ELSEVIER

11 September 1998

Chemical Physics Letters 294 (1998) 45–48

CHEMICAL
PHYSICS
LETTERS

Basis-set extrapolation

Donald G. Truhlar *

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431, USA

Received 1 June 1998; in final form 1 June 1998

Abstract

A proposal for extrapolation of correlated electronic structure calculations based on correlation-consistent polarized double- and triple-zeta basis sets is evaluated. Optimum exponents are presented for separately extrapolating the Hartree–Fock and correlation energies, and the method is shown to yield energies that are more accurate than those from straight correlation-consistent polarized sextuple-zeta calculations at less than 1% of the cost. For the test problems, the root-mean-square deviations from the complete basis limit are 1.3–2.4 kcal/mol for the extrapolated calculations and 3.0–4.4 kcal/mol for the polarized sextuple-zeta calculations. © 1998 Elsevier Science B.V. All rights reserved.

1. Introduction

There has been considerable recent interest in methods that use electronic structure calculations to predict quantitative bond energies and barrier heights [1]. Methods that extrapolate to the limit of complete configuration interaction [1] [2–11] or to a complete one-electron basis set for a given level of electron correlation [12–17] are particularly valuable since the slow convergence of correlated calculations to the limit of a complete one-electron basis set is the limiting feature in the accuracy of most electronic structure calculations.

Martin [14,16] and Halkier et al. [17] have presented very useful studies of basis-set convergence and practical suggestions for extrapolating to the limit. The emphasis was on obtaining very accurate results for small molecules by extrapolating from very large basis sets. For example, Martin [16] obtained an average accuracy of 0.2 kcal/mol in extrapolating CCSD(T) [2] calculations from aug′-cc-pVTZ [19–21], aug′-cc-pVQZ and aug′-cc-pV5Z basis sets with separate exponents [3] for the Hartree–Fock (HF) and correlated parts of the energy. Similarly, Halkier et al. [17] obtained results within a few tenths of a kcal/mol of the basis-set limit for CCSD(T) calculations by extrapolating from cc-pCV5Z [22] and cc-pCV6Z calculations. In particular they noted that inclusion of cc-pVDZ or cc-pCVDZ results in the extrapolations lowers the accuracy consistently, and they recommended omitting these calculations from the extrapolations. In the present Letter, though, we return to this question. Our moti-

---

* E-mail: truhlar@chem.umn.edu

[1] Complete CI is complete with respect to both one-electron basis and level of electron correlation; full CI is complete with respect to the level of correlation for a given one-electron basis. Basis-set convergence is concerned with the completeness of the one-electron basis set for a given level of electron correlation.

[2] CCSD(T) denotes coupled clusters method with single and double excitations and noniterative inclusion of connected triples [18].

[3] In this Letter, 'exponents' refers to the power law used for extrapolation, not to exponential parameters of individual basis functions.

PII: S0009-2614(98)00866-5

vation is economical. For many cases of interest, even cc-pVTZ basis sets are very expensive, and augmenting the basis with diffuse functions, correlating core orbitals, or going to polarized quadruple zeta are all unaffordable. Thus, for potential applications to large molecules, extrapolation from cc-pVDZ and cc-pVTZ calculations would be very useful, if it improves the accuracy. The first objective of the present Letter is to reconsider that accuracy.

Halkier et al. [17] made a significant observation that I shall build on in this Letter. They recommended that the extrapolation exponent be optimized to minimize the mean unsigned error (MUE) of the extrapolated result from the best estimate of the basis-set limit rather than by fitting calculations with three or more basis sets. I will use this idea with their estimated basis-set limits for Ne, HF, and H_2O to optimize exponents for MP2 [4], CCSD, and CCSD(T) calculations, except that we minimize the root-mean-square (RMS) error rather than the MUE, and we use the valence-only limits, not the core-correlated limits (for large molecules, changes in core correlation energy upon bond formation are usually small compared to other errors [24,25]). I combine the optimization of the extrapolation exponent with another idea [16], namely the use of separate exponents for the Hartree–Fock and correlated parts of the calculation. The second objective of the present Letter is to present the optimized exponents for these extrapolations.

2. Theory

The total energy is a sum of the HF and correlation parts:

$$E^{\text{tot}} = E^{\text{HF}} + E^{\text{cor}}. \tag{1}$$

The components of the energy are assumed to approach their basis-set limits by power laws:

$$E_X^{\text{HF}} = E_\infty^{\text{HF}} + A^{\text{HF}} X^{-\alpha} \tag{2}$$

and

$$E_X^{\text{cor}} = E_\infty^{\text{cor}} + A^{\text{cor}} X^{-\beta}. \tag{3}$$

