Short Report

Basis Set Convergence in Hartree-Fock Calculations of Some Diatomic Molecules Containing First and Second-Row Atoms

Paulo S. Sagrillo, Francisco E. Jorge, * *Pedro L. Barbieri and Paulo A. Fantin*

Departamento de Física, Universidade Federal do Espírito Santo, 29060-900 Vitória-ES, Brazil

Investiga-se a convergência de conjuntos de bases em direção ao limite numérico da energia Hartree-Fock (HF) total para as seqüências hierárquicas dos conjuntos de bases XZP e cc-pVXZ. Para as duas hierarquias, melhoramentos significativos são obtidos com cada incremento em X. Para estimar o limite do conjunto de base completo, uma forma exponencial foi usada. Entre as várias aproximações consideradas aqui, uma extrapolação exponencial de três parâmetros aplicada aos resultados TZP, QZP e 5ZP deu os limites do conjunto de bases mais precisos. Em adição, energias HF dos orbitais moleculares ocupados mais altos de algumas moléculas diatômicas foram calculadas com o conjunto 5ZP e comparadas com as correspondentes obtidas com o conjunto cc-pV5Z e com um método numérico HF.

Basis set convergence towards the numerical limit of the total Hartree-Fock (HF) energy is investigated for the hierarchical sequences of the XZP and cc-pVXZ basis sets. For both hierarchies, solid improvements are obtained with each increment in X. To estimate the complete basis set limit, an exponential form was used. Among the various approaches considered here, a three-parameter exponential extrapolation applied to the TZP, QZP, and 5ZP results yields the most accurate basis set limits. In addition, 5ZP highest occupied molecular orbital HF energies of some diatomic molecules are evaluated and compared with the corresponding ones obtained with the cc-pV5Z and numerical HF results.

Keywords: *ab initio*, calculations, diatomic molecules, Gaussian basis set, total and orbital HF energies

Introduction

The use of finite expansions in analytic basis functions lies at the heart of computational molecular electronic structure theory.[1,2] It is the choice of basis set that ultimately determines the accuracy of a calculation. The growth of central processing power, in uniprocessor and multiprocessor environments, and memory in high-performance computers facilitates the use of increasingly large flexible basis sets in molecular electronic structure calculations which in turn allows calculations of increasing accuracy by reducing the error associated with basis set truncation.

A step towards a systematic way of improving basis set for describing the correlation energy was the atomic natural orbital analysis by Almlöf and Taylor,[3] which lead Dunning and co-workers[4,5] to propose the correlation consistent basis sets of double, triple, quadruple, etc. (cc-pVXZ, X=D, T, Q, and 5) quality. We recall that these sets were constructed from the general contraction scheme of Raffenetti.[6]

Recently, Jorge and co-workers[7] presented segmented contracted double, triple and quadruple,[8] and quintuple[9] zeta valence quality plus polarization function (XZP, X=D, T, Q, and 5, respectively) basis sets for the atoms from H to Ar. At the Hartree-Fock (HF) and Møller-Plesset second-order levels, these sets were applied with success in calculations of energies, dissociation energy, harmonic vibrational frequency, and electric dipole moment of a set of diatomic molecules containing atoms of the first- and second-row[8-10] and second-row diatomic hydrides.[11] We verified[8-10] that the Jorge's sets, when compared with the corresponding sets of Dunning and co-workers,[4,5] appear to be the best compromise of accuracy and computational cost.

Basis set convergence and extrapolations are among the most important issues of contemporary molecular *ab initio* theory.[12-15] For total energies, it is known that the convergence of the correlation part is significantly slower than of the HF part. This suggests that, when studying the basis set convergence of the energy, one should treat the HF and correlation parts separately. When examining basis

*e-mail: jorge@cce.ufes.br

set convergence, it is mandatory to have a hierarchical sequence of basis sets with systematic improvements from level to level. The cc-pVXZ basis sets[4,5] constitute such a hierarchy and have been extensively used in previous works on basis set convergence studies.

In this work, total HF energies for C_2, FH, N_2, CO, F_2, PN, SC, and ClB calculated with the XZP sets[7-9] were compared with the results obtained with numerical HF (NHF) methods[15-18] and with the sets of Dunning and co-workers[4,5] Comparing with numerical solutions of the HF equations, we have also examined the complete basis set (CBS) limit of the HF method using XPZ and cc-pVXZ basis sets. For PN, SC, and ClB, 5ZP molecular orbital HF energies were also computed and compared with results obtained with other approaches. Thus, the main purpose of this short report is to compare the performance of the two hierarchical sequences of basis sets when used to achieve the CBS limit.

