Approaching the Full Configuration Interaction Low-Energy Spectrum from an Arbitrary Reference Subspace

Carlos A. Jiménez-Hoyos
Department of Chemistry, Wesleyan University, Middletown, CT, 06459
(Dated: December 10, 2020)

In a previous work [arXiv:2010.02027] we showed how the full configuration interaction (FCI) ground state energy can be obtained as a functional of an arbitrary reference wavefunction by means of a gradient descent or quasi-Newton algorithm. Here, we extend this approach and consider the optimization of the low-energy subspace of the Hamiltonian from an arbitrary reference subspace. The energies along the optimization path are obtained in terms of transition matrix elements among the states in the reference subspace. We show an application of the algorithm with a reference subspace constructed from a non-orthogonal configuration interaction (NOCI) formalism to describe the avoided crossing in LiF and the low-lying singlet and triplet spectrum of formaldehyde.

I. INTRODUCTION

Exact solutions to the electronic Schrödinger equation can only be obtained, in closed form, for very small chemical systems. Therefore, most quantum chemical calculations aim to reproduce, as closely as possible, full configuration interaction (FCI) solutions [1], where the electronic Schrödinger equation is projected onto a basis of $N$-electron wavefunctions constructed from antisymmetrized products of some suitable one-particle basis. In trying to reproduce FCI solutions, one main goal is to reduce the computational effort as much as possible. A number of methods that can yield arbitrarily accurate approximations to the ground state wavefunction are known, but the number of methods available to target excited electronic states is more limited.

In a previous work [2], hereafter referred to as paper I, we discussed gradient descent and quasi-Newton algorithms to reach the FCI ground state wavefunction starting from an arbitrary reference state $|0\rangle$. The central goal of that paper was to define systematic approximations to the ground state wavefunction, characterized by the number of steps taken in the algorithm. Along with that goal, a key message of that work was that the energies along the optimization path can be written in terms of transition matrix elements among projected configurations can yield a qualitatively correct low-energy spectrum of molecular systems. Therefore, the goal of the present paper is to explore how that reference subspace can be evolved, using gradient descent and quasi-Newton algorithms, to yield the exact FCI low-energy spectrum.

All of those advantages are still relevant to the present work where we work with a reference subspace of wavefunctions rather than a single state $|0\rangle$.

In this work, we focus our application of the algorithm to non-orthogonal configuration interaction (NOCI) expansions [3] for the low-energy spectrum of molecular systems. (We stress, however, that the algorithm is applicable to other types of wavefunctions.) Here, we use NOCI in a broad sense to refer to ground or/excited states written as linear combinations of generally non-orthogonal ($\langle \Phi_q | \Phi_p \rangle \neq 0$) determinants:

$$|k\rangle = \sum_q f_q^k |\Phi_q\rangle,$$

where $f$ are some linear coefficients determined from the corresponding generalized eigenvalue problem. Moreover, we also consider similar expansions written in terms of symmetry-projected Slater determinants [4] of the form

$$|k\rangle = \sum_q f_q^k \hat{P} |\Phi_q\rangle,$$

where $\hat{P}$ is a projection operator that restores some symmetry of the Hamiltonian. Some of our recent work [5] has shown that NOCI expansions based on symmetry-projected configurations can yield a qualitatively correct low-energy spectrum of molecular systems. Therefore, the goal of the present paper is to explore how that reference subspace can be evolved, using gradient descent and quasi-Newton algorithms, to yield the exact FCI low-energy spectrum.

The rest of this manuscript is organized as follows. In Sec. [1] we describe the optimization target as well as the

---

*Electronic address: cjimenezhoyo@wesleyan.edu
parametrization we use in carrying out the optimization. We then provide details of gradient descent (II A) and quasi-Newton (II B) optimization algorithms, providing explicit expressions for the first few iterations. In Sec. III we discuss the application of the method in a H₂ ring, in the avoided crossing of LiF, and in the low-energy spectrum of formaldehyde. Finally, in Sec. IV we provide some closing remarks.

