View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DSpace at Rice University

Comparison of self-consistent field convergence acceleration techniques

Alejandro J. Garza[1] and Gustavo E. Scuseria[2]
[1]*Department of Chemistry, Rice University, Houston, Texas 77251-1892, USA*
[2]*Department of Chemistry and Department of Physics and Astronomy, Rice University, Houston, Texas 77251-1892, USA and Chemistry Department, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia*

(Received 21 May 2012; accepted 17 July 2012; published online 3 August 2012)

The recently proposed ADIIS and LIST methods for accelerating self-consistent field (SCF) convergence are compared to the previously proposed energy-DIIS (EDIIS) + DIIS technique. We here show mathematically that the ADIIS functional is identical to EDIIS for Hartree-Fock wavefunctions. Convergence failures of EDIIS + DIIS reported in the literature are not reproduced with our codes. We also show that when correctly implemented, the EDIIS + DIIS method is generally better than the LIST methods, at least for the cases previously examined in the literature. We conclude that, among the family of DIIS methods, EDIIS + DIIS remains the method of choice for SCF convergence acceleration. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4740249]

I. INTRODUCTION

In an orthogonal basis, the self-consistent field (SCF) problem can be formulated as a nonlinear eigenvalue equation of the form

$$F(D)C = C\epsilon, \tag{1}$$

where $F(D)$ is the Fock matrix, $D = CC^\dagger$ is the density matrix, and $\epsilon = \mathrm{Diag}(\epsilon_1, \ldots, \epsilon_N)$ is the diagonal orbital energy matrix. The SCF equations appear in two fundamental approaches for tackling the electronic structure problem: the Hartree-Fock (HF) method and Kohn-Sham density functional theory (KS-DFT). Both approaches are of great practical importance in molecular applications; HF is usually a required first step for more accurate correlated methods while KS-DFT remains favorite for large scale simulations because it accounts for electron correlations at mean-field computational cost. It is therefore most desirable to have in hand a reliable, robust, and efficient "black box" method to deal with the SCF convergence problem.

The simplest procedure for solving Eq. (1) is the fixed-point algorithm, which consists of generating a sequence (D_k) in the set $\mathcal{P}_N = \{D \in \mathcal{M}_S(N_b), D^2 = D, \mathrm{Tr}(D) = N\}$ such that

$$\begin{cases} F(D_k)C_{k+1} = C_{k+1}\epsilon_{k+1}, \\ C_{k+1}^\dagger C_{k+1} = I_N, \\ D_{k+1} = C_{k+1}C_{k+1}^\dagger, \end{cases} \tag{2}$$

where N_b and N are, respectively, the number of basis functions and electrons, so that $C_k \in \mathcal{M}(N_b, N)$. However, the convergence properties of this algorithm are seldom satisfactory as it usually oscillates between states (bifurcations) that are not solutions of the SCF equations.[1,2] Furthermore, a substantial number of $F(D_k)$ diagonalizations are often required to reach convergence; this operation scales as $\mathcal{O}(N_b^3)$ and becomes the computational bottleneck in large systems.[3]

Improved SCF convergence can be achieved by retaining data from previous iterations and using it to build a better guess for the density matrix. Perhaps the most successful strategy consists in constructing a linear combination of previously iterated density matrices

$$\tilde{D}_{k+1} = \sum_{i=0}^{k+1} c_i D_i, \tag{3}$$

where $\tilde{D}$ is a pseudodensity matrix satisfying $\mathrm{Tr}(\tilde{D}) = N$. Specific methods differ by the way in which the coefficients of the linear expansion are chosen. Pulay's direct inversion of the iterative subspace,[4,5] commonly known as DIIS, is possibly the most popular SCF acceleration technique following this general philosophy. Nevertheless, the energy-DIIS (EDIIS) method of Kudin, Cancès, and Scuseria[6] improves upon DIIS when far away from convergence. A combination of both methods, EDIIS + DIIS, is usually most efficient and has been the default option for SCF convergence in the *Gaussian* suite of programs for many years.[7]

A number of alternative schemes within the general DIIS philosophy have recently been proposed in the literature. Specifically, these are the augmented-DIIS (ADIIS) (Ref. 8) and LIST methods.[9,10] These methods were found superior to the previously proposed DIIS + EDIIS procedure.[6] As shown in this paper, our results do not support these claims.

To make this article self-contained, we here present algebraic details of the ADIIS and LIST methods in light of their DIIS predecessors. We first compare the performance of the LIST methods against DIIS + EDIIS. In Refs. 9 and 10, the LIST methods were benchmarked only against DIIS. Our LIST implementation correctly reproduces the available results in the literature.[9,10] We find no evidence of LIST superiority. Quite to the contrary, EDIIS + DIIS supersedes the LIST methods in all of our tested cases. Regarding ADIIS, we mathematically prove that ADIIS and EDIIS minimize the same functional for Hartree-Fock (HF) wavefunctions and thus should be identical. We numerically verify this conclusion in actual SCF calculations. Previous benchmark comparison between ADIIS and EDIIS did not reflect this

 © 2012 American Institute of Physics

equivalence. On the other hand, our EDIIS results match the Hartree-Fock ADIIS results of Ref. 8. This forces us to conclude that EDIIS was not correctly implemented in Ref. 8.

