Comment on "Geometry optimization with an infinite basis set" [J. Phys. Chem. A 103 (1999) 651] and "Basis-set extrapolation" [Chem. Phys. Lett. 294 (1998) 45]

Asger Halkier [a,*], Trygve Helgaker [b], Wim Klopper [c], Poul Jørgensen [a], Attila G. Császár [d]

[a] *Department of Chemistry, University of Aarhus, DK-8000 Århus C, Denmark*
[b] *Department of Chemistry, University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway*
[c] *Theoretical Chemistry Group, Debye Institute, Utrecht University, Padualaan 14, NL-3584 CH Utrecht, The Netherlands*
[d] *Department of Theoretical Chemistry, Eötvös University, P.O. Box 32, H-1518 Budapest 112, Hungary*

Abstract

The extrapolation scheme for geometry optimizations presented in J. Phys. Chem. A 103 (1999) 651, which is based on a similar scheme for energies proposed in Chem. Phys. Lett. 294 (1998) 45, is examined. The scheme is unreliable in the sense that it sometimes yields results less accurate than the unextrapolated data it is aiming beyond and should be applied with caution. © 1999 Elsevier Science B.V. All rights reserved.

1. Introduction

In a recent paper entitled "Geometry optimization with an infinite basis set", Chuang and Truhlar [1] presented a scheme for geometry optimizations in which the energy and its derivatives are not simply obtained with one basis set. Instead, these quantities are determined from an extrapolation based on results obtained with the correlation-consistent polarized valence double-zeta (cc-pVDZ) and triple-zeta (cc-pVTZ) basis sets [2]. The extrapolated quantities are used in the optimization of the molecular geometry and, to the extent that the extrapolated properties

agree with the basis-set limit, this optimization produces the equilibrium molecular structure that would be obtained with a complete basis set for the given wavefunction. Therefore, such a scheme is interesting and important as it (approximately) removes one of the two errors inherent in all standard molecular electronic structure optimizations – namely, the basis-set truncation error.

Halkier et al., who have proposed a simple linear two-point extrapolation for correlation energies, have pointed out that such an extrapolation is "well suited for automatic incorporation in a broader computational scheme aimed at high-accuracy prediction of energies and molecular properties" [3]. The studies of Truhlar and co-workers [1,4] represent a realization of this procedure, using one particular extrapola-

* Corresponding author. Fax: +45-8619-6199; e-mail: asger@kemi.aau.dk

tion scheme. In the present Comment, we analyse the extrapolation introduced by Truhlar in Ref. [4] and discuss the quality and reliability of the results obtained in Ref. [1].

2. Theory

2.1. Comparison of different extrapolation schemes

The two most popular extrapolation forms for the energy / correlation energy obtained with the correlation-consistent basis sets (cc-pVXZ) are the *exponential form*

$$E_X = E_{\lim} + B \exp(-\alpha X), \tag{1}$$

which is used by Dunning and Feller (cf. Refs. [2,5,6]) and *the power form*

$$E_X = E_{\lim} + B X^{-\alpha}, \tag{2}$$

which forms the basis for the two-point extrapolation of Helgaker and co-workers [3,7] as well as for the scheme of Truhlar and co-workers [1,4]. In Eqs. (1) and (2), E_X is the energy / correlation energy obtained with the basis set with cardinal number X (D: 2, T: 3, ...), $E_{\lim}$ is the basis-set limit energy / correlation energy, and B and α are parameters, the determination of which we outline below. Note that the two-point extrapolation of Helgaker and co-workers is concerned only with the correlation energy.

As both the two-point extrapolation of Helgaker and co-workers and the extrapolation scheme of Truhlar and co-workers are based on results for two consecutive X, we shall focus on this situation, although we note that this is not the way the Dunning–Feller exponential scheme is usually employed. For two consecutive cardinal numbers $X - 1$ and X, the following closed expressions for $E_{\lim}$ are obtained for the exponential form

$$E_{\lim} = \frac{1}{1 - \exp(-\alpha)} E_X - \frac{\exp(-\alpha)}{1 - \exp(-\alpha)} E_{X-1}, \tag{3}$$

and for the power form

$$E_{\lim} = \frac{X^\alpha}{X^\alpha - (X-1)^\alpha} E_X - \frac{(X-1)^\alpha}{X^\alpha - (X-1)^\alpha} E_{X-1}. \tag{4}$$

Clearly, both expressions are of the form

$$E_{\lim} = C(X, \alpha) E_{X-1} + [1 - C(X, \alpha)] E_X$$
$$= C(X, \alpha)(E_{X-1} - E_X) + E_X. \tag{5}$$

The coefficient C depends in general on X, although we note that C is independent of X for the exponential form. Extrapolations based on Eqs. (1) and (2) which employ results for two cardinal numbers can thus be interpreted as a simple weighting of E_X and E_{X-1}.

