Accelerated Convergence of Contracted Quantum Eigensolvers through a Quasi-Second-Order, Locally Parameterized Optimization

Scott E. Smart and David A. Mazziotti*

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: A contracted quantum eigensolver (CQE) finds a solution to the many-electron Schrödinger equation by solving its integration (or contraction) to the two-electron space—a contracted Schrödinger equation (CSE)—on a quantum computer. When applied to the anti-Hermitian part of the CSE (ACSE), the CQE iterations optimize the wave function, with respect to a general product ansatz of two-body exponential unitary transformations that can exactly solve the Schrödinger equation. In this work, we accelerate the convergence of the CQE and its wave function ansatz via tools from classical optimization theory. By treating the CQE algorithm as an optimization in a local parameter space, we can apply quasi-second-order optimization techniques, such as quasi-Newton approaches or nonlinear conjugate gradient approaches. Practically, these algorithms result in superlinear convergence of the wave function to a solution of the ACSE. Convergence acceleration is important because it can both minimize the accumulation of noise on near-term intermediate-scale quantum (NISQ) computers and achieve highly accurate solutions on future fault-tolerant quantum devices. We demonstrate the algorithm, as well as some heuristic implementations relevant for cost-reduction considerations, comparisons with other common methods such as variational quantum eigensolvers, and a Fermionic-encoding-free form of the CQE.
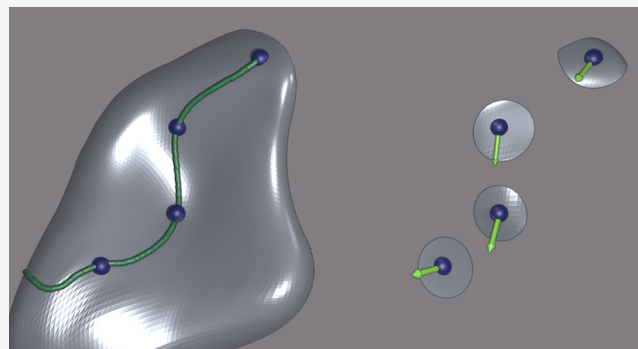
I. INTRODUCTION

The contracted Schrödinger equation (CSE)[1−3] describes the projection of the molecular Schrödinger equation for a *N*-electron system onto a two-electron space, which generates the stationary-state condition of the two-electron reduced density matrix (2-RDM), instead of the wave function. Satisfaction of the anti-Hermitian part of the contracted Schrödinger equation (ACSE) by a quantum state is equivalent to its invariance, with respect to all infinitesimal two-body unitary transformations.[4−6] Solution of the ACSE for the 2-RDM was initially challenging because the equation is dependent on both the 2-RDM and three-electron RDM (3-RDM), making it indeterminate without additional information. However, the ACSE has been practically solved by reconstructing the 3-RDM by its cumulant expansion as a functional of the 2-RDM.[4,5,7,8] While the ACSE does not strictly imply the CSE,[9] the accuracy of its solution has been observed to approach that of full configuration interaction in the absence of reconstruction approximations.[4,10−12] The ACSE has been applied to computing strongly correlated ground and excited states in both chemical reactions and conical intersections.[12−15]

Recently, the ACSE has been solved on quantum devices with applications to hydrogen chains, as well as the benzyne isomers.[10,16,17] On a quantum computer, the ACSE algorithm, known as a contracted quantum eigensolver (CQE), iteratively minimizes the residual of the ACSE, in contrast to the variational quantum eigensolvers (VQE) that minimize the energy, with respect to parameters according to the Rayleigh-Ritz variational principle. Instead of propagating only the 2-RDM as in the classical algorithm, we propagate a wave function through state preparation on a quantum computer. Thus, we avoid reconstructed RDMs, as the 2-RDM can be directly measured from the quantum state while the ACSE residual can be directly measured from an auxiliary quantum state, or a measured 3-RDM. The resulting ACSE algorithm is a potentially exact RDM approach that scales polynomially in the size of the molecular system.

In this paper, we accelerate the convergence of the CQE for the ACSE by developing quasi-second-order algorithms with superlinear convergence. Convergence acceleration is important for avoiding the accumulation of noise on near-term intermediate-scale quantum (NISQ) computers, as well as
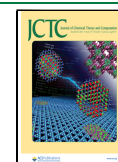
achieving highly accurate solutions on future fault-tolerant quantum devices. We draw upon research on optimization algorithms on manifolds,[18−20] which have applications across science and engineering from vision to robotics, as well as related algorithms in electronic structure for orbital optimization.[21,22] We specifically develop and implement a quasi-Newton scheme with the Broyden−Fletcher−Goldfarb−Shanno update and nonlinear conjugate gradient algorithms. The quasi-second-order algorithms avoid storage of the Hessian matrix while providing superlinear convergence. We also demonstrate convergence properties as well as some approximate implementations of the search direction and finally compare the resulting CQE algorithms with classes of common quantum algorithms, including the variational quantum eigensolver.

II. THEORY

We discuss the solution of the ACSE via CQE in section II.A, the local parametrization of the wave function in section II.B, the quasi-second-order accelerations in section II.C, and resource optimization in section II.D.

II.A. Solution of the ACSE. Given a molecular system with Hamiltonian $\hat{H}$, we can write a contraction of the Schrödinger equation onto the two-particle space, known as the contracted Schrödinger equation (CSE).[1,3]

$$\langle\Psi|\hat{H}\hat{a}_i^\dagger\hat{a}_k^\dagger\hat{a}_l\hat{a}_j|\Psi\rangle = E\,^2D_{jl}^{ik} \tag{1}$$

The CSE can be split into a Hermitian and an anti-Hermitian part, the latter of which is called the anti-Hermitian CSE, or ACSE:[4,5]

$$\langle\Psi|[\hat{H},\,\hat{a}_i^\dagger\hat{a}_k^\dagger\hat{a}_l\hat{a}_j]|\Psi\rangle = \,^2A_{jl}^{ik} \tag{2}$$

Here, $^2A_{jl}^{ik}$ is the residual of the ACSE, which is necessarily zero when $|\Psi\rangle$ is an eigenstate of the wave function. We can also obtain the ACSE in eq 2 by considering unitary transformations $\exp(\epsilon\hat{P})$ generated by a parameter ϵ and a two-body anti-Hermitian operator $\hat{P}$:

$$\hat{P} = \sum_{ij;kl} \,^2P_{kl}^{ij}\hat{a}_i^\dagger\hat{a}_k^\dagger\hat{a}_j\hat{a}_l \tag{3}$$

The derivative of the energy, with respect to the elements of the operator ^2P, yields

$$\frac{\partial E(\hat{P})}{\partial(^2P_{jl}^{ik})} + O(\epsilon^2) = \epsilon\langle\Psi|[\hat{H},\,\hat{a}_i^\dagger\hat{a}_k^\dagger\hat{a}_l\hat{a}_j]|\Psi\rangle = \epsilon\,^2A_{jl}^{ik} \tag{4}$$

which we observe is equal to the residual of the ACSE. In the solution of the ACSE,[5] the energy and 2-RDM can be expressed as a system of differential equations, in terms of a discretized, timelike parameter λ that controls the transformation of the implicit wave function to minimize the energy. As λ increases, we approach a solution of the ACSE. On a quantum computer,[10] we have a potentially exponential advantage, in terms of simulating the exact 2-RDM (or 3-RDM) and ^2A matrices.