II. SCF CONVERGENCE ACCELERATION TECHNIQUES

A. DIIS

In the DIIS algorithm, the set of coefficients $\{c_i\}$, is chosen to minimize a suitable error vector. The commutator $[F, D]$ is usually chosen as such an error vector since $[F, D] = 0$ is a necessary and sufficient condition for an SCF solution.[5,11] Mathematically,

$$\{c_i\} = \arg\inf \left\{ \left\| \sum_{i=0}^{k+1} c_i [F_i, D_i] \right\|^2, \sum_{i=0}^{k+1} c_i = 1 \right\}, \quad (4)$$

where $||\cdot||$ denotes the Frobenius norm and the constraint $\sum c_i = 1$ ensures that $\mathrm{Tr}(\tilde{D}) = N$. This particular form of DIIS is sometimes known as commutator-DIIS or CDIIS. The constrained minimization problem in Eq. (4) is solved using standard Lagrangian calculus; the result is a system of linear equations[12]

$$\begin{pmatrix} \mathbf{B} & \mathbf{1}^T \\ \mathbf{1} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}, \quad (5)$$

where the elements of $\mathbf{B}$ are $b_{ij} = [F_i, D_i] \cdot [F_j, D_j]$, $\mathbf{c} = (c_0, c_1, \ldots, c_{k+1})$, $\mathbf{1} = (1, 1, \ldots, 1)$, and λ is a Lagrange multiplier. The DIIS algorithm is often very efficient[1] and has been shown to correspond to a projected quasi-Newton method,[12,13] justifying its present relevance. Nevertheless, CDIIS does fail in some cases because the minimization of $||[F, D]||$ does not force convergence. Minimization of the energy does force convergence, and this is the motivation behind EDIIS.

B. EDIIS

The EDIIS algorithm is formally similar to DIIS, but instead of minimizing $||[F, D]||$ one minimizes the Hartree-Fock energy functional $E^{\mathrm{HF}}(D)$; see Eq. (A1). The coefficients of the linear expansion are thus given by

$$\{c_i\} = \arg\inf \left\{ E^{\mathrm{HF}} \left(\sum_{i=0}^{k+1} c_i D_i \right), \sum_{i=0}^{k+1} c_i = 1, c_i \in [0, 1] \right\}. \quad (6)$$

Note that there is an additional constraint compared to DIIS, namely, $c_i \in [0, 1]$. This restriction ensures that $\tilde{D}_k$ is in the convex set $\tilde{\mathcal{P}}_N = \{\tilde{D} \in \mathcal{M}_S(N_b), \tilde{D}^2 \leq \tilde{D}, \mathrm{Tr}(\tilde{D}) = N\}$. Because of the *aufbau* principle, all the local minima of $E^{\mathrm{HF}}(D)$ in $\tilde{\mathcal{P}}_N$ are in fact in $\mathcal{P}_N$ and hence the minimization of the energy of $\tilde{D} \in \tilde{\mathcal{P}}_N$ guarantees the convergence of EDIIS for the HF method. In DIIS, the coefficients are allowed to be negative and hence it is also possible to have $|c_i| \geq 1$ so that $\tilde{D} \notin \tilde{\mathcal{P}}_N$.

Not surprisingly, the $c_i \in [0, 1]$ restriction also comes with a disadvantage: EDIIS can only interpolate between the matrices in the linear expansion, so the chances of finding the correct minimum decrease near the convergence region. Thus, EDIIS is most efficient when starting from a poor guess since the large density jumps span greater regions of the density matrix space. In contrast, DIIS performs much better near the convergence region. Therefore, the EDIIS + DIIS scheme of Ref. 6 combines EDIIS with DIIS in a way such that EDIIS is used when the DIIS error (as defined above) is large and DIIS is used when this error is small.

C. ADIIS

The ADIIS method follows the same philosophy of EDIIS. As explained in Ref. 8, the ADIIS coefficients are chosen to minimize an energy functional that is an approximation to the augmented Roothaan Hall energy function[13,14] $E^{\mathrm{ARH}}(D)$ (see Eq. (A3)), in the convex set $\tilde{\mathcal{P}}_N$

$$\{c_i\} = \arg\inf \left\{ E^{\mathrm{ARH}} \left(\sum_{i=0}^{k+1} c_i D_i \right), \sum_{i=0}^{k+1} c_i = 1, c_i \in [0, 1] \right\}. \quad (7)$$

In analogy with EDIIS, ADIIS is an interpolation scheme and performs well when far from convergence but slows down near the solution. Thus, an ADIIS + DIIS method analogous to EDIIS + DIIS was recommended over the use of pure ADIIS in Ref. 8. Even though not obvious at first glance, we must point out that the ADIIS energy functional is mathematically identical to EDIIS for HF wave functions. A proof is presented in the Appendix. Furthermore, while the two functionals need not be identical for KS-DFT, an assumption of linearity in the derivations presented in Ref. 8 make the two functionals practically the same. Our benchmarks below demonstrate these points numerically (see also further discussion in Sec. IV B).

D. LISTi, LISTd, and LISTb

The LIST methods are derived from the corrected Hohenberg-Kohn-Sham functional,[15] $E^{\mathrm{cHKS}}(D)$, but make approximations to arrive at a system of linear equations very similar to DIIS. For LISTi, one solves a linear system of equations of the same form as Eq. (5):

$$\begin{pmatrix} \mathbf{G} & \mathbf{1}^T \\ \mathbf{1} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} \quad (8)$$

except that $\mathbf{G}$ differs from $\mathbf{B}$ in its elements $g_{ij} = \mathrm{Tr}((D_i^{\mathrm{out}} - D_i^{\mathrm{in}})(F_j^{\mathrm{out}} - F_j^{\mathrm{in}}))$. The LISTi equations are obtained by enforcing $E_0^{\mathrm{cHKS}} = E_1^{\mathrm{cHKS}} = \cdots = E_{k+1}^{\mathrm{cHKS}} = E(D_{k+1})$ and $D_i^{\mathrm{out}} = D_i^{\mathrm{in}}$, $\forall i$. In LISTd (direct LIST), the latter condition is not enforced; this leads to another linear system

$$\begin{pmatrix} \mathbf{A} & \mathbf{1}^T \\ \mathbf{1} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ -E \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}, \quad (9)$$

where $a_{ij} = E[D_i^{\mathrm{out}}] + \mathrm{Tr}((D_j^{\mathrm{out}} - D_i^{\mathrm{out}})(F_i^{\mathrm{out}} - F_i^{\mathrm{in}}))$.

