Theor Chem Acc (2005) 113: 267–273
DOI 10.1007/s00214-005-0635-2

REGULAR ARTICLE

Frank Jensen

Estimating the Hartree—Fock limit from finite basis set calculations

Received: 12 January 2005 / Accepted: 27 January 2005 / Published Online: 10 June 2005

Abstract It is demonstrated that the polarization-consistent basis sets, which are optimized for density functional methods, are also suitable for Hartree–Fock calculations, and can be used for estimating the Hartree–Fock basis set limit to within a few micro-hartree accuracy. Various two- and three-point extrapolation schemes are tested and exponential functions are found to be superior compared to functions depending on the inverse power of the highest angular momentum function in the basis set. Total energies can be improved by roughly an order of magnitude, but atomization energies are only marginally improved by extrapolation.

Keywords HF-limit, basis set, extrapolation

1 Introduction

The large majority of electronic structure calculations employ an expansion of the orbitals in a basis set, almost always of the Gaussian type and located at the nuclear positions. The basis set incompleteness is one of the factors that limits the ultimate accuracy, but the error can be controlled by performing calculations with increasingly larger basis sets, until the desired accuracy is obtained. For electron correlation methods, the inherent basis set convergence is slow, but the use of hierarchical basis sets [1] in connection with extrapolation methods [2] has made a large step towards solving this problem. The basis set convergence of the Hartree–Fock (HF) and density functional methods is significantly faster than the correlation energy [3], and for large basis sets, the resulting error is often assumed to be negligible. Nevertheless, if the target accuracy is high, for example predicting molecular stabilities to within 1 kJ/mol [4], the HF error must be reduced below this limit, which is not a trivial issue.

In recent work, we have proposed a series of basis sets designed specifically for density functional methods, denoted as polarization -consistent basis sets [5,6]. The similarity of density functional and HF methods indicates that these basis sets should also be suitable for HF calculations, or in general, methods where only a representation of the electron density is required. For atoms and diatomic systems, the limiting HF energy can be obtained by solving the integro-differential equations by finite difference or finite elements methods [7], and these results provide an absolute reference against which the results from finite basis sets can be evaluated. In the present case, we address the problem of how to estimate the HF limit from finite basis set calculations in combination with different extrapolation formulas.

2 Computational details

Table 1 contains 26 diatomic systems composed of first- and second-row elements, for which we recently have reported numerical HF results accurate to at least one micro-hartree [8], and these results will be used as reference data for evaluating the performance of various basis sets and extrapolation schemes.

Three families of standard basis sets are used in the present work:

1. The correlation-consistent basis sets (cc-pVXZ, X = D, T, Q, 5, 6) developed by Dunning et al. [9]. The X notation indicates the basis set quality, D being double zeta, T triple zeta, etc., quality. For second-row elements, the versions employing an extra tight d-function have been used.
2. The polarization-consistent basis set (pc-n, n = 0, 1, 2, 3, 4) in both their contracted and uncontracted forms. The –n notation indicates the polarization beyond the isolated atom, i.e. pc-0 contains only s- and p-functions, pc-1 contains d-functions, etc., and pc-1 is thus of double zeta quality, pc-2 is triple zeta quality etc.
3. The SVP, TZV and QZV basis sets developed by Ahlrichs et al. [10]. These are of double, triple and quadruple zeta quality, respectively.

F. Jensen
Department of Chemistry,
University of Southern Denmark,
DK-5230 Odense, Denmark
E-mail: frj@dou.dk

Table 1 Systems used in the present work

$^1CH^+$, ^3NH, ^1FH, $^1C_2$, ^2CN, $^1N_2$, ^1CO, $^1NO^+$, $^3O_2$,$^1CF^+$, ^3NF, $^1F_2$, ^1HCl, ^2CP, ^1CS, ^2SiN, ^1NP, ^3NCl,
^1SiO, ^3SO, ^3PF, ^1FCl, ^1SiS, $^1P_2$, $^3S_2$, $^1Cl_2$

Employed bond distances and numerical HF energies can be found in Ref. [8]