II.B. Local Parameterization of the Contracted Quantum Eigensolver. We can describe the generic problem in the variational quantum eigensolver[23−26] for finding the ground-state wave function of a quantum system as

$$\min_{\vec{\theta}\in\mathbb{R}^\nu}\langle\Psi_0|U^\dagger(\vec{\theta})\hat{H}U(\vec{\theta})|\Psi_0\rangle \tag{5}$$

where $\vec{\theta}$ represents a vector of ν real parameters and we assume the wave function is properly normalized. Upon convergence, the following equation is satisfied for all k,

$$\frac{\partial}{\partial\theta_k}\langle\Psi_0|U^\dagger(\vec{\theta})\hat{H}U(\vec{\theta})|\Psi_0\rangle = 0 \tag{6}$$

indicating that the gradient, with respect to all parameters, vanishes. However, there can be significant problems associated with describing the appropriate parametrization of Ψ. The exact solution of the wave function scales exponentially, which might imply that an exponential number of parameters is necessary in a variational scheme. Using an operator such as unitary coupled cluster provides an exponential ansatz, but because the mapping from the Euclidean space to the unitary space is nonlinear, there can be singularities or unphysical minima in the parameter space.[18,27,28] In addition, it has been shown that high-dimensional parametrizations in a random variational ansatz generate barren plateaus where the variance in the energy gradients vanishes as the system size increases.[29,30] Because an exponential scaling parametrization is not feasible for larger systems, and limited excitation ansatz such as UCC singles and doubles are not sufficiently accurate, a slew of iterative schemes based on the VQE and UCC schemes that deviate from the traditional CC formalism have been proposed, providing scalable approaches that generally repeat or extend upon certain ansatz fragments.[31−33] One such approach is the adaptive derivative assembled pseudotrotterized VQE method, or ADAPT-VQE, which takes elements of the ACSE (or generalized UCCSD) to generate an increasingly more-complex variational problem.[34]

In the CQE approach, we instead forego the global parametrization of the state and use an atlas of local parametrizations,[18] describing the trajectory of the state, with the parameter space being dependent on the contracted eigenvalue equation (which, here, is the ACSE). The term "local" here refers to the fact that our parameters are describing information solely in a small neighborhood around the state at the current iteration and do not contain parameters from previous iterations. Each local parametrization in the atlas is concretely generated by the exponential transformation of a two-body anti-Hermitian operator, providing a map between the Euclidean parameter space and the space of unitary transformations. Importantly, this compact mapping avoids oddities or singularities that can arise from a nonlinear mapping. Thus, the optimization is no longer defined by a fixed reference set of parameters but rather by a local parametrization at each iteration n with the $\hat{P}$ of eq 3:

$$\min_{\hat{P}_n\in\mathbb{R}^\mu}\langle\Psi_n|e^{-\hat{P}_n}\hat{H}e^{\hat{P}_n}|\Psi_n\rangle \qquad \|P_n\| < \delta \tag{7}$$

where μ denotes the dimension of the two-body operator space, and the norm on P_n indicates that we are staying within a neighborhood around the wave function (replacing the ϵ in previous formulations). Our optimization is then satisfied if we have a Ψ_n such that, for all two-body operators,

$$0 = \frac{\partial}{\partial P_n^{ij;kl}} \langle \Psi_n | e^{-\hat{P}_n} \hat{H} e^{\hat{P}_n} | \Psi_n \rangle$$

$$= \langle \Psi_n | [\hat{H}, \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l] | \Psi_n \rangle \tag{8}$$

which implies that we have fulfilled the ACSE. At each iteration n, our current state is mapped to a new state through the exponential mapping. On the classical computer, the wave function is replaced by RDMs, and the transformation requires an approximate reconstruction of the 3-RDM, in terms of the 2-RDM. On a quantum computer, the transformation can be realized through standard means of exponential computations, where we transform the Fermionic operator to the Pauli basis and implement the exponential, in terms of elementary gates through its trotterized form.

While it might be thought that the restriction of $\hat{P}$ to a set of two-body operators is too restrictive, as clearly the two-body operator space does not locally parametrize the unitary group, the current approach iteratively constructs higher-order excitations from the reference wave function[5,9,35] (see section II.E. of ref [5] for a discussion of the ACSE ansatz). As we discuss in the next section, because the ACSE is solved iteratively, we can construct quasi-second-order algorithms if we choose each $\hat{P}_n$ in the ansatz by considering not only the gradient of the current iteration but also the gradients of previous iterations, which contain information about the curvature.

II.C. Quasi-Second-Order CQE Algorithms. Previous algorithms for solving the ACSE use path-following[5] or descent[8] algorithms based on the gradient. While these algorithms are robust to reconstruction errors, gradient-descent algorithms are first-order algorithms with generally linear convergence. To accelerate convergence, we can consider choosing the search direction ^2P_k by a second-order approach such as the Newton−Raphson method:

$$^2P_k = -H_k^{-1}\, ^2A_k \tag{9}$$

where H_k is the Hessian matrix. Within a certain region of the state space, we are guaranteed quadratic convergence.[36] However, the elements of the Hessian are evaluated according to

$$H_{n\,pq;rs}^{ij;kl} = \frac{\epsilon^2}{2}(\langle \Psi_k | [\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l, [\hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s, \hat{H}]] | \Psi_k \rangle$$
$$+ \langle \Psi_k | [\hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s, [\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l, \hat{H}]] | \Psi_k \rangle) \tag{10}$$

which requires the 4-RDM or its approximation.

To address this issue, we consider the BFGS quasi-Newton method, which uses the Broyden−Fletcher−Goldfarb−Shanno (BFGS) update within Davidon's method[37,38] and is summarized in Table 1. At each step of the BFGS method, we update an approximate Hessian matrix through a secant equation where the update is designed to keep the Hessian positive definite. By including a direction based on the approximate Hessian, the BFGS method achieves a superlinear rate of convergence near the solution.

At any particular iteration, given a wave function $|\Psi_n\rangle$, the ACSE residual $\hat{A}_n$, and the inverse of the approximate Hessian B_n, we define the step direction as

$$^2P_n^{ij;kl} = -\sum_{pqrs} (B_n^{-1})_{pq;rs}^{ij;kl}\, ^2A_n^{pq;rs} \tag{11}$$

Table 1. Quasi-Newton CQE Algorithm

Quasi-Newton CQE
Set $0 \to n$
Initialize $|\Psi_0\rangle$, $^2A_0$, and B_0
Continue until $\|^2A_n\| < \delta$:
Step 1: Update $^2P_n = B_n^{-1}\, ^2A_n$
Step 2: $\min_{\alpha_n} E_n(\alpha_n \hat{P})$
$\circ$ $|\Psi_{n+1}\rangle = e^{\alpha_n \hat{P}_n}|\Psi_n\rangle$
$\circ$ $s_n = \alpha_n\, ^2P_n$
Step 3: Evaluate $^2A_{n+1}$
$\circ$ $y_n = {}^2A_{n+1} - {}^2A_n$
Step 4: Calculate B_{n+1}^{-1}
Step 5: $n + 1 \to n$

We next minimize the energy by a line search to obtain a direction α_n that satisfies the conditions of sufficient descent and curvature (Wolfe conditions). In our local frame, we have the following auxiliary BFGS functions: $s_n = \alpha_n \hat{P}_n$ and $y_n = y_n = \hat{A}_{n+1} - \hat{A}_n$. Using these, we calculate B_{n+1}^{-1}, according to the BFGS formula:

$$B_{n+1}^{-1} = \left(I - \frac{s_n y_n^{\mathrm{T}}}{y_n^{\mathrm{T}} s_n}\right) B_n^{-1} \left(I - \frac{y_n s_n^{\mathrm{T}}}{y_n^{\mathrm{T}} s_n}\right) + \frac{s_n s_n^{\mathrm{T}}}{y_n^{\mathrm{T}} s_n} \tag{12}$$

We then increase n and continue until the gradient norm $\|^2A\|$ satisfies a convergence threshold. Step 4 in Table 1 can be replaced with a suitable update replacement, and, in the results, we demonstrate the use of a limited-memory BFGS implementation, denoted l-BFGS.[38]

In practice, because of the symmetry of 2-RDM elements, we can store the ^2A matrix as a vector in a compact representation, and then store the B^{-1} matrix exactly. For larger systems, it is likely that even this would be prohibitive, and instead the limited-memory approach would be necessary, where we store only k previous steps, which is equivalent to k 2-RDMs. On a classical device, the algorithm would be very similar, although instead of updating the wave function, we would have a 2-RDM update step. Both the 2-RDM and ^2A updates would require the classical evaluation of the ACSE and hence, a reconstructed 3-RDM.

