JCTC Journal of Chemical Theory and Computation

A Comparison of Three Variants of the Generalized Davidson Algorithm for the Partial Diagonalization of Large Non-Hermitian Matrices

Marco Caricato,* Gary W. Trucks, and Michael J. Frisch

Gaussian, Inc., 340 Quinnipiac Street, Bldg. 40, Wallingford, Connecticut 06492

Received February 24, 2010

Abstract: The solution of the equation of motion coupled cluster singles and doubles problem, that is finding the lowest lying electronic transition energies and properties, is fundamentally a large non-Hermitian matrix diagonalization problem. We implemented and compared three variants of the widely diffuse generalized Davidson algorithm, which iteratively finds the lowest eigenvalues and eigenvectors of such a matrix. Our numerical tests, based on different molecular systems, basis sets, state symmetries, and reference functions, demonstrate that the separate evaluation of the left- and right-hand eigenvectors is the most efficient strategy to solve this problem considering storage, numerical stability, and convergence rate.

1. Introduction

This work originates from our attempts to implement the most efficient algorithm to find the lowest eigenvalues of the approximate Hamiltonian in the equation of motion coupled cluster singles and doubles method (EOM-CCSD).[1] The latter is one of the most accurate and yet affordable methods for the calculation of one-electron transition energies and properties. Its basic equation for the k-th excited state can be written as:

$$(\bar{H}R_k)_c|\Phi_0\rangle = \omega_k R_k|\Phi_0\rangle \tag{1}$$

where $\bar{H} = e^{-T}He^{T}$ is the similarity transformed Hamiltonian, R_k is an excitation operator toward the k-th state, Φ_0 is the reference function, ω_k is the transition energy, and the notation $(...)_c$ indicates that only connected diagrams are considered. In principle, this equation can be solved directly by diagonalization of $\bar{H}$, whose eigenvalues are the transition energies, and the eigenvectors are R_k, for all the states. In practice, this is not possible because the matrix dimension, roughly o^2v^2, with o and v being the numbers of occupied and virtual orbitals, is very large. Another complication is that the similarity transformed Hamiltonian is not Hermitian and, hence, has different left and right eigenvectors:

---

* Corresponding author. E-mail: marco@gaussian.com.

$$\langle\Phi_0|L_k\bar{H} = \omega_k\langle\Phi_0|L_k \tag{2}$$

If the transition energies are of interest, then only eq 1 (or equivalently eq 2) need be solved. However, both left and right eigenvectors are necessary in order to obtain transition properties. For example, the dipole strength between the ground and the k-th excited state is

$$\langle\Phi_0|L_0\mu R_k|\Phi_0\rangle\langle\Phi_0|L_k\mu R_0|\Phi_0\rangle \tag{3}$$

Since the dimension of the $\bar{H}$ matrix prevents a direct diagonalization, the most effective computational approach is a modified version of the Davidson algorithm to treat non-Hermitian matrices.[2–4] A brief overview of the algorithm is presented in Section 2. However, the basic concept is that the eigenvalues and eigenvectors of interest are obtained through an iterative procedure which avoids the computation, storage, and diagonalization of the complete matrix and stops when certain criteria of convergence are satisfied.

In this work, we compare three variants of the algorithm in ref 4 for the evaluation of the first k eigenvalues and left and right eigenvectors. We implemented these in the Gaussian 09 suite of programs.[5] In the first two variants, the left and right eigenvectors are converged simultaneously, in the first case expanding them in two sets of biorthonormal trial vectors and in the second expanding them in one set of orthonormal vectors. In the third variant, the right eigen-

vectors are found first and then the left eigenvectors. The results in Section 3 show that the last variant is the most efficient in terms of storage, convergence rate, and numerical stability.

