Supplementary Information for: Iterative qubit-excitation based variational quantum eigensolver

I. ENERGY GRADIENT OF AN EXCITATION EVOLUTION

The expression for the single parameter energy gradient in equation (23) of the main text, can be derived as follows:

$$\frac{\partial}{\partial\theta_p}E^{(m-1)}(\theta_p) = \frac{\partial}{\partial\theta_p}\langle\psi^{(m-1)}|\tilde{A_p}^\dagger(\theta_p)H\tilde{A_p}(\theta_p)|\psi^{(m-1)}\rangle = \frac{\partial}{\partial\theta_p}\langle\psi^{(m-1)}|e^{\theta_p\tilde{T_p}^\dagger}He^{\theta_p\tilde{T_p}}|\psi^{(m-1)}\rangle =$$

$$\langle\psi^{(m-1)}|e^{\theta_p\tilde{T_p}^\dagger}\tilde{T_p}^\dagger He^{\theta_p\tilde{T_p}}|\psi^{(m-1)}\rangle + \langle\psi^{(m-1)}|e^{\theta_p\tilde{T_p}^\dagger}H\tilde{T_p}e^{\theta_p\tilde{T_p}}|\psi^{(m-1)}\rangle = \tag{1}$$

$$\langle\psi^{(m-1)}|e^{\theta_p\tilde{T_p}^\dagger}[H,\tilde{T_p}]e^{\theta_p\tilde{T_p}}|\psi^{(m-1)}\rangle \tag{2}$$

In going from (1) to (2) we use that $\tilde{T_p}$ is skew-Hermitian, so that $\tilde{T_p}^\dagger = -\tilde{T_p}$. In the limit $\theta_p = 0$ expression (2) becomes equivalent to the expression in equation (23) of the main text.

II. STATEVECTOR CALCULATOR

Here we outline the method used to calculate the trial statevectors $|\psi(\boldsymbol{\theta})\rangle$ in the classical numerical simulations used for the results in this paper. Calculating the trial statevectors, is the most time consuming part of the numerical simulations, and optimizing it is vital.

In this work, we are concerned with states of the form

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_0\rangle = \prod_{i=N_U}^{1} e^{\theta_i S_i}|\psi_0\rangle, \tag{3}$$

where N_U is the size (the number of ansatz elements) of the ansatz $U(\boldsymbol{\theta})$, $|\psi_0\rangle$ is the initial reference state, and S_i is a skew-Hermitian operator, which in this paper corresponds to either a qubit excitation operator, a fermionic excitation operator, or a string of Pauli operators. Each of these three types of skew-Hermitian operators, also satisfies the relation

$$S_i^3 = -S_i. \tag{4}$$

To calculate the $2^{N_{\text{MO}}}$-dimensional state-vector representing $|\psi(\boldsymbol{\theta})\rangle$, classically, we need to calculate the N_U exponents $\{e^{\theta_i S_i}\}$ in equation (3), and then multiply them sequentially to $|\psi_0\rangle$. Each S_i is represented by an $2^{N_{\text{MO}}}\times 2^{N_{\text{MO}}}$-dimensional matrix. Hence, the complexity of estimating each exponent $e^{\theta_i S_i}$, directly, is $O(2^{3N_{\text{MO}}})$. However, we can make use of relation (4) and write each exponent in equation (3) as

$$e^{\theta_i S_i} = \sum_{r=0}^{\infty}\frac{\theta_i^r S_i^r}{r!} = I + \sum_{r_1=0}^{\infty}\frac{(-1)^{r_1}\theta_i^{2r_1+1}}{(2r_1+1)!}S_i + \sum_{r_2=1}^{\infty}\frac{(-1)^{r_2}\theta_i^{2r_2}}{(2r_2)!}S_i^2 = I + \sin\theta S_i + (1-\cos\theta)S_i^2. \tag{5}$$

The operators $\{S_i\}$ are fixed throughout a simulation. Therefore, if we compute in advance and store the matrix representations of each S_i and S_i^2, we can evaluate the expression in equation (5) by performing matrix addition only, which has a complexity of $O(2^{2N_{\text{MO}}})$. Hence, the calculation of $|\psi(\boldsymbol{\theta})\rangle$, requires N_U matrix-to-vector multiplications and N_U matrix additions, which have a total complexity of $O(N_U 2^{2N_{\text{MO}}})$.

The drawback of the method outlined above is that we need to store the matrices for all S_i and S_i^2 operators. For example, the most memory demanding simulation in this work, running the qubit-ADAPT-VQE for BeH_2, required around 2GB of RAM to store the matrices for all Pauli string operators, which define the ansatz element pool of the qubit-ADAPT-VQE, and their respective squares. However, in this case, a speed-up of nearly a factor of 20 was achieved, in comparison to calculating $|\psi(\boldsymbol{\theta})\rangle$ with the general IBM's Qiskit statevector simulator.