II. THEORY

We consider the optimization of the lowest-energy q FCI states. Our optimization target is the state-averaged energy

\[ E_{SA} = \frac{1}{q} (\varepsilon_0 + \varepsilon_1 + \ldots + \varepsilon_{q-1}) , \]

where \( \varepsilon_0 \) and \( \varepsilon_{q-1} \) correspond to the Hamiltonian ground and \((q - 1)\)-th excited eigenvalues, respectively. Naturally, a minimum of \( E_{SA} \) coincides with convergence of the entire low-energy subspace of dimension \( q \).

Without loss of generality, we assume that a set of orthonormal reference wavefunctions is available: \(|\{0\}, |1\}, \ldots, |q - 1\} \rangle\). We use an exponential, non-Hermitian parametrization to build states of the form

\[ |\Psi^k\rangle = \exp(\hat{Z}^k)|k\rangle, \]
\[ \hat{Z}^k = \sum_x Z^k_x |x\rangle \langle k|, \]

where \(|x\rangle\) labels an orthonormal state in the orthogonal complement of the reference subspace. The function to be optimized is then \( E_{SA}[Z] \), with \( Z \) being the coefficients of the \( \hat{Z}^k \) operators in Eq. 2. When convergence is reached, the FCI low-energy spectrum can be recovered from the solution to the generalized eigenvalue problem \( \mathcal{H} C = SC^ε \varepsilon \), with

\[ S_{kl}(Z) = \langle k| \exp(\hat{Z}^k) \exp(\hat{Z}^l)|l\rangle, \]
\[ \mathcal{H}_{kl}(Z) = \langle k| \exp(\hat{Z}^{k\dagger}) \exp(\hat{Z}^l)|l\rangle. \]

Nonetheless, the low-energy spectrum can be determined at any point in the optimization from the solution of the corresponding generalized eigenvalue problem. The matrix elements of \( \mathcal{S} \) and \( \mathcal{H} \), as a function of \( Z \), are given by

\[ S_{kl}(Z) = \delta^l_k + Z^k_x Z^{l\dagger}_x, \]
\[ \mathcal{H}_{kl}(Z) = H^k_x + H^{l\dagger}_x Z^k_x + Z^k_x H^l_x + Z^k_x H^l_y Z^{l\dagger}_y, \]

where \( H^a_\beta = \langle \alpha|H|\beta\rangle \), Einstein summation is implied and the indices \( x, y \) run only over the orthogonal complement of the reference subspace. We have assumed that the reference wavefunctions \(||\rangle\) are real and therefore we use real coefficients \( Z \), as we do throughout this work. While the state-averaged energy can be expressed in terms of the Hamiltonian eigenvalues, it can also be written as

\[ E_{SA}[Z] = \frac{1}{q} \text{Tr} \left( \mathcal{H}(Z) S^{-1}(Z) \right). \]

The gradient of \( E_{SA} \) with respect to \( Z \), evaluated at \( Z = Y \) is given by

\[ g^k_x = \frac{\partial E_{SA}[Z]}{\partial Z^k_x} \bigg|_{Z=Y} \]
\[ = \frac{2}{q} \sum_l \left( H^l_x + H^{l\dagger}_x Y^l_x \right) \left( S^{-1}(Y) \right)_{lk} \]
\[ - \frac{2}{q} \sum_l Y^l_x \left( S^{-1}(Y) \mathcal{H}(Y) S^{-1}(Y) \right)_{lk}. \]

For convenience, we shall introduce the matrices

\[ (F_1)^l_k = H^l_k, \]
\[ (F_2)^l_k = H^l_x H^l_x, \]
\[ (F_3)^l_k = H^l_x H^{l\dagger}_x H^l_y, \]