2. Algorithms Description

The algorithms presented in this section can be applied to any non-Hermitian square matrix A of dimension n. Let us start with the algorithm that finds the left and right eigenvectors simultaneously by using two sets of biorthonormal expansion vectors $\bar{B}$ and B, which satisfy:

$$\bar{B}^\dagger B = I \qquad (4)$$

where I is the unit matrix. We want to find the lowest k eigenvalues (ω_k) and left (L_k) and right (R_k) eigenvectors of the A matrix, where $k \ll n$. We start with i initial $\bar{B}$ and B vectors (which can be the same) as a guess for the eigenvectors, with $i \geq k$. A is projected onto the subspace of dimension i:

$$A^i = \bar{B}^{i\dagger} A B^i \qquad (5)$$

where the superscript i indicates the number of initial vectors. The projected matrix A^i can be diagonalized with standard techniques (its dimension is small), and the eigenvalues and eigenvectors for the subspace are found: ω_k^i, l_k^i, and r_k^i, producing approximate eigenvectors:

$$L_k \simeq L_k^{(i)} = \bar{B}^i l_k^i \\ R_k \simeq R_k^{(i)} = B^i r_k^i \qquad (6)$$

This approximation can be tested by checking the norm of the residual vectors $\bar{W}_k$ and W_k:

$$(A - \omega_k^i) R_k^{(i)} = W_k \\ (A^\dagger - \omega_k^i) L_k^{(i)} = \bar{W}_k \qquad (7)$$

Convergence is achieved when the norm of the residuals is below a certain threshold ξ; otherwise, new vectors are added to the expansion space based on the residuals:

$$Q^i = (\omega_k^i - A_D)^{-1} W_k \\ \bar{Q}^i = (\omega_k^i - A_D)^{-1} \bar{W}_k \qquad (8)$$

where A_D are the diagonal elements of the A matrix.[2] This is a good guess for a new set of vectors as long as A is diagonally dominant. This is typically the case for the similarity transformed Hamiltonian.

The vectors $\bar{Q}^i$ and Q^i are then biorthonormalized among each other and with respect to the previous vectors $\bar{B}^i$ and B^i. A set of expansion vectors is thus created, $\bar{B}^{2i}$ and B^{2i}, twice as large as the initial one. With this new set, the matrix A is projected onto a larger subspace:

$$A^{2i} = (\bar{B}^{2i})^\dagger A B^{2i} \qquad (9)$$

and the whole process is repeated until convergence. Since the eigenvectors do not converge all at the same rate, a smaller number of new vectors can be created in later iterations as more and more roots converge. More impor-

tantly, this algorithm avoids the explicit calculation and storage of the A matrix, since only products AB and $A^\dagger B$ are necessary.[6]

Although this algorithm seems a straightforward extension of the Davidson algorithm, it may encounter numerical instabilities. For instance, complex eigenvalues can be found in intermediate steps, even if the final eigenvalues are real.[7] A robust way to deal with this issue and to eliminate the intermediate complex eigenvalues in the following cycles is to create twice the number of W and $\bar{W}$ (and thus Q and $\bar{Q}$) vectors for these eigenvalues, one for the real and one for the imaginary parts, which share the same $R_k^{(i)}$ and $L_k^{(i)}$, eqs 7 and 8. We also found that it is more stable to create two distinguished projected matrices A^i and $A^{i\dagger}$ and to diagonalize them separately in order to evaluate the eigenvectors in the subspace r_k^i and l_k^i (note that the relation $A^i = (A^{i\dagger})^\dagger$ is not exactly satisfied for numerical reasons). This does not add much to the computational time, since the matrix−vector products AB^i and $A^\dagger \bar{B}^i$ are needed anyway for the calculation of the residuals in eq 7, and the final projection has $O(o^2 v^2)$ cost, which is much cheaper than the evaluation of AB^i and $A^\dagger \bar{B}^i$ (which scales as $O(o^2 v^4 + o^3 v^3)$).

Another source of instability arises when approaching convergence. At this point, the right eigenvector for one state may satisfy the convergence criterion, while the corresponding left eigenvector does not or vice versa. One might think that the convergence criteria must be satisfied by the vectors in both spaces before interrupting the creation of new vectors. However, we have found that very small residuals can generate noise that may prevent the convergence of the algorithm. A more robust strategy is to take all the new vectors, Q^i and $\bar{Q}^i$, and biorthonormalize them with respect to the left space, $\bar{B}^i$, *and* right space, B^i, expansion vectors and then among themselves. In this way, the same number of new vectors is created for both spaces, which limits the noise. The same strategy can be used, for example, when a complex eigenvalue is found for one space but not for the other or when the diagonalization of the projected matrices gives eigenvalues that differ more than a certain ratio. Although rare, the latter situation may happen due to the numerical precision of the various operations during the iterative cycles. In the following, we shall refer to this algorithm as "B-Biorth".

