

Regular article

Improving numerical integration through basis set expansion

Drew A. McCormack¹, Evert Jan Baerends¹, Erik van Lenthe², Nicholas C. Handy³

¹Theoretische Chemie, Faculteit Exacte Wetenschappen, Vrije Universiteit, De Boelelaan 1083, 1081 HV, Amsterdam, The Netherlands

²Scientific Computing and Modeling, Vrije Universiteit, De Boelelaan 1083, 1081 HV, Amsterdam, The Netherlands

³Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK

Received: 22 December 2003 / Accepted: 25 August 2004 / Published online: 22 November 2004

© Springer-Verlag 2004

Abstract. Calculations are presented to assess a theorem presented by S.F. Boys [(1969) *Proc. R. Soc. A.* 309:195], regarding the accuracy of numerical integration in quantum chemical calculations. The theorem states that the error due to numerical integration can be made proportional to the error due to basis set truncation, and thus goes to zero in the limit of a complete basis. We test this theorem on the hydrogen atom, showing that with a solution-spanning basis, the numerically exact orbital energy can indeed be calculated with a small number of integration points. Moreover, tests for H and H₂⁺ demonstrate that even when only a near-complete basis is employed, Boys' Theorem can significantly reduce integration error. However, for other systems, like the oxygen atom and the CO₂ molecule, the theorem yields no advantage for some occupied orbitals. It is concluded that the theorem would be most useful for calculations that demand large basis sets.

Keywords: Numerical integration – Boys' Theorem – Numerical error

1 Introduction

Many quantum chemical software packages utilize analytical integration schemes, which have the advantage that integrals can be quickly and accurately evaluated. However, such schemes are limited somewhat in their choice of basis sets, because analytical solutions do not always exist, or are difficult to calculate. Often Gaussian or plane-wave basis sets are utilized, because integral evaluation is simple, but they are not optimal for representing molecular wavefunctions.

Sometimes numerical integration schemes are used, particularly for density functional theory (DFT), where some form of numerical integration is unavoidable. Numerical integration allows for a more general choice of basis set, but at the cost of integration accuracy. Such programs tend to make use of basis functions better optimized for molecular systems, such as Slater-type orbitals (STOs), which allow smaller bases to be used [1]. However, numerical integration introduces an additional form of error, that due to the integration scheme itself. Understanding the balance between basis set truncation error and numerical integration error is important in such cases: expending computation time on reducing one type of error may be pointless if the other is not reduced proportionally.

Interestingly, these two types of error are not independent of one another. A study by Boys [2] as long ago as 1969 showed that in fact the numerical integration error tends to zero as the basis set is improved, independent of how many integration points are used. For a complete basis, the numerical integration error is zero. This may seem surprising at first, but is actually quite reasonable: in the limit of a complete basis, exact orbital solutions are attainable, and evaluation of an exact orbital at a single point is sufficient to yield its energy exactly. It could thus have been anticipated that for a complete basis, any numerical integration scheme, no matter how poor, would yield the orbital energies exactly.

Boys' finding is elegant and insightful, but is it of any practical benefit? How close to completeness does a basis set need to be in order for a calculation to benefit from this theorem? Can a reduction in numerical grid size be achieved with the basis sets typically used in calculations? If not, is there at least a niche of chemically relevant systems where large basis sets are a requirement, and where a significant reduction in integration points could thus be realized? These questions form the basis of this study.

Correspondence to: D. A. McCormack
e-mail: da.mccormack@few.vu.nl

Here we adapt the density-functional code Amsterdam density functional (ADF) [1] such that it satisfies the requirements necessary for Boys' Theorem to take effect. We then demonstrate the validity of the theorem by performing calculations on a hydrogen atom where the basis set utilized completely spans the solution space. This is equivalent to introducing a complete basis set, and Boys' Theorem thus predicts that the energy will be calculated exactly, regardless of the numerical grid used.

Further calculations are performed for the H atom with an incomplete basis, along with O, H₂⁺, and CO₂. For these systems, the emphasis is not on achieving the exact solution, but instead on how the results from calculations utilizing Boys' approach compare with those in the standard implementation of ADF. The question of whether significant reductions in grid size could be achieved in typical calculations by making a quantum chemistry code "Boys-conforming" is also considered.