The two-point extrapolation of Helgaker and co-workers is, for all X pairs, carried out with the fixed exponent $\alpha = 3$. This particular value of α is supported by the theoretical analysis of the partial-wave expansion of the helium atom [8] and this extrapolation thus exhibits the correct asymptotic behaviour for large X [7,9]. For this extrapolation, we prefer to use the seemingly more complicated expression (4) rather than (5) to emphasize that: (a) there is a single α parameter for *all* $(X - 1, X)$ pairs in (4) rather than a separate C parameter for *each* $(X - 1, X)$ pair as in (5); and (b) the explicit reference to X^3 serves as a reminder of the theoretical foundation of the extrapolation.

The scheme of Truhlar and co-workers is economically motivated [4] and only the X pair $(2, 3)$ is used. The Hartree–Fock and correlation parts of the energies are treated separately (i.e., there is one exponent α for the Hartree–Fock energy and another for the correlation energy) and the exponent for each part is obtained as the one that minimizes the root-mean-square error of extrapolated limits for a sample of molecules for which the basis-set limits are known either from numerical Hartree–Fock studies or explicitly correlated R12 calculations [4]. Since this scheme is purely empirical and used only for a single X pair, it appears more natural to use the form (5) rather than (4). Furthermore, as the extrapolation is empirical, a large and broad sample of systems for the determination of the optimal exponents (or coefficients) is preferable. In Ref. [4], the

exponents were determined from a sample of only three systems, and a redetermination of the parameters in the extrapolation from a larger sample of molecules is thus desirable, as mentioned also in Ref. [4].

2.2. Requirements for extrapolations

The requirements for the performance of an extrapolation scheme may be subject to personal preference. In the following, we describe the requirements we impose on an extrapolation scheme to consider it successful. As the extrapolation itself is free of cost compared with the calculations needed for obtaining the numbers used in the extrapolation, we do not require a reduction of the error by several orders of magnitude by the extrapolation. However, we do require the results to be perceptibly improved by the extrapolation. More important, the extrapolation should only rarely fail in the sense of giving a result poorer than the original numbers. If this criterion is not met, the scheme is – in our opinion – unreliable and it is better to work with the original, unextrapolated results.

Regarding the first point, we note that the scheme of Truhlar and co-workers is, by construction, biased towards those molecules in the sample that have the largest absolute basis-set errors [10]. Therefore, there is a risk of failure for molecules within the sample that have small absolute errors. Indeed, in Ref. [10], it was found that the scheme of Truhlar and co-workers overshoots the Hartree–Fock limit for H_2 by 99% compared with the error of the cc-pVTZ basis set and thus does not represent any practical improvement for this molecule. It may be argued that this does not matter much since the hydrogen molecule is rather accurately described anyway. In our view, however, such a failure signals a warning concerning the performance of the extrapolation scheme.

3. Numerical results

In the present section, we examine the results obtained with the extrapolation of Truhlar and co-workers for the equilibrium geometries of H_2O,

NH_3, and H_2O_2, which in Ref. [1] were studied at both the second-order Møller–Plesset (MP2) theory [11] and the coupled-cluster singles-and-doubles (CCSD) levels of the wavefunction [12].

3.1. Computational considerations

To increase the number of significant digits, we have recomputed all the unextrapolated cc-pVDZ, cc-pVTZ, and cc-pVQZ MP2(FC) geometries in Ref. [1]. Moreover, to enable a comparison with results that are closer to the basis-set limit than those obtained with the cc-pVQZ basis, we have performed MP2(FC) calculations with the cc-pV5Z and the aug-cc-pVXZ ($X = D - 5$) basis sets [2,13]. As in Ref. [1], only the valence electrons have been correlated in the post-Hartree–Fock treatment. All calculations have been carried out with the Gaussian 94 program [14], employing the 'tight convergence' option to ensure high numerical accuracy of the calculated geometries. The results are given in Table 1, along with the geometries obtained using the extrapolation scheme of Ref. [1]. Included in Table 1 are also MP2-R12(FC) results for H_2O obtained as a by-product of the CCSD(T)-R12(FC) calculations presented in Ref. [15]. The small discrepancies between the optimized angles in Table 1 and those in Ref. [1] are believed to arise from different convergence thresholds. Since the discrepancies are small compared with the changes observed when the basis set is increased or the extrapolation is applied, they do not affect our conclusions concerning the performance of the extrapolation scheme of Truhlar and co-workers for bond angles.

