The Journal of Chemical Physics

Approaching the full configuration interaction ground state from an arbitrary wavefunction with gradient descent and quasi-Newton algorithms

Cite as: J. Chem. Phys. **153**, 234113 (2020); https://doi.org/10.1063/5.0031958
Submitted: 05 October 2020 . Accepted: 29 November 2020 . Published Online: 18 December 2020

Carlos A. Jiménez-Hoyos

View Online Export Citation CrossMark

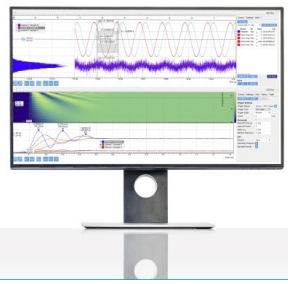
ARTICLES YOU MAY BE INTERESTED IN

Index of multi-determinantal and multi-reference character in coupled-cluster theory
The Journal of Chemical Physics **153**, 234103 (2020); https://doi.org/10.1063/5.0029339

A general implementation of time-dependent vibrational coupled-cluster theory
The Journal of Chemical Physics **153**, 234109 (2020); https://doi.org/10.1063/5.0034013

A route to improving RPA excitation energies through its connection to equation-of-motion coupled cluster theory
The Journal of Chemical Physics **153**, 234101 (2020); https://doi.org/10.1063/5.0023862

Challenge us.
What are your needs for periodic signal detection?
Watch



Zurich Instruments

© 2020 Author(s).

AIP Publishing

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Approaching the full configuration interaction ground state from an arbitrary wavefunction with gradient descent and quasi-Newton algorithms

Cite as: J. Chem. Phys. 153, 234113 (2020); doi: 10.1063/5.0031958
Submitted: 5 October 2020 • Accepted: 29 November 2020 •
Published Online: 18 December 2020

View Online Export Citation CrossMark

Carlos A. Jiménez-Hoyos[a)]

AFFILIATIONS

Department of Chemistry, Wesleyan University, Middletown, Connecticut 06459, USA

[a)]Author to whom correspondence should be addressed: cjimenezhoyo@wesleyan.edu

ABSTRACT

We consider gradient descent and quasi-Newton algorithms to optimize the full configuration interaction (FCI) ground state wavefunction starting from an arbitrary reference state $|0\rangle$. We show that the energies obtained along the optimization path can be evaluated in terms of expectation values of $|0\rangle$, thus avoiding explicit storage of intermediate wavefunctions. This allows us to find the energies after the first few steps of the FCI algorithm for systems much larger than what standard deterministic FCI codes can handle at present. We show an application of the algorithm with reference wavefunctions constructed as linear combinations of non-orthogonal determinants.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0031958

I. INTRODUCTION

In principle, all the chemical properties of molecular systems can be determined from the knowledge of the eigenstates of the Hamiltonian operator. The full-configuration interaction (FCI) wavefunction constitutes a prominent paradigm of quantum chemistry:[1] it provides the exact solution of the Schrödinger equation on the basis of suitably selected N-electron wavefunctions. While finding the FCI ground state wavefunction is algorithmically easy, the number of degrees of freedom in it increases exponentially with the system size. Therefore, one of the central goals of the quantum chemistry community is to develop approximations that can yield properties of similar quality as that from the FCI ground state while reducing significantly the computational cost entailed.

The optimization of the FCI ground state wavefunction is usually cast as an eigenvalue problem. Given a matrix representation of the Hamiltonian operator H and a vector representation of a starting reference wavefunction $|0\rangle$, potent algorithms (with that by Davidson[2] being a prominent example) have been developed to optimize the FCI ground state wavefunction $|\Psi_{\mathrm{FCI}}\rangle$ while minimizing the number of matrix-vector operations and the number of intermediate wavefunctions. Nevertheless, the dimension of the FCI problem renders this approach possible only for relatively small systems.

In this work, we obtain the FCI ground state from an unconstrained optimization of the wavefunction amplitudes in the orthogonal complement of some reference state $|0\rangle$. We carry the unconstrained optimization with gradient descent (GD) and quasi-Newton algorithms. In this way, the reference wavefunction is improved systematically with increasing number of steps. The central result obtained in this work is that the energies obtained along the optimization path can be fully expressed in terms of matrix elements of the reference $|0\rangle$. Explicitly casting the energy after some finite number of steps as a functional of the reference wavefunction allows us to consider systems for which the dimension of the Hilbert space is much larger than the disk or memory available in current computational facilities.

There is another important reason to avoid the explicit vector representation of the FCI wavefunction, which actually constitutes the motivation for this manuscript. Namely, it allows us to consider classes of wavefunctions for which a vector representation in some convenient orthonormal N-electron basis is not available,

and the computational cost of evaluating it would be proportional to the dimension of the Hilbert space. A specific example that we shall consider are reference states written as a linear combination of a few determinants that are, in general, non-orthogonal (i.e., $\langle\Phi_1|\Phi_2\rangle \neq 0$),

$$|0\rangle = \sum_q f_q|\Phi_q\rangle.$$

In that case, the orthogonal complement is not known *a priori*, and the vector representation in the basis of orthonormal Slater determinants can only be determined at a great computational expense. This class of wavefunctions occurs in non-orthogonal configuration interaction (NOCI)[3] as well as in symmetry-projected Hartree–Fock methods.[4]

The rest of this manuscript is organized as follows: In Sec. II, we describe how we parametrize the FCI ground state wavefunction and describe the gradient descent and quasi-Newton optimization algorithms considered. In Sec. III, we discuss the application of the method in an H_4 ring and in the determination of spectroscopic constants of the N_2, O_2, and F_2 molecules. Finally, in Sec. IV, we provide some closing remarks.