Note that all elements in \( F_1, F_2, \ldots \) can be evaluated in terms of matrix elements (or transition matrix elements) from the reference subspace. For instance,

\[ (F_2)^l_k = \langle k| H^2 |l\rangle - \sum_m \langle k| H |m\rangle \langle F_1 \rangle_m^l \]
\[ (F_3)^l_k = \langle k| H^2 |l\rangle - \sum_m \langle k| H^2 |m\rangle \langle F_1 \rangle_m^l \]
\[ - \sum_m \langle k| H |m\rangle \langle F_2 \rangle_m^l. \]

A. Gradient Descent

We begin at \( Z_0 = 0 \) with \( |\Psi^0_k\rangle = |k\rangle \). Naturally, \( E_{SA}^0 = 1/q \sum_k H^0_k \rangle \). The gradient at \( Z_0 \) is

\[ (g_0)^k_x = H^l_x \langle \alpha_0 \rangle_l^k, \]

with \( \langle \alpha_0 \rangle_l^k = 2/q \delta^k_\beta \).

Just as in paper I, we shall consider a full line search along \( -g_0 \). The state-averaged energy, as a function of the step size \( \sigma \), is given by

\[ E_{SA}^1 [-\sigma g_0] = \frac{1}{q} \text{Tr} \left( \mathcal{H}_1(\sigma) S_1^{-1}(\sigma) \right), \]

with

\[ \mathcal{H}_1(\sigma) = F_1 - \sigma \alpha_0 F_2 - \sigma F_2 \alpha_0 + \sigma^2 \alpha_0 F_3 \alpha_0, \]
\[ S_1(\sigma) = 1 + \sigma^2 \alpha_0 F_2 \alpha_0. \]

Note that \( \mathcal{H}_1 \) and \( S_1 \) can be assembled from matrix elements in the reference subspace, as described above. Therefore, \( E_{SA}^1 \) is itself a functional of the reference subspace that can be determined after evaluation of \( F_1, F_2, \)}
and $F_3$. A closed-form solution for $\sigma_*$ that minimizes $E_{BA}^3[-\sigma g_0]$ is possible, but it is easier in practice to carry out the minimization numerically. With $\sigma_*$ available, the states $|\Psi_k^1\rangle$ can be written as

$$|\Psi_k^1\rangle = e^{-\sigma_*(g_0)} |k\rangle.$$  \hfill (19)

We can now attempt a second step. The gradient at $Z_1 = -\sigma_0 g_0$ is

$$(g_1)_k = H_1^I(\alpha_1)_k + H_2^V H_2^R(\beta_1)_k,$$  \hfill (20)

with

$$\alpha_1)_k = \frac{2}{q} \left(S_{S_1}^{-1} + \sigma_0 S_{S_1}^{-1} H_{1,1} S_{S_1}^{-1}\right)_k,$$  \hfill (21)

$$\beta_1)_k = \frac{2}{q} \left(\sigma_0 S_{S_1}^{-1}\right)_k,$$  \hfill (22)

where $S_{S_1} \equiv S_1(\sigma_*)$ and $H_{1,1} \equiv H_1(\sigma_*)$.

Considering a line search along $-g_1$ with step size $\tau$, we can write the state-averaged energy as a function of $\tau$

$$E_{BA}^3[-\sigma_0 g_0 - \tau g_1] = \frac{1}{q} \text{Tr} \left( H_2(\tau) S_{S_1}^{-1}(\tau) \right).$$  \hfill (23)

with

$$H_2(\tau) = H_{1,1} - \tau \alpha_1 F_2 - \tau F_2 \alpha_1 - \tau \beta_1 F_3 - \tau F_3 \beta_1 + \tau^2 \alpha_1 F_3 \beta_1 + \tau \beta_1 F_3 \beta_1 + \tau^2 \alpha_1 F_3 \beta_1 + \tau \beta_1 F_3 \beta_1 + \tau \sigma_0 F_2 \alpha_0 + \tau \sigma_0 F_2 \alpha_0 + \tau \sigma_0 F_2 \alpha_0 + \tau \sigma_0 F_2 \alpha_0,$$  \hfill (24)

$$S_2(\tau) = S_{S_1} + \tau^2 \alpha_1 F_2 \alpha_1 + \tau^2 \beta_1 F_4 \beta_1 + \tau^2 \alpha_1 F_4 \beta_1 + \tau^2 \beta_1 F_4 \beta_1 + \tau \sigma_0 F_2 \alpha_0 + \tau \sigma_0 F_2 \alpha_0 + \tau \sigma_0 F_2 \alpha_0 + \tau \sigma_0 F_2 \alpha_0.$$  \hfill (25)