On a quantum computer, we have significant advantages in that we are not reconstructing the 3-RDM. This means that we generate (up to statistical and noise-related errors) pure 2-RDMs at each step, and we do not have to consider the N-representability of the 2-RDM, or the step size in the solution of the differential equations. An important note is that if we want to evaluate the ACSE's residual in the classical part of the algorithm rather than in the quantum part by tomography of an auxiliary state, we must be sure that the residual is sufficiently accurate to estimate the curvature. While an approximated 3-RDM can give good enough information to obtain chemical accuracy in many instances, for rigorous convergence, the cumulant portion of the 3-RDM should be measured by tomography. Practically, this would entail alternating evaluations of the 3-RDM (for the ACSE) and the 2-RDM (for energy evaluations).

As an alternative to the quasi-Newton approaches, we can instead use the nonlinear conjugate gradient (CG) approaches.[38] The nonlinear CG method does not require the storage of a Hessian or approximate Hessian, and instead involves only a simple update step governing the contribution

of the previous search direction $\hat{P}_k$. A description of the generic CQE algorithm with a CG solver is described in Table 2.

Table 2. Conjugate Gradient CQE Algorithm

Conjugate Gradient CQE
Set $0 \leftarrow n$
Initialize $\lvert \Psi_0 \rangle$, $^2A_0$
While $\lVert {}^2A_n \rVert > \delta$:
Step 2: $\min_{\alpha_n} E_n(\alpha_n \hat{P})$
Step 3: Evaluate $^2A_{n+1}$
Step 4: Calculate β_{n+1}
Step 5: $^2P_{n+1} = -{}^2A_{n+1} + \beta_{n+1}{}^2P_n$
Step 6: $n + 1 \rightarrow n$

There are numerous modifications to the conjugate gradient method, which we do not explore here.[39] Such modifications include preconditioning schemes and modified update coefficients, as well as additional criteria on resets and step lengths.

II.D. Resource-Optimized CQE Search Directions. In light of the current representation of the CQE algorithm, one can see that the resources that are demanded on the quantum computer will be heavily dependent on the selection and implementation on the search direction, as each term must be implemented individually. However, any modification to the search direction will be detrimental to the rate of convergence (and can potentially also negate the theoretical results). Thus, we would like to find a tradeoff between potentially reducing the number of terms and preserving a descent direction.[38]

Therefore, we focus on two approximations that still preserve the essential nature of the ideal CQE approach: p-depth and operator sparsification. First, we introduce a more formal way of describing the ACSE ansatz at a given iteration n. Let $\{\hat{\mathcal{P}}_l^{(n)}\}$ be an ordered set of anti-Hermitian two-body operators:

$$\hat{\mathcal{P}}_l^{(n)} = \sum_{ijkl} \mathcal{P}_l^{(n)ij;kl} \hat{a}_i^\dagger \hat{a}_k^\dagger \hat{a}_l \hat{a}_j \tag{13}$$

In the exact CQE approach, each iteration adds a new two-body operator to the set, and these collectively are the stored ansatz vectors. Using this, we can write the ACSE ansatz at the Nth iteration as

$$\lvert \Psi_n \rangle = \prod_l^m e^{\hat{\mathcal{P}}_l^{(n)}} \lvert \Psi_0 \rangle \tag{14}$$

where m is the total number of two-body exponential operators we are implementing. We define the p-depth as follows. Given a search direction $\hat{P}_n$, we iterate over the elements $P_n^{ij;kl}$ and the set of operators $\{\hat{\mathcal{P}}_l^{(n)}\}$ from $l = m - p$ to $l = m$. If an element $P_n^{ij;kl}$ was included in a previous operator, we update that two-body operator as

$$\mathcal{P}_{n-p}^{ijkl} \leftarrow \mathcal{P}_{n-p}^{ijkl} + P_n^{ijkl} \tag{15}$$

From the definition, the element will only be in one of the p-previous operators. If elements cannot be assigned to a previous $\mathcal{P}_l^{(n)}$, a new operator is appended. The ACSE only provides information on the gradient around the exterior of the wave function, and because, generally, each iteration does not commute with previous iterations for $p > 0$, this can be considered an approximate scheme of implementing the CQE. As an example, if all terms are included in the initial $\hat{\mathcal{P}}$, then we have a single exponential form in the approximate linear region, and any p-depth greater than 1 will be equivalent. Note that such a case is similar to a generalized UCCSD ansatz with different ordering under the first-order trotterization; however, the method of updating does not reflect the true gradient terms and, as such, provides an approximation of the operator in the linear region.
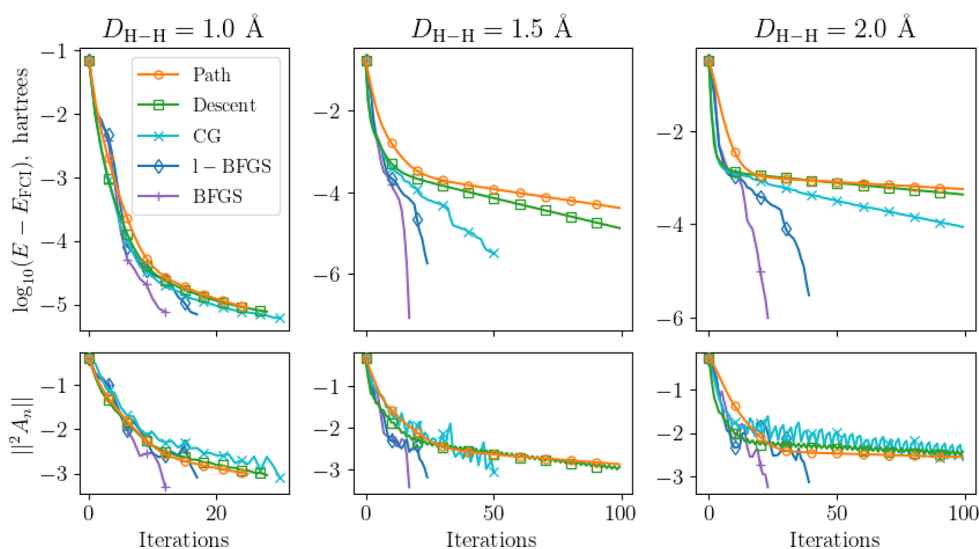
Figure 1. Comparison of methods for generating $\hat{P}$ at different H−H distances for linear H_4. Near equilibrium, similar patterns can be seen, but away from equilibrium, gradient-descent-based approaches slow down considerably. In particular, when optimizing the 1-d step size, according to a quadratic fit, an oscillatory pattern can be seen in the norm of ^2A, indicating a potential valley in a direction between the oscillating gradients. The conjugate gradient approach allows for slightly more flexibility but exhibits stronger oscillations in the gradient. The quasi-Newton approaches offer quick convergence across all regions of dissociation.