In Sect. 2 we summarize Boys' findings, and discuss the changes necessary to make ADF conform to his theorem. In Sect. 3 results are presented for H, O, H₂⁺, and CO₂, and discussed. Section 4 concludes.

2 Method

2.1 Boys' Theorem

Here we briefly summarize Boys' theorem 3 [2], which pertains to numerical integration. The theorem can be restated as follows: Approximate solutions to the eigenvalue equation

$$(\hat{H} - E_j)|\psi_j\rangle = 0 \quad (1)$$

can be obtained by solving the secular equations

$$\sum_i \langle \phi_k | \hat{H} - E'_j | \phi_i \rangle c_{ij} = 0 \quad (2)$$

for the approximate eigenvectors $|\psi'_j\rangle = \sum_i c_{ij} |\phi_i\rangle$, and approximate eigenvalues E'_j , where the $|\phi_i\rangle$ form an incomplete set of basis functions. Equation (2) can be further approximated by evaluating the integrals numerically:

$$\sum_i \langle \phi_k | \hat{Q} (\hat{H} - E_j^Q) | \phi_i \rangle c_{ij}^Q = 0 \quad (3)$$

giving approximate eigenvectors $|\psi_j^Q\rangle = \sum_i c_{ij}^Q |\phi_i\rangle$, and approximate eigenvalues E_j^Q . (The label Q is used here to denote 'numerically evaluated'.) The operator \hat{Q} is of the form

$$\hat{Q} = \sum_k \delta(\mathbf{r} - \mathbf{r}_k) w_k \quad (4)$$

where the \mathbf{r}_k are integration points, and the w_k are the corresponding weights.

Boys' showed, using perturbation theory, that the leading term in the error of the energy (E_j^Q), e_j , is given by

$$e_j = E_j^{11} (e_j^B + e_j^Q) = E_j^{11} (\mu_j^2 + \mu_j^Q \mu_j), \quad (5)$$

where

$$E_j^Q = E_j + e_j. \quad (6)$$

(Note that where Boys' treatment was quite general, we are only treating the case where \hat{H} is Hermitian, and left and right eigenvectors are thus identical.) μ_j^Q is the largest fractional error due to numerical integration that occurs in any of the integrals of Eq. (3), and E_j^{11} is a perturbation coefficient. μ_j is the least-squares error obtained by fitting the exact solution, ψ_j , with the basis functions $\{\phi_j\}$. That is

$$\mu_j = \|\psi_j - \tilde{\psi}_j\|, \quad (7)$$

with

$$\tilde{\psi}_j = \sum_k c_{jk} \phi_k, \quad (8)$$

the best least-squares approximation to ψ_j .

What Eq. (5) demonstrates is that as the error due to basis set truncation, $e_j^B \equiv \mu_j^2$, goes to zero, so too does the total error in the energy, regardless of the magnitude of the error due to the numerical integration scheme. The error introduced by numerical integration is $e_j^Q \equiv \mu_j^Q \mu_j$, not μ_j^Q as you might expect. The accuracy of a program utilizing a numerical integration scheme can thus be increased by utilizing a basis set that can accurately represent the solution (i.e., that reduces the error μ_j).

It should be stressed that Boys' theorem applies only to the orbital energies derived from the secular equations, and not to any expression for the total energy, or indeed any other numerical integration (e.g., normalization, moments of the density). Any practical application of Boys' theorem would require a second integration grid in order to accurately evaluate such quantities.

It is also important to emphasize that the theorem is only valid if all of the integrals that arise in Eq. (3) are evaluated using the same numerical integration scheme. If this is not the case, the error arising from using numerical integration is μ_j^Q , not $\mu_j^Q \mu_j$, and any advantage is negated [2]. Most of the software utilizing numerical integration will not comply with this requirement, and would need to be modified to garner any advantage from Boys' theorem. ADF is no exception in this regard, and the modifications that were required are described in the next section.