II. THEORY

We parametrize the FCI ground state $|\Psi_{\text{FCI}}\rangle$ using an exponential non-Hermitian ansatz of the form

$$|\Psi\rangle = \exp(\hat{Z})|0\rangle, \tag{1}$$

$$\hat{Z} = \sum_x Z_x|x\rangle\langle 0|, \tag{2}$$

where $|0\rangle$ is the reference wavefunction and $|x\rangle$ labels an orthonormal state in the orthogonal complement of $|0\rangle$. The FCI ground state can be reached after an unconstrained optimization in the parameters Z. (We note that the same parametrization was used in Ref. 5, also casting FCI as an unconstrained minimization problem.) An important ingredient of this work is that an explicit construction of the states $\{|x\rangle\}$ is not necessary. We assume in what follows that $\langle 0|\Psi_{\text{FCI}}\rangle \neq 0$.

Given Z (as a vector of amplitudes), the energy of $|\Psi\rangle$ can be evaluated as

$$E[Z] \equiv \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle} = \frac{\langle 0|\exp(\hat{Z}^\dagger)H\exp(\hat{Z})|0\rangle}{\langle 0|\exp(\hat{Z}^\dagger)\exp(\hat{Z})|0\rangle} = \frac{H_0^0 + H_0^x Z_x + Z^x H_x^0 + Z^x H_x^y Z_y}{1 + Z^x Z_x}, \tag{3}$$

where $H_\alpha^\beta = \langle\alpha|H|\beta\rangle$, Einstein summation is implied, and the indices x and y run only over the orthogonal complement of $|0\rangle$. Here, we have used the fact that $\hat{Z}^2|0\rangle = 0$. We have also assumed a real wavefunction $|0\rangle$ and real coefficients Z, as we do throughout this work.

The energy gradient with respect to the amplitudes in Z evaluated at $Z = Y$ is given by

$$g_x \equiv \left.\frac{\partial E[Z]}{\partial Z^x}\right|_{Z=Y} = \frac{2H_x^0 + 2H_x^y Y_y - 2E[Y]Y_x}{1 + Y^y Y_y}. \tag{4}$$

A necessary and sufficient condition for $|\Psi\rangle$ to be an eigenstate of H is that the gradient g vanishes in all directions. This is equivalent to the fulfillment of all k-particle Brillouin conditions.[6]

Along the optimization paths described in Subsections II A and II B, a line search is needed to minimize the energy of $E[Y + sX]$, with Y being the vector of current amplitudes and X the search direction, with respect to the step size s. Explicitly, $E[Y + sX]$ becomes

$$E[Y+sX] = \frac{E[Y](1 + Y^x Y_x) + s(H_0^x + Y^y H_y^x)X_x + sX^x(H_x^0 + H_x^y Y_y) + s^2 X^x H_x^y X_y}{(1 + Y^x Y_x) + sY^x X_x + sX^x Y_x + s^2 X^x X_x}. \tag{5}$$

This is a rational equation in s of the form

$$\frac{c + bs + as^2}{f + es + ds^2}.$$

Let s_* be the value of s that extremizes $E[Y + sX]$; this takes the form

$$s_* = \frac{(cd - af) \pm \sqrt{(cd - af)^2 - (ae - bd)(bf - ce)}}{(ae - bd)}.$$

For convenience, we shall introduce the quantities $f_1 = H_0^0$, $f_2 = H_0^y H_y^0$, $f_3 = H_0^y H_y^z H_z^0$, etc. Note that all those matrix elements can be evaluated in terms of expectation values from the reference wavefunction $|0\rangle$. For instance,

$$f_2 = \langle 0|H^2|0\rangle - f_1\langle 0|H|0\rangle, \tag{6}$$

$$f_3 = \langle 0|H^3|0\rangle - f_1\langle 0|H^2|0\rangle - f_2\langle 0|H|0\rangle. \tag{7}$$

A. Gradient descent

We begin at $Z_0 = 0$ with $|\Psi_0\rangle = |0\rangle$. The gradient at Z_0 is

$$(g_0)_x = \alpha_0 H_x^0, \tag{8}$$

with $\alpha_0 = 2$.