Table 2 Basis sets used in the present work

Basis set	$L_{\max}$	Contracted functions		Primitive *sp*-functions	
		1-Row	2-Row	1-Row	2-Row
pc-0	1	3s2p	4s3p	5s3p	8s6p
pc-1	2	3s2p1d	4s3p1d	7s4p	11s8p
pc-2	3	4s3p2d1f	5s4p2d1f	10s6p	13s10p
pc-3	4	6s5p4d2f1g	6s5p4d2f1g	14s9p	17s13p
pc-4	5	8s7p6d3f2g1h	7s6p6d3f2g1h	18s11p	21s16p
pc-5[a]	6	22s14p9d5f3g2h1i	25s18p9d5f3g2h1i	22s14p	25s18p
cc-pVDZ	2	3s2p1d	4s3d2d	9s4p	12s8p
cc-pVTZ	3	4s3p2d1f	5s4p3d1f	10s5p	15s9p
cc-pVQZ	4	5s4p3d2f1g	6s5p4d2f1g	12s6p	16s11p
cc-pV5Z	5	6s5p4d3f2g1h	7s6p5d3f2g1h	14s8p	20s12p
cc-pV6Z	6	7s6p5d4f3g2h1i	8s7p6d4f3g2h1i	16s10p	21s14p
SVP	2	3s2p1d	4s3p1d	7s4p	10s7p
TZV	3	5s3p2d1f	5s4p2d1f	11s6p	14s9p
QZP	4	7s4p3d2f1g	9s6p4d2f1g	15s8p	20s14p

[a] No contraction defined

Table 2 contains a description of these basis sets in terms of the basis set composition and the highest angular momentum ($L_{\max}$) function included.

Since the pc-*n* basis sets have been optimized at a density functional level, we also performed a full HF exponent optimization of the uncontracted pc-*n* basis sets, denoted opt-pc-*n*, and extended the series to also include a (fully optimized) pc-5-type basis set. Analysis for N_2 and P_2 indicate that the consistent polarization at this level is 9d5f3g2h1i, and this is used in connection with 22s14p and 25s18p basis sets for first- and second-row elements, respectively. No attempt has been made to define a contraction of the pc-5 basis sets. These results primarily serve to probe the errors related to using standard (fixed exponent) basis sets. Calculations and exponent optimizations have been performed with the Gaussian [11] and Dalton [12] program packages. Open shell species have been treated within the restricted open shell formulation. The quality of the calculated results is quantified in terms of the mean ($\langle\Delta E\rangle$) and standard deviation (σ) relative to the numerical reference data ($\Delta E = E_{\text{actual}} - E_{\text{reference}}$) [8].

3 Basis set results

Table 3 shows the average errors in absolute energies relative to the numerical HF results for the basis sets in Table 2. The (uncontracted) pc-*n* results display an exponential convergence as a function of the highest angular momentum function included. The corresponding contracted basis sets (pc-*n*c) converge to a result a few milli-hartree above the limit, which is due to the use of contraction coefficients derived from density functional calculations, as discussed elsewhere [13]. The cc-pV*XZ* series display a slower convergence and the cc-pV6Z ($L_{\max}$ = 6) basis set contains errors of the order of one milli-hartree, which is of comparable accuracy to the pc-4 ($L_{\max}$ = 5) basis set. The results labeled opt-pc-*n* refer to energies obtained by fully optimizing all basis set exponents for each individual molecular system, and are thus the best results that can be obtained with the given basis set composition. The improvements relative to the pc-*n* results are rather marginal, indicating that the density functional optimized exponents are suitable also for HF calculations. Note also that the opt-pc-5 results are of microhartree accuracy, which is essentially at the same level of accuracy as the reference data.

4 Extrapolating functions

The use of extrapolation methods for electron correlation energies has been able to significantly improve the raw calculated results. One of the most successful approaches is based on the perturbation analysis of Kutzelnigg and Morgan [14], which indicates that the leading term should be proportional to $L_{\max}^{-3}$. Neglecting all higher order terms leads to a simple two-point extrapolation formula, which has been used extensively in connection with the cc-pV*XZ* basis sets.

$$\Delta E_{\text{corr}}\,(L_{\max}) = A + BL_{\max}^{-3}. \tag{1}$$

The basis set convergence at the HF level has only been analyzed theoretically for the hydrogen atom, where Kutzelnigg and Klopper showed that the convergence is exponential with respect to the square root of the number of *s*-functions included [15]. Numerical results for a few molecular systems indicate also an exponential convergence with respect to $L_{\max}$ [3,16].

Table 3 Average errors and standard deviations for calculated total energies

$L_{\max}$	opt-pc-*n*		pc- *n*		pc-*n*c		cc- pV*XZ*		Ahlrichs	
	$\langle\Delta E\rangle$	σ	$\langle\Delta E\rangle$	σ	$\langle\Delta E\rangle$	σ	$\langle\Delta E\rangle$	σ	$\langle\Delta E\rangle$	σ
1	414.3	134.7	436.6	138.6	446.6	140.3				
2	62.0	25.2	69.3	28.0	81.9	30.4	42.8	18.9	200.1	72.5
3	6.2	2.5	8.3	3.3	14.6	6.3	10.4	4.4	13.2	8.1
4	0.27	0.15	0.43	0.25	4.9	4.0	2.3	1.1	0.87	0.32
5	0.018	0.012	0.046	0.027	3.9	4.0	0.42	0.16		
6	*0.0011*	*0.0008*					0.088	0.039		