Extrapolation to the complete basis set

In a previous study, Halkier *et al.*[15] used different equations for extrapolation and came to the conclusion that an exponential form, which was at first time used by Feller[19] for estimating the CBS limit, is the best among others to estimate the HF energy, namely:

$$E_R(L) = E_R(\infty) + A\ e^{-BL} \qquad (1)$$

In this equation $E_R(L)$ is the total energy computed at the inter-nuclear distance R and L denotes the highest angular functions of the basis sets used in extrapolation. In XZP and cc-pVXZ hierarchies L ranges from 2 (DZP and cc-pVDZ) to 5 (5ZP and cc-pV5Z). Also $E_R(\infty)$ is the total energy in CBS limit. The quantities A and B are fitting parameters without physical significance. Since there are three unknown quantities in this equation [$E_R(\infty)$, A, and B], at least three consecutive basis sets are needed for extrapolation. Restricting ourselves to this requirement, we have used three member groups (L, L+1, L+2) in each hierarchy. In this regard, the two hierarchies can be extrapolated with {(2,3,4), (3,4,5)}collections. Halkier *et al.*[15] noted that inclusion of cc-pVDZ result in the extrapolations lowers the accuracy consistently, and they recommended omitting this calculation from the extrapolations. Thus, we consider in this work only the (3,4,5) collection.

Results and Discussion

The HF calculations reported in this section were carried out with the GAUSSIAN 03 program.[20] For each

molecule, the calculation was carried out with the same bond distance used in the NHF calculation and spherical harmonic Gaussian-type functions were employed.

The XZP, cc-pVXZ, and CBS HF energy errors (in mhartree) for C_2, FH, N_2, CO, F_2, PN, SC, and ClB are listed in Table 1. The NHF energies[15-18] are also shown.

A brief look at Table 1 offers some general trends. In all cases enlargement of the basis set causes improvement of E_{HF}. For the hierarchical sequences of basis sets, the largest total HF energy decrease occurs from DZP to TZP. For the HF calculations, the convergence towards the basis set limit is monotonic, smooth, and fast. The error in the energy is reduced approximately by a factor of three each time X is incremented (*i.e.*, linear convergence).

The cc-pVXZ basis sets give in general better energies than the corresponding ones evaluated with the XZP sets. It is not surprising, since the atomic SCF energy loss in a general contraction is, in general, smaller than that observed in a segmented contraction, and this result is reflected in the total HF energy of a molecule, whose wave function is formed from atomic basis sets. On the other hand, we verified that when comparing the second order correlation energies computed with the cc-pVXZ and XZP (X = 2, 3, 4, and 5) sets, the opposite occurred, *i.e.*, the correlation energies calculated with the sets of Jorge and co-workers[7-9] were in general better. The largest errors obtained with the 5ZP and cc-pV5Z basis sets are respectively 5.2 and 4.3 mhartree for ClB.

As demonstrated above, it is possible to reach millihartree and even near microhartree precision easily with direct computation for the molecular collection under study. Although the rate of convergence of basis set derived total energies relative to the CBS limit are remarkable, it is evident that for larger systems with more electrons the same accuracy could not be reached with direct computation. A reliable estimation of HF energies is even more important for large systems. This is particularly vital in the case of isodesmic reactions,[21] which could be used for a reliable estimation of thermodynamic quantities of large polyatomic molecules without considering the correlation energy contribution directly.

For all molecules displayed in Table 1, the errors of the extrapolated values of the XZP/(3,4,5) model are smaller than those obtained with the 5ZP basis set. For C_2, the 5ZP error is about 50 times larger than that observed with the extrapolated value. The interesting trend is that the CBS limits are in general overestimates with this model. Except for PN and ClB, the exact total energy is somewhere between the 5ZP and extrapolated value of the XZP/(3,4,5) model. Similar specifications also hold for the extrapolated total energies from the correlation

Table 1. Total HF energy errors (in mhartree) for the ground states of some diatomic molecules. NHF energies are in hartree