2.2 Adjustments made to ADF

Several adjustments were needed to ADF in order to make it conform to the requirements of Boys' theorem. Because the kinetic energy matrix should be symmetric, but is not owing to the numerical integration used to calculate its elements, ADF averages corresponding elements, forming a symmetric matrix. This averaging breaks the requirements of Boys' theorem.

In the Boys-conforming version of ADF—to be known here as ADF-B—this averaging of kinetic energy matrix elements does not occur. Instead, the full non-symmetric matrix is used in the calculation. This means, of course, that the Fock matrix is no longer symmetric, and cannot be diagonalized with a standard algorithm for real-symmetric matrices. A diagonalization routine for general real matrices was introduced for this purpose.

A real nonsymmetric matrix can potentially have complex eigenvalues, occurring in complex-conjugate pairs. Because the Fock matrix calculated by ADF-B is almost symmetric, the eigenvalues remained real throughout our calculations, and the problem of handling complex eigenvalues was avoided. However, any practical application of the method would undoubtedly have to address the issue of when complex eigenvalues of the Fock matrix could arise, and how they should be treated in the event.

In setting up the generalized-eigenvalue problem to be solved, ADF uses an analytically-calculated basis set overlap matrix. In order to conform to Boys' theorem, all integrals must be performed numerically, including the overlap matrix integrals. The ADF code was thus modified so that the overlap matrix was calculated utilizing numerical integration of the basis function products. This matrix was used, along with the non-symmetric Fock matrix discussed earlier, to calculate the orbital energies.

2.3 Tests on the hydrogen atom with a solution-spanning basis

To test the validity and practicality of Boys' theorem, several systems were considered. The hydrogen atom provides a good initial testing ground for Boys' theorem, because it is the only system for which a simple basis can be generated that completely spans the solution space. This system not only acts as a test of Boys' theorem, but also of the changes made in ADF in order to meet the requirements set by the theorem.

Because the exchange–correlation and electron–electron coulomb potentials do not cancel exactly in DFT calculations on hydrogen, as they should, each was excluded from the calculations. We are thus dealing with the solution in the simple $-1/r$ coulomb potential. The solutions are known analytically, and can be spanned completely by sets of STO basis functions.

2.4 Tests on hydrogen with non-spanning bases

Having tested whether Boys' theorem leads to numerically-exact results when a solution-spanning basis is used, we carried further tests out to ascertain whether results for non-spanning basis sets also benefit from the theorem, and how complete a set need be before that benefit is realized. The hydrogen atom was again adopted, because the exact orbitals are known analytically, which means the basis set error can be quantitatively established.

Calculations were performed on hydrogen, with a basis set consisting of a single STO function, with radial dependence proportional to

$$\exp(-\zeta r). \quad (9)$$

The function's exponent, ζ , was varied systematically, to gauge the relationship between basis completeness, and numerical error in the energy of the $1s$ orbital. The least-squares error in the approximation of an orbital is given by

$$\mu = \|\psi - \tilde{\psi}\| = \sqrt{2 - S}, \quad (10)$$

where $S = \langle \psi | \tilde{\psi} \rangle$, with ψ the exact orbital, and $\tilde{\psi}$ the best *normalized* least-squares fit to the exact solution for the basis set given. For an orthonormal basis set,

$$|\tilde{\psi}\rangle = \frac{\sum_i |\phi_i\rangle \langle \psi | \phi_i \rangle}{\sqrt{\sum_i \langle \psi | \phi_i \rangle^2}}. \quad (11)$$

For the $1s$ state the basis set error was easily evaluated with only a single basis function, because evaluating Eqs. (10) and (11) amounts to little more than determining the overlap between the basis function and the exact solution, i.e., between two STOs.

To ensure that results obtained for hydrogen with a single basis function were not uncharacteristic of more general cases, tests were performed as described earlier, but with two STO basis functions, instead of one. When using multiple STO basis functions, $\{\chi_i\}$, the $\{\phi_i\}$ in Eq. (11) should represent an orthogonalized set such as that generated by the symmetrical orthonormalization $\phi = \chi \mathbf{S}^{-1/2}$.

2.5 Tests on the oxygen atom

Given the simplicity of hydrogen, it was decided that calculations should also be performed for a multi-electron atom with a more general central potential. The oxygen atom was chosen for this purpose.