Note that $H_2$ and $S_2$ can be assembled with $F_1$, $F_2$, ..., $F_5$ available. Therefore, $E_{BA}^3$ is still a functional of the reference subspace. Let $\tau_*$ be the minimizer of $E_{BA}^3[-\sigma_0 g_0 - \tau g_1]$; the states $|\Psi_2^k\rangle$ can then be written as

$$|\Psi_2^k\rangle = e^{-\sigma_*(g_0)\tau_* - \tau_*(g_1)\tau} |k\rangle.$$  \hfill (26)

If a third step is attempted, by simple inspection one can readily realize that the gradient at $Z_2 = -\sigma_0 g_0 - \tau_* g_1$ takes the form

$$(g_2)_k = H_2^I(\alpha_2)_k + H_2^V H_2^R(\beta_2)_k + H_2^R H_2^R(\gamma_2)_k.$$  \hfill (27)

The corresponding $E_{BA}^3$ would also be a functional of the reference subspace that can be assembled from $F_1$, $F_2$, ..., $F_7$. Subsequent steps require the evaluation of higher order $F_k$ matrices.

B. Quasi-Newton

We now consider a quasi-Newton approach [21] in order to improve the rate of convergence of the state-averaged energy. 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 gradient descent. We perform a full line search along $p_k$ as in gradient descent. Following our previous work, we choose to set $B_0 = I$ as this allows us to fully define the quasi-Newton method as a functional of the reference subspace.

With the choice $B_0 = I$, the first step coincides with that from gradient descent and $E_{BA}^3$ remains unchanged. While the gradient $g_1$ is the same as in gradient descent (see Eq. 20), the search direction $p_1$ is determined from $p_1 = -B_1 g_1$, with $B_1$ constructed using a quasi-Newton update formula.

As shown in appendix [A], $p_1$ determined from a Broyden-Fletcher-Goldfarb-Shanno (BFGS) update formula, takes the form

$$(p_1)_k = H_1^I(\alpha_1')_k + H_2^V H_2^R(\beta_1')_k,$$  \hfill (28)

with $\alpha_1'$ and $\beta_1'$ being some matrices that are numerically different from $\alpha_1$ and $\beta_1$. Given that $p_1$ takes the same functional form as $g_1$, we conclude that $E_{BA}^3$ determined from the BFGS approach is also a functional of $F_1$, $F_2$, $F_3$, $F_4$, $F_5$, $F_6$, $F_7$. Namely, $E_2$ would take the same form as Eq. 23 with $\alpha_1 \rightarrow \alpha_1'$ and $\beta_1 \rightarrow \beta_1'$ in the definitions of $H_2(\tau)$ and $S_2(\tau)$. Further quasi-Newton steps can also be cast as functionals of the reference subspace, with higher order $F_k$ matrices required.

III. RESULTS AND DISCUSSION

We proceed to discuss the application of the optimization algorithms described above in a $H_3$ ring, in the avoided crossing of LiF, and in the low-lying singlet and triplet spectrum of formaldehyde.

A. $H_4$

We begin by revisiting the $H_4$ system discussed in paper I and shown in Fig. 1. While in paper I we focused on the ground state, there are in fact two low-lying singlet states at large $\theta$. Those two states can be reasonably well described using two different unrestricted Hartree–Fock (UHF) configurations, whose character is depicted in Fig. 1.