	C_2	FH	N_2	CO	F_2	PN	SC	ClB
DZP[a]	16.508	26.86	33.715	31.652	49.268	42.32	34.741	26.0
TZP[b]	6.567	10.04	11.464	10.641	17.096	19.61	14.954	15.3
QZP[b]	1.536	2.57	2.769	2.612	5.184	7.34	6.060	8.2
5ZP[c]	0.355	0.62	0.611	0.607	1.160	2.82	2.419	5.2
CBS[d]	−0.007	−0.07	−0.101	−0.060	−0.893	0.19	−0.107	3.1
cc-pVDZ[e]	19.544	51.40	39.234	41.595	87.659	45.49	32.717	29.6
cc-pVTZ[e]	5.188	12.80	9.673	10.526	21.276	13.32	9.692	10.5
cc-pVQZ[e]	0.900	3.13	2.029	2.022	5.046	4.09	2.854	5.5
cc-pV5Z[e]	0.144	0.38	0.344	0.296	0.580	1.11	0.674	4.3
CBS[f]	−0.017	−0.71	−0.133	−0.166	−1.115	−0.31	−0.347	3.9
NHF[g]	−75.406565[h]	−100.07082[i]	−108.993826[j]	−112.790906[j]	−198.773323[h]	−395.18864[i]	−435.362420[k]	−484.1665[i]

In all calculations the nuclear distances used were: C_2 (1.248 Å), FH (0.917 Å), N_2 (1.094 Å), CO (1.128 Å), F_2 (1.412 Å), PN (1.491 Å), SC (1.534 Å), and ClB (1.716 Å). [a]HF energies calculated in this work with the set of reference 7. [b]HF energies calculated in this work with the sets of reference 8. [c]HF energies calculated in this work with the set of reference 9. [d]These values were obtained from 3-point fits (TZP, QZP, and 5ZP) to equation (1). [e]HF energies calculated in this work with the sets of Dunning and co-workers.[4,5] [f]These values were obtained from 3-point fits (cc-pVTZ, cc-pVQZ, and cc-pV5Z) to equation (1). [g]The nuclear distances used in the calculations were 1.248, 0.917, 1.094, 1.128, 1.412, 1.491, 1.534, and 1.716 Å, respectively. [h]Numerical HF energy from reference 15. [i]Numerical HF energy from reference 17. [j]Numerical HF energy from reference 16. [k]Numerical HF energy from reference 18.

consistent hierarchy, although in this case the extrapolated values are not always better than those evaluated with the cc-pV5Z.

From Table 1, one can also verify that except for F_2, the smallest errors are obtained with the extrapolations based on XZP (X=T, Q, 5) energies. Overall, the mean absolute deviations among the results obtained with the 5ZP, cc-pV5ZP, XZP/(3,4,5) and cc-pVXZ/(3,4,5) models from the corresponding NHF energies are 1.724, 0.979, 0.566, and 0.837 mhartree, respectively. From these results, it is clear that the best mean deviation is obtained with XZP/(3,4,5).

We believe that all the errors in the XZP (X=T, Q, 5) extrapolated energies are smaller than the corresponding ones obtained from the cc-pVXZ (X=T, Q, 5) extrapolated energies because in general the total HF energies calculated with the Jorge´s sets are larger (see Table 1). We recall that the basis sets of Jorge and co-workers[7-9] were implemented by employing the segmented contraction scheme, whereas the sets of Dunning and co-workers[4,5] use the general contraction scheme of Raffenetti.[6] As a consequence to use the segmented contraction scheme, the extrapolated energies obtained from the basis sets constructed by Jorge and co-workers are closer to the corresponding HF limits than the extrapolations derived from the correlation consistent basis sets. An extensive discussion about the main differences between segmented and general contraction schemes was done in references 1 and 10.

It is also interesting to consider the A and B values of equation (1) in the fitting procedure. A vast difference has been found among A and B values corresponding to different molecules. This variance is more pronounced for A than B. So no general predictable pattern seems to exist for these quantities in fitting the collection studied.

Table 2 displays the highest occupied molecular orbital energies of PN, SC, and ClB computed with the 5ZP, cc-pV5Z basis sets, and with a NHF method.[17] For any molecule, the agreement among corresponding results obtained with the three different approaches is very good.

Conclusions

It is shown that the XPZ basis sets provide a systematic series of basis sets increasing accuracy and completeness. The HF energy showed a trend to converge to well-defined limits as the basis set is systematically enlarged, from double to 5 zeta valence quality.

Table 2. Highest Occupied Molecular Orbital HF energies (in hartree) for the ground states of some diatomic molecules

Molecule[a]	Basis sets					
	5ZP		cc-pV5Z		NHF[b]	
	$\varepsilon(\sigma)$	$\varepsilon(\pi)$	$\varepsilon(\sigma)$	$\varepsilon(\pi)$	$\varepsilon(\sigma)$	$\varepsilon(\pi)$
PN	−0.48560	−0.44318	−0.48567	−0.44314	−0.48590	−0.44333
SC	−0.47034	−0.46301	−0.47029	−0.46295	−0.47056	−0.46299
ClB	−0.38030	−0.52446	−0.38027	−0.52446	−0.38040	−0.52470

[a]In all calculations the nuclear distances used were 1.491, 1.534, 1.716 Å, respectively. [b]Numerical HF energies from reference 17.

