Computation of precise two-electron correlation energies with imprecise Hartree–Fock orbitals

Edward F. Valeev [a,b]

[a] *Center for Computational Molecular Science and Technology, Georgia Institute of Technology, Atlanta, GA 30332, USA*
[b] *Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

Abstract

Recently developed linear R12-type explicitly correlated methodology promises few-percent basis set errors with small, double-ζ quality, Hartree–Fock basis sets. We present an MP2-R12 method which achieves the two-electron basis limit as closely as desired using any finite Hartree–Fock basis set. We found that the two-electron correlation energy can be computed precisely from a very approximate Hartree–Fock function. The error in the two-electron correlation energy due to the finite Hartree–Fock basis set was averaged over a set of 20 small molecules. The mean error was found to be small, at most 1% for a double-ζ basis set, and quickly decaying.
© 2005 Elsevier B.V. All rights reserved.

1. Introduction

Explicitly correlated wave function methods [1] have only recently reached the efficiency necessary for chemical applications on molecules with more than a few atoms. These methods promise to reduce the basis set error in the electron correlation energy to below 1 kcal/mol[1]. Such level of *precision* is completely out of reach of conventional methods, primarily due to the inefficient description of the wave function at short interelectronic distances (Coulomb hole). Valence MP2 energy of the water molecule, for example, is in error by 25 and 11 kcal/mol with standard correlation consistent cc-pVTZ and cc-pVQZ basis sets, respectively [2]. The energy extrapolation, e.g., due to Helgaker et al. [3], can reduce the basis set error to below 1 kcal/mol. Such corrections, however, must be calibrated for each property, have limited accuracy [4], and require a quadruple-ζ basis set calculation [5] to attain better than 1 kcal/mol precision for the energy.

The explicitly correlated methods allow smaller basis set errors than the conventional methods because the two-elec-

tron basis includes terms with explicit dependence on r_{12}. Such basis set can model the Coulomb hole more efficiently than the products of one-particle functions [6]. Unfortunately, four- and, sometimes, five-electron integrals appear in explicitly correlated methods. Linear R12 methods of Kutzelnigg, Klopper, and co-workers are currently most practical explicitly correlated methods for chemical computations due to the ingenious use of the resolution of the identity (RI) and other approximations to avoid computation of three- and higher-electron integrals [7,8]. Remaining four-center two-electron integrals, although non-standard, can be readily evaluated in Gaussian basis using standard (recursive) techniques [9,10]. Density fitting can lower the cost of integrals even further [11,12], by analogy with the conventional methodology.

Linear R12 methods which do not use the Hartree–Fock basis set for RI (ABS [13], CABS [14], and numerical quadratures [15]) can be used with an arbitrary set of Hartree–Fock orbitals. Unfortunately, the basis set error is still significant with small Hartree–Fock basis sets, e.g., 10% for valence cc-pVDZ MP2-R12 energy of Ne atom [13]. The origin of this error is manifold: (1) resolution of the identity for the integrals, (2) extended and generalized Brillouin conditions, (3) exchange commutator approximation, and (4) residual two-particle basis incompleteness. Recent

---

[1] 1 kcal/mol is known as *chemical* accuracy, because it corresponds roughly to the value of kT at room temperature.

analysis of these contributions [16] identified (4) as the primary source of the error.

The two-particle basis incompleteness error can be reduced by using Gaussian-type ($\exp(-\alpha r_{12}^2)$), Slater-type ($\exp(-\alpha r_{12})$), or other types of geminals [17,18,16,19]. Non-linear correlation factors can be far superior with small Hartree–Fock basis sets, e.g., few percent basis set errors were observed at the MP2 level with double-ζ quality basis sets using a single Slater-type geminal [18,16,19]. A simple justification goes along the following lines. The MP1-R12 pair correlation functions are expanded in terms of conventional bi-products of virtual orbitals and r_{12}-multiplied bi-products of occupied orbitals:

$$|\phi_{ij}\rangle = \sum_{a<b} t_{ab}^{ij}|ab\rangle + \sum_{k<l} c_{kl}^{ij}\hat{Q}_{12}r_{12}|kl\rangle, \quad (1)$$

where $\hat{Q}_{12}$ ensures the strong orthogonality. When the Hartree–Fock basis set is too small, the reference MP1 description of the Coulomb hole is not adequate and the r_{12} terms must compensate for the deficiency. Unfortunately, the *global* r_{12} term cannot efficiently reproduce *local* features such as the Coulomb hole. Slater and Gaussian geminals are more effective with small, double-ζ quality basis sets because these factors are more localized.