Because the exact solutions for oxygen are not known analytically, calculating the least-squares error for a given basis set presented more of a problem than for

hydrogen. A utility called ‘DIRAC’, which can be found in the ADF package [1], was used to calculate numerical, one-dimensional orbitals on a logarithmic radial grid. The basis truncation error, μ , was calculated as for the hydrogen atom with multiple basis functions (see Sect. 2.4), except that the overlaps between the basis functions and the ‘exact’ solutions, ψ , were determined by performing numerical integrals on the radial grid generated by DIRAC.

To allow for more direct comparison between calculations with different basis sets and integration grids, all calculations for the oxygen atom made use of a fixed potential, eliminating the need for a self-consistent-field cycle. The potential used in each case was the one generated by DIRAC, corresponding to the numerical orbitals calculated. This potential was read into ADF, and interpolated onto the ADF integration grid, in calculating the Fock matrix. The calculations in ADF and DIRAC were performed for the $X\alpha$ functional, with an α value of 0.7.

2.6 Tests on the H_2^+ molecule

It is more difficult to approach a solution-spanning basis set for noncentral potentials, such as those found in molecules, than for the single-center potentials of atoms. The atom-centered basis sets used in calculations on molecules are less capable of describing solutions in the bonding region than they are in the immediate vicinity of an atom.

The simplest molecule conceivable was chosen to test Boys’ theorem for multi-center potentials: H_2^+ . Being a one electron system, the exchange–correlation and electron–electron Coulomb potentials were removed from the calculations, as described previously for the hydrogen atom. Spin-unrestricted calculations were performed, with a bond length of 2.0 bohr.

Since analytical exact solutions of the H_2^+ wavefunction are not available, the size of each basis set—in terms of the number of functions—was used as a rough measure of basis set error. Standard ADF basis sets were used for the calculations, in addition to larger basis sets that included bond-centered STOs. For the latter, diffuse functions were included at the bond center, because they are known to complement the atom-centered functions [3].

2.7 Tests on the CO_2 molecule

CO_2 was chosen to test Boys’ theorem on a multi-electron molecule. No effort was made to develop solution-spanning bases; instead, standard basis sets taken from the ADF package were utilized. Calculations were performed with the Becke–Perdew generalized gradient approximation functional [4]. A fixed potential was used for each basis set. This potential was calculated for a given basis set by performing a standard ADF run with high integration accuracy, corresponding to an integration accuracy parameter (*accint*) of 12. The CO bond lengths were each set to 2.2120 bohr.

3 Results and discussion

3.1 Hydrogen atom with solution-spanning basis

Calculations were performed for the hydrogen atom with a basis set consisting of three STOs: two $1s$ functions, with exponents (ζ) of -1 and -0.5 (in atomic units), and one $2s$ function, with exponent -0.5 . This basis spans the solution space for the $1s$ and $2s$ atomic orbitals of hydrogen.

With the exchange–correlation and electron–electron Coulomb potential terms removed, ADF-B could be made to calculate the $1s$ and $2s$ orbital energies to within machine precision (Table 1), for integration grids ranging in size from 5 to 30 points. As would be expected, the accuracy of the standard version of ADF increased with the number of integration points, but did not reach the numerically-exact energy for any of the grids tested. These results fully confirm Boys’ theoretical findings.

3.2 Hydrogen atom with nonspanning bases

Results for a single basis function with varying exponent are shown in Figs. 1 and 2. Figure 1 shows the error in the $1s$ orbital energy calculated by ADF, relative to that calculated by ADF-B, as a function of basis completeness, and for different integration grids. The relative error is simply the absolute value of the ADF error divided by the absolute value of the ADF-B error. For values of the error greater than 1.0, ADF-B is more