All values in milli-hartree relative to numerical HF data for the systems in Table 1
Data in italic exclude the seven triplet species in Table 1

Both exponential and inverse power functions have been used for extrapolating to the basis set limit. Feller suggested a three-point exponential form, which has also been used by others [16–19].

$$E\left(L_{\max}\right) = A + B\mathrm{e}^{-CL_{\max}}. \tag{2}$$

The A parameter provides the limiting energy, i.e. $A = E$ $(L_{\max} = \infty)$. Peterson et al. [20] have used the three-parameter exponential form shown in Eq. (3) for the H_3 system.

$$E\left(L_{\max}\right) = A + B\mathrm{e}^{-(L_{\max}-1)} + C\mathrm{e}^{-(L_{\max}-1)^2}. \tag{3}$$

We have proposed an alternative exponential form for use in connection with the pc-*n* basis sets, which depends also on the number of s-functions (N_s).

$$E\left(L_{\max}\right) = A + B\left(L_{\max}+1\right)\mathrm{e}^{-C\sqrt{N_s}}. \tag{4}$$

The argumentation behind this formula is that the atomic convergence is expected to depend on the square root of the number of s-functions [15]. The pc-*n* basis sets are constructed such that each sub-shell included provide similar errors, i.e. the total error relative to the basis set limit should have ($L_{\max}$+1) terms of equal magnitude, with remaining terms from higher L-values being significantly smaller and thus negligible.

Martin has proposed several formulas based on an inverse power dependence [21].

$$\begin{aligned} E\left(L_{\max}\right) &= A + B\left(L_{\max}+\tfrac{1}{2}\right)^{-4} \\ E\left(L_{\max}\right) &= A + B\left(L_{\max}+\tfrac{1}{2}\right)^{-4} + C\left(L_{\max}+\tfrac{1}{2}\right)^{-6} \\ E\left(L_{\max}\right) &= A + B\left(L_{\max}+\tfrac{1}{2}\right)^{-C}. \end{aligned} \tag{5}$$

The offset factor of half was chosen as a compromise between values of 0 for hydrogen and 1 for first- and second- row elements. An exponent of five was used in latter work [22], and this is also used in the CBS-APNO model [23].

$$E\left(L_{\max}\right) = A + B\left(L_{\max}+\tfrac{1}{2}\right)^{-5}. \tag{6}$$

In connection with the W*n* methods for estimating the infinite basis set, infinite correlation limit, a formula with no offset parameter was used [4].

$$E\left(L_{\max}\right) = A + BL_{\max}^{-5} \tag{7}$$

Truhlar has used the same form with a variable exponent for fitting the results of small basis sets to estimate the basis set limit [24].

$$E\left(L_{\max}\right) = A + BL_{\max}^{-C} \tag{8}$$

The optimized effective exponent has a value around four for the HF energy.

When discussing extrapolation formulas, it may be useful to distinguish between functional forms motivated by theoretical arguments, and more heuristic expressions [16,24]. The *atomic* correlation energy has been shown to converge as an inverse power series in $L_{\max}$ for basis sets *saturated up to angular momentum* $L_{\max}$. This is not the case for *molecular* calculations using e.g. the cc-pV*XZ* basis sets, but Eq. (1) has nevertheless been very useful for improving the raw calculated results. The theoretical background for the convergence rate for HF energies is less well developed, but the available numerical evidence suggest an exponential convergence for basis sets saturated up to angular momentum $L_{\max}$[3,16]. Again, this is not the case for standard basis sets, but exponential extrapolations are expected to perform better than inverse power forms. From a practical point of view, however, any extrapolation method capable of providing a reliable estimate of the basis set limit will be useful, and inverse power forms may fit this bill.

5 Three-point extrapolations

In the present case, we investigate the following three-point extrapolation formulas for estimating the HF limit from results generated by the cc-pV*XZ*, pc-*n* or Ahlrichs type of basis sets.

$$E\left(L_{\max}\right) = A + B\mathrm{e}^{-CL_{\max}} \tag{9}$$

$$E\left(L_{\max}, N_s\right) = A + B\left(L_{\max}+1\right)\mathrm{e}^{-C\sqrt{N_s}} \tag{10}$$

$$E\left(L_{\max}\right) = A + B\left(L_{\max}+a\right)^{-C}. \tag{11}$$

The offset parameter a can have values of 0 and ½.

