NISQ algorithm for the matrix elements of a generic observable

Rebecca Erbanni,[1] Kishor Bharti,[2] Leong-Chuan Kwek,[3, 4, 5] and Dario Poletti[1, 6, 7, 5]

[1]*Science, Mathematics and Technology Cluster, Singapore University*
of Technology and Design, 8 Somapah Road, 487372 Singapore
[2]*Joint Center for Quantum Information and Computer Science and Joint Quantum*
Institute, NIST/University of Maryland, College Park, Maryland 20742, USA
[3]*Centre for Quantum Technologies, National University of Singapore 117543, Singapore*
[4]*National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616*
[5]*MajuLab, CNRS-UNS-NUS-NTU International Joint Research Unit, UMI 3654, Singapore*
[6]*Engineering Product Development Pillar, Singapore University of Technology and Design, 8 Somapah Road, 487372 Singapore*
[7]*The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy*

The calculation of off-diagonal matrix elements has various applications in fields such as nuclear physics and quantum chemistry. In this paper, we present a noisy intermediate scale quantum algorithm for estimating the diagonal and off-diagonal matrix elements of a generic observable in the energy eigenbasis of a given Hamiltonian. Several numerical simulations indicate that this approach can find many of the matrix elements even when the trial functions are randomly initialized across a wide range of parameter values without, at the same time, the need to prepare the energy eigenstates.

I. Introduction

The landscape of quantum computing technology has shifted dramatically over the previous four decades. Once considered a theoretical endeavour, quantum computing is today a vibrant experimental pursuit. With the advent of noisy intermediate scale quantum (NISQ) devices [1, 2], considerable effort has been directed into developing algorithms that can run in the presence of noise on quantum computing systems with 50–70 qubits and restricted qubit connectivity. These algorithms are known as NISQ algorithms, and common examples include the variational quantum eigensolver (VQE) [3–5] and the quantum approximate optimization algorithm (QAOA) [6, 7].

There has been an explosion of recent work on NISQ algorithms to address problems such as finding ground state of Hamiltonians [3–5, 8–19] quantum simulation [20–33], combinatorial optimization [6, 7], quantum metrology [34–36], and machine learning [37–41]. Notwithstanding these research efforts, the practical application of NISQ devices is still a long way off. To attain practical quantum advantage in the NISQ era, it is crucial to examine previously untapped or poorly understood applications. One such application area that may need to receive more attention is calculating off-diagonal matrix elements of generic observables.

Many problems in nuclear physics and other sciences make extensive use of off-diagonal matrix elements. Consider the Brueckner-Bethe-Goldstone equation solution; the nuclear potential is computed from the off-diagonal matrix elements of the Brueckner reaciton matrix [42]. The generalized eigenvalue problems for obtaining energy levels of time-dependent Euclidean correlators [43] in lattice quantum chromodynamics (QCD) rely heavily on information acquired from off-diagonal elements of the matrix of correlators. The rotational-vibrational coupling in quantum chemistry considers the off-diagonal block in the matrix of kinematic coefficients [44]. Off-diagonal element calculations have also been used to determine the diamagnetic susceptibility and form factor of atoms such as Helium [45]. Given the importance of finding the off-diagonal matrix elements of generic observables, it is pertinent to explore the potential of NISQ devices for the aforementioned task. Despite the hope that noisy intermediate scale quantum devices will enable the solution of problems in a variety of fields (including quantum chemistry and nuclear physics), the subject of noisy intermediate scale quantum algorithms for off-diagonal matrix element calculation remains poorly understood. One reason for this is that an observable's off-diagonal matrix elements can be a complex number, yet nearly all noisy intermediate scale quantum algorithms are designed to compute real values.

In this paper, we provide a NISQ algorithm for hybrid classical-quantum computation of the matrix elements (both diagonal and off-diagonal) of a given observable W in the energy eigenbasis. Indeed, variational principles have been applied successfully in several NISQ algorithms. The Rayleigh-Ritz variational method is used in variational eigensolvers to seek the ground state of a Hamiltonian of a system[3]. The McLachlan principle proves useful for dynamical solutions of open quantum systems[46]. By deriving its motivations from Ref. [47–49], here we propose a variational principle to find off-diagonal (and diagonal) elements of observables or operators. Our algorithm uses Lagrange multipliers to encode the constraints for the underlying problem into the refined objective. For our model problems, we discuss approaches for both exact and iterative evaluation of Lagrange multipliers. Various numerical simulations suggest that our approach manages to find many of the matrix elements even when one initializes randomly the trial functions over a very broad range of parameters. Given the importance of finding off-diagonal matrix elements for a variety of fundamental problems, we believe that

our approaches can be applied to a variety of interesting practical contexts.

We would like to point out that a quantum variational approach for calculating matrix elements was proposed recently in Ref. [50], which relied on the preparation of the energy eigenstates corresponding to which the overlap for the given observable is to be calculated. Unlike this work, our approach does not involve the preparation of the energy eigenstates, hence avoiding the accompanying quantum resource requirements. While our choice of ansatz may seem to resembles existing works in the literature [8, 12, 18], none of these results work for the off-diagonal matrix elements of a generic observable. Moreover, our approach is fundamentally different and uses Lagrange multipliers to encode the constraints for the underlying problem into the refined objective.

II. The classical variational algorithms for off-diagonal elements

Given a system with Hamiltonian H with different eigenenergies E_i and corresponding eigenfunctions $|\phi_i\rangle$ and an observable W, what we aim to find are the elements $F_{i,j} = \langle \phi_i | W | \phi_j \rangle$.

One natural approach would be to find the different eigenfunctions of the Hamiltonian and then evaluate the elements, including the off-diagonal ones. However one may not need to do this. In [47, 48] the authors showed a variational approach to find such elements which we summarize in the following. First we can use a variational ansatz for the eigenfunctions $|\phi_i\rangle \approx |\phi_{i,t}(\vec{\eta}_i)\rangle$ parametrized by the $\vec{\eta}_i$. We can then write a variational function $F_{i,j}^v$ which has zero derivative respect to the $\vec{\eta}_i$ when $F_{i,j}^v = F_{i,j}$. Such approach can readily give both diagonal and off-diagonal elements. In the following we consider a normalized parametrization of the trial eigenfunctions $|\phi_{i,t}(\vec{\eta}_i)\rangle$, and an extension to the case of non normalized trial eigenfunctions is discussed in the App.C.