For all molecules studied, solid improvements in the total HF energy are obtained with each increment in the cardinal number X, and the largest difference among the 5ZP and NHF energies is equal to 5.2 mhartree for ClB. The XZP energies are in general worse than those evaluated with the cc-pVXZ basis sets. We recall that our sets were mainly designed to generate accurate correlated molecular wave functions; see references 7-9.

For current hardware technology and high performance computer code, in the case of small molecules, it is possible to use direct computation as the best method to reach the CBS limit, but in the case of larger systems an extrapolation scheme seems mandatory. The simple exponential type functions like equation (1) work relatively well, but it seems that in the case of the increase size basis set hierarchies studied here a general slight overestimation exist. So modification of equation (1) opens the door for a better estimation of the CBS limit. Among the ten schemes used in this work (see Table 1) to estimate the basis set limit of a set of diatomic molecules, the extrapolated total energies of the XZP (X = 3, 4, and 5) hierarchy gave the best values. We attribute this result to the segmented contraction scheme used to construct these basis sets.

For PN, SC, and ClB, the agreement among the 5ZP, cc-pV5Z, and NHF highest occupied molecular orbital HF energies is very good.

Finally, it is important to say that the development of the XZP basis sets has been partly motivated by the expectation that it is more efficient for programs not designed for general contractions employed in cc-pVXZ.

The complete set of the s, p, d, f, g, and h parameters of the XZP basis sets for hydrogen, helium, and first- and second-row atoms are available through the internet at http://www.cce.ufes.br/qcgv/pub/.

Acknowledgments

References

1. Feller, D.; Davidson, E. R.; *Rev. Comput. Chem*. **1990**, *1*, 1.

2. Wilson, S.; *Basis Sets for Electronic Structure Calculations (Methods in Computational Chemistry),* Plenum: New York, 1994.

3. Almlöf , J.; Taylor, P. R.; *Adv. Quantum Chem.* **1991**, *22*, 301.

4. Dunning Jr., T. H.; *J. Chem. Phys*. **1989**, *90*, 1007.

5. Peterson, K. A.; Wilson, A. K.; Woon, D. E.; Dunning Jr., T. H.; *Theor. Chem. Acc*. **1997**, *97*, 251.

6. Raffenetti, R. C.; *J. Chem. Phys*. **1973**, *58*, 4452.

7. Canal Neto, A.; Muniz, E. P.; Centoducatte, R.; Jorge, F. E.; *THEOCHEM* **2005**, *718*, 219.

8. Barbieri, P. L.; Fantin, P. A.; Jorge, F. E.; *Mol. Phys.* **2006**, *104*, 2945.

9. Jorge, F. E.; Sagrillo, P. S.; Oliveira, A. R.; *Chem. Phys. Lett.* **2006**, *432*, 558.

10. Muniz, E. P.; Jorge, F. E.; *Int. J. Quantum Chem.* **2006**, *106*, 943.

11. Jorge, F. E.; Bernardo, L. M.; Muniz, E. P.; *J. Theor. Comput. Chem.* **2006**, *5*, 223.

12. Feller, D.; *J. Chem. Phys*. **1992**, *96*, 6104.

13. Martin, J. M. L.; Taylor, P. R.; *J. Chem. Phys*. **1997**, *106*, 8620.

14. Varandas, A. J. C.; *J. Chem. Phys*. **2000**, *113*, 8880 and references therein.

15. Halkier, A.; Helgaker, T.; Jorgenson, P.; Klopper, W.; Olsen, J.; *Chem. Phys. Lett.* **1999**, *302*, 437.

16. Kobus, J.; *Chem. Phys. Lett.* **1993**, *202*, 7.

17. Sundholm, D.; Pyykko, P.; Laaksonen, L.; *Mol. Phys.* **1985**, *56*, 1411; Pyykko, P.; Dierscken, G. H. F.; Muller-Plathe, F.; Laaksonen, L.; *Chem. Phys. Lett.* **1987**, *134*, 575.

18. Kobus, J.; Moncrieff, D.; Wilson, S.; *J. Phys. B: At., Mol. Opt. Phys.* **1994**, *27*, 2867.

19. Feller, D.; *J. Chem. Phys*. **1993**, *98*, 7059.

20. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; *Gaussian 03, Revision A.1*; Gaussian, Inc.: Pittsburgh, 2003.

21. Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, J. A.; *Ab Initio Molecular Orbital Theory,* Wiley: New York, 1986.