The "better" LIST (LISTb) method derives from the transposition of the LISTd equations. Thus, the LISTb system is also described by Eq. (9) except that $\mathbf{A}$ is substituted by $\mathbf{A}^T$. This transposition leads to a different set of coefficients since $\mathbf{A} \neq \mathbf{A}^T$.

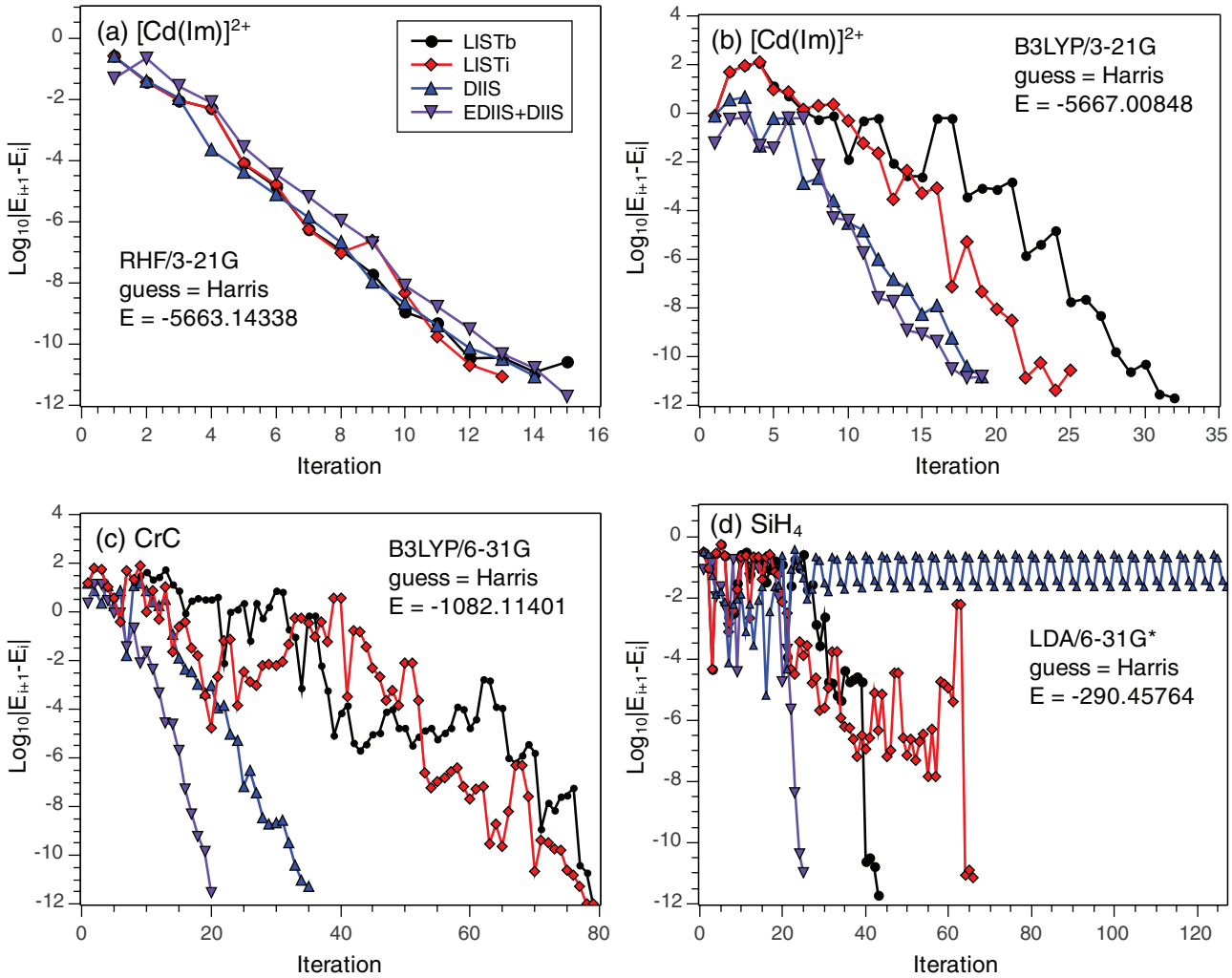

FIG. 1. Comparison of SCF energy convergence for LISTi, LISTb, DIIS, and EDIIS + DIIS in challenging cases. All calculations retained four matrices in the iterative process to conform with benchmarks in Refs. 9 and 10. The legend in panel (a) applies to all the panels in this figure.