Table 4 shows the errors relative to the basis set limit using either of the extrapolation Eqs. (9)–(11) in connection with the basis sets in Table 1. For the uncontracted pc-*n* basis sets, Eq. (10) performs better than Eq. (9), with Eq. (11)

Table 4 Average errors and standard deviations for total energies derived from three-point extrapolated results

Fitting function Data		Eq. (9)			Eq. (10)			Eq. (11), $a = 0$			Eq. (11), $a = 1/2$		
	$L_{\max}$	$\langle \Delta E \rangle$	σ	$\langle C_{\text{fit}} \rangle$	$\langle \Delta E \rangle$	σ	$\langle C_{\text{fit}} \rangle$	$\langle \Delta E \rangle$	σ	$\langle C_{\text{fit}} \rangle$	$\langle \Delta E \rangle$	σ	$\langle C_{\text{fit}} \rangle$
opt-pc-0,1,2	3	−5.0	6.5	1.9	2.0[a]	3.4[a]	6.5[a]	−33.6	22.9	2.3	−24.2	17.4	3.3
opt-pc-1,2,3	4	−0.63	0.62	2.3	0.076	0.15	7.0	−1.84	1.54	5.2	−1.60	1.35	6.4
opt-pc-2,3,4	5	0.007	0.005	3.3	0.002	0.004	6.3	−0.005	0.007	11.2	−0.003	0.006	12.9
opt-pc-3,4,5	6	*−0.0002*	*0.0006*	*2.7*	*−0.0006*	*0.0008*	*6.0*	*−0.001*	*0.001*	*11.9*	*−0.001*	*0.001*	*13.2*
pc-0,1,2	3	−4.7	8.3	1.8	3.4[a]	4.5[a]	6.4[a]	−37.8	28.3	2.2	−26.9	21.5	3.2
pc-1,2,3	4	−1.2	1.6	2.1	0.09	0.25	6.5	−3.5	4.0	4.7	−3.0	3.5	5.8
pc-2,3,4	5	0.024	0.014	3.1	0.015	0.011	6.0	−0.0004	0.01	10.6	0.003	0.01	12.2
pc-0c,1c,2c	3	−1.5	12.2	1.7	8.5[a]	8.0[a]	6.2[a]	−42.8	36.9	2.1	−29.1	28.2	3.0
pc-1c,2c,3c	4	2.9	2.9	1.9	4.4	4.1	6.2	0.16	2.2	4.3	0.7	2.2	5.3
pc-2c,3c,4c	5	3.8	4.0	2.3	3.8	4.0	4.6	3.7	4.0	7.7	3.7	4.0	8.8
cc-pVDZ,TZ,QZ	4	−0.47	0.39	1.4	−152.9	166.1	5.7	−4.0	1.4	2.9	−3.2	1.1	3.6
cc-pVTZ,QZ,5Z	5	−0.15	0.33	1.5	0.04	0.6	11.0	−0.63	0.65	4.9	−0.56	0.60	5.6
cc-pVQZ,5Z,6Z	6	−0.001	0.03	1.7	−9.1	10.4	3.9	−0.06	0.06	7.1	−0.05	0.06	7.9
Ahlrichs-S,T,Q	4	−0.14	0.78	2.9	0.35	0.28	5.2	−1.7	2.0	7.0	−1.4	1.8	8.5

All values in milli-hartree relative to numerical HF data for the systems in Table 1

[a] Fitting using $N_s = 10$ for the pc-1 basis set for second-row systems

Data in italic exclude the seven triplet species in Table 1

being significantly worse, except for the extrapolation based on pc-2,3,4 results. The contraction errors for the pc-*n*c basis sets (Table 3) render these unsuitable for estimating the HF limit. The performance for the opt-pc-*n* basis sets is only marginally better than the fixed exponent pc-*n* basis sets, again showing that the use of exponents optimized by a density functional method is not a major limitation. The mean error of the extrapolated opt-pc-3,4,5 results based on Eq. (9) or (10) in Table 4 is a few tenths of one micro-hartree, which is at the limit of the accuracy of the reference data. For the cc-pV*XZ* basis sets, Eq. (9) performs best, with Eq. (10) providing an unstable fitting. This is due to the use of the number of *s*-functions as the main fitting parameter, and the non-uniform variation between $L_{\max}$ and N_s, as discussed below. Eq. (11) with either of the offset values performs poorly and in most cases leads to results that are worse than the unextrapolated results.

The Eqs. (9)–(11) employ $L_{\max}$ as a variable, which lead to a slight ambiguity for systems having elements from different rows in the periodic table. When hydrogen is present, the maximum angular momentum function for hydrogen is one less than for first- or second-row elements. This ambiguity is the reason for the offset factor in Eq. (5), with half being a heuristic compromise between 0 and 1. For the present systems, there is little difference in the performance of Eq. (11) with an offset of either 0 or 1/2, and no improvement is observed if the hydrogen containing systems are omitted.