Table 1. Orbital energies calculated for the hydrogen atom, with exchange–correlation and electron–electron Coulomb potentials excluded. Results are given for the standard implementation of Amsterdam density functional (ADF), and the implementation conforming to Boys’ theorem (ADF-B). The basis set used completely spanned the solution space of the $1s$ and $2s$ atomic orbitals of hydrogen. ‘*accint*’ refers to a parameter, stipulated in the input to ADF, that controls the density of the numerical integration grid used. Energies are given in atomic units

accint	Number of integration points	Atomic orbital	ADF energy	ADF-B energy	Exact energy
0.001	5	$1s$	−0.46560953370147	−0.5000000000000000	−0.5
		$2s$	−0.11710059227062	−0.1250000000000000	−0.125
4	19	$1s$	−0.50000020326244	−0.5000000000000000	−0.5
		$2s$	−0.12473182102444	−0.1250000000000000	−0.125
8	30	$1s$	−0.49999999999864	−0.5000000000000000	−0.5
		$2s$	−0.12498386637249	−0.1250000000000000	−0.125

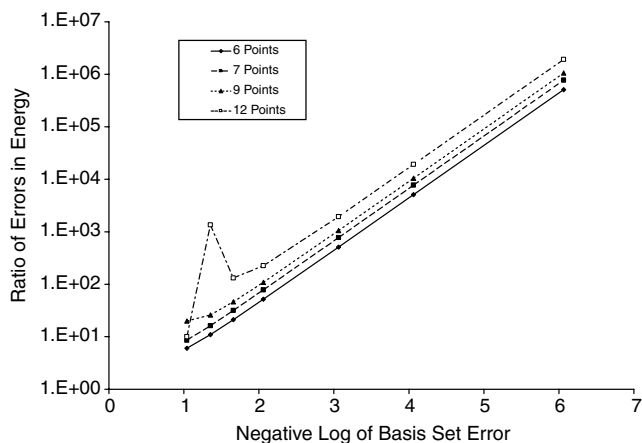


Fig. 1. Error in the $1s$ orbital energy for hydrogen of Amsterdam density functional (ADF) relative to the Boys-conforming ADF (ADF-B) versus the negative logarithm (base 10) of the least-squares basis set error (μ). The results are for a single Slater-type orbital (STO) basis function. Plots are shown for integration grids of various sizes, with the number of integration points used in each grid indicated in the legend. The ratio of errors is shown on a logarithmic scale. Values greater than 1.0 occur when the ADF error is larger than that of ADF-B, and values less than 1.0 when the ADF error is smaller

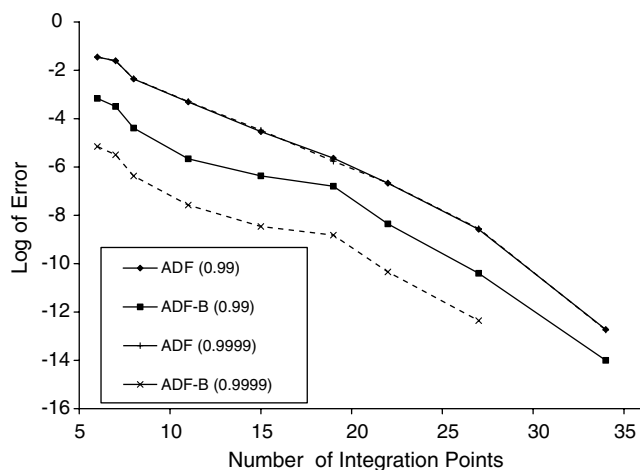


Fig. 2. Logarithm (base 10) of the absolute error in the $1s$ orbital energy of hydrogen versus the number of integration points used. The basis sets consisted of a single $1s$ STO basis function. Each plot is for a particular basis function exponent, ζ , and methodology (i.e., ADF or ADF-B). The methodology used is indicated, with the exponent given in *parentheses* in atomic units

accurate, and for values less than 1.0, ADF is more accurate.

The plots in Fig. 1 show an almost linear trend. (Though this is generally true, there are aberrations, since the error is not required to vary systematically with the completeness metric.) This upward trend is in agreement with the expectation that as the least-squares basis set error (μ) goes to zero [i.e., $-\log(\mu)$ increases], ADF-B should approach the exact orbital energy. The plots run approximately parallel, and are shifted upward by increasing the number of integration points, indi-

cating that ADF-B is relatively more accurate than ADF for larger integration grids. It is not clear why this should be the case.