III. COMPUTATIONAL DETAILS

All calculations reported in this paper were carried out using the GAUSSIAN 09 suite of programs.[7] The LISTi, LISTd, and LISTb algorithms were implemented there and benchmarked against the DIIS and EDIIS + DIIS methods already available in the package. For simplicity, default options of GAUSSIAN 09 were used (unless otherwise indicated); this includes the Harris functional diagonalization as an initial guess, and tight convergence criterion which stops the SCF when the RMS deviation between elements of successive density matrices is smaller than 10^{-8} and the change in energy is below 10^{-6}. The only parameter varied from its default option was the dimension of the DIIS subspace (i.e., the number of matrices included in the linear expansion); the *Gaussian* default of 20 matrices was reduced to 4 and 6 to permit comparison with literature results.

In the default EDIIS + DIIS procedure in *Gaussian*, EDIIS is used when the largest DIIS error (errMax) is greater than 10^{-1} a.u. but DIIS is employed when this error goes below 10^{-4} a.u. In between these values, the EDIIS and DIIS coefficients are weighted such that $\mathbf{c} = 10(\text{errMax})\mathbf{c}^{\text{EDIIS}} + [1 - 10(\text{errMax})]\mathbf{c}^{\text{DIIS}}$; however, if the error of the last cycle is 10% greater than the minimum error, pure EDIIS is used. Also, the default algorithm does not use any Fermi broadening[16] and utilizes damping only during the first iteration.

In all implemented methods, the Fock matrix, rather than the density matrix, is extrapolated (or interpolated). The extrapolated Fock matrix is then diagonalized to obtain an idempotent density matrix. This approach is exact for HF but not for KS-DFT, where the Fock matrix is not linear with respect to the density matrix. However, this approximation has already been shown to be quite reasonable[6,17] and it has become standard in the implementation of DIIS-like methods.

The benchmarks presented in this paper are deemed representative of challenging SCF convergence cases and were taken from previous work of our research group and others. The examples include high-symmetry cases containing transition metal and actinide species where multiple atomic states compete in bonding and are prone to spontaneous symmetry breaking. Specifically, the geometries of the $[\text{Cd(Im)}]^{2+}$ (cadmium-imidazole) and $Ru_4(CO)$ complexes were taken from Ref. 8; the SiH_4 molecule has three bonds with length of 1.48 Å and an extended bond of 4.00 Å as

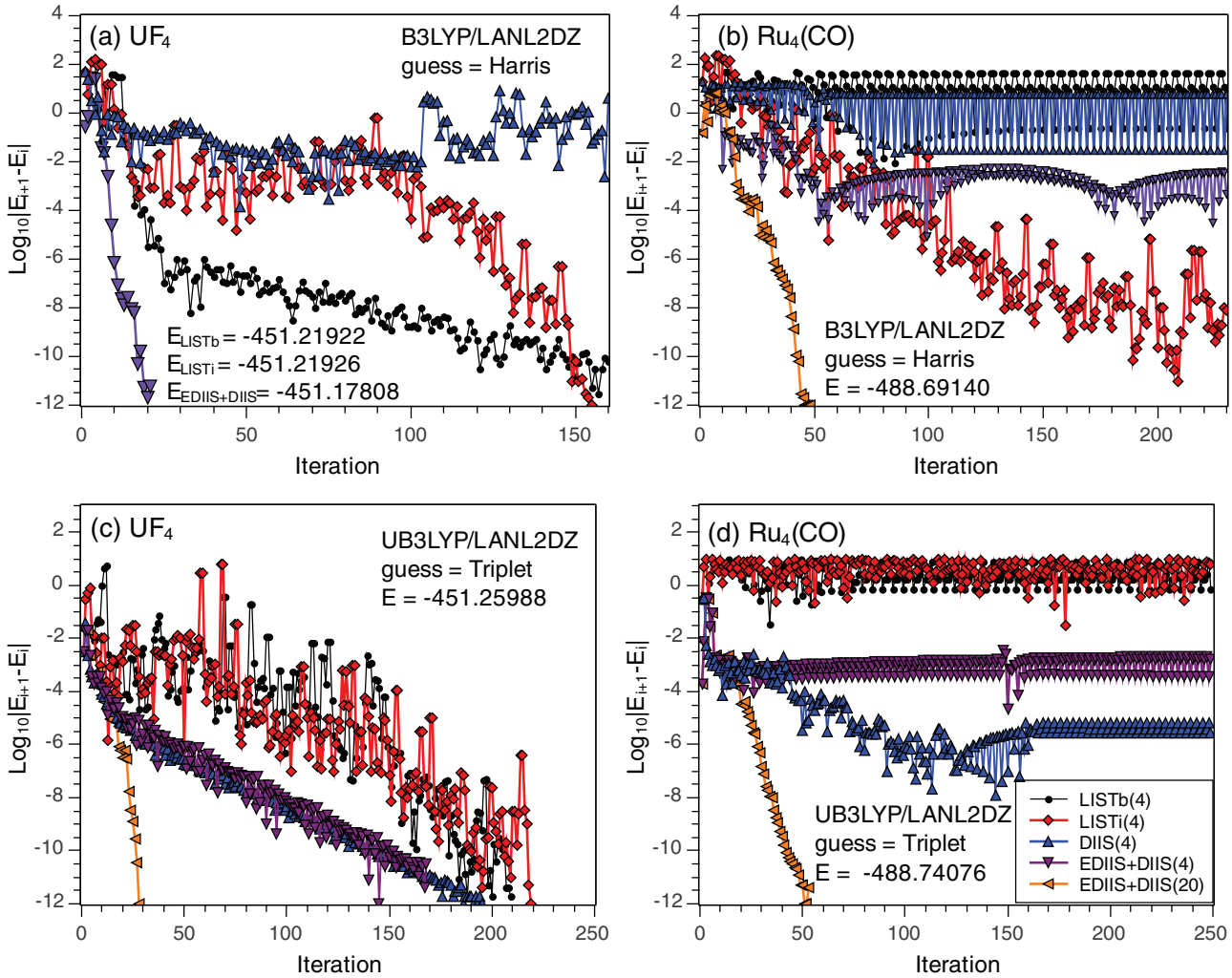

FIG. 2. Energy convergence for the restricted B3LYP and broken symmetry ($m_s = 0$) UB3LYP solutions of tetrahedral UF_4 and $Ru_4(CO)$. The initial guess for UB3LYP is obtained from a converged calculation of the triplet ($m_s = 1$) states. The legend in panel (d) applies to all the figures; the dimension of the DIIS subspace is indicated in parenthesis.