The second approach is an operator sparsification scheme, which effectively reduces the number of new terms appended at each step. The method of ordering is of particular importance, and we investigate two options. First, we can sort the elements according to their absolute value, which is important for implementation purposes (i.e., we can remove the smallest elements first). Alternatively, we can sort elements according to the energy contribution in the descent direction. This value is obtained element-wise as the product $A_n^{ij;kl} P_n^{ij;kl}$. For gradient-descent approaches, these two criteria are equivalent, and, in previous work, only the former approach was used.

We can also control the number of terms that are removed through a constant $c \in [0, 1]$. This is a constant scaling factor where $c = 1$ (strict truncation) indicates only the largest term in $\hat{P}$ is included and $c = 0$ (no truncation) indicates that all terms in $\hat{P}$ are included. Another potentially useful control instead of a constant scaling factor (which we do not explore here) would be to limit explicitly the number of terms included in each term. The last approximation is a boolean INCLUDE = {True, False} option, whereby elements of the search direction that would be assigned to the previous p-operators due to the p-depth specified are included in the sparsification scheme. The controls here can be likened to techniques used in trotterization or compact forms of the two-body operator, with a goal of efficiently implementing the two-body ansatz.[40−42] They can be applied broadly to any optimizer instance, and could have different effects with different optimizers.

III. APPLICATIONS AND RESULTS

In this section, we look at applications of the optimized CQE scheme. First, we investigate the role of different optimizers. Second, We present discuss schemes related to practical implementations and modifications to the search direction. Third, we show results for linear hydrogen chains and lithium hydride with a CQE utilizing the unencoded ACSE, and finally, we compare our results to variational quantum eigensolvers, including the ADAPT-VQE algorithm.

III.A. Implementation of Optimized CQE. We begin by comparing the exact (up to a first-order trotterization) implementations of different methods for the H_4 system at different bond lengths, corresponding with differing degrees of electron correlation. In this work, our convergence criteria is typically taken to be the Frobenius norm of the ^2A matrix. For the conjugate gradient approach, we do not utilize any preconditioning, and we use the update strategies of Fletcher and Reeves.[38] The limited BFGS strategy utilizes three previously stored steps. Figure 1 displays our results. Further computational details are included in the Appendix.

While for the bond distance near equilibrium $D = 1$ Å, the different optimizers in the ACSE appear to have no apparent advantage, the more correlated distances show strong deviations between the different approaches. Particularly, in the 1.5 Å case, the convergence flattens when gradient-descent-based approaches are used. For the 2.0 Å case, this is accentuated, as we appear to have entered a region where the gradient is quite shallow. The quadratic step appears to be even slightly worse in optimizing the norm of the ^2A matrix than a simple gradient descent. The conjugate gradient approach appears to be better but shows strong oscillations in the gradient norm (note if we loosen the update parameter, these oscillations decrease, but we do not observe a significant

increase in convergence). The most successful approaches are the quasi-Newton methods, which are able to achieve high accuracy results in only a few iterations. The l-BFGS offers a reliable approach as well, with a quality between the conjugate gradient and full BFGS methods.

While this approach is the closest to the ideal implementation and offers some advantages for a classical quantum simulation (i.e., similar to the classical ACSE, except instead of the 2-RDM, we only need to store the state vector), for near-term applications, there are several constraints. In a noiseless regime (note that the importance of noise is also relevant, but, since this also changes the optimization strategy, it is not addressed directly here), addressing the compactness or efficiency is an important problem. In particular, we would however like to know if we can reduce the amount of terms that are added at each iteration, using the ideas mentioned above. These result in an appproximate search direction, which we constrain to represent a descent direction.

We use the 1.5 Å case, which contains nontrivial electron correlation and starts to differentiate between different optimizers, to look at ways that we can modify the search direction. Table 3 explores the number of iterations and CNOT gate cost for a variety of options with the BFGS optimizer. In particular, we examine the absolute norm or energy contribution for the sparsity operator acting on the search direction, as well as the inclusion or exclusion of terms that appear in the p-depth addition scheme (specified by include). For each of these criteria, we look at different p-depths and values of the sparse scheme.

Many interesting trends emerge. First, there is a difference in application of the sparsification operator acting on elements according to their energy contribution or absolute value. Namely, when using large c for the absolute value, problems in the optimization can occur. These are instances where the search direction has little overlap with the gradient, and the largest term selected is ordered in such a way that it is not strictly increasing. However, the descent condition is able to converge across every configuration, albeit at different rates of convergence. This also leads to a stratification in the rates of convergence for the descent condition, which can be seen in Figure 2. While the absolute value condition leads to accelerated convergence in almost all instances, it does seem more sensitive to using restricted operators. Of course, the inclusion of previous terms does seem to alleviate this problem, and because it seems to occur when the search direction is nearly orthogonal to the gradient, resetting the optimization might allow the optimization to continue. It is also possible that the order of magnitude for the descent condition should be lower than the absolute value condition, although this could vary significantly based on the system.

Second, the INCLUDE option has a strong impact on the rate of convergence. For both selection criteria, inclusion of previous terms clearly helps in assisting the overall convergence. As this can be considered as a way of increasing the pool of operators at each step, depending on both c and the p-depth, the advantage here makes sense. Third, the p-depth appears to have a 2-fold role. More generally, it serves to reduce the total number of terms needed in the ansatz. That is, as the p-depth increases, the number of total terms in the ansatz decreases. In addition, for a given c, we do see numerous instances where the total number of iterations decreases as the p-depth increases when INCLUDE = True. When INCLUDE = False, the trends are somewhat unclear, and the optimization

Table 3. Reported Number of Iterations and CNOT Gates (in brackets, × 10³) for Approximate Implementation Schemes for the CQE with BFGS Optimizer with Linear H₄ at a Distance of 1.5 Å^{a}

	$\|P_n^{ijkl}\|$										$(A_n P_n)^{klij}$									
	INCLUDE = False					INCLUDE = True					INCLUDE = False					INCLUDE = True				
	p-depth					p-depth					p-depth					p-depth				
sparse[c]	9	7	5	3	1	9	7	5	3	1	9	7	5	3	1	9	7	5	3	1
0.9	*	*	*	*	*	76 [1.8]	73 [2.5]	*	65 [4.0]	117 [8.9]	×	×	283 [18]	×	×	47 [1.6]	48 [2.2]	60 [4.0]	80 [4.0]	107 [8.1]
0.5	*	*	*	*	*	25 [0.90]	25 [0.90]	42 [2.6]	*	*	71 [0.90]	139 [5.4]	123 [6.6]	119 [7.6]	75 [13]	26 [0.91]	35 [1.9]	35 [2.5]	39 [3.3]	63 [7.6]
0.25	26 [1.6]	28 [3.4]	28 [4.6]	27 [5.8]	31 [11]	24 [1.3]	24 [1.3]	24 [2.2]	28 [6.5]	29 [11]	63 [0.93]	67 [0.93]	70 [3.7]	81 [6.1]	76 [15]	26 [1.2]	26 [1.2]	30 [3.3]	36 [5.1]	46 [8.8]
0.125	22 [1.5]	22 [1.5]	22 [1.5]	23 [7.0]	24 [13]	21 [1.5]	21 [1.5]	22 [5.6]	21 [6.5]	24 [13]	47 [1.4]	56 [3.9]	74 [6.3]	44 [7.8]	54 [16]	24 [1.5]	26 [3.2]	26 [3.7]	29 [5.5]	38 [8.8]