An alternative to the previous algorithm is to orthonormalize the Q^i and $\bar{Q}^i$ with respect to the previous series of B^i and $\bar{B}^i$ and with respect to each other. Since for the first guess $\bar{B}^i = B^i$, the same set of expansion vectors is used for both spaces for all iterations. We shall call this variant "B-Orth". Note that twice the number of vectors is created for both spaces at each iteration for B-Orth in comparison to B-Biorth. For symmetric matrices, it is known that creating a larger number of expansion vectors than number of target roots helps to increase the convergence rate.[8] An advantage of this variant compared to the previous algorithm is that it is intrinsically more stable, since the same number of new vectors is created at each iteration for both spaces. However, the possibility of complex eigenvalues at intermediate steps holds, and it is dealt with the same strategy as above.

A third variant, that we shall refer to as "B-1Space" is to evaluate the eigenvectors separately. In this way, only one set of orthonormal B^i vectors is used. This approach can encounter intermediate complex eigenvalues, but does not raise the issue of unbalanced description of the subspaces, as they are spanned separately. Furthermore, the second diagonalization can start with the converged eigenvectors from the first diagonalization, which is usually much better than the initial CIS-based guess. The B-1Space variant does not guarantee that the same eigenvalues are found in the two separate diagonalizations, and a check is necessary after convergence in both spaces. A final biorthonormalization of the converged eigenvectors is performed once the equivalence of both spaces is verified.

For all the algorithms, a maximum subspace dimension can be set according to the details of the calculation and the machine setup. If convergence is not achieved before the subspace limit is reached, then the diagonalization can be restarted by using the last updated eigenvectors as a starting guess. We choose a limit equal to $20 \times n_{states}$, where n_{states} is the number of states (eigenvalues) to be computed.

If one set of eigenvectors is found, say the right-hand ones and the corresponding eigenvalues, then the left eigenvectors can be also evaluated by solving a linear equation for each root k, eq 2.[1] Standard iterative algorithms can be used to solve these equations.[1,9] As for EOM-CCSD, eq 2 is very similar to the ground-state Λ vector equation of gradient theory:[10,11]

$$\langle \Phi_0 | (1 + \Lambda) \bar{H} = 0 \qquad (10)$$

in the sense that the same matrix−vector products are involved. Technically, eq 10 is an $Ax = b$ problem with b corresponding to the $\langle ij \| ab \rangle$ integrals, whereas eq 2 is an $A'x = 0$ problem; nevertheless, the algorithms employed in the solution of both problems are similar. We shall refer to this alternative as "B-LinSys". The cost of each step of the iterative diagonalization and the iterative solution of the linear system is comparable. The matrix−vector product $A^\dagger \bar{B}^i$ is the same, and the extra work (evaluation of the residuals in B-1Space) is much cheaper than the building of $A^\dagger \bar{B}^i$. Thus, the difference in efficiency of the two approaches arises from the different rate of convergence of the iterations.

B-Biorth requires the largest amount of storage with two sets of matrices used for the two spaces: $\bar{B}$ and B, $A^\dagger \bar{B}$ and AB, $\bar{W}$ and W, and $\bar{Q}$ and Q. The largest of those are $\bar{B}$, B, $A^\dagger \bar{B}$, and AB, given that the number of vectors is as large as the maximum dimension of the subspace. $\bar{W}$, W, $\bar{Q}$, and Q only require a number of vectors corresponding to the number of target roots. B-Orth is much less demanding, as only one set of B vectors is necessary. B-1Space is the least demanding, as only one set of B, AB, W, and Q matrices are necessary, and the solution of the left-hand problem reuses the same storage. B-LinSys is equivalent to B-1Space. Both storage requirement and computational time can be reduced by exploitation of the equivalence of α and β electrons for closed shell calculations and of Abelian molecular point group symmetry.

The recommended convergence criteria found in the literature is that the norm of the residual vectors must be below a certain threshold ξ. For EOM-CCSD, Stanton and Bartlett[1] proposed that $\xi = 10^{-5}$ is sufficient to obtain convergence for transition properties. We prefer to use slightly more conservative criteria, thus we check: (i) the norm of the residual vectors; (ii) the change in the eigenvalues ($< \xi \times 10^{-2}$); and (iii) the absolute change in the current eigenvectors ($< \xi$).