in Ref. 9; the CrC and UF_4 bond lengths are 2.00 Å and 1.98 Å as in Ref. 6. All calculations (except when noted) were carried out for closed shell configurations with methods and bases specified in the figure panels, where converged energies (in Hartrees) and chosen initial guesses are also reported.

IV. RESULTS AND DISCUSSION

A. LIST vs. EDIIS + DIIS

Figures 1 and 2 compare the SCF energy convergence for LISTi, LISTb, DIIS, and EDIIS + DIIS. Four Fock matrices are kept in the iterative process of all calculations in Fig. 1; this number was previously found to be adequate for LIST.[9,10] In Fig. 2, we have included 20 Fock matrices in EDIIS + DIIS to illustrate its impact on the convergence pattern. We have not included the data for LISTd as we found it to be inferior to the other LIST methods, in agreement with the observations of Refs. 9 and 10.

Based on the results presented in Figs. 1 and 2, EDIIS + DIIS appears to be the most efficient of the convergence acceleration methods. For $[Cd(Im)]^{2+}$ at the RHF/3-21G level, all four methods display a similar rate of convergence. However, EDIIS + DIIS is faster for $[Cd(Im)]^{2+}$ at the B3LYP/3-21G level, as well as for CrC at B3LYP/6-31G, and SiH_4 at LDA/6-31G*.

The benchmark calculations on UF_4 and $Ru_4(CO)$, shown in Fig. 2, deserve some special comment. For UF_4, the LIST and EDIIS + DIIS methods converge to three different solutions when using the Harris guess. Further inspection reveals that the lowest energy state for tetrahedral UF_4 is a broken symmetry UB3LYP $m_s = 1$ state and that there is also a broken symmetry $m_s = 0$ solution much lower in energy than any of the restricted singlet solutions. We loosely refer to these broken symmetry solutions as "singlet" ($m_s = 0$) and "triplet" ($m_s = 1$) even though technically, they are both contaminated by higher spin states. More importantly, the existence of these unrestricted solutions means that the states quoted in Fig. 2(a) are high-energy stationary points. Figure 2(c) shows the energy convergence for the unrestricted UF_4 calculations using the broken-symmetry triplet state as the initial guess, a choice that facilitates locating the lowest singlet UB3LYP solution for this case. Here, all methods converge to the same energy, with EDIIS + DIIS being the fastest.