In a standard gradient descent implementation, a line search would be performed along $-g_0$. It is common practice to accept a step size that satisfies Wolfe[7,8] conditions to avoid the potentially expensive full line search. In our case, we aim to perform a full line search, as we may only be able to afford a few steps. The optimal step size σ_* can be easily found given the rational form of $E[Z]$ discussed above. Therefore, $Z_1 = -\sigma_* g_0$ and $|\Psi_1\rangle$ become

$$|\Psi_1\rangle = e^{-\sigma_* g_0}|0\rangle. \tag{9}$$

Published under license by AIP Publishing

An explicit expression for E_1 is given by

$$E_1 \equiv \frac{\langle\Psi_1|H|\Psi_1\rangle}{\langle\Psi_1|\Psi_1\rangle} = \frac{f_1 - 2\sigma_*\alpha_0 f_2 + \sigma_*^2\alpha_0^2 f_3}{1 + \sigma_*^2\alpha_0^2 f_2}, \tag{10}$$

with

$$\sigma_* = \frac{(4f_3 - 4f_1f_2) \pm \sqrt{64f_2^3 + (4f_3 - 4f_1f_2)^2}}{-16f_2^2}. \tag{11}$$

Note that E_1 is a functional of $|0\rangle$, which can be determined after the evaluation of f_1, f_2, and f_3. (This is an interesting functional in itself. One may consider optimizing E_1 with respect to the reference wavefunction parameters $|0\rangle$, which we have done for one of the examples considered.)

We now proceed to another step. The gradient at $Z_1 = -\sigma_* g_0$ is given by

$$(g_1)_x = \alpha_1 H_x^0 + \beta_1 H_x^y H_y^0, \tag{12}$$

with

$$\alpha_1 = \frac{2(1 + \sigma_*\alpha_0 E_1)}{1 + \sigma_*^2\alpha_0^2 f_2}, \tag{13}$$

$$\beta_1 = \frac{-2\sigma_*\alpha_0}{1 + \sigma_*^2\alpha_0^2 f_2}. \tag{14}$$

After looking for the optimal step size τ_* (i.e., minimizing $E[-\sigma_* g_0 - \tau g_1]$ with respect to τ), $|\Psi_2\rangle$ becomes

$$|\Psi_2\rangle = e^{-\sigma_* g_0 - \tau_* g_1}|0\rangle. \tag{15}$$

An explicit expression for E_2 is given as follows:

$$E_2 \equiv \frac{\langle\Psi_2|H|\Psi_2\rangle}{\langle\Psi_2|\Psi_2\rangle} = \frac{E_1(1 + \sigma_*^2\alpha_0^2 f_2) - 2\tau_*(\alpha_1 f_2 + \beta_1 f_3) + 2\sigma_*\tau_*\alpha_0(\alpha_1 f_3 + \beta_1 f_4) + \tau_*^2(\alpha_1^2 f_3 + 2\alpha_1\beta_1 f_4 + \beta_1^2 f_5)}{(1 + \sigma_*^2\alpha_0^2 f_2) + 2\sigma_*\tau_*\alpha_0(\alpha_1 f_2 + \beta_1 f_3) + \tau_*^2(\alpha_1^2 f_2 + 2\alpha_1\beta_1 f_3 + \beta_1^2 f_4)}. \tag{16}$$

A closed-form expression for τ_* can be deduced from the rational form of E_2 as a function of τ, as described before. E_2 is also a functional of $|0\rangle$ that can be assembled after the evaluation of $f_1, \ldots, f_5$.

A third step would require the evaluation of g_2. Given the form of the gradient [Eq. (4)], it can be shown that g_2 takes the form

$$(g_2)_x = \alpha_2 H_x^0 + \beta_2 H_x^y H_y^0 + \gamma_2 H_x^y H_y^z H_z^0. \tag{17}$$

Therefore, E_3 would be also a functional of $|0\rangle$ that can be determined after the evaluation of $f_1, \ldots, f_7$. Subsequent steps require the evaluation of higher order f_k values. It is a common practice in numerical optimization to use the norm of the gradient as a convergence criterion. While in this work, we have suggested that the explicit construction and storage of the gradient can be avoided, we note that the two-norm of the gradient is accessible from matrix elements of $|0\rangle$. For instance, at $Z_0 = 0$, with the gradient given by Eq. (8), it follows that $|g_0|^2 = \alpha_0^2 f_2$.

At this point, we note that the gradient descent approach described above can, in principle, be used to converge to the true eigenfunction of the Hamiltonian $\mathcal{H}$ as opposed to its representation in some finite N-electron basis H. It would involve replacing the expectation values $\langle 0|H^k|0\rangle$ with $\langle 0|\mathcal{H}^k|0\rangle$, with everything else holding. This, however, requires further work given that individual integrals appearing in $\mathcal{H}^2$ (also higher powers) diverge when evaluated with atomic Gaussian basis functions (see Ref. 9).