3.2. Equilibrium geometries of H_2O, NH_3, and H_2O_2

In the discussion section of Ref. [1], it is stated that ''the extrapolated results are much closer to the quadruple-zeta values than to the triple-zeta ones''. Although this conclusion is correct for the angles and holds also for the mean absolute percentage deviation of all the geometrical parameters, it does not hold for all the bond distances. For both the O–O bond distance in H_2O_2 and the bond distance in NH_3 at the CCSD level (see Table 2 of Ref. [1]), the extrapolated results are closer to the cc-pVTZ results than to the cc-pVQZ results. Moreover, con-

Table 1

Optimized equilibrium geometries of H_2O, NH_3, and H_2O_2 obtained at the MP2(FC) level with the cc-pVXZ and aug-cc-pVXZ, ($X = D - 5$), basis sets and the extrapolation scheme of Truhlar and co-workers. Included are also the MP2-R12(FC) results for H_2O. Bond lengths are in pm and angles in degrees

Basis	H_2O		NH_3		H_2O_2			
	r_{OH}	$\angle_{HOH}$	r_{NH}	$\angle_{HNH}$	r_{OH}	r_{OO}	$\angle_{HOO}$	dihedral
cc-pVDZ	96.49	101.90	102.35	103.90	97.04	145.68	98.78	118.44
cc-pVTZ	95.91	103.52	101.14	105.95	96.44	145.05	99.32	114.26
cc-pVQZ	95.77	104.02	100.98	106.47	96.31	144.64	99.66	112.73
cc-pV5Z	95.79	104.29	100.95	106.84	96.34	144.54	99.79	113.01
aug-cc-pVDZ	96.59	103.87	102.02	106.30	97.16	147.02	99.14	112.92
aug-cc-pVTZ	96.14	104.11	101.21	106.77	96.68	145.37	99.61	112.57
aug-cc-pVQZ	95.89	104.27	101.02	106.89	96.42	144.71	99.80	112.67
aug-cc-pV5Z	95.84	104.33	100.97	106.93	96.38	144.57	99.82	112.92
R12[a]	95.86	104.43						
Truhlar scheme[b]	95.7	103.96	100.7	106.59	96.0	144.9	99.54	112.66

[a] From Ref. [15].
[b] From Ref. [1].

sidering the title of Ref. [1], we should ask how close the cc-pVQZ results are to the basis-set limit.

An inspection of the convergence of the bond distances obtained within the aug-cc-pVXZ series indicates that the aug-cc-pV5Z bond distances are close to the basis-set limit. This is supported by the excellent agreement between the R12 and aug-cc-pV5Z results for r_{OH} in H_2O. We further observe that the bond distances extrapolated by Truhlar and co-workers are *further away* from the basis-set limit (as represented by the aug-cc-pV5Z bond distances) than are the direct cc-pVQZ results. This is also seen from Table 2, where we have listed the mean errors, mean absolute errors, maximum absolute errors, and root-mean-square errors of the calculated bond distances and angles relative to the aug-cc-pV5Z level. In Ref. [1], it is stated that ''it appears reasonable to

assume that the extrapolated geometries are actually closer to the infinite-basis results than are the cc-pVQZ geometries'', but given the results in Tables 1 and 2, this assumption seems too optimistic.

In the discussion section of Ref. [1], it is stated that ''the alternate to extrapolated results for larger molecules will usually be unextrapolated cc-pVTZ results, which are much less accurate''. As the most accurate estimates of the basis-set limit are the aug-cc-pV5Z results, we shall compare here the cc-pVTZ and extrapolated numbers with these results. Again, for the mean absolute percentage deviation of *all* geometrical parameters, the extrapolated results represent an improvement on the cc-pVTZ results: 0.3% (extr.) vs. 0.5% (cc-pVTZ). The improvement is not uniform, however. First, whereas the root-mean-square error for the bond angles is significantly

Table 2

Mean errors, mean absolute errors, maximum absolute errors, and root-mean-square errors relative to the aug-cc-pV5Z level for the cc-pVXZ basis sets and the extrapolation scheme of Truhlar and co-workers. The errors have been calculated based on the numbers in Table 1