The complete two-electron basis can, in principle, be approached as closely as desired using linear R12-type approaches with general correlation factors, while the Hartree–Fock basis is kept fixed. This goal can be achieved via any of the following routes, among others:

(1) the set of virtual orbitals is expanded systematically toward completeness;
(2) the set of explicitly correlated terms is expanded systematically by increasing the number of geminals per pair function;
(3) explicitly correlated terms are constructed as $\hat{Q}_{12}g(r_{12})|xy\rangle$ and the basis set supporting x and y is expanded systematically.

Any two-electron approximation to the *N*-electron Schrödinger equation (e.g., first-order Møller–Plesset or singles and doubles coupled cluster equations) can thus be solved exactly if we use exact Hartree–Fock orbitals. It is not practical to obtain 'exact' Hartree–Fock orbitals using global basis sets for arbitrary molecules. Thus the error will be introduced due to the finite Hartree–Fock basis set. The magnitude of the error will depend on the Hartree–Fock basis and the correlation model. The following question looms: how large is this error for typical basis sets?

Many previous studies provide useful hints. Adamowicz and Sadlej have demonstrated that optimal parameters of the two-electron basis of Gaussian geminals can be obtained safely with crude reference SCF orbitals [20]. Several groups have formulated Slater determinant-based methods to use a separate basis set for virtual orbitals [21–23], following basic ideas of Huzinaga et al. [24]. Following the approach of [21], Jurgens-Lutovski and Almlöf

demonstrated that the MP2 energy computed with a small orbital basis set (OBS) and a large virtual basis set (VBS) is nearly identical to the standard MP2 value obtained with the large basis set. The small-large OBS differences in absolute second-order energies for water decreased from 0.88 to 0.12 m$E_{\rm h}$ as the VBS was increased. Wolinski and Pulay drew similar conclusions [23].

The goal of this study is to analyze the error in the second-order Møller–Plesset energy due to the OBS as the two-electron basis set limit is approached as closely as possible.

2. Computational method

The first-order wave function was expanded according to Eq. (1). The MP2-R12 energy expression remains identical to previous work[2]. Standard approximation A′ was utilized, i.e. terms $[K_1 + K_2, r_{12}]$ were discarded [8]. Generalized and extended Brillouin conditions [8,16] were also assumed. The intermediates were computed according to the CABS approach modified accordingly: Eq. (15) of [14] becomes

$$\Gamma_{ij}^{\alpha\beta}\Omega_{\alpha\beta}^{kl} = (\Gamma\Omega)_{ij}^{kl} - \Gamma_{ij}^{mn}\Omega_{mn}^{kl} - \Gamma_{ij}^{ma}\Omega_{ma}^{kl} - \Gamma_{ij}^{am}\Omega_{am}^{kl}$$
$$- \Gamma_{ij}^{ab}\Omega_{ab}^{kl} - \Gamma_{ij}^{ma'}\Omega_{ma'}^{kl} - \Gamma_{ij}^{a'm}\Omega_{a'm}^{kl}. \quad (2)$$

The CABS basis $\{a'\}$ was constructed using the CABS approach, ensuring orthogonality to occupied and virtual orbitals.

Standard correlation consistent [25] and polarization consistent [26] families of basis sets were obtained from the EMSL Gaussian Basis Set database [27]. The K19s2i basis set is described in Ref. [13]. The CABS for each calculation was constructed from the union of the OBS, the VBS, and the K19s2i basis. Virtual orbitals were constructed from an AO VBS as follows: (1) the AO basis set was orthogonalized; (2) the occupied SCF orbitals were projected out using the SVD-based procedure described in [14]; (3) the resulting orbitals were rotated to diagonalize the Fock matrix. This procedure does not require OBS to be a subset of VBS, unlike the approach of [22,23].