As a second remark, we note that the evaluation of other expectation values can be done in the same way as the energy. For instance, the expectation value of W, after the first gradient descent step, is given by

$$X_1 \equiv \frac{\langle\Psi_1|W|\Psi_1\rangle}{\langle\Psi_1|\Psi_1\rangle} = \frac{f_1^w - 2\sigma_*\alpha_0 f_2^w + \sigma_*^2\alpha_0^2 f_3^w}{1 + \sigma_*^2\alpha_0^2 f_2}, \tag{18}$$

with $f_1^w = W_0^0$, $f_2^w = \frac{1}{2}(W_0^y H_y^0 + H_0^y W_y^0)$, and $f_3^w = H_0^y W_y^z H_z^0$. This is also a functional of $|0\rangle$ since

$$f_2^w = \tfrac{1}{2}\langle 0|WH + HW|0\rangle - f_1\langle 0|W|0\rangle, \tag{19}$$

$$f_3^w = \langle 0|HWH|0\rangle - \tfrac{1}{2}f_1\langle 0|WH + HW|0\rangle - f_2^w\langle 0|H|0\rangle. \tag{20}$$

B. Quasi-Newton

Because the convergence of the gradient descent algorithm is slow, we now consider a quasi-Newton optimization. In a quasi-Newton approach,[10] it is customary to start from an initial inverse Hessian B_0, which needs only to be an approximation to the true Hessian. At each step along the optimization, the search direction p_k is determined from $B_k p_k = -g_k$, rather than setting $p_k = -g_k$ as in the gradient descent approach. In this work, we choose to set $B_0 = I$ to keep things as simple as possible: under such conditions, a quasi-Newton algorithm can still be fully defined in terms of $|0\rangle$. Naturally, if a better B_0 is used, the number of iterations required to reach convergence is expected to decrease.

Just as in the gradient descent approach, after defining a search direction p_k, a line search is performed along it. Commonly, a step size that satisfies Wolfe conditions is accepted. We choose, however, to carry a full line search as in the gradient descent algorithm.

Because $B_0 = I$, it follows that the first step coincides with that from the gradient descent approach, and E_1 remains unchanged. The gradient g_1 is also the same as in gradient descent [see Eq. (12)]. The direction p_1 is determined from $p_1 = -B_1 g_1$, with B_1 constructed using a quasi-Newton update formula.

We show in Appendix A that p_1, determined from a Broyden–Fletcher–Goldfarb–Shanno (BFGS)[11–14] update formula, takes the

Published under license by AIP Publishing

form

$$(p_1)_x = \alpha_1' H_x^0 + \beta_1' H_x^y H_y^0, \tag{21}$$

with α_1' and β_1' being some constants that are numerically different than α_1 and β_1. Given that p_1 takes the same functional form as g_1, we conclude that E_2 determined from the BFGS approach is also a functional of $f_1, \ldots, f_5$. Specifically, E_2 would take the same form as Eq. (16), with $\alpha_1 \to \alpha_1'$ and $\beta_1 \to \beta_1'$. Further quasi-Newton steps can also be cast as functionals of $|0\rangle$, in the same way as in the gradient descent algorithm.

III. RESULTS AND DISCUSSION

We proceed to discuss the application of the optimization algorithms described above in a H_4 ring and in the N_2, O_2, and F_2 diatomics.

A. H_4 ring

We consider a system of 4 H atoms placed along a ring of radius $r = 3.3$ bohrs,[15] with the arrangement depicted in Fig. 1. For $\theta \approx 24°$, the system consists of two weakly interacting H_2 molecules near their equilibrium geometry. Conversely, for $\theta = 90°$, the 4 H atoms form a square and the system has a strong multireference character. In scanning θ in the range between 20° and 90°, the system evolves from a weak to a strong correlation regime. Our calculations in this system were performed by explicitly constructing the vector representation of the considered wavefunctions; this allows us to study the convergence behavior of the gradient descent and quasi-Newton algorithms.

We show in Fig. 2 (left panel) the energy of the H_4 system as a function of θ, evaluated with the 6-31G basis set. In particular, we show restricted and unrestricted Hartree–Fock (RHF and UHF), spin-projected UHF (SUHF), and FCI ground state energies. In addition, we show the energy after the first and second gradient descent (GD) steps as well as the energy after the second quasi-Newton (QN) step starting from the RHF wavefunction. In the right panel, we show the errors with respect to the FCI ground state calculated with each method, where we now show also the corrections starting from UHF and SUHF. The unusual profile displayed by the UHF curves is associated with the collapse back to the RHF wavefunction at $\theta \approx 40°$.

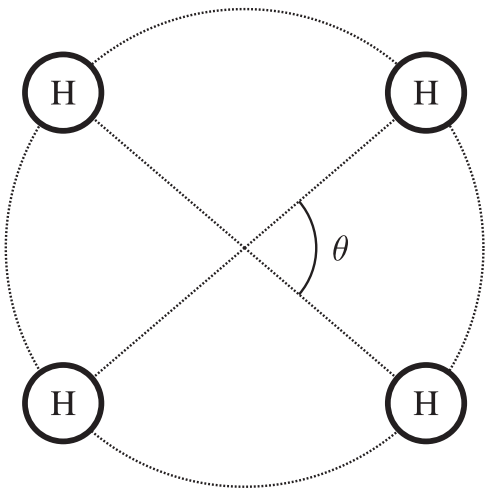


FIG. 1. The H_4 system described in Ref. 15 consists of four H atoms placed along a ring of radius $r = 3.3$ bohrs controlled by an angle θ. At $\theta \approx 24°$, the system corresponds to two weakly interacting H_2 units near their equilibrium geometry.