^{a}We varied the inclusion of terms in the selection of P (sorted by either $|P^{ijkl}|$, or by the descent condition), the manner of truncation of P, as well as varying the sparsity and the p-depth for each condition. The asterisk symbol (*) represents a run that did not converge, and the cross symbol (×) indicates runs that reached the maximum iterations (300).
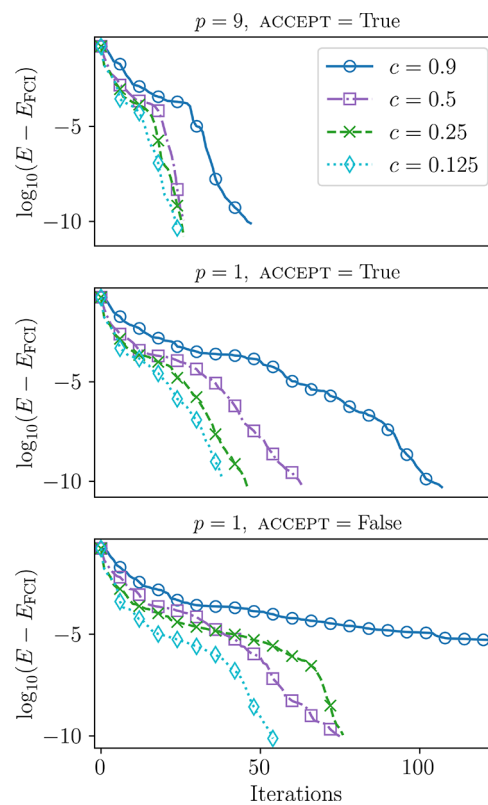

Figure 2. Low threshold ($\|^2 A\| = 10^{-5}$) results for varying values of sparsity with different p-depths and acceptance criteria, utilizing the BFGS optimizer for molecular H₄ at the intermediate distance ($D = 1.5$ Å), with the descent condition. In the top two plots, we use ACCEPT = True, where elements below the sparsity threshold are automatically included if they are in included in one of the previous p-terms. The bottom plot has ACCEPT = False. Only the first 125 iterations are shown.

has a tendency to be more sensitive. Interestingly, for $c = 0.5$, 0.25, and 0.125, with INCLUDE = False, the total iterations appear to increase as the p-depth increases, and then decrease. More sparse (see $c = 0.9$, 0.5) truncations result in slower convergence, which can be aided with the INCLUDE option, but not completely mitigated.

While these results are obviously not generalizable to every system, it is likely that some of these trends can be seen elsewhere. We expect that the p-depth can lead to lower circuit depths. In addition, while sparser operators are desirable from a NISQ perspective, optimization, with respect to a single parameter, is clearly detrimental to the rate of convergence. This can be mitigated through expanding the pool directly with more terms and the sparsification operator, or indirectly through the p-depth.

III.B. Encoding-Free CQE. The encoding-free (or un-encoded CQE) approaches for preparing states as an alternative to Fermionic state preparation have recently been explored,[33,43−45] within the VQE framework as well as in attempting to understand the success of heuristic and nonfermionic ansatz preparation. Recent works by the present authors showed that the Fermionic 2-RDM can be function-alized from a qubit-particle wave function.[46] As long as Fermionic tomography is performed on a N-qubit particle state, the 2-RDM represents a valid Fermionic 2-RDM. In addition, in recent work, we introduce an encoding-free CQE algorithm that evaluates the anti-Hermitian component of the

two-qubit-particle contraction onto the Schrödinger equation. Table 4 displays calculations for several bond distances of H_5 for both the encoded and unencoded CQE cases with the BFGS algorithm as an example system.

Table 4. Comparisons of Total Iterations, Total CNOT Gate Count, and Average CNOT Gates Per Iteration for Unencoded and Encoded CQE, Using the BFGS Optimizer for H_5 at Various Bond Distances from Equilibrium in the Minimal STO Basis[a]

| $D - D_{eq}$ (Å)	iterations (CQE,UCQE)	total CNOT $\times 10^4$ (CQE,UCQE)	$\langle CNOT_k \rangle \times 10^2$ (CQE,UCQE)
−0.25	26, 33	1.7, 1.5	6.4, 4.5
+0.00	29, 36	2.1, 1.8	7.4, 5.0
+0.25	47, 49	3.8, 2.7	8.2, 5.5
+0.50	40, 34	4.3, 2.6	11, 7.8
+0.75	39, 28	5.7, 2.8	15, 9.9
+1.00	42, 43	6.9, 4.4	16, 10.
+1.25	47, 80	8.0, 7.4	17, 9.3

[a]The accuracy of both approaches is largely similar across the dissociation curve, although, for +1.00 and +1.25 distance separation, the unencoded CQE requires more iterations.

The unencoded CQE under optimization matches the Fermionic case in most instances, and it consistently has a smaller average number of CNOT gates per iteration. For the two longest separation lengths, the number of iterations required does increase, leading to similar CNOT counts for the total ansatz. A future goal would be to incorporate compilation schemes or adjust the set of ACSE or unencoded ACSE excitations to favor a largely commuting pool.

III.C. Comparison with VQE. Finally, we compare the CQE approach utilizing a BFGS optimization with other known quantum algorithms. While in previous work,[16] similarities between iterative nature of the ACSE and ADAPT-VQE were discussed, here, we provide example calculations of VQE, ADAPT-VQE, and the ACSE that demonstrate fundamental differences in these algorithms. These are included in Figure 3, as well as in Table 5.

While the VQE results in Figure 3 are not that surprising, based on the use of the unitary coupled cluster ansatz, we still can see some interesting comparisons. For equilibrium distances, UCCSD provides a good ansatz, and there are numerous methods exploring the UCC ansatz.[31,47] By comparison, with ADAPT, we are able to obtain seemingly arbitrary convergence, matching previous work. However, note that the iterative cost of the ADAPT is much higher than either VQE or the CQE. The CQE, on the other hand, performs quite well in a variety of instances, with the most challenging case being dissociated H_6, where higher-order excitations dominate and the system is strongly correlated. While the number of macro iterations for the ADAPT procedure might look only slightly worse than the CQE or VQE approaches, when taken into account with the VQE cost, (i.e., micro + macro iterations), the length of the ADAPT procedure is somewhat unwieldy, due to the VQE subroutine. The total number of gradient and residual evaluations for each of these instances is seen in Table 5.

We can also look at qubit implementation of the ADAPT scheme, which here follows the qubit-particle excitation-based scheme of Yordanov et al.,[45] and not the quasi-particle
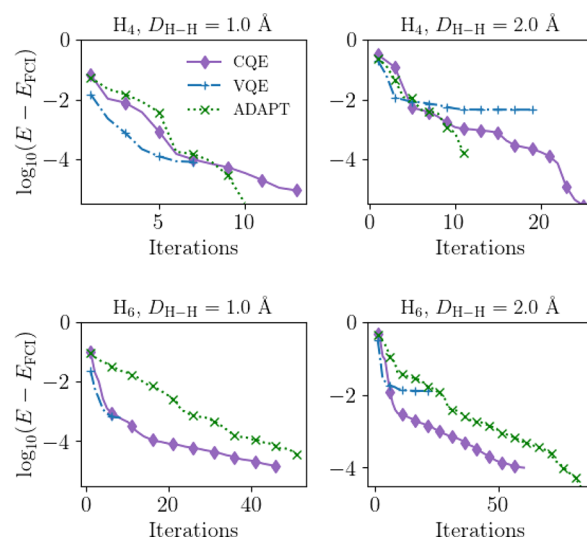
Figure 3. Simulations of molecular H_4 and H_6 for separations of 1 Å, and 2 Å utilizing the CQE, VQE, and ADAPT-VQE with a BFGS optimizer. For H_4 a minimal ansatz is reached in the ADAPT through the VQE subroutine, and for H_6 the CQE algorithm slightly outperforms the ADAPT. VQE, here using the UCCSD anastz, provides rapid convergence for near equilibrium states, but has significant errors at larger separations, which are overcome by the iterative ADAPT-VQE and CQE algorithms.