A set of 20 molecules, suggested by Helgaker, Olsen and Jørgensen, was used to estimate the average errors in total and pair second-order energies due to the finite OBS. Twelve orbital and 4 virtual basis sets of varying sizes were used for each molecule. The relative error for a given VBS is defined as

$$\delta E(\text{OBS}, \text{VBS}) = \frac{E(\text{OBS}, \text{VBS}) - E(\text{pc-3-aug}, \text{VBS})}{E(\text{pc-3-aug}, \text{VBS})}, \quad (3)$$

---

[2] The off-diagonal (OBS-VBS, OBS-complete basis, and VBS-complete basis) blocks of the Fock operator must be formally included in the perturbation operator [20]. Contribution from the second-order wave function to the energy then becomes zero. One-particle (non-Brillouin) contributions to the first-order wave function also vanish.

where E is the total or spin-adapted pair correlation energy.

All computations were performed with the MPQC package, version 2.3 [28].

3. Results and discussion

Statistical averages for the OBS errors are presented in Tables 1–4. The errors were computed as described in the previous section.

The first, and most important, observation is that the mean and RMS errors due to the OBS are small, but not negligible. The RMS errors are largest with the smallest, pc-1 and pc-1-aug, basis sets: 1% for the total energy and less than 2% for the pair energies. These errors are significantly larger than the error due to the residual incompleteness of the two-particle basis with the largest VBS. The latter error is very conservatively estimated at 0.4%, which is the computed RMS average of the E(OBS,aug-cc-pV5Z) − E(OBS,K19s2i) difference. Standard approximations of linear R12-type theories can be responsible for similar (1–2%) errors [16]. It is clear that the OBS error cannot be neglected always.

The OBS error, however, decreases rapidly as the basis set is expanded, e.g., triple-ζ quality basis sets result in under 0.5% RMS errors. The rate of decay is approximately exponential. The OBS error changes little as VBS is varied. This is a clear indication that the one- and two-particle parameters of the wave function are only weakly coupled.

Comparison of correlation consistent and polarization consistent families reveals a few surprises. The double-ζ correlation consistent basis sets yield smaller errors than the pc-1 sets and thus seem to be a better choice. Differences between cc-pVTZ and pc-2 sets are less pronounced but still hint that the correlation consistent series is preferred for the orbital expansion.

There is no reason to expect a normal (Gaussian) distribution of the OBS error. Further statistical analysis of the computed sample of OBS errors suggests strongly non-Gaussian distributions when large OBS are employed. The OBS error distributions for small OBS are significantly closer to Gaussian, e.g., Skew and Kurt for the OBS errors in (OBS = aug-cc-pVDZ, VBS = K19s2i) pair energies are −0.196 and 0.693, respectively. It is therefore reasonable to assume that the 'true' distribution of the OBS error is

Table 1
Mean relative OBS errors (%) in the MP2-R12 correlation energies for the test set of molecules

OBS	VBS			
	aug-cc-pVTZ	aug-cc-pVQZ	aug-cc-pV5Z	K19s2i
cc-pVDZ	0.253	0.228	0.243	0.340
cc-pVTZ	0.122	0.108	0.132	0.151
cc-pVQZ	0.043	0.042	0.044	0.063
aug-cc-pVDZ	−0.346	−0.322	−0.302	−0.192
aug-cc-pVTZ	−0.065	−0.078	−0.054	−0.028
aug-cc-pVQZ	−0.022	−0.030	−0.029	−0.007
pc-1	−0.650	−0.727	−0.776	−0.683
pc-2	−0.012	−0.081	−0.118	−0.117
pc-3	0.003	0.003	0.003	0.003
pc-1-aug	−0.732	−0.819	−0.867	−0.770
pc-2-aug	−0.020	−0.084	−0.122	−0.119
pc-3-aug	0.000	0.000	0.000	0.000

Table 3
Mean relative OBS errors (%) in the MP2-R12 pair energies for the test set of molecules