3. Results

We report results for four molecular systems: formaldehyde (C_{2v}), ethene (D_{2h}), acetone (C_{2v}), and *trans*-1,3-butadiene (C_{2h}) and six basis sets: 6-31G*, 6-31+G*, 6-31++G**, 6-311++G**, aug-cc-pVDZ, and aug-cc-pVTZ. We computed 3 states for each irrep, thus 12 states for formaldehyde, acetone, and butadiene and 24 for ethene. Restricted closed shell and unrestricted open shell (with a +1 charge) Hartree−Fock (HF) wave functions were considered as reference functions. This range of options allows to test the behavior of the algorithms in a variety of different conditions. The geometries of all the systems were optimized at MP2/6-311+G** level of theory and used for all the excited state calculations.[12]

The results for the closed shell calculations are reported in Table 1. The B-1Space algorithm requires the smallest number of matrix−vector products in all cases, in part because of the use of the converged right eigenvectors as a starting guess for the left eigenvectors.

B-Orth is the least-efficient algorithm, since orthonormalizing the new vectors for both spaces in order to create a single set of vectors seems not to help the convergence. The iterative procedure is also restarted several times, as the expansion subspace is more quickly filled; this further increases the number of cycles necessary to reach convergence and is the reason why in many cases the number of matrix−vector products is more than twice as large as the B-1Space variant. A fairer comparison would be to use a subspace limit twice as large as for the other two algorithms, since twice the number of vectors is added at each cycle, but B-Orth would still be the least efficient choice, and the storage requirement would become even larger than for B-Biorth.

B-Biorth is much closer to B-1Space for the right-hand diagonalization than B-Orth. A slightly larger number of iterations is usually required even in well-behaved cases, since close to convergence some vectors in one space can satisfy the convergence criteria, while the corresponding ones in the other space are slightly off for numerical reasons. Thus, a step where the new vectors of both spaces are orthonormalized to each other is performed in order to add the same number of vectors to both spaces as discussed in Section 2. A larger number of vectors than for B-1Space is, on the other hand, always necessary for the left-hand diagonalization, since the same starting point (CIS eigenvectors) is used for both spaces for B-Biorth. Furthermore, numerical instabilities prevented the convergence for the A_1 irrep of acetone with the 6-31+G* basis set and for the B_u irrep of butadiene with the 6-311++G** basis set. Such instabilities arose close to convergence when numerical noise in the creation of new

Large Non-Hermitian Matrices

J. Chem. Theory Comput., Vol. 6, No. 7, 2010 **1969**

Table 1. Number of Matrix−Vector Products for the Ground State (GS) and for Right + Left Eigenvectors with the Three Variants of the Diagonalization Algorithm[a]

	GS	B-1Space	B-Orth	B-Biorth	GS	B-1Space	B-Orth	B-Biorth
			6-31 G*				6-31+G*	
formaldehyde (12)	14	159 + 152	323 + 323	214 + 214	15	178 + 157	364 + 364	429 + 429
ethene (24)	12	292 + 252	579 + 579	308 + 308	13	299 + 254	615 + 615	314 + 314
acetone (12)	16	199 + 176	451 + 451	205 + 205	16	208 + 171	467 + 467	nc +156
butadiene (12)	17	204 + 179	485 + 485	606 + 606	17	201 + 172	449 + 449	203 + 203
			6-31++G**				6-311++G**	
formaldehyde (12)	15	174 + 147	357 + 357	177 + 177	15	174 + 152	363 + 363	182 + 182
ethene (24)	13	324 + 267	660 + 660	335 + 335	13	323 + 266	665 + 665	327 + 327
acetone (12)	16	218 + 177	468 + 468	221 + 221	17	220 + 179	492 + 492	220 + 220
butadiene (12)	17	208 + 168	478 + 478	211 + 211	17	206 + 173	491 + 491	149 + nc
			aug-cc-pVDZ				aug-cc-pVTZ	
formaldehyde (12)	15	178 + 159	388 + 388	183 + 183	15	190 + 165	420 + 420	210 + 210
ethene (24)	13	339 + 289	733 + 733	355 + 355	13	347 + 286	734 + 734	348 + 348
acetone (12)	17	227 + 180	502 + 502	228 + 228	17	228 + 188	521 + 521	515 + 515
butadiene (12)	17	215 + 173	523 + 523	225 + 225	17	208 + 165	486 + 486	213 + 213

[a] Reference wave function is the restricted HF. Total number of excited states is indicated in parentheses next to the molecule, and nc indicates that the diagonalization is not converged for one of the irreps.