For an observable W which is only real or only imaginary, the variational function $F_{i,j}^v$ is given by

$$
\begin{aligned}
F_v = & \langle \phi_{i,t} | W | \phi_{j,t} \rangle \\
& + \langle L_{i,a} | (H - E_i) | \phi_{i,t} \rangle + \langle \phi_{i,t} | (H - E_i)^\dagger | L_{i,b} \rangle \\
& + \langle L_{j,a} | (H - E_j) | \phi_{j,t} \rangle + \langle \phi_{j,t} | (H - E_j)^\dagger | L_{j,b} \rangle \\
& + \lambda \left[\langle \phi_{i,t} | W | \phi_{j,t} \rangle \mp \langle \phi_{j,t} | W | \phi_{i,t} \rangle \right],
\end{aligned}
\tag{1}
$$

where in the last line one uses the sign $-$ or $+$ depending on whether W is real or imaginary respectively. We note that for a given matrix W we can always write $W_R = (W + W^T)/2$ and $W_I = (W - W^T)/2$, wher A^T is the transposition of A, respectively for the real components and for the imaginary ones. In Eq.(1) we have used the Lagrange multipliers vectors $|L_{i,\nu}\rangle$ (which, generally, are not normalized) and the scalar λ. It is thus clear that the terms on the right-hand side of Eq.(1) are, apart from the first one, zero for the exact solution, which then results in

obtaining the correct evaluation of the off-diagonal term $F_{i,j}$.

In order to get the expressions for the Lagrange multipliers L_i and λ, one can expand the first order variation of the functional and set the coefficients of $\langle \delta \phi_1 |$, $|\delta \phi_1 \rangle, \langle \delta \phi_2 |, |\delta \phi_2 \rangle$ to zero. Details of such computations are in App.A

We now take a small variation $|\delta \phi_i\rangle$ to the exact eigenfunctions $|\phi_i^{ex}\rangle$, which gives $|\phi_{i,t}\rangle = |\phi_i^{ex}\rangle + |\delta \phi_i\rangle$, and set that the first order corrections to the exact result of the function $F_{i,j}^v$ are zero. This gives expressions for the Lagrange multipliers which are

$$\lambda = -1/2 \tag{2}$$

$$(H - E_i)|L_{i,\nu}\rangle = -\xi_{R,I}^\nu W_{R/I}|\phi_j\rangle/2. \tag{3}$$

with $\xi_{R,I}^\nu = \pm 1$. More precisely, for the real case $\xi_R^\nu = 1$ whether $\nu = a, b$ while for the imaginary case $\xi_R^a = 1$ and $\xi_R^b = -1$. What is important to state, though, is that in principle we do not know the value E_i and thus this will be approximate by the expectation value of the Hamiltonian for that wave function, i.e. $E_i \approx \langle \phi_i | H | \phi_i \rangle$.

At this point we should solve Eq.(3) for $|L_{i,\nu}\rangle$, but this is not straightforward because $H - \langle \phi_i | H | \phi_i \rangle$ is not invertible. For the purpose of computing these Lagrange multipliers we thus use a modified Hamiltonian

$$H_{mod,i} = H - \frac{H|\phi_i\rangle \langle \phi_i | H}{\langle \phi_i | H | \phi_i \rangle} \tag{4}$$

such that the matrix $H_{mod,i} - \langle \phi_i | H | \phi_i \rangle$ is not singular.

When it is difficult to evaluate the inverse of $H_{mod,i} - \langle \phi_i | H | \phi_i \rangle$ exactly, it is also possible to implement an iterative approach. As shown in [48], we can find $|L_{i,\nu}\rangle$ by minimizing

$$
\begin{aligned}
M(|L_{i,\nu}\rangle) = & \langle L_{i,\nu} | (H_{mod,i} - \langle \phi_i | H | \phi_i \rangle) | L_{i,\nu} \rangle \\
& + \langle \phi_j | W_R | L_{i,\nu} \rangle
\end{aligned}
\tag{5}
$$

for the real case, and

$$M(|L_{i,\nu}\rangle) = \langle L_{i,\nu} | (H_{mod,i} - \langle \phi_i | H | \phi_i \rangle) | L_{i,\nu} \rangle \tag{6}$$

for the imaginary one. More details can be found in App.B.

A. Scaling analysis for the overlap calculation

We consider Hamiltonians H and observables W which can be written as linear combination of poly(n) unitaries where n is the number of qubits over which the Hamiltonian is defined. A typical example is a local spin Hamiltonian which can be written as a sum of polynomially many n-qubit Pauli matrices. Furthermore, we consider a number of ansatz states and Lagrange multipliers that scales polynomially with the system size (number of qubits) as this is often sufficient to obtain accurate results, e.g. using a Krylov basis [18].

We thus highlight that there are three types of overlaps that need to be computed for the successful implementation of our NISQ algorithm,

- $\langle\phi_i|H|\phi_i\rangle$: Since the number of terms in the Hamiltonian and the number of ansatz states $|\phi_i\rangle$ are polynomially many in number of qubits, the overlaps $\langle\phi_i|H|\phi_i\rangle$ can be calculated efficiently [51].

- $\langle L_{i,\nu}|H_{mod,i}|L_{i,\nu}\rangle$: the estimation of these overlaps requires the calculation of the terms of the form $\langle\phi_i|H|\phi_i\rangle$, $\langle L_{i,\nu}|H|L_{i,\nu}\rangle$ and $\langle L_{i,\nu}|H|\phi_i\rangle$. Since the number of ansatz states and Lagrange multipliers scale polynomially with the system size, also the aforementioned overlaps can be evaluated efficiently.

- $\langle\phi_j|W_R|L_{i,\nu}\rangle$: Since the operator W can be expressed as a linear combination of polynomially many unitaries, it is easy to see that also the overlaps $\langle\phi_j|W_R|L_{i,\nu}\rangle$ can be calculated efficiently.

III. Results

We will now show examples which elucidate the effectiveness of an hybrid classical-quantum implementation of this variational approach. We will consider both a single and a two-qubit Hamiltonian and we will use both the exact and the iterative approaches to get the Lagrange multiplier.