III. ENERGY VECTOR GRADIENT

When using a gradient-descent minimizer, e.g. the BFGS, we have the option to supply a function that returns the gradient vector of the minimized function. If we are close to the global minimum, supplying a gradient vector function guarantees a faster optimization of the variational parameters. In the case of minimizing the Hamiltonian expectation value $E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle$, the ith component of the energy gradient vector, $\boldsymbol{\nabla}E(\boldsymbol{\theta})$, is given by

$$\boldsymbol{\nabla}_i E(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_i}\langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle = \frac{\partial}{\partial\theta_i}\langle\psi_0|U^\dagger(\boldsymbol{\theta})HU(\boldsymbol{\theta})|\psi_0\rangle = \frac{\partial}{\partial\theta_i}\langle\psi_0|\prod_{1}^{k_1=N_U} e^{\theta_{k_1}S_{k_1}^\dagger} H \prod_{k_2=N_U}^{1} e^{\theta_{k_2}S_{k_2}}|\psi_0\rangle = \tag{6}$$

$$\langle\psi(\boldsymbol{\theta})|H\prod_{k_1=N_u}^{i+1} e^{\theta_{k_1}S_{k_1}} S_i \prod_{k_1=i}^{1} e^{\theta_{k_2}S_{k_2}}|\psi_0\rangle + \langle\psi_0|\prod_{k_1=1}^{i} e^{\theta_{k_1}S_{k_1}^\dagger} S_i^\dagger \prod_{k_1=i+1}^{N_U} e^{\theta_{k_2}S_{k_2}^\dagger}|H|\psi(\boldsymbol{\theta})\rangle = 2\langle\alpha_i(\boldsymbol{\theta})|S_i|\beta_i(\boldsymbol{\theta})\rangle, \tag{7}$$

where

$$|\beta_i(\boldsymbol{\theta})\rangle = \prod_{k=i}^{1} e^{\theta_k S_k}|\psi_0\rangle \quad \text{and} \tag{8}$$

$$|\alpha_i(\boldsymbol{\theta})\rangle = \prod_{k=i+1}^{N_u} e^{\theta_k S_k^\dagger} H|\psi(\boldsymbol{\theta})\rangle, \tag{9}$$

and N_U is the size (the number of ansatz elements) of the ansatz $U(\boldsymbol{\theta})$.

For the numerical simulations presented in this paper, the N_U components of $\boldsymbol{\nabla}E(\boldsymbol{\theta})$ can be calculated with minimum number of matrix multiplications by updating $|\beta_i(\boldsymbol{\theta})\rangle$ and $|\alpha_i(\boldsymbol{\theta})\rangle$ in the following way:

1. For $i = N_U$, initiate

$$|\alpha_{N_U}(\boldsymbol{\theta})\rangle = H|\psi(\boldsymbol{\theta})\rangle \quad \text{and} \tag{10}$$

$$|\beta_{N_U}(\boldsymbol{\theta})\rangle = |\psi(\boldsymbol{\theta})\rangle \tag{11}$$

2. For $1 < i < N_U$, update

$$|\alpha_{i-1}(\boldsymbol{\theta})\rangle = e^{\theta_i S_i^\dagger}|\alpha_i(\boldsymbol{\theta})\rangle \quad \text{and} \tag{12}$$

$$|\beta_{i-1}(\boldsymbol{\theta})\rangle = (e^{\theta_i S_i})^{-1}|\beta_i(\boldsymbol{\theta})\rangle = e^{\theta_i S_i^\dagger}|\beta_i(\boldsymbol{\theta})\rangle, \tag{13}$$

where in equation (13), we use that S_i is skew-Hermitian, so $S_i^\dagger = -S_i$.

Assuming that we have already computed and stored the matrices of the exponentials $\{e^{\theta_i S_i}\}$, when calculating $|\psi(\boldsymbol{\theta})\rangle$, to calculate each component of $\boldsymbol{\nabla}E(\boldsymbol{\theta})$ we need to perform 3 matrix-to-vector multiplications. Thus, overall to calculate $\boldsymbol{\nabla}E(\boldsymbol{\theta})$ we need to perform $3N_U$ matrix-to-vector multiplications, resulting in a total cost of $O(3N_U 2^{2N_{\text{MO}}})$ operations.

The cost of calculating $\boldsymbol{\nabla}E(\boldsymbol{\theta})$ is about 3 times the cost of calculating $|\psi(\boldsymbol{\theta})\rangle$. However, we find that using the energy gradient vector in the optimization subroutine of the VQE, reduces the number of VQE iterations by at least an order of magnitude, which justifies the use of the gradient vector.