The improvement upon the RHF energy obtained with the first GD step is quite significant: the error is reduced by nearly a third at $\theta = 90°$ and by an order of magnitude in the equilibrium region. The second step (either GD or QN) reduces the energy even further, with the QN step being considerably better than the GD step. In the case of UHF and SUHF, the improvement after the first and second steps is very modest near $\theta = 90°$, although the energy lowering near the equilibrium region is still substantial. Around $\theta \approx 50°$, the first step correction to RHF leads to a lower energy than the first step correction to UHF even when the reference wavefunction itself is lower in energy. As more steps are taken, the expectation is that the QN algorithm will outperform the GD algorithm as the former converges linearly, while the latter should approach quadratic convergence.[10]

We show in Fig. 3 the convergence profile of the GD and QN algorithms on top of the RHF, UHF, and SUHF wavefunctions at $\theta = 24°$ (top) and $\theta = 90°$ (bottom). At $\theta = 24°$, convergence is relatively fast: with GD, 10 steps are sufficient to converge the energy to 10^{-5} hartree, while 5 steps are enough with the QN algorithm. At $\theta = 90°$, the profiles are very different. While convergence starting from the SUHF reference wavefunction is quite fast, a starting RHF or UHF wavefunction leads to a much slower convergence. In particular, the GD algorithm from UHF converges extremely slowly. The QN algorithms do recover a faster convergence rate after 15 or so iterations.

B. N_2, O_2, and F_2

We now consider the dissociation profile of N_2 with a cc-pVDZ basis set. Reference FCI (with a frozen-core approximation) results for this system are available from Ref. 16. In Fig. 4, we show the profiles obtained with RHF, UHF, and SUHF, as well as the corresponding profiles after a single GD step is performed (RHF+1, UHF+1, and SUHF+1). Complementary to the RHF and UHF results, we show configuration interaction singles and doubles (CISD) results starting from RHF and UHF, calculated using Gaussian 16.[17] We note that it can be shown that the RHF+1 and UHF+1 energies should be above the corresponding RCISD and UCISD ones, as the latter involve optimization of the amplitudes in the orthogonal complement.

We emphasize that, in this case, we have produced the energies after a single GD step without a formal vector representation of the starting reference wavefunctions (in a basis of Slater determinants, the dimension of the FCI vector is $\approx 1.4 \times 10^{12}$). In evaluating the energy after the single GD step, we computed $\langle 0|H^3|0\rangle$ for a reference wavefunction that is either a single Slater determinant or a linear combination of non-orthogonal determinants. In both cases, the evaluation can be completed, by making use of Wick's theorem, with a computational effort of $\mathcal{O}(N^3M^3)$ (with N being the number of electrons and M being the number of virtual orbitals). For this system, we do not show results after a second step as we do not currently have a code that can evaluate up to $\langle 0|H^5|0\rangle$.

Published under license by AIP Publishing

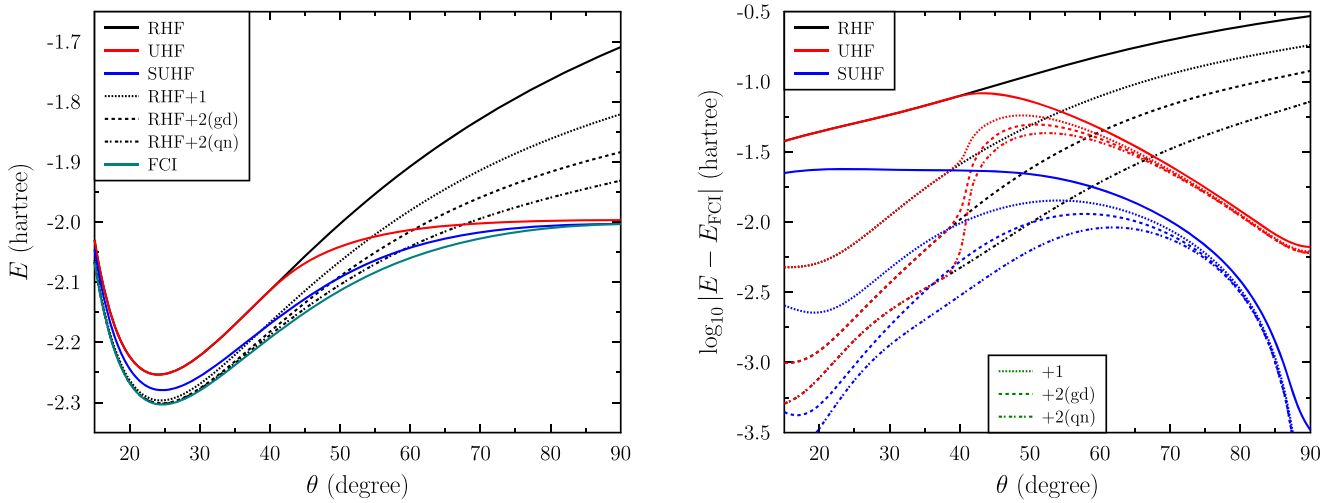


FIG. 2. Left panel: energy (in hartree) of the H_4 system of Fig. 1, as a function of θ, computed with various methods and the 6-31G basis set. We show the energies after 1 and 2 gradient descent (GD) steps, as well as the energy after 2 quasi-Newton (QN) steps. Right panel: error in the energy, with respect to the FCI ground state, after 1 and 2 GD or QN steps starting from RHF, UHF, and SUHF wavefunctions.