A. Models

We consider two scenarios. The first scenario is a two level system with Hamiltonian

$$H_1 = X \tag{7}$$

and $W_1 = H_d W_1^D H_d$ where

$$W_1^D = \begin{pmatrix} 5 & 2-2j \\ 2+2j & 3 \end{pmatrix}. \tag{8}$$

W_1^D is thus the matrix in the energy eigenbasis, as H_d diagonalizes the Hamiltonian H_1.

In the second case we consider the following $2-$ qubit Hamiltonian

$$H_2 = 2X \otimes \mathbb{I} + \mathbb{I} \otimes X + 2Z \otimes X \tag{9}$$

while we take an W_2^D (therefore in the energy eigenbasis) as the following Hermitian complex matrix

$$W_2^D = \begin{pmatrix} 1 & 3+1j & 5-3j & 13+8j \\ 3-1j & 4 & 20+5j & 25+10j \\ 5+3j & 20-5j & 7 & 6-15j \\ 13-8j & 25-10j & 6+15j & 10 \end{pmatrix}. \tag{10}$$

The matrix W_2 in the computational basis, which we use in the computations, can be obtained from W_2^D and the eigenvalues of H_2.

B. Implementation for hybrid classical-quantum computation

We test the usefulness of the variational principle from Eq.(1) on a hybrid classical-quantum algorithm. Since the solutions can be found where the derivatives are zero, we use a classical optimization algorithm which performs a gradient descent after having evaluated the gradients of $F_{i,j}^v$ over the parameters $\vec{\eta}_i$. The quantum part of the algorithm helps with the evaluation of the gradients. In practice, one can evaluate all the relevant overlaps once, and then use such knowledge to evaluate the gradients for any given value of the $\vec{\eta}_i$. To evaluate the overlaps we write the states $|\phi_{i,t}\rangle$ as

$$|\phi_{i,t}(\theta)\rangle = \cos(\theta_i)|0\rangle + \sin(\theta_i)|1\rangle. \tag{11}$$

for the one-qubit case, and for the two-qubit case we consider the parametrization

$$|\phi_{i,t}(\alpha_i,\beta_i,\gamma_i)\rangle = \cos(\alpha_i)|00\rangle + \sin(\alpha_i)\cos(\beta_i)|01\rangle$$
$$+ \sin(\alpha)\sin(\beta_i)\cos(\gamma_i)|10\rangle + \sin(\alpha_i)\sin(\beta_i)\sin(\gamma_i)|11\rangle. \tag{12}$$

Note that here we only consider real trial functions, which is sufficient for our examples, and a generalization to complex ones is straightforward. Furthermore, we currently use a variational representation of the eigenfunctions which scales linearly with the size of the Hilbert space. In practice, for systems with large Hilbert space, for which a quantum computer would come in handy, one would have to resort to a much smaller parameter space, for example using a limited Krylov basis [52, 53]. For the Lagrange multipliers $|L_{i,\nu}\rangle$ we use an unnormalized ansatz of the form

$$|L_{i,\nu}\rangle = c_{i,1}|0\rangle + d_{i,1}|1\rangle \tag{13}$$

for the one qubit case and

$$|L_{i,\nu}\rangle = c_{i,2}|00\rangle + d_{i,2}|01\rangle + e_{i,3}|10\rangle + f_{i,4}|11\rangle \tag{14}$$

for the qubit case, where the parameters of Eq.(13,14) are real numbers.

All quantum computations are implemented on the IBM Belem QPU simulator using error mitigation. The overlaps are evaluated each by averaging 50 estimates of the overlaps each done with 1000 shots.

C. Single qubit using an exact evaluation of the Lagrange multipliers

We now consider the model with a single qubit, with Hamiltonian H_1 from Eq.(7) and the matrix W_1 derived from the matrix in the energy eigenbasis Eq.(8). We use a classical optimization algorithm by deriving analytically the derivatives of Eq. (1) with respect to the angles of Eqs. (11) and the parameters of Eq. (13); the expectation
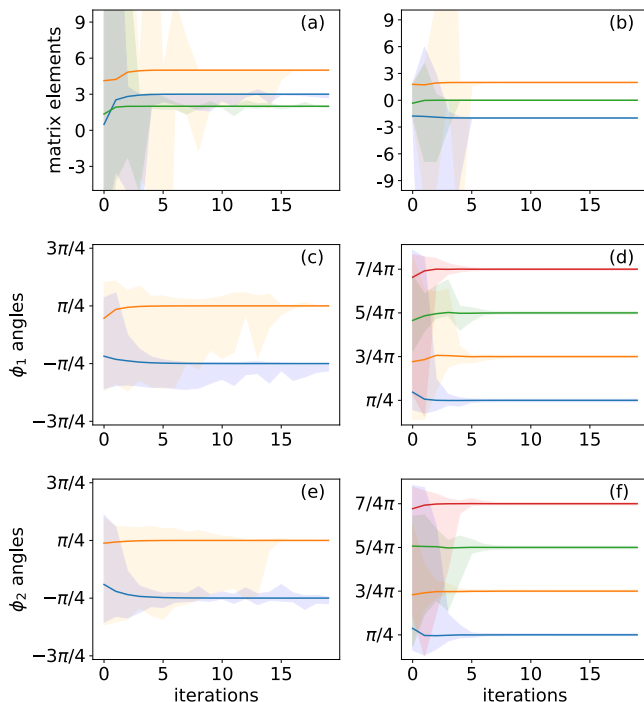
FIG. 1. Plots for the real and imaginary parts of the matrix elements versus iterations of the classical optimization algorithm, through exact calculation of $|L_{1,\nu}\rangle$ and $|L_{2,\nu}\rangle$ in the 1-qubit case and with error mitigation. Panels (a,c,e) are the results For the real part of W_1 in the energy eigenbasis, while (b,d,f) for the imaginary part. Panels (a,b) show the matrix elements while panels (c) to (f) show the angles α_i of the trial eigenfunctions. Each panel shows the results from a total of 150 runs, where the angles have been randomly initialized between 0 and 2π. In all panels, the lines represent the medians of the runs converging to a specific value and the error intervals include 92% of the corresponding runs.

values and gradients are evaluated using overlaps computed on the IBM Belem quantum processor's simulator. In Fig. 1 we show the estimated value of the element of $W_{R/I}$ versus number of iterations of the classical optimization procedure, panels (a) and (b), and the values of the angles α_i which parametrize the trial eigenfunctions, panels (c) to (f). In the left panels, (a, c, e), we consider the real part of the observable W, i.e. W_R, while in the right panels (b, d, f), the imaginary part, W_I. In each of the two cases we show the results from 150 runs of the protocol with initial angles for the two eigenfunctions α_1 and α_2 chosen uniformily between $-\pi$ and π. The line in each of the panels is obtained by the median value of the expectation value or angle between 50 runs which end in the same vicinity. The colored background reflects the value taken by the middle 92% of the corresponding realizations. In panels (a) and (b) we observe that while in the initial steps the angles cover the all range from $-\pi$ to π, and the expectation values take a very large range of possible values, within 20 iterations the prediction of the matrix elements is very accurate, both for the real,

panel (a), and imaginary part, panel (b). We note that the accuracy in the angles may not be as good as that on the matrix elements. This is one advantage of using this variational approach which is tailored to give the matrix elements directly.