Table 5. Comparison of Gradient Element (VQE) and Residual Element (ACSE) Evaluations for the CQE, VQE, and ADAPT-VQE Methods, Corresponding to Simulations for Molecular H_4 and H_6 at Separations of 1 and 2 Å[a]

method	quantity	H_4, 1 Å	H_4, 2 Å	H_6, 1 Å	H_6, 2 Å
CQE	Iterations	13	25	46	60
	Residuals	1950	3750	38640	50400
VQE	Iterations	7	19	9	22
	Gradients	182	494	1053	2574
ADAPT-VQE	Macro	10	12	51	84
	Micro	71	167	1221	5854
	Gradients	362	1226	43112	354988
	Residuals	660	792	16830	27720

[a]Here, note tha the ADAPT-VQE has a symmetry-adapted pool of operators, which are not implemented here in the CQE or VQE approaches. The VQE tolerance is also taken to be quite low, i.e., 10^{-3} in the norm of the parameter vector (whereas the VQE subroutine in the ADAPT procedure is generally lower).

approach taken by the original authors. We find a similar result to previous work, presented in Figure 4. Namely, for stretched LiH, conserving the particle number and projected spin leads to similar qubit-based excitations. The main difference seen (as a result of the relative scale mostly) in these results is for the CNOT cost of the unencoded CQE approach, although a similar decrease in the CNOT cost of the IQEB approach exists as well. While the approaches appear to be inversely related in the rate of convergence (through total iterations), and the CNOT count (where the CQE schemes are more costly), the number of parameter evaluations differs substantially. While here we do not distinguish between the residuals of the ACSE as parameters and the VQE parameters, since these can be obtained many different ways (for instance, the ACSE residuals can be taken from either the 2-RDM with a quantum solver or the 3-RDM on the quantum computer, and

numerous methods of measuring VQE gradients exist as well), for larger system limits on the number of parameters should be considered.

Despite achieving quicker convergence, from all of these examples, we see that the main drawback to the CQE algorithm is the iteratively increasing CNOT cost, which is due to the use of additional gates at each step. While this should be reasonable for high performing quantum devices, for near-term devices, further reduction of the CNOT count is important. However, the gain in performance we see by performing a quasi-second order optimization in the local parameter space is quite significant. In addition, when compared to CNOT gates of the gradient-descent-based methods (as in Figure 1), the optimized-ACSE allows for more flexibility in constructing the compact ansatz.

IV. DISCUSSION

While generic algorithms have been known for approaching the problem of quantum simulation for awhile, calculations involving increasingly complex systems have only recently begun to emerge. These require the advancement and development of new quantum algorithms, similar to the past century of classical quantum chemistry algorithms. The CQE offers an approach that is potentially applicable in the near-term, and provides a strong alternative to VQE.

The present work improves the convergence of the CQE method for quantum simulation by introducing a quasi-second-order, local parametrization of the state space. While existing algorithms for solving the ACSE employ a local, first-order parametrization in that the state is iteratively updated by a series of gradient-based, unitary transformations,[4,5,8] here we present a local, quasi-second-order parametrization in which the state is iteratively updated by a series of unitary transformations that approximately incorporate energy-curvature information. Generally, this allows for more rapid convergence in the vicinity of the solution as well as more flexibility in the construction of CQE algorithms. The current work focusing on the BFGS optimizer[37,38] is an example of potential applications, and also is comparable to published work with other methods, but other optimizers and techniques can be readily implemented.

The largest limitation when translating this work to NISQ devices is the presence of noise. Both sampling and device errors can negatively impact the state. These are present as errors in pure-state N-representability, since the state no longer represents a pure quantum state, or in misapplications of the target operator. Device errors are especially challenging since they corrupt the gate sequences underlying quantum simulation algorithms. Both additional sampling and noisy optimization techniques can be employed to mitigate the effects of noise. Because the CQE is iterative, it can also, in principle, adjust for some errors generated by noise in future iterations. Future work will further examine the effect of noise in the CQE algorithms. As the fidelity of quantum devices increases, the CQE will become more accurate and thereby, more applicable to larger molecular systems.

From the discussion in section II.B, we can see that the variational principle used in the VQE and in the current CQE algorithm are similar in that they solve an optimization problem, but differ in the goal of the minimization. In the VQE, we are often trying to minimize the energy of a state through a global parametrization of some wave function. In addition, barren plateaus or regions where the optimization fail

become likely with an increasingly large parameter space and hence, the suitability (i.e., over or under parametrization) of the state is often in question. As an example, the UCCSD or generalized UCCSD ansatz, which both have approximately $O(r^4)$ parameters, are not sufficient to parametrize the state. Iterative constructions of the ansatz, such as the iterative generalized singles pair-cluster CC,[31] or the ADAPT-VQE, offer fewer parameters initially, but no bounds on the number of required parameters. In the ACSE algorithm, the local parametrization leads to a locally updated optimization, which is constant, with respect to the number of parameters. Importantly, in a VQE, the optimization converges *toward a solution of the VQE problem*, which is not the ACSE. As the VQE subproblem becomes larger and larger (i.e., in adaptive or iterative schemes), eventually the VQE solution can (but by design will not necessarily) satisfy the ACSE. In the CQE approach presented here, we have convergence *toward our contracted eigenvalue problem*, and not a variational subproblem.
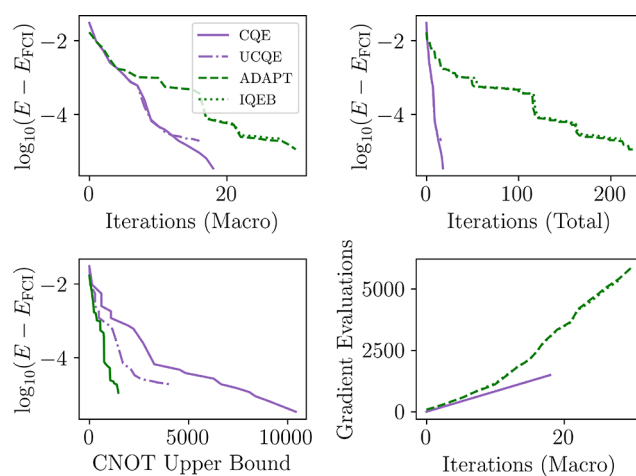
Figure 4. Simulations of molecular lithium hydride at a separation of 2 Å utilizing CQE, VQE, Fermionic ADAPT-VQE and the iterative qubit-excitation based ADAPT-VQE (IQEB), essentially a unencoded APAPT-VQE algorithm. The upper-left shows the macro iterations of each scheme, which are nearly identical for the encoded and unencoded forms, and are only slightly slower for the ADAPT. The upper-right shows the total iterations, including micro iterations in the VQE procedure with a threshold of 10^{-3} in the parameter vector. The lower-left compares the energy convergence with the ansatz cost, showing that the ADAPT does produce a more compact representation of the ansatz. The lower-right shows the number of parameter evaluations (a lower bound on the number of gradients evaluated at each optimizer step), which is linear scaling with the ACSE procedure but for ADAPT becomes quadratic, with respect to the number of iterations.