We show in the left panel of Fig. 2 the energy of the two different UHF solutions, as well as the energy of the two spin-projected UHF (SUHF) solutions that have a similar character. We also show the energy obtained in NOCI-UHF where we use the two UHF solutions plus their spin-flipped counterparts and consider the resulting two eigenvectors with “singlet” character. [15] We also
technically, the later is not a spin singlet state, but the most
UHF determinant and its spin-flipped counterpart (techni-
cally, the later is not a spin singlet state, but the most

our previous work [14]. Explicitly, we describe the two
this case, we work with a NOCI description, following

crossing requires a multi-reference description [13]. In
potential energy curve. A correct description of the avoided
resulting NOCI states provide a qualitatively correct de-
scription of the avoided crossing, but the quantitative
aspects are not quite correct. We have evaluated the LiF
potential energy curve using the same basis set as that
in Ref. [13] where frozen-core FCI results were published.

We have calculated the correction in the state-averaged
energy after 1 step starting from the NOCI reference sub-
space: sa-NOCI+1. Here, we emphasize that this com-
putation was done by evaluation of the $F_2$ and $F_3$ ma-
trices, without an explicit vector representation of the
two states: in this case, the dimension of the FCI vector,
in a basis of $n_s = 0$ Slater determinants is $1.42 \times 10^{11}$,
which would render storage of the FCI vector impossible
in most common computational facilities.

We show in the left panel of Fig. 4 the potential en-
ergy curves of the ground and excited state obtained with
NOCI as well as sa-NOCI+1. Additionally, we show in
the right panel the dipole moment for the $3\Sigma$ states.
As shown in the left panel, only a fraction of the missing cor-
relation energy is captured by a single gd step, but sig-
nificantly better results could be obtained if more steps
were taken (not done in this work). As shown in the right
panel, the FCI ground state dipole moment peaks near
10.5 bohr or so, at which point the avoided crossing oc-
curs. NOCI predicts an avoided crossing near 6.7 bohr or
so (the crossing of the two black curves), a reflection of
the poor quantitative agreement with FCI. After one gd
step, the avoided crossing in sa-NOCI+1 shifts by about
0.5 bohr or so in the right direction. While this is only a
modest improvement (consistent with only a fraction of
the missing correlation energy recovered), we still find it
effortanding.

It is interesting to compare sa-NOCI+1 with NOCI+1,
where a single gd step is used to improve just the NOCI
ground state. In NOCI+1, the dipole moment of the
ground state is nearly identical to that of NOCI itself.
This suggests that the eigenvector obtained after the di-
gonalization of the Hamiltonian matrix in sa-NOCI+1
has significant mixing between the ground and the ex-
cited state. It also implies that the internal contraction
used in NOCI+1 (i.e., the eigenvector from NOCI) leads
to larger qualitative errors in the ground state dipole mo-
moment versus sa-NOCI+1.

C. Formaldehyde low-energy spectrum

As a last example we consider the low-energy spec-
trum of the formaldehyde molecule. In a recent paper
[6], we showed how a state-averaged resonating Hartree–
Fock approach (sa-ResHF) can yield a good description of
the low-energy spectrum of formaldehyde. Upon revisit-
ning those results, we realized that most of the states are
well described by a single SUHF determinant [16], with
the only exception being the $1^1A_1$ and $3^1A_1$ states. In
those two states (the ground and $\pi \rightarrow \pi^*$ states), the
mixing between the ground RHF-like determinant and

show NOCI-SUHF curves obtained using the two differ-
ent SUHF solutions, as well as the two FCI lowest-lying
singlet states.

The right panel of Fig. 2 shows the improvement to
both low-lying singlet states, as a function of $\theta$, after
one or two gradient descent (gd) or quasi-Newton (qn)
steps have been taken, starting from the NOCI-SUHF
reference subspace. The energy of both states is improved
substantially even after just one iteration. Using a qn
algorithm yields better results after two steps than using
a gd algorithm, but in both cases the energy of both low-
lying singlet states is within a mHartree of the exact FCI
results.