As a technical, but important, detail, for W_R we had to fix the global phase of the trial eigenfunctions, otherwise the expectation value may show the wrong sign. For this reason when plotting the angles and evaluating the corresponding matrix elements, we mapped them between the angles $-\pi/2$ and $\pi/2$ so that the cosine of the angle would be positive. Such procedure is not needed for the imaginary elements of W, and for this reason the plotted range of angles in panels (d) and (f) is between 0 and 2π.

IV. Analysis of errors for the single qubit case
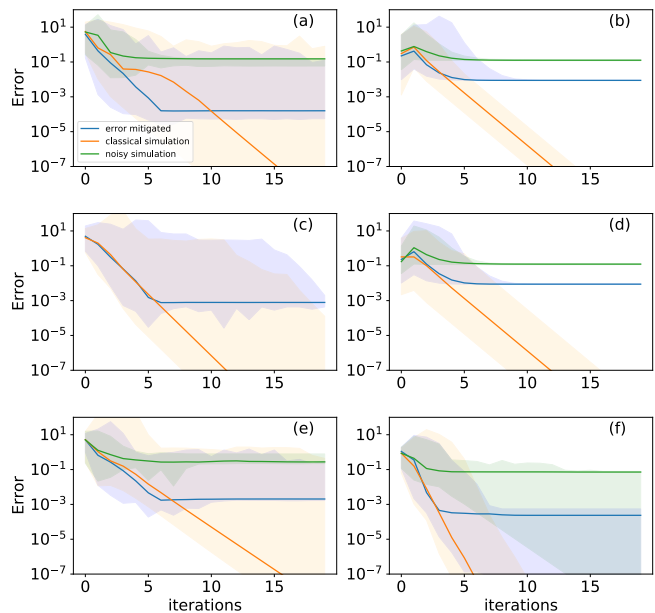
FIG. 2. Absolute value of the errors for each matrix element along 20 iterations, for classical, noisy and error mitigated simulations, along with the 92% confidence level of the points. Both columns refer to experiments where the Lagrange multipliers $|L_{1,\nu}\rangle$ and $|L_{2,\nu}\rangle$ are computed exactly as discussed in Sec. II. Each panel (a-f) corresponds respectively to the matrix elements (in the energy eigenbasis) 2, 2i, 5, −2i, 3 and 0. The orange continous line correspons to classical simulations, the blue dashed line to hybrid classical-quantum simulations with error mitigation, and the green dot-dashed line to hybrid classical-quantum simulations with no error mitigation. The lack of a green dot-dashed line in panel (c) implies that this element did not appear in our attempts hybrid classical-quantum simulations with no error mitigation.

For the single qubit case, both for the real and imaginary part we expect to obtain three different numbers from the variational approach. In Fig. 2 we consider, in

each panel, the error from each of these six possible values. More specifically we consider the real values 2, 5 and 3 in panels (a), (c) and (e) and the values from the imaginary part 2i, −2i and 0 in panels (b), (d) and (f). In each panel we show the median of the 50 runs approaching that value for three different cases: completely classical simulations (continuous orange lines), hybrid classical-quantum simulations without error mitigation (dashed green lines) and with error mitigation (dot-dashed blue lines). Also in this case the shadowing represent the 92% confidence interval of the corresponding runs (i.e. between the 4−th and the 96−th percentile). We observe that only the fully classical approach is able to reach very small errors and continuously decrease as the iterations of the optimization routine increase. At the same time, also for the classical approach the process shows a non-negligible error bar. For the hybrid approach, instead, we observe that the error reaches a plateau after about 10 iterations. This error is reduced when employing error-mitigation techniques. We thus associate this performance to an erroneous evaluation of the overlaps for the gradients. In some cases, as for the value 5, the errors are so important that the hybrid classical-quantum algorithm is not able to converge to this solution when one does not implement error mitigation.

A. Single qubit using an approximate evaluation of the Lagrange multipliers

Fig. 3 is completely analogous to Fig. 1, and thus it shows results for real and imaginary elements of W and the corresponding angles versus the number of iterations. However in this case we used the iterative algorithm from finding the minima of Eq.(5) and Eq.(6). Since in this case the Lagrange multipliers $|L_{i,\nu}\rangle$ are not exact, the optimization is not as accurate and hence while the medians of the expectation values and of the angles approach the exact values, the error bars, here also indicated by the 92% confidence level of the runs ending close to one solution, are larger.