From our calculations, we can also understand some elements of the ADAPT-VQE algorithm, since it is related to the CQE algorithm. The ADAPT-VQE method chooses the largest ACSE residual at each macro iteration. This leads to a flexible and efficient ansatz, that, when not restarted, will, by construction, improve the energy in the VQE. However, the reoptimized state is often not close to the previous state, highlighting the strong variational nature of the ADAPT algorithm. Restarting the VQE optimization, which has been done in some ADAPT work, can lead to a suboptimal solution of the ACSE, or for the VQE subroutine to fail. Recent work by Liu et al.[48] used a reconstructed 3-RDM in the ACSE to

obtain approximate residuals to seed the ADAPT algorithm. Because these approximate gradients can differ in a substantial way from the exact gradients, more terms are needed, which significantly increases the variational flexibility of the ansatz. As a result, these calculations exhibited faster convergence and required fewer iterations than traditional ADAPT-VQE.

However, unlike Liu et al.,[48] it is important *not* to use ACSE residuals from reconstructed 3-RDMs in the CQE, because these will likely lead to convergence issues. The 3-RDM can still be directly obtained on a quantum computer through a variety of techniques. Work involving qubit-particle approaches also shows promise, with another advantage being the increased number of commuting terms that exists between qubit-particle excitations, as opposed to Fermionic-particle excitations.

In terms of memory storage, the CQE primarily stores the $\mathcal{P}$ vectors of eq 13. The set of $\mathcal{P}$ vectors is composed of l 2-RDMs, or $O(lr^4)$ vectors at most, although, as seen in the preceding discussion, each vector can be taken to be quite sparse. The optimization occurs over the dimension of $\hat{P}$, which here is $O(r^4)$ and not unreasonable for highly accurate electronic structure methods. This is equivalent to the number of parameters in a UCC generalized singles and doubles variational ansatz. However, the gradient evaluations in VQE and the CQE are quite different. For the ACSE, the gradients are evaluated at worst case using the $O(r^6)$ 3-RDM, and which can be measured more efficiently through circuit tomography. In contrast, VQE gradients scale with the number of parameters and the energy evaluation, which, in some instances, is $O(r^8)$. The ADAPT-VQE is difficult to analyze in terms of the number of parameters required, but because each iteration involves an evaluation of the entire pool (which can range from a minimal set to an entire two-body pool), in the standard Fermionic application the total number of gradient evaluations are strictly worse than the CQE.

Another element that can be overlooked is that the ACSE is not necessarily equivalent to the CSE, except when all higher-order excitations are included. Despite this limitation, in the exponential form of the ACSE, higher-order excitations can be seen to emerge naturally through products of exponential two-body operations. In addition, by considering information on the curvature of the space beyond the gradient, we also should include contributions from triple and higher excitations in our selection of operators to propagate the wave function. In practice, the use of the CQE for solving the ACSE leads to a highly accurate solution.

The primary drawback of the CQE when compared to an algorithm such as the ADAPT-VQE is the large number of CNOT gates. Even with low-error CNOT gates, efficient and noise-robust means of obtaining accurate gradients and 2-RDMs will be necessary. All of these also affect the success or failure of the underlying optimization algorithm, and so exploring noise-tolerant approaches will also be critical for near-term applications.

V. CONCLUSION

In this work, we address the convergence of the contracted quantum eigensolver using tools from optimization theory. By using methods beyond traditional gradient descent, we achieve superlinear convergence, allowing us to propagate the wave function rapidly toward a solution of the ACSE. Practical implementations where the search direction is modified to conserve quantum resources show promising reductions in the cost of the algorithm, and we expect further simplification schemes to be attempted aimed at improving the efficiency of the CQE approach. In addition, the present work provides a basis for understanding approaches that use the ACSE in pool selection, and could lead to further hybrid optimization schemes for use in NISQ applications.

■ APPENDIX A. COMPUTATIONAL DETAILS

All calculations were performed using the HQCA (v22.4)[49] set of tools, which utilizes QISKIT (v0.29.0)[50] and PYSCF (v1.7.6)[51] for interfacing with quantum simulators and obtaining electron integrals for circuit-based simulations. Each simulation utilizes a minimal-basis Slater-type orbital representation (i.e., STO-3G). The Jordan–Wigner (or qubit-particle) mapping was utilized, with parity $\mathbb{Z}_2$ symmetries removed for the majority of examples. State vector or unitary simulations with no noise were used for each run. The ADAPT-VQE results in Figure 3 were obtained using code from the respective publication,[34] where analytical recursive solutions of the VQE gradients are used. The threshold for the VQE subroutine in those instances was 10^{-6}. In the lithium hydride case, code from HQCA was used. Parameters for the VQE optimization in the ADAPT-VQE optimizations were not reset between runs—a single parameter is essentially appended to the parameter vector.

The line-search implementation used in the BFGS, nonlinear CG, and l-BFGS optimizations follows from the Nocedal algorithms,[38] which is present in the scipy implementation. For some steps (notably the first few steps), often $\alpha_0 = 1$ is too large, and $\alpha_0 = 0.5$ (or a quadratic step) is preferred. While a dynamic step-size is not necessary for BFGS, after the first step, we interpolate the α_0 with a quadratic based on the current energy, previous energy, and previous gradient information, and then constrain $\alpha_0 \in [0.5, 1]$, which we expect, in some instances, leads to a more appropriate step size.

■ AUTHOR INFORMATION

Corresponding Author

    David A. Mazziotti − *Department of Chemistry and The James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States;* ⊙ orcid.org/0000-0002-9938-3886; Email: damazz@uchicago.edu

Author

    Scott E. Smart − *Department of Chemistry and The James Franck Institute, The University of Chicago, Chicago, Illinois 60637, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.2c00446

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Mazziotti, D. A. Contracted Schrödinger equation: Determining quantum energies and two-particle density matrices without wave functions. *Phys. Rev. A* **1998**, *57*, 4219−4234.

(2) Alcoba, D. R.; Valdemoro, C. Family of modified-contracted Schrödinger equations. *Phys. Rev. A* **2001**, *64*, 7.

(3) Nakatsuji, H. Equation for the direct determination of the density matrix. *Phys. Rev. A* **1976**, *14*, 41−50.

(4) Mazziotti, D. A. Anti-Hermitian contracted Schrödinger equation: Direct determination of the two-electron reduced density matrices of many-electron molecules. *Phys. Rev. Lett.* **2006**, *97*, 143002.

(5) Mazziotti, D. A. Anti-Hermitian part of the contracted Schrödinger equation for the direct calculation of two-electron reduced density matrices. *Phys. Rev. A* **2007**, *75*, 1−12.

(6) Mukherjee, D.; Kutzelnigg, W. Irreducible Brillouin conditions and contracted Schrödinger equations for n -electron systems. I. The equations satisfied by the density cumulants. *J. Chem. Phys.* **2001**, *114*, 2047−2061.

(7) Mazziotti, D. A. Approximate solution for electron correlation through the use of Schwinger probes. *Chem. Phys. Lett.* **1998**, *289*, 419−427.

(8) Sand, A. M.; Mazziotti, D. A. Enhanced computational efficiency in the direct determination of the two-electron reduced density matrix from the anti-Hermitian contracted Schrödinger equation with application to ground and excited states of conjugated π-systems. *J. Chem. Phys.* **2015**, *143*, 134110.

(9) Mazziotti, D. A. Exactness of wave functions from two-body exponential transformations in many-body quantum theory. *Phys. Rev. A* **2004**, *69*, 012507.

(10) Smart, S. E.; Mazziotti, D. A. Quantum Solver of Contracted Eigenvalue Equations for Scalable Molecular Simulations on Quantum Computing Devices. *Phys. Rev. Lett.* **2021**, *126*, 070504.

(11) Mazziotti, D. A. Multireference many-electron correlation energies from two-electron reduced density matrices computed by solving the anti-Hermitian contracted Schrödinger equation. *Phys. Rev. A* **2007**, *76*, 052502.