We show in Fig. 5 the convergence profile of the gd and
qn algorithms using the NOCI-UHF and NOCI-SUHF
reference subspaces at $\theta = 90$ deg. It is again evident that
the qn algorithm reaches convergence significantly faster
than the gd algorithm, as expected. The NOCI-SUHF
reference subspace is significantly better than the NOCI-
UHF one and convergence (with $\mu$Hartree accuracy) is
reached after a handful of iterations. Using a qn algo-
rithm, both in the case of NOCI-UHF and NOCI-SUHF,
convergence of the energy for both low-lying states can
be reached with a similar number of iterations.

B. LiF Avoided Crossing

We now consider the avoided crossing in the LiF po-
tential energy curve. A correct description of the avoided
crossing requires a multi-reference description [13]. In
this case, we work with a NOCI description, following
our previous work [14]. Explicitly, we describe the two
$1\Sigma$ states by using the restricted Hartree–Fock (RHF) de-
terminant and the symmetric linear combination of the
UHF determinant and its spin-flipped counterpart (tech-
nically, the later is not a spin singlet state, but the most
significant triplet contaminant has been removed). The
model.
the excited $\pi \rightarrow \pi^*$ SUHF determinant is quite significant, such that the state-averaged resonating Hartree–Fock description, where the orbital optimization is done targeting the state-averaged energy, is required. The calculations in this work therefore use a single SUHF determinant for each state, except for the $1^1A_1$ and $3^1A_1$ states.\cite{17} Our calculations use the same aug-cc-pVDZ basis set used previously. For the triplet states we have used $m_s = 0$ UHF determinants in SUHF solutions, although we emphasize that the spin projection was done to a triplet state.

We show in Tab.\cite{11} the vertical excitation energies obtained by sa-ResHF and sa-(sa-ResHF)+1, where in the latter case we have carried out a single gd step in the state-averaged formalism described in this work. In this case we have also directly evaluated the $F_2$ and $F_3$ matrices without an explicit vector representation of the reference states: the dimension of the Hilbert space for $m_s = 0$ determinants is $2.0 \times 10^{19}$. We note that the calculations for each symmetry sector were carried out independently.

As shown in Tab.\cite{11} the ground state energy is lowered by $> 5$ eV after a single gd step. Nonetheless, we only see small differences (the largest differences are about 0.6 eV) between the reference spectrum and that obtained after the gd step. Both of the $1^1B_1$ states are shifted by around 0.55 eV; part of that shift is likely because the calculations on each symmetry sector were done independently. If we focus on the relative energy shift within each symmetry sector, all the shifts are below 0.2 eV, with the single outlier being the $2^1A_1$ state.

It is instructive, for the case of the $1^1A_1$ states, to look at the $S_{1s}$ and $H_{1s}$ matrices. (Recall that the solution to the generalized eigenvalue problem using $H_{1s}$ and $S_{1s}$ yields the low-energy spectrum after one gd step.) They are given by (with the $H_{1s}$ matrix expressed in a.u.)

\[
S_{1s} = \begin{pmatrix}
1.0269 & -0.0000 & -0.0020 \\
-0.0000 & 1.0240 & 0.0001 \\
-0.0020 & 0.0001 & 1.0275
\end{pmatrix},
\]

\[
H_{1s} = \begin{pmatrix}
-149.4399 & 0.0046 & 0.3123 \\
0.0046 & -148.6996 & -0.0116 \\
0.3123 & -0.0116 & -149.1454
\end{pmatrix}.
\]
The reader may convince himself that the resulting eigenvectors have considerable mixing between the ground and the \( \pi \rightarrow \pi^* \) states. Note that this is beyond the mixing present in sa-ResHF itself, as the \( S_{1\pi} \) and \( H_{1\pi} \) matrices are expressed in the basis of the orthonomal reference states from sa-ResHF. This significant mixing implies that the character of those two states is being adjusted in the presence of the correlation captured by the single gd step.