Basis	Bond lengths (pm)				Angles (deg)			
	mean	mean abs.	max. abs.	r.m.s.	mean	mean abs.	max. abs.	r.m.s.
cc-pVDZ	0.95	0.95	1.38	1.00	−0.24	3.00	5.52	3.41
cc-pVTZ	0.20	0.20	0.48	0.26	−0.24	0.91	1.34	0.96
cc-pVQZ	−0.02	0.05	0.07	0.06	−0.28	0.28	0.46	0.30
Truhlar scheme	−0.11	0.28	0.38	0.29	−0.31	0.31	0.37	0.32

reduced from 0.96 to 0.32° by the extrapolation, it *increases* from 0.26 to 0.29 pm for the bond distances. Second, even though the *mean error* for the bond distances is reduced from 0.20 to -0.11 pm by the extrapolation, for the bond distances in H_2O and NH_3 and in particular for the O–H bond distance in H_2O_2, the difference between the cc-pVTZ and aug-cc-pV5Z results is *smaller* than the difference between the extrapolated result and the aug-cc-pV5Z result.

According to our requirements on extrapolation schemes given above, the scheme of Ref. [1] is thus unreliable and it is, in our opinion, better to work with the cc-pVTZ basis set – despite its limitations with respect to completeness – than to apply this particular extrapolation. Currently, the most accurate equilibrium geometries are obtained by performing standard calculations in the cc-pVXZ and aug-cc-pVXZ basis sets to as high an X as possible, although this may change with the development of more reliable extrapolation schemes. It should finally be noted that the convergence behaviour of the cc-pVXZ and aug-cc-pVXZ geometric data is rather different. Although computationally less economical, the latter basis sets seem better suited for extrapolations for the systems chosen in Ref. [1], as they include functions that make a proper description of the lone-pairs and bond polarity possible, making the extrapolation to the basis-set limit easier.

4. Conclusions

The extrapolation scheme for energies and geometry optimizations recently proposed by Truhlar and co-workers has been examined. Although the extrapolation in average absolute percentage terms represents an improvement on the unextrapolated data, it sometimes yields results that are poorer than those it is aiming beyond. In this sense, the scheme is unreliable and should be employed with caution.

References

[1] Y. Chuang, D.G. Truhlar, J. Phys. Chem. A 103 (1999) 651.
[2] T.H. Dunning, J. Chem. Phys. 90 (1989) 1007.
[3] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A.K. Wilson, Chem. Phys. Lett. 286 (1998) 243.
[4] D.G. Truhlar, Chem. Phys. Lett. 294 (1998) 45.
[5] D. Feller, J. Chem. Phys. 96 (1992) 6104.
[6] D. Feller, K.A. Peterson, J. Chem. Phys. 108 (1998) 154.
[7] T. Helgaker, W. Klopper, H. Koch, J. Noga, J. Chem. Phys. 106 (1997) 9639.
[8] R.N. Hill, J. Chem. Phys. 83 (1985) 1173.
[9] W. Klopper, K.L. Bak, P. Jørgensen, J. Olsen, T. Helgaker, J. Phys. B: At. Mol. Opt. Phys. 32 (1999) R103.
[10] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, J. Olsen, Chem. Phys. Lett. 302 (1999) 437.
[11] C. Møller, M.S. Plesset, Phys. Rev. 46 (1934) 618.
[12] G.D. Purvis, R.J. Bartlett, J. Chem. Phys. 76 (1982) 1910.
[13] R.A. Kendall, T.H. Dunning, R.J. Harrison, J. Chem. Phys. 96 (1992) 6796.
[14] M.J. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.A. Robb, J.R. Cheeseman, T. Keith, G.A. Petersson, J.A. Montgomery, K. Raghavachari, M.A. Al-Laham, V.G. Zakrzewski, J.V. Ortiz, J.B. Foresman, C.Y. Peng, P.Y. Ayala, W. Chen, M.W. Wong, J.L. Andres, E.S. Replogle, R. Gomperts, R.L. Martin, D.J. Fox, J.S. Binkley, D.J. Defrees, J. Baker, J.P. Stewart, M. Head-Gordon, C. Gonzalez and J.A. Pople, Gaussian 94, Revsion E2, Gaussian, Inc., Pittsburgh, PA, 1995.
[15] V. Bakken, T. Helgaker, W. Klopper, K. Ruud, Mol. Phys. 96 (1999) 653.