For Eq. (10), there is an additional ambiguity, since the number of *s*-functions is used as an extrapolation parameter. For the second-row elements, the fitting of the pc-0,1,2 results to Eq. (10) leads to multiple solutions, neither of which provide reliable results. A stable fitting could be obtained by using a value of 10 for N_s in the pc-1 basis set, rather than the value of 11 from Table 2. This choice is at least partly motivated by the basis set construction, since the 10th and 11th *s*-function contributes to the 1*s* and 3*s*-orbitals, respectively, and consequently have almost the same energy contribution.

For mixed first- and second-row systems, one has a choice of using the N_s values from either the first- or second-row elements (Table 2). For the pc-*n* basis sets, there is virtually no difference between either choice, which is closely related to the fact that the increment in N_s by each step up in $L_{\max}$ is almost the same for first- and second-row elements. Given that the total energy primarily is determined by the elements with highest atomic number, we recommend that the N_s value for the heaviest element is used in the extrapolation for systems with elements from more than one row in the periodic table, and this has been used for generating the data in Table 4.

For the pc-*n* basis sets, there is a close correlation between $L_{\max}$ and N_s, but this is not the case for the cc-pV*XZ* basis sets, which results in instabilities when fitting to Eq. (10). For the first-row elements, the number of *s*-functions increase by two for $X = \text{T} - 6$, but for second-row elements, the increment is rather irregular (12, 15, 16, 20, 21 *s*-functions for $X = \text{D} - 6$). Clearly Eq. (9) provides a much more stable extrapolation procedure when used with cc-pV*XZ* data, and it is also slightly better than Eq. (10) for extrapolation of the data generated by the Ahlrichs basis sets.

From a practical point of view, any fitting function which reliably can improve the results is useful, and we have considered whether Eqs. (9)–(11) could be turned into heuristic fitting functions. Using the offset parameter a in Eq. (11) as a tuning parameter for reproducing the basis set limiting value is not a viable option, as no sensible value can be chosen to provide a vanishing mean error. We have similarly considered choosing an effective power α for either $L_{\max}$ or N_s in Eq. (9) and Eq. (10) (i.e. $\exp(L_{\max}^{\alpha})$ or $\exp(N_s^{\alpha})$), but the optimum α value depends significantly on the data set. The standard deviation is furthermore insensitive to α, indicating that the improvement is merely data fitting. We conclude that

α values of 1 and 0.5 for Eq. (9) and Eq. (10), respectively, are close to optimum, and the quality of the final results are not particularly sensitive to the exact value used.

6 Two-point extrapolations

From the results in Table 3, it is seen that each step-up in basis set quality improves the energies by almost an order of magnitude. A three-point extrapolation thus employs energies that are more than two orders of magnitude removed from the HF-limit. Alternatively, Eqs. (9)–(11) can be made into two-point extrapolations by choosing a suitable fixed C value. This, however, is only a viable approach if the C value is relatively constant for a variety of systems and basis sets. A near constant value for C furthermore indicates that the functional form captures some of the physics behind the data, rather than just being a mathematical fitting function. Table 4 shows that the effective inverse power C in Eq. (11) varies significantly with the basis set data, and the excellent results by extrapolation of the pc-2,3,4 data is most likely accidental. The C value for Eq. (10), on the other hand has a near constant value for the pc-n basis sets. The corresponding parameter for Eq. (9) increases slightly with $L_{\max}$, but is clearly much less sensitive to the data set than in Eq. (11).

Table 5 shows the results for two-point fittings for Eq. (9) and Eq. (10) with suitable average C values for each basis set chosen from the results in Table 4. We have also included data for $C = 5$, $a = 0$ in Eq. (11), which then becomes equivalent to Eq. (7), since this is the formula used in the Wn methods [4]. For the pc-n basis sets, the simple two-point extrapolation performs remarkably well, as anticipated from the near-constant C value in the three point fittings in Table 4. The mean error for the extrapolated values based on the pc-0,1 results (i.e. using at most a polarized double zeta basis set) is a remarkable 3 milli-hartree, although with a standard deviation of 28 milli-hartree. The two-point extrapolation in all cases provides results which are as accurate as the corresponding three-point extrapolations for a given $L_{\max}$. Note again that extrapolations based on the opt-pc-4,5 basis sets provide results that are at the limit of the accuracy of the reference data. Extrapolations using either of Eqs. (9), (10) or (7) with the cc-pVXZ or Ahlrichs basis sets do not provide a systematic improvement, in agreement with previous findings of Halkier et al [19].