### IV. CONCLUSIONS

We have generalized the gradient descent and quasi-Newton algorithms presented in paper I to the optimization of a low-energy spectrum instead of just the ground state. The method presented optimizes the state-averaged energy thereby recovering the exact low-energy spectrum as the algorithm reaches convergence. The state-averaged energies along the optimization path are fully expressed in terms of transition matrix elements among the states in the reference subspace. This allows us to avoid an explicit vector representation of the intermediate wavefunctions which is crucial for systems where the dimension of the Hilbert space becomes intractable. Moreover, the algorithm defines a systematic approximation to the exact low-energy spectrum.

We have shown an application of the algorithm using a reference subspace written as a non-orthogonal configuration interaction in the case of LiF and formaldehyde. While we only carried out a single step of the algorithm in those cases, the results can be improved by carrying out a few more steps or using an improved reference subspace.

### TABLE I: Vertical excitation energies (in eV) of several low-lying singlet and triplet states of formaldehyde evaluated using the aug-cc-pVDZ basis set.

| state | character | sa-ResHF | sa-(sa-ResHF)+1 |
|-------|-----------|----------|-----------------|
| singlet states | | | |
| 1 \(^3\)A\(_1\) | \(\pi \rightarrow \pi^*\) | 5.94 | 6.16 |
| 2 \(^3\)A\(_1\) | \(n \rightarrow 3pb_2\) | 8.46 | 8.56 |
| 3 \(^3\)A\(_2\) | \(n \rightarrow \pi^*\) | 4.01 | 3.98 |
| 2 \(^3\)A\(_2\) | \(n \rightarrow 3pb_1\) | 9.23 | 9.32 |
| 1 \(^3\)B\(_1\) | \(\sigma \rightarrow \pi^*\) | 8.98 | 8.96 |
| 3 \(^3\)B\(_2\) | \(n \rightarrow 3sa_1\) | 7.52 | 7.57 |
| 2 \(^3\)B\(_2\) | \(n \rightarrow 3pa_1\) | 8.52 | 8.54 |

\(^a\)The ground state energy is \(-113.940521\) a.u.

\(^b\)The ground state energy is \(-114.127074\) a.u.

Our calculations in LiF and formaldehyde showed that in some cases there is significant mixing between the states in the reference subspace in the presence of the missing correlation captured by the algorithm. That is, even when the reference subspace was deemed as qualita-
tively correct, the weights of the reference configurations adjust as they evolve towards the FCI states. As a consequence of this, if the method is used to target the ground state exclusively this can lead to larger qualitative errors compared to the state-averaged description.

As presented, the method can use other type of reference subspaces such as complete active space (CAS) or more general multi-configurational self-consistent field (MC-SCF) solutions. We plan to explore the utility of the systematic approximation here presented using those wavefunctions in the near future.

Acknowledgments

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

Appendix A: BFGS Update

We discuss in this appendix the form of the search direction \( p_1 = -B_1 g_1 \), with \( B_1 \) constructed from a Broyden-Fletcher-Goldfarb-Shanno (BFGS) \cite{Shanno1970, Fletcher1970} update formula (starting from \( B_0 = I \)). Let

\[
\begin{align*}
{s_0} &= Z_1 - Z_0 = Z_1, \quad (A1) \\
{y_0} &= g_1 - g_0, \quad (A2)
\end{align*}
\]

which yields \( s_0 = -\sigma_y g_0 \) and

\[
(y_0)^k_x = H_l^i (\alpha_1 - \alpha_0)^k + H_l^i (\beta_1)^k. \quad (A3)
\]

Defining \( \rho_0 \equiv 1/[(s_0)^k_x (y_0)^k_x] \), the BFGS update takes the form

\[
\begin{align*}
\begin{array}{l}
B_1^k \{ s_0 \}^k_x = p_1 \{ s_0 \}^k_x \\
- \rho_0 \{ s_0 \}^k_x (y_0)^m_x \{ s_0 \}^k_x - \rho_0 \{ s_0 \}^k_x (y_0)^m_x \{ s_0 \}^k_x \\
+ \rho_0^2 \{ \rho_0^{-1} + (y_0)^n_x \{ s_0 \}^k_x \} \{ s_0 \}^k_x
\end{array}
\end{align*}
\]

