task learning Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD'04* (New York, USA: Association for Computing Machinery) pp. 109–117
[50] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (MIT Press)
[51] Kübler J M, Arrasmith A, Cincio L and Coles P J 2020 *Quantum* **4** 263