7 Relative energies

While absolute energies are the ultimate test of the predictive power of a given basis set and extrapolation method, the main focus in most applications is on relative energies. For the present systems, we have investigated the convergence for the atomization energies of the systems in Table 1. Table 6 shows the raw calculated results as well as the results from two- and three-point extrapolations of the total energies, and Fig. 1 shows the logarithmic error depending on the number of contracted basis functions for a second-row element. Note that the opt-pc-n and pc-n results in Fig. 1 are actually generated using the uncontracted basis sets. Only Eq. (9) has been used for the cc-pVXZ and Ahlrichs basis sets, while only Eq. (10) has been used with the pc-n basis sets, since these provide the best results in Tables 4 and 5. It can be noted that 1 milli-hartree corresponds to 2.6 kJ/mol, i.e. predicting molecular stabilities to within 1 kJ/mol requires errors well below 0.4 milli-hartree in Table 6.

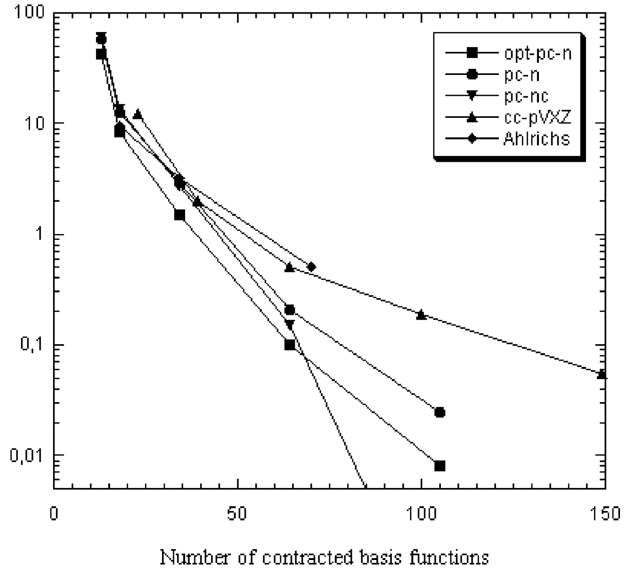

Fig. 1 Logarithm of the average deviation (milli-hartree) for atomization energies (Table 6) as a function of the number of contracted basis functions for a second-row element.

The atomization energy benefits from error cancellation between the molecular and atomic energies, and is therefore expected to converge faster to the basis set limit than the total energy. The improvement relative to the errors for the total energies in Table 3 approaches an order of magnitude for the smaller basis set (pc-0, pc-1, cc-pVDZ, SVP), but is only a factor of two for the larger basis sets. One difference, however, is that the contracted pc-nc basis sets now provide results of comparable accuracy as the uncontracted versions, i.e. the contraction error of a few milli-hartrees is effectively concentrated in the inner core orbitals and cancels out when calculating atomization energies.

The various three-point extrapolation formulas provide only marginal improvements of the atomization energies, and a two-point extrapolation of the pc-n results using Eq. (9) appears as stable as the corresponding three-point formula. The most noticeable result is the two-point extrapolation of the pc-0 and pc-1 results, which gives an order of magnitude improvement, and provides a remarkable accuracy of a few milli-hartree relative to the basis set limit, for a computational cost of a double-zeta basis set. The improvement by extrapolation for the atomization energy is not as large as expected, with the errors being only slightly smaller than for the corresponding extrapolated total energies in Table 4. We note again that extrapolation of the results obtained with the

Table 5 Average errors and standard deviations for total energies derived from two-point extrapolated results

Fitting function Data	$L_{\max}$	Eq. (9)[a]		Eq. (10)[b]		Eq. (7)	
		$\langle\Delta E\rangle$	σ	$\langle\Delta E\rangle$	σ	$\langle\Delta E\rangle$	σ
opt-pc-0,1	2	30.5	16.9	−2.1	26.7	50.6	22.1
opt-pc-1,2	3	1.2	2.4	−2.8	4.9	−2.3	3.2
opt-pc-2,3	4	−0.26	0.10	0.0018	0.039	−1.6	0.6
opt-pc-3,4	5	−0.0029	0.0025	0.0026	0.0032	−0.096	0.053
opt-pc-4,5	*6*	*−0.0004*	*0.0005*	*−0.0001*	*0.0004*	*−0.0099*	*0.0069*
pc-0,1	2	36.4	19.3	3.0	27.7	57.4	24.8
pc-1,2	3	2.8	3.3	−1.5	5.2	−1.0	4.2
pc-2,3	4	−0.27	0.096	0.086	0.080	−2.0	0.7
pc-3,4	5	0.011	0.010	0.020	0.012	−0.14	0.08
cc-pVDZ,TZ	3	1.1	0.7	−5.6	14.7	5.5	2.3
cc-pVTZ,QZ	4	−0.06	0.24	−4.0	4.6	−0.25	0.20
cc-pVQZ,5Z	5	−0.11	0.21	0.11	0.32	−0.48	0.39
cc-pV5Z,6Z	6	0.42	0.16	−0.23	0.21	−0.13	0.05
Ahlrich-S,T	3	2.3	5.3	7.5	5.2	−15.2	5.4
Ahlrich-T,Q	4	0.16	0.36	0.66	0.27	−3.0	2.3