Even with μ as high as 10^{-1} , ADF-B has a significant advantage over ADF, leading to an improvement of roughly a factor of 10. The relative improvement increases from there in logarithmic proportion to μ . With μ at 10^{-3} , the ratio of errors is 10^3 , and with μ at 10^{-6} , it is around 10^6 .

To put these results into perspective, a value of 10^{-1} for μ , leading to an order of magnitude reduction in the error of ADF-B over ADF, requires the overlap between basis set and exact solution—as defined by S in Eq. (10)—to be 0.995. For μ of 10^{-3} , leading to a 1,000-fold improvement, an overlap of 0.9999995 is required. These results would seem to imply that where a basis could be devised to achieve such overlaps, significant gains in accuracy could be achieved by invoking Boys' Theorem.

Figure 2 shows plots of the absolute error in the $1s$ energy as a function of integration grid size. Each plot is for a particular combination of methodology and basis function exponent. The results for ADF are relatively independent of exponent, with the two ADF curves lying practically on top of one another. For the exponents shown, ADF-B is considerably more accurate than ADF over all grid sizes. The accuracy of ADF-B increases as the basis set error approaches zero: the results for ADF-B with an exponent of 0.9999 lie below those with an exponent of 0.99.

An exponent of 0.99 corresponds to a value of 0.9999621224 for the basis-solution overlap, S , and 0.0087 for μ . An exponent of 0.9999 gives 0.000087 and 0.9999999962, respectively.

At this point we may ask the question: How many integration points are needed for ADF-B to achieve the accuracy generally accepted for calculations performed with ADF? The default value for *accint*, the integration parameter in ADF, is 4 for a single point calculation like this. With *accint* equal to 4, the integration grid has between 17 and 19 points in these particular calculations. If we trace a horizontal line from the value of the ADF curves in Fig. 2 corresponding to 19 integration points, back to the curves for ADF-B, we obtain an approximate answer to our question.

With exponent 0.99, ADF-B achieves the same error as ADF at 19 points with around 10–12 points. With an exponent of 0.9999, only approximately five points are needed. So, in this example, it is possible to make savings of 2–4 in the number of integration points required to achieve a satisfactory error, by using a Boys-conforming methodology.

Results analogous to those in Fig. 2, but for basis sets with two STO functions, are shown in Fig. 3. The exponents of the basis functions in each basis set were chosen to be symmetrical about 1.0, the exponent of the exact solution for the $1s$ orbital. The basis sets used had exponents of 0.5 and 1.5, corresponding to a basis truncation of $\mu=0.161915$; 0.9 and 1.1, giving

$\mu=0.006139$; and 0.975 and 1.025, giving $\mu=0.000383$. Results are shown for ADF and ADF-B, but only one plot is given for ADF, because the results for ADF are insensitive to the basis set used, as was seen in Fig. 2.

The findings for single-function basis sets carry over quite well to the two-function bases. Absolute errors for ADF-B are considerably smaller than those of ADF for all but the poorest basis shown. Errors for ADF-B tend to decrease (i.e., plots move downward) as μ goes to zero. This is true at most integration grid sizes, but not all.

We again wish to determine the integration grid size required for ADF-B to give the same error as ADF with 19 integration points, corresponding to an accint parameter value of 4. From Fig. 3, we estimate around 12 points would be needed for the most complete basis shown. This is a significant reduction, but less than was seen in Fig. 2. However, the basis truncation is larger for the most complete basis in Fig. 3 ($\mu=0.000383$) than it was in Fig. 2 ($\mu=0.000087$), which could partially explain the discrepancy.

Figures 1, 2, and 3 indicate that utilizing bases that are almost solution-spanning leads to dramatically reduced error in orbital energies when using a Boys-conforming methodology, for single- and multiple-function basis sets. Whether such bases are practically attainable in more complex systems occupies much of our attention for the rest of this study.

3.3 Oxygen atom

Figure 4 shows the dependence of the absolute energy error on the number of integration points for oxygen.