In the case of RHF, a single GD step recovers ≈50% of the correlation energy near equilibrium. Starting from the SUHF reference wavefunction, we again recover ≈50% of the missing energy after a single GD step. As expected, RCISD and UCISD energies are significantly lower than RHF+1 and UHF+1, although the curves are quite parallel to each other.

As we have discussed above, E_1 [Eq. (10)] is explicitly a functional of $|0\rangle$. We can therefore choose to optimize the parameters in $|0\rangle$ to minimize E_1, instead of E_0. We have done so with a UHF functional form, as shown in Fig. 4 under the label UHF+1 (opt). We note that, in setting up the optimization process, we evaluate the derivative of E_1 analytically, as explained in Appendix B. The gradient evaluation can still be carried out with a computational effort of $\mathcal{O}(N^3M^3)$, i.e., the same scaling as for the energy. While in this case, we only see a modest improvement associated with the minimization of E_1 directly, it can still be a valuable approach as this can

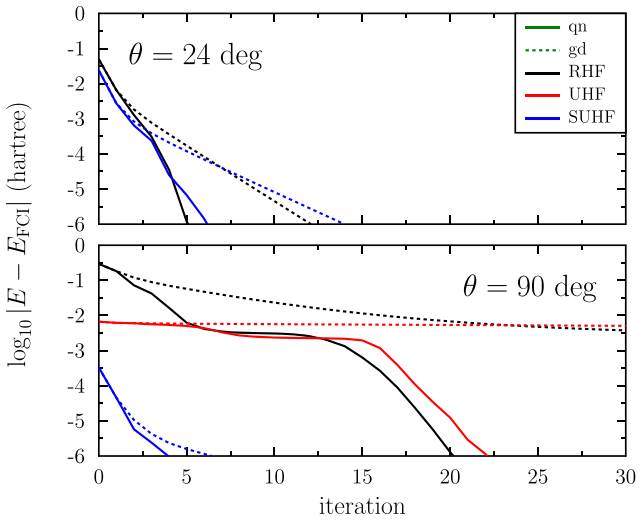


FIG. 3. Convergence of the energy, at θ = 24° (top panel) and θ = 90° (bottom panel), as a function of iteration, using the gradient descent (GD) and quasi-Newton (QN) algorithms starting from the RHF, UHF, and SUHF reference wavefunctions.

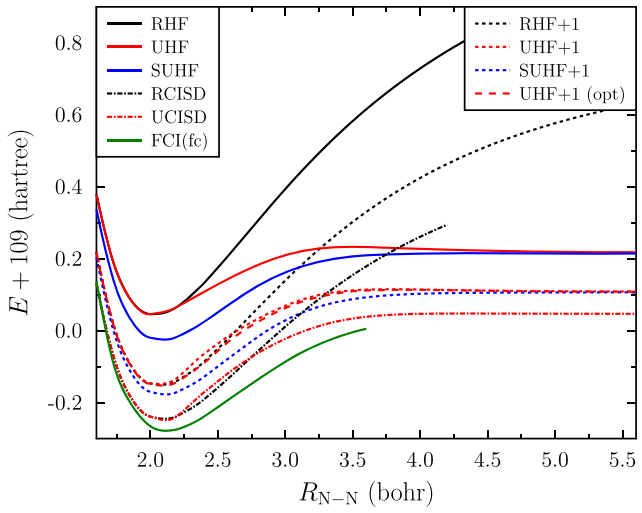


FIG. 4. Dissociation profile of the N_2 molecule evaluated with various methods and the cc-pVDZ basis set. RHF+1 indicates the energy after a single GD step starting from the RHF wavefunction. UHF+1 (opt) indicates that the functional E_1 of Eq. (10) is minimized using a trial wavefunction of UHF form. The FCI results, using the frozen-core approximation, are from Ref. 16.

Published under license by AIP Publishing

TABLE I. Spectroscopic constants (E_e in hartree, D_e in kcal/mol, r_e in Å, and ω_e in cm^{-1}) of N_2, O_2, and F_2 computed with various methods using the cc-pCVTZ basis set.

	N_2				O_2				F_2			
Method	E_e	D_e	r_e	ω_e	E_e	D_e	r_e	ω_e	E_e	D_e	r_e	ω_e
HF	−108.987 698	116.7	1.0660	2727.7								
HF+1	−109.081 335	124.2	1.0642	2719.9								
SUHF	−109.057 362	158.2	1.0909	2417.8	−149.675 867	30.8	1.1566	1980.4	−198.851 568	14.1	1.4803	663.9
SUHF+1	−109.146 989	163.2	1.0880	2435.5	−149.777 920	36.0	1.1541	1971.8	−198.961 698	15.1	1.4709	676.6
IC-MRCI[a]	−109.464 001	220.0	1.1007	2350.7	−150.209 357	112.6	1.2099	1575.0	−199.376 661	32.3	1.4165	893.3
Expt.[b]		228.4	1.0977	2358.6		120.6	1.2075	1580.2		39.0	1.4119	916.6

[a] All-electron internally contracted multi-reference configuration interaction results from Ref. 18.
[b] Experimental results as compiled in Ref. 18.

simplify the evaluation of molecular properties expressed as energy derivatives.