All values in milli-hartree relative to numerical HF data for the systems in Table 1

Data in italic exclude the seven triplet species in Table 1

[a] C = 2.5 for the pc-n and opt-n basis sets, C = 1.5 for cc-pVXZ and C = 2.9 for the Ahlrichs basis sets

[b] $C = 6.3$

Table 6 Average errors and standard deviations for calculated atomization energies

	$L_{\max}$	opt-pc-n		pc-n		pc-nc		cc-pVXZ		Ahlrichs	
		$\langle\Delta E\rangle$	σ	$\langle\Delta E\rangle$	σ	$\langle\Delta E\rangle$	σ	$\langle\Delta E\rangle$	σ	$\langle\Delta E\rangle$	σ
	1	42.3	35.1	58.0	38.1	60.2	40.0				
	2	8.3	5.5	12.6	5.7	13.6	6.3	12.2	5.3	9.4	6.0
	3	1.5	1.3	2.9	2.1	2.7	2.4	2.0	0.9	3.2	2.1
	4	0.10	0.08	0.21	0.16	0.15	0.20	0.51	0.26	0.51	0.22
	5	0.008	0.008	0.025	0.018	0.0002	0.05	0.19	0.09		
	6	*0.0005*	*0.0004*					0.055	0.028		
xpol-1,2	2	−4.5	8.7	−1.8	7.5	−0.9	8.5				
xpol-1,2,3	3	1.3	1.4	2.4	2.3	2.0	2.6				
xpol-2,3	3	−2.7	4.5	−1.8	5.1	−2.3	5.4	−0.91	0.71	2.8	2.0
xpol-2,3,4	4	0.08	0.11	0.09	0.23	0.12	0.25	0.62	0.41	0.20	0.28
xpol-3,4	4	0.014	0.038	0.070	0.066	0.003	0.10	0.08	0.20	0.35	0.16
xpol-3,4,5	5	0.0017	0.0032	0.007	0.009	0.008	0.075	0.12	0.08		
xpol-4,5	5	0.0023	0.0031	0.012	0.010	−0.014	0.058	0.10	0.06		
xpol-4,5,6	6	*−0.0004*	*0.0008*					0.005	0.025		
xpol-5,6	6	*−0.00009*	*0.0004*					0.19	0.09		

All values in milli-hartree relative to numerical HF data for the systems in Table 1

xpol-n,m indicates a two-point extrapolated result based on $L = n$ and m results, xpol-n,m,l indicates a three point extrapolated result based on $L = n$, m and l results

Extrapolation for the opt-pc-n and pc-n basis set are done using Eq. (10) with C = 6.3 for two-point extrapolations. Extrapolation for the cc-pVXZ and Ahlrichs basis sets are done using Eq. (9) with C = 1.5 and 2.9, respectively for two-point extrapolations

Data in italic exclude the seven triplet species in Table 1

cc-pVXZ basis sets is not very successful, and sub-milli-hartree accuracy is difficult to obtain with these basis sets. For the pc-n basis sets, however, the extrapolation in all cases leads to an improvement, and display an exponential convergence towards the basis set limit. The opt-pc-5 basis set again provides results of an accuracy which rivals the reference data.

8 Summary

We have shown that the polarization-consistent basis sets in their uncontracted forms can be used for estimating the Hartree–Fock basis set limit to within a few micro-hartree accuracy. A comparison with results from fully optimized basis sets indicate that the pc-n basis sets, although optimized for density functional calculations, are also near optimum for Hartree–Fock calculations. For absolute energies, the uncontracted forms should be used, but for atomization energies, the contracted version provides results as accurate as the uncontracted forms. The pc-n basis sets tend to be more efficient than the cc-pVXZ basis sets, especially if sub-milli-hartree accuracy is desired.

It is shown that total energies can be improved by an exponential extrapolation, while extrapolation based on an inverse power of the highest angular momentum included in the basis set is inferior. Based on the present results, we recommend the use of Eq. (10) with the pc-n basis sets and Eq. (9) with the cc-pVXZ and Ahlrichs-type basis sets. Both of these formulae are three-point extrapolation schemes, but two-point schemes derived by choosing suitable effective C-constants can be used with essentially no loss of accuracy, at least for the present data. Given that real applications will cover a more diverse set of molecules, we recommend a three-point procedure when results from three basis sets are available. If only results from the two lowest quality basis sets are available, a two-point extrapolation is strongly recommended. It is noted that extrapolation of the results from unpolarized and polarized double-zeta type basis sets (pc-0 and pc-1) can give results within a few milli-hartree of the basis set limit. Unfortunately the extrapolation of total energies does not appear to significantly improve the results for atomization energies.