We now carry an explicit evaluation of \( p_1 = -B_1 g_1 \). We note that

\[
\rho_0^{-1} = -\sigma_y \text{Tr} \{ \alpha_0 F_2 (\alpha_1 - \alpha_0) + \alpha_0 F_3 \beta_1 \},
\]

\[
[y_0 y_0] = (y_0)^k_x (y_0)^k_x
\]

\[
= \text{Tr} \{ (\alpha_1 - \alpha_0) F_2 (\alpha_1 - \alpha_0) + (\alpha_1 - \alpha_0) F_3 \beta_1 + \beta_1 F_3 (\alpha_1 - \alpha_0) + \beta_1 F_4 \beta_1 \},
\]

\[
[s_0 g_1] = (s_0)^k_x (g_1)^k_x
\]

\[
= -\sigma_y \text{Tr} \{ \alpha_0 F_2 \alpha_1 + \alpha_0 F_3 \beta_1 \},
\]

\[
[y_0 g_1] = (y_0)^k_x (g_1)^k_x
\]

\[
= \text{Tr} \{ (\alpha_1 - \alpha_0) F_2 \alpha_1 + (\alpha_1 - \alpha_0) F_3 \beta_1 + \beta_1 F_3 \alpha_1 + \beta_1 F_4 \beta_1 \}.
\]

Therefore, \( p_1 \) takes the form

\[
(p_1)^k_x = H_l^i (\alpha_1)^k + H_l^i (\beta_1)^k, \quad (A5)
\]

with

\[
\alpha_1' = \alpha_1 - \sigma_y \rho_0^{-1} \{ y_0 y_0 \} \{ s_0 g_1 \} \alpha_0
\]

\[
- \rho_0 \{ s_0 g_1 \} \{ (\alpha_1 - \alpha_0) + \sigma_y \rho_0 \{ y_0 g_1 \} \alpha_0 \}, \quad (A6)
\]

\[
\beta_1' = \beta_1 + \rho_0 \{ s_0 g_1 \} \beta_1. \quad (A7)
\]

[1] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (McGraw-Hill, New York, 1989).
[2] C. A. Jiménez-Hoyos, *Approaching the full configuration interaction ground state from an arbitrary wavefunction with gradient descent and quasi-Newton algorithms* (2020), arXiv:2010.02027.
[3] E. J. Sundstrom and M. Head-Gordon, J. Chem. Phys. 140, 114103 (2014).
[4] K. T. Jensen, R. L. Benson, S. Cardamone, and A. J. W. Thom, J. Chem. Theory Comput. 14, 4621 (2018).
[5] C. A. Jiménez-Hoyos, T. M. Henderson, T. Tsuchimochi, and G. E. Scuseria, J. Chem. Phys. 136, 164109 (2012).
[6] J. Nite and C. A. Jiménez-Hoyos, J. Chem. Theory Comput. 15, 5343 (2019).
[7] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, New York, 2006), 2nd ed.
[8] C. G. Broyden, IMA J. Appl. Math. 6, 76 (1970).
[9] R. Fletcher, Comput. J. 13, 317 (1970).
[10] D. Goldfarb, Math. Comp. 24, 23 (1970).
[11] D. F. Shanno, Math. Comp. 24, 647 (1970).
[12] Y. Qiu, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. 146, 184105 (2017).
[13] C. W. Bauschlicher and S. R. Langhoff, J. Chem. Phys. 89, 4246 (1988).
[14] J. Nite and C. A. Jiménez-Hoyos, *Efficient multi-configurational wavefunction method with dynamical correlation using non-orthogonal configuration interaction singles and doubles (NOCISD)* (2019), ChemRxiv:11369641.
[15] Those are the eigenvectors which are symmetric under flipping all of the spins.
[16] Naturally, the SUHF determinants describing higher energy states do not correspond to the lowest energy SUHF solution but rather to higher-energy ones.
[17] This accounts for the small differences between the results here reported and those in Ref. [3].