V. Two qubits

To understand how this variational approach would perform on larger systems, we now consider a two-qubit case. We will use the Hamiltonian H_2 from Eq.(9) and the observable W_2 which in the eigenbasis of the Hamiltonian has the values W_2^D from Eq.(10). The trial eigenfunctions are parametrized as shown in Eq.(12). We consider 300 different initializations and we show, in Fig. 4 a density plot of the resulting matrix elements vs the number of iterations. By density plot we mean that for each of the 20 iterations, we consider a vertical range and divide it into small bins of length 0.2 and count how many points fall into those intervals. The initial conditions are prepared such that near the exact angles for the
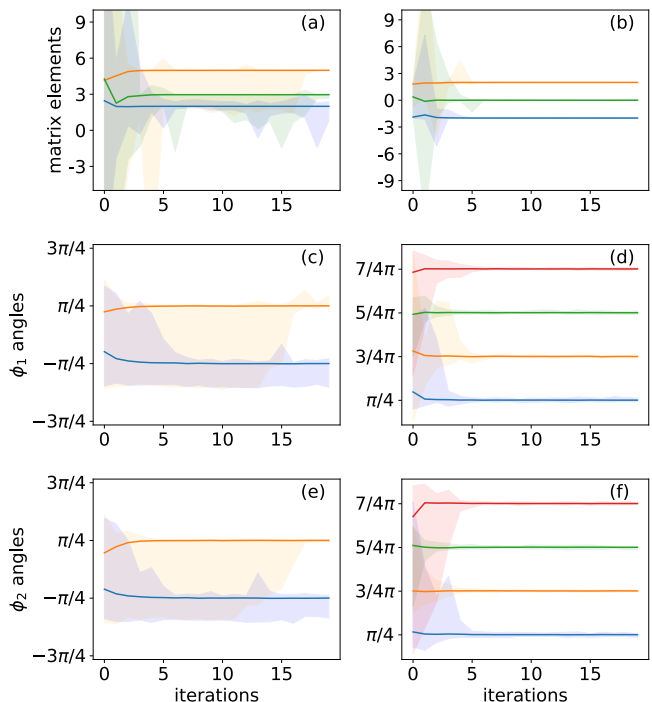
FIG. 3. Exactly the same description as Fig. 1 with the only difference that in the optimization procedure we have used approximated Lagrange multipliers from minimization of Eq.(5) or Eq.(6).
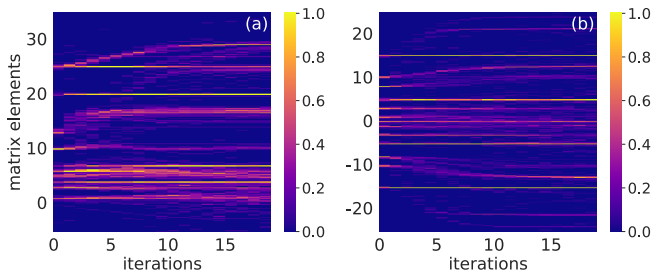
FIG. 4. Heatmaps for the two-qubit case and for the real (a) and imaginary (b) elements of W_2 in energy eigenbasis, through exact calculation of the Lagrange multipliers. Both subplots show the convergence to the matrix elements for 300 runs of 20 iterations each, initialized in intervals of radius 0.15 of the optimal angles.

solutions of the eigenfunctions in Eq.(12) with an error ±0.15 from the exact α_i, β_i and γ_i. Furthermore, noises from the quantum machine have not been considered in these two-qubit calculations, and only statistical errors, i.e. shot-noise, are considered. From Fig. 4 we find that some matrix elements are much more stable than others. For instance the elements 20, 25 and 6 converge in few iterations and they have a large probability of appearing. Other values, like 13, are unstable and they do not appear unless one chooses initial conditions for the parameters of the trial eigenfunctions very close to the

exact ones. This is true also for completely classical simulations, hinting at the fact that the landscape of this optimization problem is particularly difficult and better classical optimization routines should be used. What is possibly more striking is that there are also converged results to values which do not belong to W_2, as for instance the value ≈ 30. From this we deduce that this method can give a good number of matrix elements, and depending on the amount of noise in the machine, they can be fairly accurate, but it may also not find some values and possibly return a few which are not correct. It would thus be important to cross-check these values with other methods.

VI. Conclusions

We have considered a variational approach for hybrid classical-quantum computation of the matrix elements of an observable W in the energy eigenbasis. The diagonal elements of local observables typically relaxes to its long time asymptotic value. The fluctuations around this value as well as it relaxation times are however dictated by the off-diagonal elements [54]. Ths our approach has implications on non-equilibrium statistical studies. At the beginning, we have implemented the method with one qubit to learn some of its basic features, and then we proceed to two qubits to study how the method performs when the size of the systems studied increases. We have found that in general the method can perform well, meaning that it finds many of the matrix elements even when one initializes randomly the trial functions over a very broad range of parameters. The performance of the method are limited by the errors in the estimation of the gradients and overlaps from the quantum computer. These errors can be mitigated but they are still important, due to the fact that the variational function is not very stable close to some matrix elements. Furthermore, the variational function may also return some values that are not elements of the observable W. This occurs also in a fully classical implementation of the algorithm, indicating that improvements can be sought after with the landscape of the variational function and also with the classical optimizer. We highlight that one can use a much smaller basis, which represent states closer to those of interest, thus limiting the emergence of spurious incorrect results. In this case the optimization procedure will also be better controlled. Another way to improve the performance is to re-initialize the angles from the results after a certain number of iterations. Here we used normalized trial functions, but the method can be generalized, as we show in App. C, to non-normalized trial states. A relevant direction for future investigation is to find ways to stabilize the less stable matrix elements.

Acknowledgments

K.B. acknowledges funding by U.S. Department of Energy Award No. DE-SC0019449, DoE ASCR Accelerated Research in Quantum Computing program (award No. DE-SC0020312), DoE QSA, NSF QLCI (award No. OMA-2120757), NSF PFCQC program, DoE ASCR Quantum Testbed Pathfinder program (award No. DE-SC0019040), AFOSR, ARO MURI, AFOSR MURI, and DARPA SAVaNT ADVENT. KLC acknowledges support from the Ministry of Education and the National Research Foundation, Singapore. D.P. acknowledges support from NRF-ISF grant NRF2020-NRF-ISF004-3528.

[1] J. Preskill, Quantum **2**, 79 (2018).

[2] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, et al., Reviews of Modern Physics **94**, 015004 (2022).

[3] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. Obrien, Nature communications **5**, 4213 (2014).

[4] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New Journal of Physics **18**, 023023 (2016).

[5] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature **549**, 242 (2017).

[6] E. Farhi, J. Goldstone, and S. Gutmann, arXiv:1411.4028 (2014), 10.48550/ARXIV.1411.4028.

[7] E. Farhi and A. W. Harrow, arXiv preprint arXiv:1602.07674 (2016), 10.48550/ARXIV.1602.07674.

[8] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. De Jong, Physical Review A **95**, 042308 (2017).

[9] O. Kyriienko, npj Quantum Information **6**, 1 (2020).

[10] R. M. Parrish and P. L. McMahon, arXiv preprint arXiv:1909.08925 (2019), 10.48550/ARXIV.1909.08925.

[11] T. A. Bespalova and O. Kyriienko, PRX Quantum **2** (2021), 10.1103/prxquantum.2.030318.