Here $X = 2$ for the cc-pVDZ basis and $X = 3$ for the cc-pVTZ basis. Then the basis-set limit for the total energy is obtained by

$$E_\infty^{\text{tot}} = E_\infty^{\text{HF}} + E_\infty^{cor}. \tag{4}$$

Combining these equations yields

$$E_\infty^{\text{tot}} = \frac{3^\alpha}{3^\alpha - 2^\alpha} E_3^{\text{HF}} - \frac{2^\alpha}{3^\alpha - 2^\alpha} E_2^{\text{HF}} + \frac{3^\beta}{3^\beta - 2^\beta} E_3^{\text{cor}} - \frac{2^\beta}{3^\beta - 2^\beta} E_2^{\text{cor}}. \tag{5}$$

Table 1
RMS deviations (kcal/mol) from the complete basis limit for the total energy

X	MP2	CCSD	CCSD(T)
2	102.6	95.6	99.9
3	37.5	32.4	33.8
5	7.6	5.3	5.5
6	4.4	3.0	3.1
2,3	1.3	1.9	2.4

3. Results

Optimizing the exponents as discussed in Section 1 yields $\alpha = 3.4$, $\beta_{\text{MP2}} = 2.2$, and $\beta_{\text{CCSD}} = \beta_{\text{CCSD(T)}} = 2.4$.

Table 1 gives RMS deviations from the complete-basis limit [17] for Ne, HF, and H_2O of single-level calculations with cc-pVDZ, cc-pVTZ, cc-pV5Z, and cc-pV6Z as well as by the dual-level calculations based on Eq. (5).

4. Discussion

The method works remarkably well. The RMS errors are smaller than those for unextrapolated results from cc-pV6Z calculations, and they are 2.3–5.8 times smaller than unextrapolated results from cc-pV5Z calculations.

The number N of basis functions per atom in a cc-pVXZ calculation scales as [17]

$$N = (X+1)\left(X+\frac{3}{2}\right)(X+2)/3. \tag{6}$$

Furthermore the computer time for MP2, CCSD, and CCSD(T) calculations scales as n^5N^4, n^6N^4, and

---

[4] MP2 denotes second-order Møller–Plesset perturbation theory [23].

n^7N^4, respectively, where n is the number of atoms [26]. Thus N^4 is the relevant scaling for all three correlated methods when they are applied to a single system with a sequence of basis-set sizes. These values are presented in Table 2, where the final row corresponds to the sum for $X = 2$ and $X = 3$. Table 2 shows that the extrapolated calculation is nominally only 5% more expensive than performing only the cc-pVTZ calculation. Thus it would be a shame not to extrapolate whenever cc-pVTZ is affordable. Furthermore the single-level cc-pV6Z calculation is nominally 453 times more expensive than the extrapolation based on $X = 2$ and 3, which has a smaller RMS error. Clearly the unextrapolated, 'pure' single-level approach cannot compete with even very simple extrapolation schemes.

One could also apply this method to extrapolate other methods of calculating correlation energy, e.g., density functional theory. However, the CCSD(T) extrapolations are liable to be most useful because CCSD(T) is often very close to full CI when a single reference configuration is adequate. Notice that, unlike some extrapolation schemes and some forms of DFT that have parameters fit to experiment, the present method is totally ab initio because the data [17] used for parameterization is all ab initio data.

Eq. (5) demonstrates that the $X = 2, 3$ result is a linear combination of four components. Thus, if analytic gradients are available for these four components, as they often are, one can even optimize geometries at the extrapolated level. Thus will be especially useful for kinetics, where the saddle point location is typically strongly correlated to the barrier height (this correlation is a consequence of Hammond's postulate [27,28]). For example, in an exothermic reaction the barrier is usually early, and lowering the barrier (which is what will typically occur when one extrapolates) is expected to make the barrier occur even earlier along the reaction path [29]. Analytic gradients at the extrapolated level could also be used for ab initio molecular dynamics calculations [30–32].

Table 2
Computational effort according to scaling laws

X	N^4	Ratio[a]
2	3.8×10^4	0.05
3	8.1×10^5	1.00
5	9.2×10^6	11.3
6	3.8×10^8	474.3
2,3	8.5×10^5	1.05

[a] Ratio to effort for $X = 3$.

The extrapolated Hessians are also linear combinations of the Hessians of the four components. Furthermore, because of the linearity of Eq. (5) one may directly extrapolate relative energies as well as absolute energies. In these respects the new approach shares some of the advantages of the integrated molecular orbital–molecular orbital (IMOMO) [33,6] and SAC-type [2–8] methods, which are also linear combinations of components.

There is another converged-basis correlation energy in the literature that can be used as a check, namely the MP2 correlation energy of HF calculated by Klopper [34]. Wilson and Dunning [35] found errors of 7.4 and 3.0 kcal/mol for cc-pVDZ and cc-pVTZ calculations, respectively. Eq. (5) yields an error of only 0.04 kcal/mol, which is another remarkable success and an order of magnitude smaller error than is obtained [35] by using the cc-pV6Z basis set without extrapolation.