OBS	VBS			
	aug-cc-pVTZ	aug-cc-pVQZ	aug-cc-pV5Z	K19s2i
cc-pVDZ	0.319	0.255	0.263	0.374
cc-pVTZ	0.161	0.148	0.172	0.195
cc-pVQZ	0.055	0.053	0.058	0.078
aug-cc-pVDZ	−0.411	−0.428	−0.412	−0.289
aug-cc-pVTZ	−0.081	−0.092	−0.068	−0.038
aug-cc-pVQZ	−0.027	−0.036	−0.034	−0.009
pc-1	−0.984	−1.086	−1.151	−1.037
pc-2	−0.098	−0.186	−0.225	−0.223
pc-3	0.004	0.005	0.004	0.005
pc-1-aug	−0.966	−1.081	−1.149	−1.034
pc-2-aug	−0.097	−0.176	−0.219	−0.214
pc-3-aug	0.000	0.000	0.000	0.000

Table 2
RMS relative OBS errors (%) in the MP2-R12 correlation energies for the test set of molecules

OBS	VBS			
	aug-cc-pVTZ	aug-cc-pVQZ	aug-cc-pV5Z	K19s2i
cc-pVDZ	0.670	0.668	0.677	0.716
cc-pVTZ	0.245	0.229	0.241	0.253
cc-pVQZ	0.089	0.089	0.092	0.101
aug-cc-pVDZ	0.404	0.414	0.405	0.308
aug-cc-pVTZ	0.084	0.086	0.065	0.039
aug-cc-pVQZ	0.026	0.034	0.032	0.011
pc-1	0.930	0.982	1.033	0.955
pc-2	0.087	0.118	0.147	0.142
pc-3	0.006	0.006	0.006	0.006
pc-1-aug	0.907	0.981	1.038	0.942
pc-2-aug	0.053	0.096	0.133	0.128
pc-3-aug	0.000	0.000	0.000	0.000

Table 4
RMS relative OBS errors (%) in the MP2-R12 pair energies for the test set of molecules

OBS	VBS			
	aug-cc-pVTZ	aug-cc-pVQZ	aug-cc-pV5Z	K19s2i
cc-pVDZ	1.438	1.429	1.427	1.434
cc-pVTZ	0.422	0.418	0.420	0.430
cc-pVQZ	0.175	0.172	0.173	0.176
aug-cc-pVDZ	0.702	0.678	0.661	0.600
aug-cc-pVTZ	0.133	0.129	0.107	0.088
aug-cc-pVQZ	0.041	0.050	0.047	0.026
pc-1	1.713	1.706	1.737	1.672
pc-2	0.379	0.402	0.415	0.411
pc-3	0.021	0.020	0.020	0.020
pc-1-aug	1.404	1.413	1.454	1.377
pc-2-aug	0.307	0.326	0.343	0.335
pc-3-aug	0.000	0.000	0.000	0.000

Gaussian for small OBS. The OBS error for more complex systems can therefore be estimated from the computed averages.

The computed OBS error estimates are relevant to the recent and future work on explicitly correlated methods. The linear R12-type methods with general geminal factors (Slater-type and others) are capable of recovering more than 95% of correlation energy even with a double-ζ basis sets [17,18,16,19]. Present results demonstrate that (1) a significant part of this error should be attributed to the effect of the finite OBS basis and (2) below 1% error may not be achievable with a double-ζ quality OBS. We can also conclude that a triple-ζ basis set is more than sufficient to achieve a 99% precision for the two-body correlation energy. These rules can be summarized as follows: the improvement of the two-particle basis must be accompanied by increasing the orbital basis set to keep the OBS error comparable or below the two-particle basis incompleteness error.

4. Conclusions

A generalization of the MP2-R12 method which uses a separate basis for virtual orbitals can be used to approach systematically the two-particle basis set limit as closely as desired while keeping the Hartree–Fock basis fixed. This approach was used to compute the error in the second-order energy due to the finite OBS, statistically averaged for a set of 20 small molecules. The OBS error is small, on the order of 1%, and decays quickly as OBS becomes complete. For small, double-ζ quality basis sets, the error is not negligible relative to the other sources of the error in correlation energy.

The most important conclusion of this study is that very precise two-electron correlation energies ($\sim$98–99%) can be computed from Hartree–Fock orbitals expanded in small, double-ζ quality, basis sets. Although the error due to the finite Hartree–Fock basis is small, it is not entirely negligible compared to the error due to approximations in explicitly correlated methods. The dominant source of error, however, is still the incomplete two-particle basis.