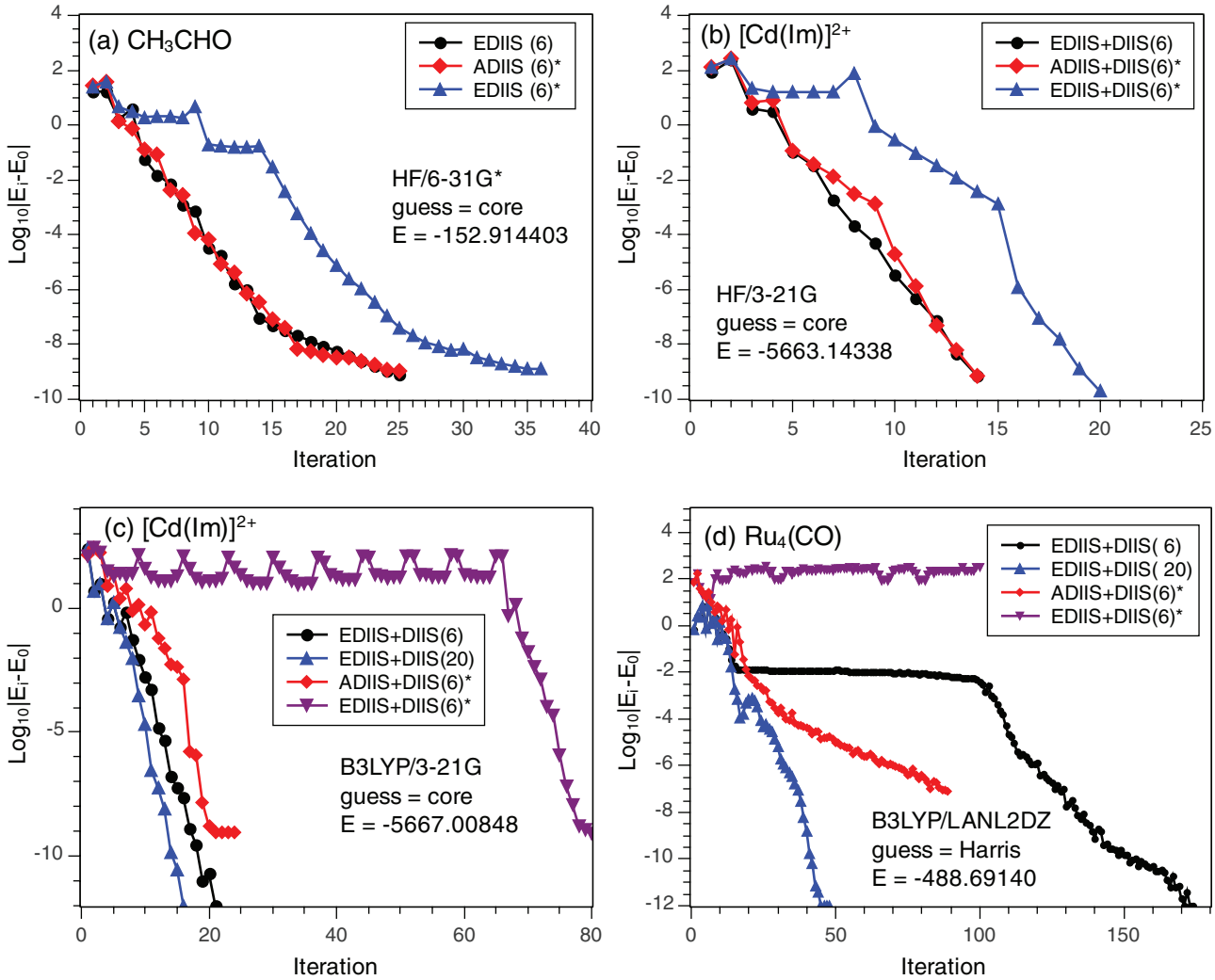

FIG. 3. Comparison of SCF energy convergence for EDIIS, ADIIS, EDIIS + DIIS, and ADIIS + DIIS in a set of challenging cases. The data marked with an asterisk was extracted from Ref. 8. The number of matrices in the iterative process is indicated in parenthesis.

The situation is similarly complicated for $Ru_4(CO)$. Starting from the Harris guess (Fig. 2(b)), only EDIIS + DIIS with 20 matrices is able to converge this case (B3LYP/LANL2DZ) to a restricted solution, which is not the lowest energy one among restricted singlets. Here too, a broken symmetry triplet is the lowest energy state. As shown in Fig. 2(d), when singlet UB3LYP calculations are started from the broken-symmetry triplet, only EDIIS + DIIS manages to converge. Our lowest RB3LYP solution ($E = -488.70477$) is obtained using the triplet as an initial guess. Starting from a "superposition of atomic densities" guess, a lower energy solution ($E = -488.71068$) was reported in Refs. 8 and 10 . It is likely that this small energy difference arises from different versions of the B3LYP functional. Evidently, a word of caution is in order when trying to converge triplet unstable cases; EDIIS + DIIS can easily get trapped on high-energy solutions and most other methods fail to converge.

The above examples also indicate that an increase of the DIIS subspace from 4 to 20 matrices greatly improves the rate of convergence for EDIIS + DIIS. In fact, while EDIIS + DIIS benefits from a larger DIIS subspace, this is not true for LIST. For example, changing the DIIS subspace from 4 to 5 matrices in LISTi can increase the number of iterations to convergence from 60 to more than 760,[10] a fact that obviously casts doubts on the robustness of the method.

Another problem encountered with LIST, especially when using tight convergence criteria, is its tendency of leading to systems of linear equations that are ill-conditioned. This weakness was noted for LISTd (Ref. 9) and it was claimed that LISTb alleviated it.[10] However, LISTb only transposes the **A** matrix of LISTd in Eq. (9), an operation that neither changes the condition number of **A** nor its singular values. Hence, the LISTd and LISTb systems are equally ill-conditioned. Furthermore, there is no theoretical argument for transposing **A** and there is currently no satisfactory explanation for the better performance of LISTb over LISTd.

Also, in contrast to EDIIS or ADIIS, the LIST algorithms do not minimize the energy and do not impose the condition $\tilde{D} \in \tilde{\mathcal{P}}$. Therefore, LIST does neither guarantee convergence for HF nor for KS-DFT. LIST is an extrapolation scheme bearing resemblance to DIIS; in fact, it can be formulated in the same way as the original DIIS proposed by Pulay[4] by choosing an alternative error vector instead of the commutator $[F, D]$.

We conclude this section noting two things. The first one is that for the test cases in the original LIST papers,[9,10] the E^{cHKS} functional,[15] rather than the actual energy, was evaluated at each iteration. Because E^{cHKS} is designed to approximate the converged energy, LIST methods appear to converge faster than they actually do when measuring convergence on the actual energy or density matrix. The second thing to note is that LIST was previously benchmarked only against DIIS; the authors of Refs. 9 and 10 concluded that LIST rivaled ADIIS + DIIS and surpassed EDIIS + DIIS based on the claim of Ref. 8 that ADIIS was superior to EDIIS. However, as shown in Sec. IV B, we find no evidence that ADIIS + DIIS is superior to EDIIS + DIIS. As a matter of fact, the methods should be identical for HF wave functions.

B. ADIIS vs. EDIIS

The calculations using EDIIS and EDIIS + DIIS in this section were carried out as described above; only the dimension of the DIIS subspace was adjusted in some cases. The data marked with an asterisk (*) in Fig. 3 were extracted from Ref. 8 for comparison purposes with our own results. Figure 3(a) shows that for acetaldehyde with HF/6-31G*, the rates of convergence of EDIIS and ADIIS are virtually identical, as they should be. However, the EDIIS data taken from Ref. 8 differ significantly. Because the EDIIS and ADIIS methods are mathematically equivalent for HF wavefunctions (see the Appendix), we have no explanation for the EDIIS curves reported in Ref. 8 other than that the algorithm not being correctly implemented.