[12] W. J. Huggins, J. Lee, U. Baek, B. O'Gorman, and K. B. Whaley, New Journal of Physics (2020), 10.1088/1367-2630/ab867b.

[13] T. Takeshita, N. C. Rubin, Z. Jiang, E. Lee, R. Babbush, and J. R. McClean, Physical Review X **10**, 011004 (2020).

[14] N. H. Stair, R. Huang, and F. A. Evangelista, Journal of Chemical Theory and Computation **16**, 2236 (2020).

[15] M. Motta, C. Sun, A. T. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. Brandão, and G. K.-L. Chan, Nature Physics **16**, 205 (2020).

[16] K. Seki and S. Yunoki, PRX Quantum **2**, 010333 (2021).

[17] K. Bharti, arXiv preprint arXiv:2009.11001 (2020), 10.48550/ARXIV.2009.11001.

[18] K. Bharti and T. Haug, Physical Review A **104**, L050401 (2021).

[19] A. Cervera-Lierta, J. S. Kottmann, and A. Aspuru-Guzik, PRX Quantum **2**, 020329 (2021).

[20] Y. Li and S. C. Benjamin, Physical Review X **7**, 021050 (2017).

[21] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Quantum **3**, 191 (2019).

[22] M. Benedetti, M. Fiorentini, and M. Lubasch, Physical Review Research (2021), 10.1103/physrevresearch.3.033083.

[23] K. Bharti and T. Haug, Phys. Rev. A **104**, 042418 (2021).

[24] S. Barison, F. Vicentini, and G. Carleo, Quantum **5**, 512 (2021).

[25] B. Commeau, M. Cerezo, Z. Holmes, L. Cincio, P. J. Coles, and A. Sornborger, arXiv:2009.02559 (2020), 10.48550/ARXIV.2009.02559.

[26] K. Heya, K. M. Nakanishi, K. Mitarai, and K. Fujii, arXiv preprint arXiv:1904.08566 (2019), 10.48550/ARXIV.1904.08566.

[27] C. Cirstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, npj Quantum Information **6**, 1 (2020).

[28] J. Gibbs, K. Gili, Z. Holmes, B. Commeau, A. Arrasmith, L. Cincio, P. J. Coles, and A. Sornborger, arXiv preprint arXiv:2102.04313 (2021), 10.48550/ARXIV.2102.04313.

[29] J. W. Z. Lau, K. Bharti, T. Haug, and L. C. Kwek, arXiv preprint arXiv:2101.07677 (2021), 10.48550/ARXIV.2101.07677.

[30] T. Haug and K. Bharti, Quantum Science and Technology **7**, 045019 (2022).

[31] M. Otten, C. L. Cortes, and S. K. Gray, arXiv preprint arXiv:1910.06284 (2019), 10.48550/ARXIV.1910.06284.

[32] K. H. Lim, T. Haug, L. C. Kwek, and K. Bharti, Quantum Science and Technology **7**, 015001 (2021).

[33] J. W. Z. Lau, T. Haug, L. C. Kwek, and K. Bharti, arXiv preprint arXiv:2103.05500 (2021), 10.48550/ARXIV.2103.05500.

[34] J. J. Meyer, J. Borregaard, and J. Eisert, npj Quantum Information **7**, 1 (2021).

[35] J. J. Meyer, arXiv preprint arXiv:2103.15191 (2021), 10.22331/q-2021-09-09-539.

[36] K. Bharti, Quantum Views **5**, 61 (2021).

[37] M. Schuld and N. Killoran, Phys. Rev. Lett. **122**, 040504 (2019).

[38] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209 (2019).

[39] T. Kusumoto, K. Mitarai, K. Fujii, M. Kitagawa, and M. Negoro, arXiv:1911.12021 (2019).

[40] E. Farhi and H. Neven, arXiv:1802.06002 (2018).

[41] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Physical Review A **98**, 032309 (2018).

[42] X.-L. Shang, J.-M. Dong, W. Zuo, P. Yin, and U. Lombardo, Physical Review C **103**, 034316 (2021).

[43] T. Harris, H. B. Meyer, and D. Robaina, arXiv preprint arXiv:1611.02499 (2016), 10.48550/ARXIV.1611.02499.

[44] L. Gribov, Journal of Structural Chemistry **59**, 272 (2018).

[45] R. Shakeshaft, L. Rosenberg, and L. Spruch, Physical Review A **14**, 1989 (1976).

[46] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, et al., Nature Reviews Physics (2020), 10.1038/s42254-021-00348-9.

[47] E. Gerjuoy, A. R. P. Rau, and L. Spruch, Phys. Rev. A **8**, 662 (1973).

[48] E. Gerjuoy, A. R. P. Rau, and L. Spruch, Rev. Mod. Phys. **55**, 725 (1983).

[49] J. Wadehra, L. Spruch, and R. Shakeshaft, Physical Review A **18**, 344 (1978).

[50] Y. Ibe, Y. O. Nakagawa, N. Earnest, T. Yamamoto, K. Mitarai, Q. Gao, and T. Kobayashi, Physical Review Research **4**, 013173 (2022).

[51] K. Mitarai and K. Fujii, Phys. Rev. Research **1**, 013006 (2019).

[52] C. Lanczos, Journal of Research of the National Bureau of Standards **45** (1950), 10.6028/JRES.045.026.

[53] Y. Saad, SIAM Journal on Numerical Analysis **29**, 209 (1992).

[54] W. Beugeling, R. Moessner, and M. Haque, Physical Review E **91**, 012144 (2015).

Appendix

A. Variational principle

The variational principle is best gleaned from examples. Here, we restrict ourselves to the determination of off-diagonal matrix elements in quantum mechanics. In order to do so, we write the expression of the variational principle starting from the trial objective and the constraints, i.e.

$$F_v = \langle \phi_{i,t} | W | \phi_{j,t} \rangle \tag{A1}$$

and

$$B_1 = (H - E_{i,j}) | \phi_{i,j} \rangle = 0 \text{ and } B_1^\dagger = \langle \phi_{i,j} | (H - E_{i,j}) = 0. \tag{A2}$$

The final expression is obtained by promoting the constraints to the objective by multiplication by some Lagrange multipliers L and λ. The goal is then to find expressions for L and λ either exactly or iteratively, and to check which ones are best suited for the variational principle.