Table 2. Number of Matrix−Vector Products for the Ground State (GS) and for the Right + Left Eigenvectors with the Three Variants of the Diagonalization Algorithm[a]

	GS	B-1Space	B-Orth	B-Biorth	GS	B-1Space	B-Orth	B-Biorth
			6-31 G*				6-31+G*	
formaldehyde (12)	17	169 + 154	366 + 366	190 + 190	17	183 + 156	390 + 390	219 + 219
ethene (24)	11	370 + 283	846 + 846	nc+303	12	357 + 299	775 + 775	376 + 376
acetone (12)	22	181 + 152	409 + 409	215 + 215	22	182 + 153	416 + 416	198 + 198
butadiene (12)	17	166 + 141	356 + 356	173 + 173	17	171 + 143	369 + 369	171 + 171
			6-31++G**				6-311++G**	
formaldehyde (12)	17	182 + 155	394 + 394	229 + 229	17	180 + 157	395 + 395	202 + 202
ethene (24)	12	367 + 301	776 + 776	374 + 374	12	360 + 298	776 + 776	368 + 368
acetone (12)	22	185 + 153	420 + 420	191 + 191	22	185 + 156	434 + 434	191 + 191
butadiene (12)	17	172 + 143	369 + 369	175 + 175	17	172 + 145	373 + 373	174 + 174
			aug-cc-pVDZ				aug-cc-pVTZ	
formaldehyde (12)	17	182 + 157	393 + 393	194 + 194	19	179 + 159	394 + 394	206 + 206
ethene (24)	12	367 + 299	790 + 790	377 + 377	13	356 + 293	782 + 782	365 + 365
acetone (12)	22	184 + 155	433 + 433	703 + 703	24	183 + 159	438 + 438	nc + 138
butadiene (12)	17	172 + 144	373 + 373	177 + 177	18	175 + 147	374 + 374	176 + 176

[a] Reference wave function is the unrestricted HF, and the total charge for each molecule is +1. Total number of excited states is indicated in parentheses next to the molecule, and nc indicates that the diagonalization is not converged for one of the irreps.

vectors unevenly propagated in the two subspaces and the algorithm failed in correcting this effect.

The results for the open shell calculations are reported in Table 2. The same trend as in the closed shell case is observed. We note that the B-Biorth algorithm showed numerical instability for the A_g irrep of ethene with the 6-31G* basis and for the A_2 irrep of acetone with the aug-cc-pVTZ basis (which did not converge). Tables 1 and 2 also report the number of cycles necessary for the convergence of the ground state (GS) CCSD equations. These numbers are close to the average number of cycles per state for the EOM-CCSD equations at least for the right space with the B-1Space algorithm, although the convergence criterion for the ground state equations is one order of magnitude tighter than for the excited state calculation. This is due to the increasing difficulty in converging higher states.

As mentioned in Section 2, when the solution for the two spaces is sought separately, the left eigenvectors can be found by solving a linear system of equations (B-LinSys algorithm) once the eigenvalues for the right space are found with the diagonalization. The number of matrix−vector products for this case are reported in Table 3. A larger number of vectors

Table 3. Number of Matrix−Vector Products for the Left Eigenvectors with the B-LinSys Algorithm for the Closed Shell Case

	6-31 G*	6-31+G*	6-31++G**	6-311++G**	aug-cc-pVDZ
formaldehyde	193	215	211	212	223
ethene	342	360	384	379	396
acetone	197	213	232	230	239
butadiene	204	197	209	210	218

than for B-1Space is required. However, the convergence criterion for this algorithm is the root-mean-square (rms) of the norm of the new vectors created in the orthogonalization step ($<\xi \times 10^{-4}$). This is the same as the orthogonalization step for the diagonalization algorithms, but since in B-LinSys no residuals are created at each cycle, there is no other criteria to decide when convergence is reached. In order to test the efficiency of B-LinSys, we reduced the threshold for the formaldehyde case, see Table 4. Although the number of matrix−vector products decreases, the transition properties are not converged for thresholds smaller than $\xi \times 10^{-3}$, with differences of the order of $10^{-4}-10^{-3}$ for the oscillator

1970 *J. Chem. Theory Comput., Vol. 6, No. 7, 2010*

Caricato et al.