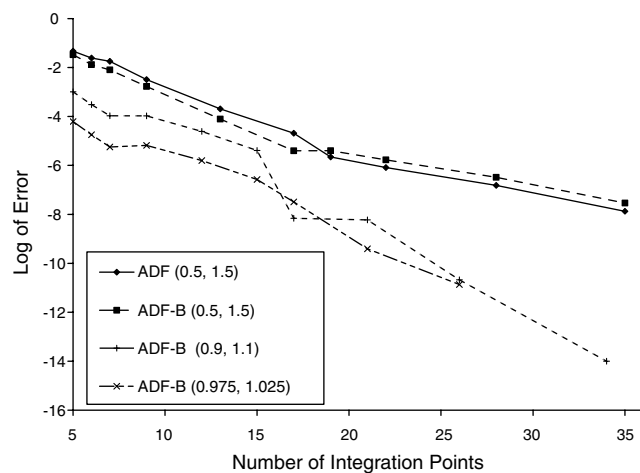


Fig. 3. Logarithm (base 10) of the absolute error in the 1s orbital energy of hydrogen *versus* the number of integration points used, for two-function basis sets. The basis sets used consisted of 1s STO basis functions. Each plot is for a particular pair of basis function exponents, ζ , and methodology (i.e., ADF or ADF-B). The methodology used is indicated, with the exponents given in *parentheses* in atomic units

Because the exact orbital energies are not known, the error is taken with respect to an ADF calculation with a very high integration accuracy. The integration parameter,