The situation is similar in Fig. 3(b), which shows energy convergence for the $[\mathrm{Cd(Im)}]^{2+}$ complex at the HF/3-21G level. The first six points for ADIIS + DIIS and EDIIS + DIIS correspond to pure ADIIS and EDIIS, respectively, and are practically identical. The region in which these curves differ corresponds to the range where ADIIS + DIIS uses only DIIS, while EDIIS + DIIS weights the EDIIS and DIIS coefficients based on the DIIS error, as discussed above. The EDIIS + DIIS data from Ref. 8 are again incongruent with our results and with the demonstrated equivalence between ADIIS and EDIIS for HF wave functions.

EDIIS and ADIIS are not equivalent for the KS-DFT calculations in Figs. 3(c) and 3(d). We observe that for B3LYP/3-21G $[\mathrm{Cd(Im)}]^{2+}$ (Fig. 3(c)), with either 6 or 20 matrices in the linear expansion, EDIIS + DIIS converges faster than ADIIS + EDIIS. For $Ru_4(CO)$ with B3LYP/LANL2DZ (Fig. 3(d)), ADIIS + DIIS can converge faster than EDIIS + DIIS when retaining 6 matrices. Nonetheless, EDIIS + DIIS benefits greatly from a larger DIIS subspace ($n = 20$) and easily surpasses ADIIS + DIIS when this recommendation is followed.[17] Once more, the poorly performing EDIIS + DIIS results of Ref. 8 are not reproduced here. We note that the initial guess for $Ru_4(CO)$ in Ref. 8 was constructed from a superposition of atomic densities, while we are here diagonalizing the Harris functional.

Finally, we must point out that it was argued in Ref. 8 that ADIIS was accurate for KS-DFT if the quasi-Newton condition imposed to obtain a suitable approximate expression for E^{ARH} was sufficient. At the same time, these authors noted that EDIIS is just an approximation for KS-DFT because the exchange-correlation term is not linear in the density matrix. However, the expression minimized by ADIIS

$$E^{\mathrm{ARH}}\left(\sum c_i D_i\right) = E(D_n) + 2\sum_{i=1}^{n} c_i \mathrm{Tr}((D_i - D_n)F_n) + \sum_{i,j=1}^{n} c_i c_j \mathrm{Tr}((D_i - D_n)(F_j - F_n)) \tag{10}$$

is exact *if and only if* the Fock matrix has a linear dependence in the density matrix (see Eq. (A3)), which is not true for KS-DFT. Thus, ADIIS does not exactly minimize E^{ARH} and might not be accurate for KS-DFT even if the quasi-Newton condition is sufficient. For KS-DFT, the Fock matrix is

$$F \approx h + J(D) + \frac{\delta E_{xc}(D)}{\delta D}, \tag{11}$$

where h and $J(D)$ contain, the core and two-electron Coulomb interaction terms, respectively. As noted in Ref. 17, $E_{xc}(D)$ and $\delta E_{xc}(D)/\delta D$ are numerically smaller than the other terms contributing to the KS-DFT energy and matrices. Given that ADIIS and EDIIS are equivalent for HF (see the Appendix), they would differ only by these smaller $E_{xc}(D)$ and $\delta E_{xc}(D)/\delta D$ terms. Our results indicate that EDIIS + DIIS, with a suitable dimension for the DIIS subspace and a reasonable initial guess, can generally perform similarly or better than ADIIS + DIIS for KS-DFT.

V. CONCLUSIONS

The recently proposed ADIIS and LIST procedures do not appear to outperform the previously proposed EDIIS + DIIS method. In fact, ADIIS is mathematically equivalent to EDIIS for HF wave functions; previous benchmarks and claims in the literature do not agree with our conclusions. A correctly implemented EDIIS + DIIS method, with a suitable dimension for the DIIS subspace and a reasonable initial guess, is similar or better than ADIIS + DIIS or any of the LIST methods, at least for all the cases examined in this paper, which are the same set of benchmarks used in previous studies. We conclude that, among the family of DIIS methods, EDIIS + DIIS remains the method of choice for SCF convergence acceleration.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation CHE-1110884 and the Welch Foundation (C-0036). We thank Dr. Konstantin Kudin for valuable suggestions.

APPENDIX: PROOF OF THE EQUIVALENCE BETWEEN ADIIS AND EDIIS FOR HF

The EDIIS and ADIIS methods minimize, respectively, $E^{\mathrm{HF}}(\sum c_i D_i)$ and $E^{\mathrm{ARH}}(\sum c_i D_i)$ as a function of c_i constrained to satisfy $\sum c_i = 1$ and $c_i \geq 0$. Thus, we can

prove the equivalence between ADIIS and DIIS by showing that $E^{\mathrm{HF}}(\sum c_i D_i) = E^{\mathrm{ARH}}(\sum c_i D_i)$ when $E(D) = E^{\mathrm{HF}}(D)$.

We start by writing down the closed-shell restricted HF energy functional (extension to the general case is straightforward)

$$E^{\mathrm{HF}}(D) = 2\mathrm{Tr}(hD) + \mathrm{Tr}(G(D)D), \tag{A1}$$

where $h \in \mathcal{M}_S(N_b)$ is the matrix of the core Hamiltonian and $D \in \mathcal{M}_S(N_b)$ is the density matrix. The elements of the $G(D)$ matrix are

$$\begin{aligned} G(D)_{\mu\nu} &= \sum_{\kappa,\lambda=1}^{N_b} (\mu\nu|\kappa\lambda) D_{\kappa\lambda} - \sum_{\kappa,\lambda=1}^{N_b} (\mu\lambda|\nu\kappa) D_{\kappa\lambda}, \\ &= \sum_{\kappa,\lambda=1}^{N_b} (\mu\nu||\kappa\lambda) D_{\kappa\lambda}, \end{aligned} \tag{A2}$$

where we denote $(\mu\nu||\kappa\lambda) = (\mu\nu|\kappa\lambda) - (\mu\lambda|\nu\kappa)$.