We show in Table I the spectroscopic constants of N_2, O_2, and F_2, evaluated using the cc-pCVTZ basis set, from HF and SUHF calculations with and without the single GD step. In the case of N_2, the length of the FCI vector expressed in a basis of Slater determinants is $\approx 2.9 \times 10^{19}$, far exceeding what is currently feasible with any current deterministic FCI code. We do not show results starting from a UHF reference wavefunction for O_2 (as UHF does not dissociate correctly) or F_2 (as the molecule is unbound with UHF). The SUHF calculations on O_2 use a reference determinant with $m_s = 0$ (while still projecting to a triplet state) in order to reach the best possible dissociation limit.

The calculations shown in Table I are compared to high-quality internally contracted multi-reference CI (IC-MRCI) results from Ref. 18, as well as the experimental values compiled in the same reference. In all cases, the single GD step yields improved spectroscopic constants over the results obtained with HF or SUHF. While the amount of the missing correlation recovered is modest in all cases (around 30%), it is still noteworthy given that a single GD step was performed. A better reference wavefunction or a further GD or QN step can significantly improve the results presented in Table I.

IV. CONCLUSIONS

In this work, we have considered gradient descent and quasi-Newton algorithms to reach the FCI ground state energy starting from an arbitrary reference wavefunction. The FCI ground state is systematically approached with increasing number of steps taken using either algorithm.

A central ingredient of this work is to avoid an explicit representation of the wavefunction, opting to write the energies, at each step in the optimization, as functionals of the reference wavefunction. The resulting functionals have some useful properties:

- They are independent of the form of the reference wavefunction. Therefore, they provide an unbiased way to compare how different wavefunctions evolve toward the FCI ground state along the optimization path.
- The functional forms are independent of the specific wavefunction parameters. That is to say, if the reference wavefunction is chosen as a single Slater determinant, then *any* Slater determinant (as long as it is not orthogonal to the FCI ground state) can be used with the functional forms provided.

While in this work we have focused on single determinant wavefunctions as well as linear combinations of non-orthogonal determinants, the evaluation of the required $\langle 0|H^k|0\rangle$ matrix elements can also be efficiently done (in polynomial time) for other types of wavefunctions such as multi-configuration self-consistent field (MC-SCF) or generalized valence-bond (GVB) expansions. An efficient evaluation of the $\langle 0|H^k|0\rangle$ matrix elements avoids the storage of intermediate wavefunctions as explicit vectors over the Hilbert space, thus allowing the computation of energies along the optimization path for systems where the dimension of the Hilbert space is larger than the storage resources available.

Given that the evaluation of the matrix elements $\langle 0|H^k|0\rangle$ can become quite expensive as k increases, we realize that practical applications may be limited to the first few steps. In this case, a better starting initial inverse Hessian (as opposed to $B = I$) can lead to substantially improved results. We are currently investigating this possibility.

ACKNOWLEDGMENTS

This work was supported by a generous start-up package from Wesleyan University.

APPENDIX A: SEARCH DIRECTION FROM BFGS UPDATE

We discuss the form of the search direction $p_1 = -B_1 g_1$, with B_1 constructed from a Broyden–Fletcher–Goldfarb–Shanno (BFGS)[11–14] update formula (starting from $B_0 = I$).

Let

$$s_0 \equiv Z_1 - Z_0 = Z_1, \tag{A1}$$

$$y_0 \equiv g_1 - g_0, \tag{A2}$$

which would yield $s_0 = -\sigma_* g_0$ and $(y_0)_x = (\alpha_1 - \alpha_0)H_x^0 + \beta_1 H_x^y H_y^0$. With $\rho_0 = 1/[(s_0)^x (y_0)_x]$, the BFGS update takes the form

Published under license by AIP Publishing

$$(B_1)_s^t = (B_0)_s^t - \rho_0 (B_0)_s^x (y_0)_x (s_0)^t - \rho_0 (s_0)_s (y_0)^x (B_0)_x^t + \rho_0^2 \left[\rho_0^{-1} + (y_0)^x (B_0)_x^y (y_0)_y\right] (s_0)_s (s_0)^t. \quad \text{(A3)}$$

We now carry an explicit evaluation of $p_1 = -B_1 g_1$. We note that

$$\begin{aligned}
\rho_0^{-1} &= -\sigma_* \alpha_0 (\alpha_1 - \alpha_0) f_2 - \sigma_* \alpha_0 \beta_1 f_3, \\
[y_0 y_0] &\equiv (y_0)^x (y_0)_x \\
&= (\alpha_1 - \alpha_0)^2 f_2 + 2(\alpha_1 - \alpha_0)\beta_1 f_3 + \beta_1^2 f_4, \\
[s_0 g_1] &\equiv (s_0)^x (g_1)_x \\
&= -\sigma_* \alpha_0 \alpha_1 f_2 - \sigma_* \alpha_0 \beta_1 f_3, \\
[y_0 g_1] &\equiv (y_0)^x (g_1)_x \\
&= \alpha_1 (\alpha_1 - \alpha_0) f_2 - \alpha_0 \beta_1 f_3 + \beta_1^2 f_4.
\end{aligned}$$