Acknowledgements This work was supported by grants from the Danish Center for Scientific Computation and the Danish Natural Science Research Council.

References

1. (a) Dunning TH Jr (1989) J Chem Phys 90:1007; (b) Peterson KA, Dunning TH Jr (2002) J Chem Phys 117:10548
2. Halkier A, Helgaker T, Jorgensen P, Klopper W, Koch H, Olsen J, Wilson AK (1998) Chem Phys Lett 286:243
3. Jensen F (2000) Theo Chem Acc 104:484
4. (a) Martin JML, Oliveira G (1999) J Chem Phys 111:1843; (b) Parthiban S, Martin JML (2001) J Chem Phys 114:6014; (c) Boese AD, Oren M, Atasoylu O, Martin JML, Kallay M, Gauss J (2004) J Chem Phys 120:4129
5. Jensen F (2001) J Chem Phys 115:9113; (2002) 116:3502
6. (a) Jensen F (2002) J Chem Phys 116:7372; (b) Jensen F (2002) J Chem Phys 117:9234; (c) Jensen F (2003) J Chem Phys 118:2459; (d) Jensen F, Helgaker T (2004) J Chem Phys 121:3462.
7. (a) Kobus J (1997) Adv Quant Chem 28:1; (b) McCullough Jr EA (1998) Encyclopedia of computational chemistry, vol 3, P 1941
8. Jensen F (2005) Theo Chem Acc 113:187
9. Dunning, TH Jr, Peterson KA, Wilson AK (2001) J Chem Phys 114:9244
10. (a) Schafer A, Horn H, Ahlrichs R (1992) J Chem Phys 97:2571; (b) Schafer A, Huber C, Ahlrichs R (1994) J Chem Phys100:5829; (c) Weigend F, Furche F, Ahlrichs R (2003) J Chem Phys 119:12753
11. Gaussian 98, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery Jr JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MV, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Salvador P, Dannenberg JJ, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CP, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian, Inc., Pittsburgh
12. Dalton, a molecular electronic structure program, experimental version, written by Helgaker T, Jensen HJAa, Jørgensen P, Olsen J, Ruud K, Ågren H, Auer AA, Bak KL, Bakken V, Christiansen O, Coriani S, Dahle P, Dalskov EK, Enevoldsen T, Fernandez B, Hättig C, Hald K, Halkier A, Heiberg H, Hettema H, Jonsson D, Kirpekar S, Kobayashi R, Koch H, Mikkelsen KV, Norman P, Packer MJ, Pedersen TB, Ruden TA, Sanchez A, Saue T, Sauer SPA, Schimmelpfennig B, Sylvester-Hvid KO, Taylor PR, Vahtras O
13. Jensen F (2005) Chem Phys Lett 402:510
14. Kutzelnigg W, Morgan JD III (1992) J Chem Phys 96:4484; J Chem Phys 97: 8821(E)
15. Klopper W, Kutzelnigg W (1986) J Mol Struct 135:339
16. Schwenke DW (2005) J Chem Phys 122:014107
17. Feller D (1993) J Chem Phys 98:7059
18. (a) Wilson AG, Dunning TH Jr (1997) J Chem Phys 106:8718; (b) Helgaker T, Klopper W, Koch H, Noga J (1997) J Chem Phys 106:9639; (c) Schuurman MS, Muir SR, Allen WA, Schaefer HF III (2004) J Chem Phys 120:11586; (d) Tajti A, Szalay PG, Csaszar AG, Kallay M, Gauss J, Valeev EF, Flowers BA, Vazquez J, Stanton JF (2004) J Chem Phys 121:11599
19. Halkier A, Helgaker T, Jørgensen P, Klopper W, Olsen J (1999) Chem Phys Lett 302:437
20. Peterson KA, Woon DE, Dunning TH Jr (1994) J Chem Phys 100:7410
21. Martin JML (1996) Chem Phys Lett 259:669
22. Martin JML, Taylor PR (1997) J Chem Phys 106:8620
23. Montgomery JA, Jr, Ochterski JW, Petersson GA (1994) J Chem Phys 101:5900
24. (a) Truhlar DG (1998) Chem Phys Lett 294:49; (b) Patton LF, Sanchez, ML, Truhlar DG (1999) J Chem Phys 111:2921