Concretely, the trial quantities of interest and the trial Lagrange multipliers are defined as

$$| \phi_{i/j,t} \rangle = | \phi_{i/j} \rangle + | \delta\phi_{i/j} \rangle, \tag{A3}$$

$$\lambda_t = \lambda + \delta\lambda, \tag{A4}$$

and

$$| L_{i/j,t} \rangle = | L_{i/j} \rangle + | \delta L_{i/j} \rangle. \tag{A5}$$

From this, we write the variational form of F, F_v, as

$$\begin{aligned} F_v = & \langle \phi_{i,t} | W | \phi_{j,t} \rangle \\ & + \langle L_{i,a,t} | (H - E_i) | \phi_{i,t} \rangle + \langle \phi_{i,t} | (H - E_i)^\dagger | L_{i,b,t} \rangle \\ & + \langle L_{j,a,t} | (H - E_j) | \phi_{j,t} \rangle + \langle \phi_{j,t} | (H - E_j)^\dagger | L_{j,b,t} \rangle \\ & + \lambda \left[\langle \phi_{i,t} | W | \phi_{j,t} \rangle \mp \langle \phi_{j,t} | W | \phi_{i,t} \rangle \right], \end{aligned} \tag{A6}$$

and by replacing the trial quantities with equations A3-A5, one can get the error δF_v as

$\delta F_v = F_v - \langle\phi_{i,t}| W |\phi_{j,t}\rangle = (\langle\phi_i| + \langle\delta\phi_i|)W(|\phi_j\rangle + |\delta\phi_j\rangle) - \langle\phi_i| W |\phi_j\rangle + (\langle L_{i,a}| + \langle\delta L_{i,a}|)[(H - E_i)(|\phi_i\rangle + |\delta\phi_i\rangle)] + [(\langle\phi_i| + \langle\delta\phi_i|)(H - E_i)^\dagger](|\delta L_{i,b}\rangle + |L_{i,b}\rangle) + (\langle L_{j,a}| + \langle\delta L_{j,a}|)[(H - E_j)(|\phi_j\rangle + |\delta\phi_j\rangle)] + [(\langle\phi_j| + \langle\delta\phi_j|)(H - E_j)^\dagger](|\delta L_{j,b}\rangle + |L_{j,b}\rangle) + (\lambda + \delta\lambda)((\langle\phi_i| + \langle\delta\phi_i|)W(|\phi_j\rangle + |\delta\phi_j\rangle) \mp (\langle\phi_j| + \langle\delta\phi_j|)W(|\phi_i\rangle + |\delta\phi_i\rangle)) = 0$ By using constraints A2 and A3,discarding terms of second order and putting equal to 0 the coefficients of $|\delta\phi\rangle$ and $\langle\delta\phi|$, δF_v vanishes for all allowed $|\delta\phi\rangle$ and $\langle\delta\phi|$, and we get, for the real case

$$\langle\delta\phi_i| \rightarrow (H - E_i) |L_{i,b}\rangle = -(\lambda + 1)W |\phi_j\rangle, \quad (A7)$$

$$\langle\delta\phi_j| \rightarrow (H - E_j) |L_{j,b}\rangle = \lambda W |\phi_i\rangle, \quad (A8)$$

$$|\delta\phi_i\rangle \rightarrow \langle L_{i,a}| (H - E_i) = \lambda \langle\phi_j| W \quad (A9)$$

and

$$|\delta\phi_j\rangle \rightarrow \langle L_{j,a}| (H - E_j) = -(\lambda + 1) \langle\phi_i| W. \quad (A10)$$

Multiplying on the left of equation A7 by $\langle\phi_i|$ and on the left of equation A8 by $\langle\phi_j|$, and applying constraint A2, we get $\lambda = -1/2$. $|L\rangle$ can be obtained by $|L\rangle = c_1 |\phi\rangle$ and made unique by $\langle\phi|L\rangle = c_2 = 1$. We can see that for the real case $|L_{i,a}\rangle = |L_{i,b}\rangle$ and $|L_{j,a}\rangle = |L_{j,b}\rangle$.

For pure imaginary matrix elements, the procedure is the same, and we get

$$\langle\delta\phi_i| \rightarrow (H - E_i) |L_{i,b}\rangle = -(\lambda + 1)W |\phi_j\rangle, \quad (A11)$$

$$\langle\delta\phi_j| \rightarrow (H - E_j) |L_{j,b}\rangle = -\lambda W |\phi_i\rangle, \quad (A12)$$

$$|\delta\phi_i\rangle \rightarrow \langle L_{i,a}| (H - E_i) = -\lambda \langle\phi_j| W \quad (A13)$$

and

$$|\delta\phi_j\rangle \rightarrow \langle L_{j,a}| (H - E_j) = -(\lambda + 1) \langle\phi_i| W. \quad (A14)$$

Again, $\lambda = -1/2$ but this time $|L_{i,a}\rangle = -|L_{i,b}\rangle$ and $|L_{j,a}\rangle = -|L_{j,b}\rangle$.

B. Variational principle for $L_{i,\nu}$

The next step is to derive an expression, exact or approximated, for computing the Lagrange multiplier $|L_{i,\nu}\rangle$. This could in theory be done exactly by inverting $(H - E_i)$ in A7. In practice, a near-singularity problem arises because the operator $H - E_1$, where E_1 is the ground state energy, has a zero eigenvalue and therefore cannot be inverted; this can be solved by replacing said operator by a shifted one that does not have a zero eigenvalue as discussed in the main text. In addition, in order to get an extremum principle that would give us an approximation to $|L_{i,\nu}\rangle$ without inverting any matrix, this operator must be positive definite.