Therefore, p_1 takes the form

$$(p_1)_x = \alpha_1' H_x^0 + \beta_1' H_x^y H_y^0, \quad \text{(A4)}$$

with

$$\alpha_1' = \alpha_1 - 2\sigma_* \rho_0^2 (\rho_0^{-1} + [y_0 y_0])[s_0 g_1] - \rho_0 (\alpha_1 - \alpha_0)[s_0 g_1] + 2\sigma_* \rho_0 [y_0 g_1], \quad \text{(A5)}$$

$$\beta_1' = \beta_1 - \rho_0 \beta_1 [s_0 g_1]. \quad \text{(A6)}$$

APPENDIX B: OPTIMIZATION OF E_1 WITH A SLATER DETERMINANT

In this section, we consider the optimization of the E_1 functional of Eq. (10) using a Slater determinant trial wavefunction. A Slater determinant wavefunction $|\Phi\rangle$ can be parametrized in terms of Thouless rotations (see Ref. 19) as

$$|\Phi\rangle = \mathcal{N} \exp\left(\sum_i T_{ai} c_a^\dagger c_i\right) |\Phi_0\rangle, \quad \text{(B1)}$$

where $\mathcal{N}$ is a normalization factor, $|\Phi_0\rangle$ is some initial reference Slater determinant, and T_{ai} are Thouless amplitudes to be optimized. Here, indices a and i run over particles and holes, respectively.

The functional E_1 is fully expressed in terms of the expectation values $\langle 0|H|0\rangle$, $\langle 0|H^2|0\rangle$, and $\langle 0|H^3|0\rangle$. In assembling the gradient of E_1, it is sufficient to consider the gradients of those matrix elements. Consider the matrix element $\langle 0|H^k|0\rangle$, with $|0\rangle = |\Phi\rangle$ and $|\Phi\rangle$ written in the form of Eq. (B1). The matrix element is given by

$$\frac{\langle \Phi|H^k|\Phi\rangle}{\langle \Phi|\Phi\rangle} \equiv \frac{\langle \Phi_0|\exp(T^\dagger)\, H^k \exp(T)|\Phi_0\rangle}{\langle \Phi_0|\exp(T^\dagger)\, \exp(T)|\Phi_0\rangle}. \quad \text{(B2)}$$

Explicit differentiation of Eq. (B2) yields

$$\frac{\partial}{\partial T_{ai}^*} \frac{\langle \Phi|H^k|\Phi\rangle}{\langle \Phi|\Phi\rangle}\bigg|_{T=0} = \frac{\langle \Phi_0|c_i^\dagger c_a H^k|\Phi_0\rangle}{\langle \Phi_0|\Phi_0\rangle}. \quad \text{(B3)}$$

We finally note that the evaluation of $\langle \Phi|H^k|\Phi\rangle$ and its derivatives, with $|0\rangle$ chosen as a Slater determinant, can be carried out by the use of Wick's theorem.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (McGraw-Hill, New York, 1989).
[2] E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).
[3] E. J. Sundstrom and M. Head-Gordon, J. Chem. Phys. **140**, 114103 (2014).
[4] C. A. Jiménez-Hoyos, T. M. Henderson, T. Tsuchimochi, and G. E. Scuseria, J. Chem. Phys. **136**, 164109 (2012).
[5] Z. Wang, Y. Li, and J. Lu, J. Chem. Theory Comput. **15**, 3558 (2019).
[6] W. Kutzelnigg, Chem. Phys. Lett. **64**, 383 (1979).
[7] P. Wolfe, SIAM Rev. **11**, 226 (1969).
[8] P. Wolfe, SIAM Rev. **13**, 185 (1971).
[9] H. Nakatsuji, Acc. Chem. Res. **45**, 1480 (2012).
[10] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, New York, 2006).
[11] C. G. Broyden, IMA J. Appl. Math. **6**, 76 (1970).
[12] R. Fletcher, Comput. J. **13**, 317 (1970).
[13] D. Goldfarb, Math. Comput. **24**, 23 (1970).
[14] D. F. Shanno, Math. Comput. **24**, 647 (1970).
[15] Y. Qiu, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **146**, 184105 (2017).
[16] J. W. Krogh and J. Olsen, Chem. Phys. Lett. **344**, 578 (2001).
[17] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji *et al.*, Gaussian 16, Revision C.01, Gaussian, Inc., Wallingford, CT, 2016.
[18] K. A. Peterson, A. K. Wilson, D. E. Woon, and T. H. Dunning, Jr., Theor. Chem. Acta **97**, 251 (1997).
[19] C. A. Jiménez-Hoyos, R. Rodríguez-Guzmán, and G. E. Scuseria, Phys. Rev. A **86**, 052102 (2012).

Published under license by AIP Publishing