Similarly, the ARH energy functional (as defined in Ref. 8, which is not strictly the same as in Refs. 13 and 14) is

$$\begin{aligned} E^{\mathrm{ARH}}(D) = E^{\mathrm{HF}}(D_n) &+ 2\mathrm{Tr}((D - D_n)F(D_n)) \\ &+ \mathrm{Tr}((D - D_n)(F(D) - F(D_n)), \end{aligned} \tag{A3}$$

where $F(D_n)$ is the Fock matrix

$$F(D) = h + G(D). \tag{A4}$$

We now expand $E^{\mathrm{ARH}}(D)$ using the definitions of $E^{\mathrm{HF}}(D)$ and $F(D)$ in Eqs. (A1) and (A4). Trivial manipulations yield

$$\begin{aligned} E^{\mathrm{ARH}}(D) &= 2\mathrm{Tr}(hD) + \mathrm{Tr}(G(D)D) \\ &+ \mathrm{Tr}(DG(D_n)) - \mathrm{Tr}(D_n G(D)). \end{aligned} \tag{A5}$$

Since D is a linear combination of previously iterated density matrices, $\tilde{D}_n = \sum_{i=1}^n c_i D_i$, ADIIS minimizes

$$\begin{aligned} E^{\mathrm{ARH}}\left(\tilde{D}_n\right) &= 2\sum_{i=1}^n c_i \mathrm{Tr}(hD_i) + \sum_{i,j=1}^n c_i c_j \mathrm{Tr}(G(D_i)D_j) \\ &+ \sum_{i=1}^n c_i \left[\mathrm{Tr}(D_i G(D_n)) - \mathrm{Tr}(D_n G(D_i))\right]. \end{aligned} \tag{A6}$$

However, one should note that

$$\begin{aligned} \mathrm{Tr}(G(D_n)D_i) &= \sum_{\mu,\nu=1}^{N_b} G(D_n)_{\mu\nu}(D_i)_{\mu\nu}, \\ &= \sum_{\mu,\nu=1}^{N_b} \sum_{\kappa,\lambda=1}^{N_b} (\mu\nu||\kappa\lambda)(D_n)_{\kappa\lambda}(D_i)_{\mu\nu}, \\ &= \mathrm{Tr}(G(D_i)D_n) \end{aligned} \tag{A7}$$

and hence Eq. (A6) becomes

$$\begin{aligned} E^{\mathrm{ARH}}\left(\tilde{D}_n\right) &= 2\sum_{i=1}^n c_i \mathrm{Tr}(hD_i) + \sum_{i,j=1}^n c_i c_j \mathrm{Tr}(G(D_i)D_j), \\ &= E^{\mathrm{HF}}\left(\tilde{D}_n\right), \end{aligned} \tag{A8}$$

which is precisely the same functional minimized in EDIIS.

[1] E. Cancès, M. Defranceschi, W. Kutzelnigg, C. Le Bris, and Y. Maday, *Handbook of Numerical Analysis: Special Volume: Computational Chemistry*, edited by P. G. Ciarlet and C. Le Bris (Elsevier, Amsterdam, 2003), Vol. X, pp. 116–130.
[2] M. A. Natiello and G. E. Scuseria, Int. J. Quantum Chem. **26**, 1039 (1984).
[3] G. E. Scuseria, J. Phys. Chem. **A 103**, 4782 (1999).
[4] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[5] P. Pulay, J. Comput. Chem. **3**, 556 (1982).
[6] K. N. Kudin, G. E. Scuseria, and E. Cancès, J. Chem. Phys. **116**, 8255 (2002).
[7] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 09, Revision B.1, Gaussian, Inc., Wallingford, CT, 2009.
[8] X. Hu and W. Yang, J. Chem. Phys. **132**, 054109 (2010).
[9] Y. A. Wang, C. Y. Yam, Y. K. Chen, and G. Chen, J. Chem. Phys. **134**, 241103 (2011).
[10] Y. K. Chen and Y. A. Wang, J. Chem. Theory Comput. **7**, 3045 (2011).
[11] R. McWeeny and B. T. Sutcliffe, *Methods of Molecular Quantum Mechanics* (Academic, London, 1969).
[12] T. Rohwedder and R. Schneider, J. Math. Chem. **49**, 9 (2011).
[13] S. Høst, J. Olsen, B. Jansik, L. Thøgersen, P. Jørgensen, and T. Helgaker, J. Chem. Phys. **129**, 124106 (2008).
[14] S. Høst, B. Jansik, J. Olsen, P. Jørgensen, S. Reine, and T. Helgaker, Phys. Chem. Chem. Phys. **10**, 5344 (2008).
[15] Y. A. Zhang and Y. A. Wang, J. Chem. Phys. **130**, 144116 (2009).
[16] A. Rabuck and G. E. Scuseria, J. Chem. Phys. **110**, 695 (1999).
[17] K. N. Kudin and G. E. Scuseria, ESAIM, Math. Model. Numer. Anal. **41**, 281 (2007).