The presented MP2-R12 method is clearly not the best practical route towards precise two-electron correlation energies. The r_{12} factor is not the optimal factor far from the asymptotic regime, i.e. it only makes sense to use the r_{12} factor with a large space of virtual orbitals. Other geminal factors are much more efficient in the small-basis regime. Generalization of the present MP2-R12 method to other geminal factors is straightforward and has also been recently implemented. Initial results for open-shell systems will be published shortly.

References

[1] J. Rychlewski (Ed.), Explicitly Correlated Wave Functions in Chemistry and Physics. Theory and Applications, Progress in Theoretical Chemistry and Physics, vol. 13, Springer, Berlin, 2004.
[2] A.K. Wilson, T.H. Dunning, J. Chem. Phys. 106 (1997) 8718.
[3] T. Helgaker, W. Klopper, H. Koch, J. Noga, J. Chem. Phys. 106 (1997) 9639.
[4] E.F. Valeev, W.D. Allen, R. Hernandez, C.D. Sherrill, H.F. Schaefer, J. Chem. Phys. 118 (2003) 8594.
[5] K.L. Bak, P. Jørgensen, J. Olsen, T. Helgaker, W. Klopper, J. Chem. Phys. 112 (2000) 9229.
[6] T.L. Gilbert, Rev. Mod. Phys. 35 (1963) 491.
[7] W. Kutzelnigg, Theor. Chim. Acta 68 (1985) 445.
[8] W. Kutzelnigg, W. Klopper, J. Chem. Phys. 94 (1991) 1985.
[9] W. Klopper, R. Röhse, Theor. Chim. Acta 83 (1992) 441.
[10] E.F. Valeev, H.F. Schaefer, J. Chem. Phys. 113 (2000) 3990.
[11] F.R. Manby, J. Chem. Phys. 119 (2003) 4607.
[12] S. Ten-no, F.R. Manby, J. Chem. Phys. 119 (2003) 5358.
[13] W. Klopper, C.C.M. Samson, J. Chem. Phys. 116 (2002) 6397.
[14] E.F. Valeev, Chem. Phys. Lett. 395 (4–6) (2004) 190.
[15] S. Ten-no, J. Chem. Phys. 121 (2004) 117.
[16] A.J. May, E. Valeev, R. Polly, F.R. Manby, Phys. Chem. Chem. Phys. 7 (2005) 2710.
[17] B.J. Persson, P.R. Taylor, J. Chem. Phys. 105 (1996) 5915.
[18] S. Ten-no, Chem. Phys. Lett. 398 (2004) 56.
[19] D.P. Tew, W. Klopper, J. Chem. Phys. 123 (2005) 074101.
[20] L. Adamowicz, A.J. Sadlej, Chem. Phys. Lett. 48 (1977) 305.
[21] S. Havriliak, H.F. King, J. Am. Chem. Soc. 105 (1983) 4.
[22] R. Jurgens-Lutovski, J. Almlöf, Chem. Phys. Lett. 178 (1991) 451.
[23] K. Wolinski, P. Pulay, J. Chem. Phys. 118 (2003) 9497.
[24] S. Huzinaga, C. Arnau, Phys. Rev. A 1 (1970) 1285.
[25] T.H. Dunning, J. Chem. Phys. 90 (1989) 1007.
[26] F. Jensen, J. Chem. Phys. 115 (2001) 9113.
[27] Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 02/25/04, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory which is part of the Pacific Northwest Laboratory, P.O. Box 999, Richland, WA 99352, USA, and funded by the US Department of Energy. The Pacific Northwest Laboratory is a multi-program laboratory operated by Battelle Memorial Institute for the US Department of Energy under contract DE-AC06-76RLO 1830. Contact Karen Schuchardt for further information.
[28] C.L. Janssen, I.B. Nielsen, M.L. Leininger, E.F. Valeev, E.T. Seidl, The Massively Parallel Quantum Chemistry Program (MPQC): Version 2.3, Sandia National Laboratories, Livermore, CA, USA, 2005. Available from: http://www.mpqc.org/.