An operator that satisfies this condition is $H_{mod,i} - E_{1,t} = H - \frac{HP_{1,t}H}{E_{1,t}} - E_{1,t}$ where $P_{1,t}$ is the trial projection operator to the ground state $P_{1,t} = |\phi_{1,t}\rangle \langle\phi_{1,t}|$ and $E_{1,t}$ is the trial ground state energy $E_{1,t} = \langle\phi_{1,t}| H |\phi_{1,t}\rangle$. The idea is to find another variational principle for $|L_t\rangle$ (where we omit the indexes i and ν to lighten the notation) that would replace Eq. (A11-A14) , and that could then be used in the variational principle of Eq. (A6) to find better approximations of the trial functions. This variational principle for $|L_t\rangle$ is of the form

$$M(|X_{tt}\rangle) = \langle X_{tt}| A |X_{tt}\rangle - \langle X_{tt}|q_t\rangle - \langle q_t|X_{tt}\rangle \quad (B1)$$

where A will be the shifted non-negative operator, $|X_t\rangle$ is the trial Lagrange multiplier $|L_{tt}\rangle$, $|q_t\rangle = A |X_t\rangle$ a known function and we write $|X_{tt}\rangle = |X_t\rangle + |\delta X_t\rangle$. By using this last equivalence, eq B1 becomes

$$M(|X_t\rangle + |\delta X_t\rangle) = M(|X_t\rangle) + \langle\delta X_t| A |\delta X_t\rangle \quad (B2)$$

where the quadratic term on the far right is strictly convex since A is positive definite. This expression has a minimum for $|X_{tt}\rangle = |X_t\rangle$, i.e. $|L_{tt}\rangle = |L_t\rangle$.
Setting $|q_t\rangle = A |X_t\rangle = (H_{mod,i} - E_{1,t}) |L_t\rangle = [(\langle\phi_{1t}| W |\phi_{1t}\rangle) - E_{1t}c_{1t}] |\phi_{1t}\rangle - W |\phi_{1t}\rangle$, the variational principle for $|L_t\rangle$ becomes

$$M(L_{tt}) = \langle L_{tt}| (H^1_{mod,t} - E_{1t}) |L_{tt}\rangle - \langle L_{tt}|q_t\rangle - \langle q_t|L_{tt}\rangle \quad (B3)$$

and with $|L_{tt}\rangle = |L_t\rangle + |\delta L_t\rangle$,

$$M(|L_t\rangle + |\delta L_t\rangle) = M_{11}(|L_t\rangle) + \langle\delta L_t| (H^1_{mod,t} - E_{1t}) |\delta L_t\rangle \quad (B4)$$

which has its minimum at $|L_{tt}\rangle = |L_t\rangle$.
For the real case, eq B3 becomes:

$$M(L_{i/j}) = \langle L_{i/j}| (H_{mod,i/j} - E_{i/j}) |L_{i/j}\rangle + \langle\phi_{j/i}| W |L_{i/j}\rangle \quad (B5)$$

while for the imaginary one, it is

$$M(L_{ib/jb}) = \langle L_{ib/jb}| (H_{mod,i/j} - E_{i/j}) |L_{ib/jb}\rangle \quad (B6)$$

C. Variational principle for non normalized states

The variational principle for non-normalized wave functions requires the introduction of the following additional constraints

$$\langle\phi_{i/j}|\phi_{i/j}\rangle - 1 = 0 \quad (C1)$$

and the respective Lagrange multipliers $\lambda_{i/j}$. The variational principle then becomes

$$\begin{aligned}
F_v &= \langle\phi_{i,t}| W |\phi_{j,t}\rangle \\
&+ \langle L_{i,a,t}| (H - E_i) |\phi_{i,t}\rangle + \langle\phi_{i,t}| (H - E_i)^\dagger |L_{i,b,t}\rangle \\
&+ \langle L_{j,a,t}| (H - E_j) |\phi_{j,t}\rangle + \langle\phi_{j,t}| (H - E_j)^\dagger |L_{j,b,t}\rangle \\
&+ \lambda_i(\langle\phi_{i,t}|\phi_{i,t}\rangle - 1) + \lambda_j(\langle\phi_{j,t}|\phi_{j,t}\rangle - 1) \\
&+ \lambda [\langle\phi_{i,t}| W |\phi_{j,t}\rangle \mp \langle\phi_{j,t}| W |\phi_{i,t}\rangle]
\end{aligned}$$
$$(C2)$$

and we get the following equations for the real case

$$\langle\delta\phi_i| : (H - E_i)|L_{i,b}\rangle = -\lambda_i|\phi_i\rangle - (\lambda+1)W|\phi_j\rangle, \tag{C3}$$

$$\langle\delta\phi_j| : (H - E_j)|L_{j,b}\rangle = -\lambda_j|\phi_j\rangle + \lambda W|\phi_i\rangle, \tag{C4}$$

$$|\delta\phi_i\rangle : \langle L_{i,a}|(H - E_i) = -\lambda_i\langle\phi_i| + \lambda\langle\phi_j|W, \tag{C5}$$

and

$$|\delta\phi_j\rangle : \langle L_{j,a}|(H - E_j) = -\lambda_j\langle\phi_j| - (\lambda+1)\langle\phi_i|W \tag{C6}$$

which give $\lambda = -1/2$ and $\lambda_i = \lambda_j = -(\langle\phi_i|W|\phi_j\rangle)/2$. For the iterative approach, the function to optimize is

$$M(L_{i/j}) = \langle L_{i/j}|(H_{mod,i/j} - E_{i/j})|L_{i/j}\rangle + 2\lambda_{i/j}\langle\phi_{i/j}|L_{i/j}\rangle + \langle\phi_{j/i}|W|L_{i/j}\rangle. \tag{C7}$$

For imaginary elements, we get

$$\langle\delta\phi_i| : (H - E_i)|L_{i,b}\rangle = -\lambda_i|\phi_i\rangle - (\lambda+1)W|\phi_j\rangle, \tag{C8}$$

$$\langle\delta\phi_j| : (H - E_j)|L_{j,b}\rangle = -\lambda_j|\phi_j\rangle - \lambda W|\phi_i\rangle, \tag{C9}$$

$$|\delta\phi_i\rangle : \langle L_{i,a}|(H - E_i) = -\lambda_i\langle\phi_i| - \lambda\langle\phi_j|W, \tag{C10}$$

and

$$|\delta\phi_j\rangle : \langle L_{j,a}|(H - E_j) = -\lambda_j\langle\phi_j| - (\lambda+1)\langle\phi_i|W \tag{C11}$$

with $\lambda = -1/2$ and $\lambda_i = \lambda_j = -(\langle\phi_i|W|\phi_j\rangle)/2$. For the iterative approach for the imaginary elements, one needs to optimize

$$M(L_{ib/jb}) = \langle L_{ib/jb}|(H_{mod,i/j} - E_{i/j})|L_{ib/jb}\rangle + 2Re(\lambda_{i/j})\langle\phi_{i/j}|L_{ib/jb}\rangle \tag{C12}$$