Although we used the correlation-consistent basis sets here, one could presumably attempt similar correlations with other kinds of basis sets, but one would expect such correlations to be less successful because other basis-set sequences are less systematic.

One could imagine various improvements on the scheme presented here. Foremost among these would be parameterizing the method against a greater number of basis-set-limit data when such data become available. Meanwhile the parameterization presented here should be useful for a variety of applications.

Acknowledgements

This work was supported in part by the National Science Foundation.

References

[1] K.K. Irikura, D.J. Frurip (Eds.), ACS Symposium Series 677: Computational Thermochemistry, American Chemical Society, Washington, DC, 1998.

[2] F.B. Brown, D.G. Truhlar, Chem. Phys. Lett. 117 (1985) 307.

[3] M.S. Gordon, D.G. Truhlar, J. Am. Chem. Soc. 108 (1986) 5412.

[4] M.S. Gordon, K.A. Nguyen, D.G. Truhlar, J. Phys. Chem. 93 (1989) 7356.

[5] I. Rossi, D.G. Truhlar, Chem. Phys. Lett. 234 (1995) 64.

[6] J.C. Corchado, D.G. Truhlar, in: J. Gao, M.A. Thompson (Eds.), Hybrid Quantum Mechanical and Molecular Mechanical Models, Am. Chem. Soc., Washington, DC (in press).

[7] P.E.M. Siegbahn, M.R.A. Blomberg, M. Svensson, Chem. Phys. Lett. 223 (1994) 35.

[8] M.R.A. Blomberg, P.E.M. Siegbahn, Am. Chem. Soc. Symp. Ser. 677 (1998) 197.

[9] G.A. Petersson, A.K. Lee, A. Bennett, J. Chem. Phys. 83 (1985) 5105.

[10] J.A. Montgomery, J.W. Ochterski, G.A. Petersson, J. Chem. Phys. 101 (1994) 5900.

[11] G.A. Petersson, Am. Chem. Soc. Symp. Ser. 677 (1998) 237.

[12] D. Feller, J. Chem. Phys. 96 (1992) 6104.

[13] K.A. Peterson, T.H. Dunning Jr., J. Phys. Chem. 99 (1995) 3898.

[14] J.M.L. Martin, Chem. Phys. Lett. 259 (1996) 669.

[15] J.M.L. Martin, P.R. Taylor, J. Chem. Phys. 105 (1997) 8620.

[16] J.M.L. Martin, Am. Chem. Soc. Symp. Ser. 677 (1998) 212.

[17] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olson, A.K. Wilson, Chem. Phys. Lett. 286 (1998) 243.

[18] J.A. Pople, M. Head-Gordon, K. Raghavachari, J. Chem. Phys. 87 (1987) 5968.

[19] T.H. Dunning Jr., J. Chem. Phys. 90 (1989) 1007.

[20] R.A. Kendall, T.H. Dunning, R.J. Harrison, J. Chem. Phys. 96 (1992) 6796.

[21] J.E. Del Bene, J. Phys. Chem. 97 (1993) 107.

[22] K.A. Peterson, A.K. Wilson, D.E. Woon, T.H. Dunning Jr., Theor. Chem. Acc. 97 (1997) 251.

[23] J.A. Pople, J. Binkley, R. Seeger, Int. J. Quant. Chem. Symp. 10 (1976) 1.

[24] K.A. Peterson, T.H. Dunning Jr., J. Chem. Phys. 106 (1997) 4119.

[25] K.A. Peterson, S.S. Xantheas, D.A. Dixon, T.H. Dunning Jr., J. Chem. Phys. A 102 (1998) 2449.

[26] K. Raghavachari, J.B. Anderson, J. Phys. Chem. 100 (1996) 12960.

[27] G.S. Hammond, J. Am. Chem. Soc. 77 (1955) 334.

[28] C.A. Parr, D.G. Truhlar, J. Phys. Chem. 75 (1971) 1844.

[29] T.N. Truong, D.G. Truhlar, K.K. Baldridge, M.S. Gordon, R. Steckler, J. Chem. Phys. 90 (1989) 7137.

[30] M. Parrinello, in: E. Clementi (Ed.), Modern Techniques in Computational Chemistry, MOTECC-91, ESCOM, Leiden, 1991, p. 833.

[31] B. Hartke, E.A. Carter, Chem. Phys. Lett. 189 (1992) 358.

[32] D.G. Truhlar, in: D. Heidrich (Ed.), The Reaction Path in Chemistry, Current Approaches and Perspectives, Kluwer, Dordrecht, 1995, p. 229.

[33] S. Humbel, S. Sieber, K. Morokuma, J. Chem. Phys. 105 (1996) 1959.

[34] W. Klopper, J. Chem. Phys. 102 (1995) 6168.

[35] A.K. Wilson, T.H. Dunning Jr., J. Chem. Phys. 106 (1997) 8718.