Table 4. Number of Matrix−Vector Products for the Left Eigenvectors with the B-LinSys Algorithm for the Closed Shell Formaldehyde by Changing the Convergence Threshold ($\xi = 10^{-5}$) in the Orthonormalization of the New Vectors

	6-31 G*	6-31+G*	6-31++G**	6-311++G**	aug-cc-pVDZ
$\xi \times 10^{-4}$	193	215	211	212	223
$\xi \times 10^{-3}$	170	187	187	187	190
$\xi \times 10^{-2}$	146	155	157	154	153
$\xi \times 10^{-1}$	122	123	122	124	125
$\xi \times 10^{0}$	95	87	87	86	94

strength. With a threshold of $\xi \times 10^{-3}$, B-1Space and B-LinSys are basically equivalent. Therefore, there seems not to be a particular advantage in using the solution of the system of equations over the diagonalization. The latter choice requires the construction of the residuals at each cycle, but this is $O(o^2 v^2)$ work for EOM-CCSD and negligible compared to the $A^\dagger \bar{B}^i$ work, which scales as $O(o^2 v^4 + o^3 v^3)$. For the reasons above and for the practical advantage of using the same code for both spaces, we prefer the B-1Space algorithm to the B-LinSys one.

4. Conclusions

In this paper we present a comparison of three variants of the generalized Davidson algorithm for the iterative diagonalization of large non-Hermitian matrices applied to the EOM-CCSD equations. Two variants seek the right and left eigenvectors simultaneously by using one set of orthonormal trial vectors, B-Orth, or two sets of biorthonormal trial vectors, B-Biorth. A third variant, B-1Space, diagonalizes the matrix from both sides separately and biorthonormalizes the final eigenvectors.

Our numerical tests indicate that the three variants provide the same final results (EOM-CCSD transition energies and properties). The B-1Space option is the most efficient in terms of storage, numerical stability, and convergence rate. The same trend is consistently obtained by varying molecular system, basis set, and reference function (restricted or unrestricted HF). Therefore, the separate left- and right-hand iterative diagonalization is the preferred strategy to find the lowest eigenvalues and eigenvectors of a large non-Hermitian matrix.

Supporting Information Available: Transition energies and oscillator strengths for all the systems are reported in Tables 1−8. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Stanton, J. F.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 7029–7039.

(2) Davidson, E. R. *J. Comput. Phys.* **1975**, *17*, 87–94.

(3) Rettrup, S. *J. Comput. Phys.* **1982**, *45*, 100–107.

(4) Hirao, K.; Nakatsui, H. *J. Comput. Phys.* **1982**, *45*, 246–254.

(5) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Norm, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.

(6) Some elements of the *A* matrix are never explicitly calculated because the terms in the matrix−vector contractions can be properly organized in order to maintain $O(N^6)$ scaling, where N is the number of basis functions, and storage of at most four indexes quantities.[1]

(7) Here we assume that the target roots of the similarity transformed Hamiltonian are real. Although a generic non-Hermitian matrix may have complex eigenvalues, the above assumption is justified in the context of the EOM-CCSD method because the eigenvalues are excitation energies, which are real quantities. If complex excitation energies are found among the target roots, this is an indication that there is a problem with the description of the wave function (for example the reference function may not be stable), and the rate of convergence of the diagonalization algorithm is, therefore, not relevant.

(8) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. *J. Phys. Chem.* **1998**, *109*, 8218–8224.

(9) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. In *Numerical Recipes in Fortran*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 1992; pp 22−98.

(10) Salter, E. A.; Trucks, G. W.; Bartlett, R. J. *J. Chem. Phys.* **1989**, *90*, 1752–1766.

(11) Gauss, J.; Stanton, J. F.; Bartlett, R. J. *J. Chem. Phys.* **1991**, *95*, 2623–2638.

(12) Caricato, M.; Trucks, G. W.; Frisch, M. J.; Wiberg, K. B. *J. Chem. Theory Comput.* **2010**, *6*, 370–383.

CT100111W