(12) Alcoba, D. R.; Valdemoro, C.; Tel, L. M.; Pérez-Romero, E. The correlation contracted Schrödinger equation: An accurate solution of the G -particle-hole hypervirial. *Int. J. Quantum Chem.* **2009**, *109*, 3178−3190.

(13) Gidofalvi, G.; Mazziotti, D. A. Direct calculation of excited-state electronic energies and two-electron reduced density matrices from the anti-Hermitian contracted Schrödinger equation. *Phys. Rev. A* **2009**, *80*, 1−11.

(14) Foley, J. J.; Rothman, A. E.; Mazziotti, D. A. Strongly correlated mechanisms of a photoexcited radical reaction from the anti-Hermitian contracted Schrdinger equation. *J. Chem. Phys.* **2011**, *134*, 034111.

(15) Valdemoro, C.; Alcoba, D. R.; Tel, L. M.; Pérez-Romero, E. Theoretical and applicative properties of the correlation and G -particle-hole matrices. *Int. J. Quantum Chem.* **2009**, *109*, 2622−2638.

(16) Smart, S. E.; Boyn, J.-N.; Mazziotti, D. A. Resolving correlated states of benzyne with an error-mitigated contracted quantum eigensolver. *Phys. Rev. A* **2022**, *105*, 022405.

(17) Boyn, J.-N.; Lykhin, A. O.; Smart, S. E.; Gagliardi, L.; Mazziotti, D. A. Quantum-classical hybrid algorithm for the simulation of all-electron correlation. *J. Chem. Phys.* **2021**, *155*, 244106.

(18) Taylor, C. J.; Kriegman, D. J. *Minimization on the Lie Group and Related Manifolds Minimization on the Lie Group SO (3) and Related Manifolds.* Technical Report No. 9405, 1994, 10.

(19) Absil, P. A.; Mahony, R.; Sepulchre, R. *Optimization Algorithms on Matrix Manifolds*; Princeton University Press, 2007.

(20) Huang, W.; Gallivan, K. A.; Absil, P.-A. A Broyden Class of Quasi-Newton Methods for Riemannian Optimization. *SIAM J. Optimiz.* **2015**, *25*, 1660−1685.

(21) Arute, F.; Arya, K.; Babbush, R.; Bacon, D.; Bardin, J. C.; Barends, R.; Boixo, S.; Broughton, M.; Buckley, B. B.; Buell, D. A. e. a.; et al. Hartree-Fock on a superconducting qubit quantum computer. *Science* **2020**, *369*, 1084−1089.

(22) Shepard, R.; Brozell, S. R.; Gidofalvi, G. The Representation and Parametrization of Orthogonal Matrices. *J. Phys. Chem. A* **2015**, *119*, 7924−7939.

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

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

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

(26) Endo, S.; Cai, Z.; Benjamin, S. C.; Yuan, X. Hybrid Quantum-Classical Algorithms and Quantum Error Mitigation. *J. Phys. Soc. Jpn.* **2021**, *90*, 032001.

(27) Stuelpnagel, J. On the Parametrization of the Three-Dimensional Rotation Group. *SIAM Rev.* **1964**, *6*, 422−430.

(28) Evangelista, F. A.; Chan, G. K.-L.; Scuseria, G. E. Exact parameterization of fermionic wave functions via unitary coupled cluster theory. *J. Chem. Phys.* **2019**, *151*, 244112.

(29) McClean, J. R.; Boixo, S.; Smelyanskiy, V. N.; Babbush, R.; Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **2018**, *9*, 1−6.

(30) Wang, S.; Fontana, E.; Cerezo, M.; Sharma, K.; Sone, A.; Cincio, L.; Coles, P. J. Noise-Induced Barren Plateaus in Variational Quantum Algorithms. *arXiv* **2020**, 1−23.

(31) Lee, J.; Huggins, W. J.; Head-Gordon, M.; Whaley, K. B. Generalized Unitary Coupled Cluster Wave functions for Quantum Computation. *Chem. Theory Comput.* **2019**, *15*, 311−324.

(32) Chen, J.; Cheng, H.-P.; Freericks, J. K. Low-Depth Unitary Coupled Cluster Theory for Quantum Computation. *J. Chem. Theory Comput.* **2022**, *18*, 2193−2198.

(33) Ryabinkin, I. G.; Yen, T. C.; Genin, S. N.; Izmaylov, A. F. Qubit Coupled Cluster Method: A Systematic Approach to Quantum Chemistry on a Quantum Computer. *J. Chem. Theory Comput.* **2018**, *14*, 6317−6326.

(34) Grimsley, H. R.; Economou, S. E.; Barnes, E.; Mayhall, N. J. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nat. Commun.* **2019**, *10*, 3007.

(35) Mazziotti, D. A. Exact two-body expansion of the many-particle wave function. *Phys. Rev. A* **2020**, *102*, 030802.

(36) Smith, S. T. *Geometric Optimization Methods for Adaptive Filtering*, 1993.

(37) Davidon, W. C. Variable Metric Method for Minimization. *SIAM J. Opt.* **1991**, *1*, 1−17.

(38) Nocedal, J.; Wright, S. J. Numerical Optimization. In *Springer Series in Operations Research and Financial Engineering*; Springer: New York, 2006.

(39) Hager, W. W. W.; Zhang, H. A Survey of Nonlinear Conjugate Gradient Methods. *Pac. J. Opt.* **2006**, *2*, 35−58.

(40) Tranter, A.; Love, P. J.; Mintert, F.; Wiebe, N.; Coveney, P. V. Ordering of Trotterization: Impact on Errors in Quantum Simulation of Electronic Structure. *Entropy* **2019**, *21*, 1218.

(41) Motta, M.; Ye, E.; McClean, J. R.; Li, Z.; Minnich, A. J.; Babbush, R.; Chan, G. K. L. Low rank representations for quantum simulation of electronic structure. *npj Quantum Inf.* **2021**, *7*, 83.

(42) Rubin, N. C.; Lee, J.; Babbush, R. Compressing Many-Body Fermion Operators under Unitary Constraints. *J. Chem. Theory Comput.* **2022**, *18*, 1480−1488.

(43) Xia, R.; Kais, S. Qubit coupled cluster singles and doubles variational quantum eigensolver ansatz for electronic structure calculations. *Quant. Sci. Technol.* **2021**, *6*, 015001.

(44) Tang, H. L.; Shkolnikov, V.; Barron, G. S.; Grimsley, H. R.; Mayhall, N. J.; Barnes, E.; Economou, S. E. Qubit-ADAPT-VQE: An Adaptive Algorithm for Constructing Hardware-Efficient Ansätze on a Quantum Processor. *PRX Quantum* **2021**, *2*, 020310.

(45) Yordanov, Y. S.; Armaos, V.; Barnes, C. H. W.; Arvidsson-Shukur, D. R. M. Qubit-excitation-based adaptive variational quantum eigensolver. *Commun. Phys.* **2021**, *4*, 228.

(46) Mazziotti, D. A.; Smart, S. E.; Mazziotti, A. R. Quantum simulation of molecules without fermionic encoding of the wave function. *New J. Phys.* **2021**, *23*, 113037.

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

(48) Liu, J.; Li, Z.; Yang, J. An efficient adaptive variational quantum solver of the Schrödinger equation based on reduced density matrices. *J. Chem. Phys.* **2021**, *154*, 244112.

(49) Smart, S. E.; Mazziotti, D. A. *hqca - hybrid quantum computing algorithms for quantum chemistry*. Available via the Internet at: https://github.com/damazz/HQCA (accessed June 8, 2022).

(50) Qiskit:*An Open-Source Framework for Quantum Computing*. 2021.

(51) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework. *WIREs Comput. Mol. Sci.* **2018**, *8*, 1−35.