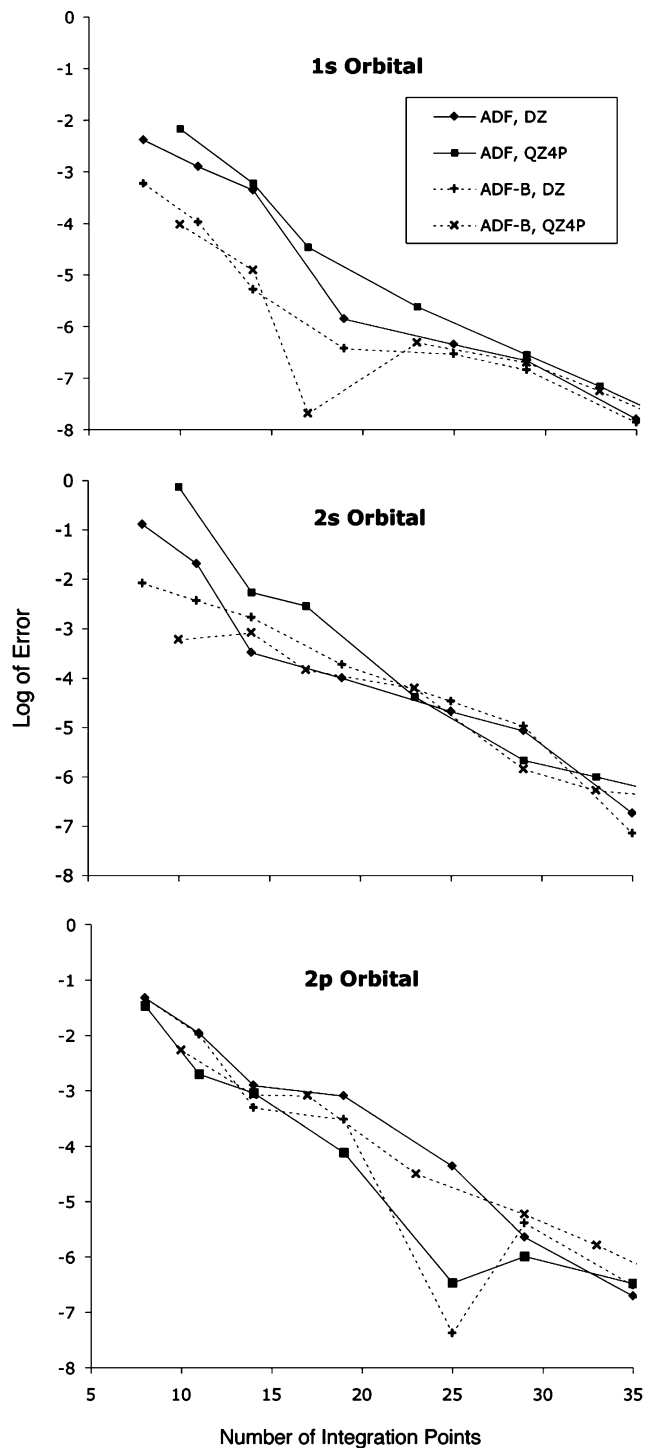


Fig. 4. Dependence of the logarithm of the absolute integration error in the orbital energy on the number of integration points, for oxygen. Separate panels are shown for each of the orbitals of oxygen. Each panel gives results calculated using ADF and ADF-B, for two different basis sets: double-zeta (DZ) and quadruple-zeta with four polarization functions (QZ4P). These basis sets are described in the text. The error itself is also defined in the text

accint, was set to 12 for this purpose, which corresponds approximately to the limit of machine precision.

Separate plots for each orbital are shown, for ADF and ADF-B with two different standard ADF basis sets. The double zeta (DZ) basis included the following STOs, with the exponent given in parentheses: $1s$ (9.80), $1s$ (6.80), $2s$ (1.70), $2s$ (2.82), $2p$ (3.06), $2p$ (1.30). The quadruple-zeta with four polarization functions (QZ4P) basis included $1s$ (14.800), $1s$ (8.750), $1s$ (7.100), $1s$ (6.650), $2s$ (3.200), $2s$ (2.050), $2s$ (1.500), $2s$ (0.750), $2p$ (5.700), $2p$ (3.050), $2p$ (1.650), $2p$ (1.000). STOs not contributing to the occupied orbitals, such as d and f STOs, are not given here. The basis error metrics, S and μ , corresponding to these bases, are given on a per orbital basis in Table 2.

The results in Fig. 4 are not nearly as impressive as they were for hydrogen. The results for the $1s$ orbital are good, with the error for ADF-B 1–2 orders of magnitude less than for ADF with small integration grids. Remarkably, this improvement with small integration grids is obtained for both the small DZ basis and the large QZ4P basis. To achieve approximately the same accuracy as ADF with around 30 integration points (accint approximately equal to 4), ADF-B only requires around half as many points.

The results for ADF-B deteriorate moving to higher orbitals, such that for $2p$ there is virtually no advantage in using ADF-B over ADF. ADF-B does show some improvement over ADF for the $2s$ orbital, but only with very few integration points. There is certainly no gain to be had for either of these orbitals when using the standard integration grids commonly adopted for ADF calculations (approximately 30 integration points).

The least-squares basis set errors in Table 2 increase going from $1s$ to $2p$, as would be anticipated. However, in light of the magnitude of these errors, the results are somewhat disappointing when compared with those for the hydrogen atom. The values in Table 2 demonstrate that the basis set error for $1s$ is in the range $[-\log\mu = 3.2(\text{DZ})-3.6(\text{QZ4P})]$, which for hydrogen was sufficient to reduce the integration error by factors in the range 100–5,000. Reductions in the integration error for O $1s$ are on the order of 100, for a small grid (in the range 5–15 points). The basis set error for $2p$ is larger $[-\log\mu = 1.6(\text{DZ})-2.9(\text{QZ4P})]$, but reductions by factors of 10–100 would still be expected on the basis of the hydrogen results. The results in Fig. 4 do not bear this out. Clearly the basis set error alone is not adequate for establishing if a significant improvement in integration accuracy can be achieved through Boys' theorem.

3.4 H_2^+ molecule

Orbital energy errors for calculations on H_2^+ are presented in Fig. 5. Results are given for different basis sets, and integration grid parameters (i.e., accint). Two of the basis sets shown, single zeta (SZ) and QZ4P, are stan-

Table 2. Basis set error metrics for the ADF basis sets used in calculations on the oxygen atom. The basis sets and metrics are defined in the text. Values are given for each of the occupied orbitals of oxygen: $1s$, $2s$, and $2p$. The values were obtained utilizing numerical orbitals taken from calculations with the utility DIRAC

Basis set	Metric	$1s$	$2s$	$2p$
DZ	S	0.99999977	0.99999779	0.99972973
	μ	0.00067212	0.00210164	0.02324944
QZ4P	S	0.99999997	0.99999996	0.99999933
	μ	0.00025229	0.00029381	0.00115951

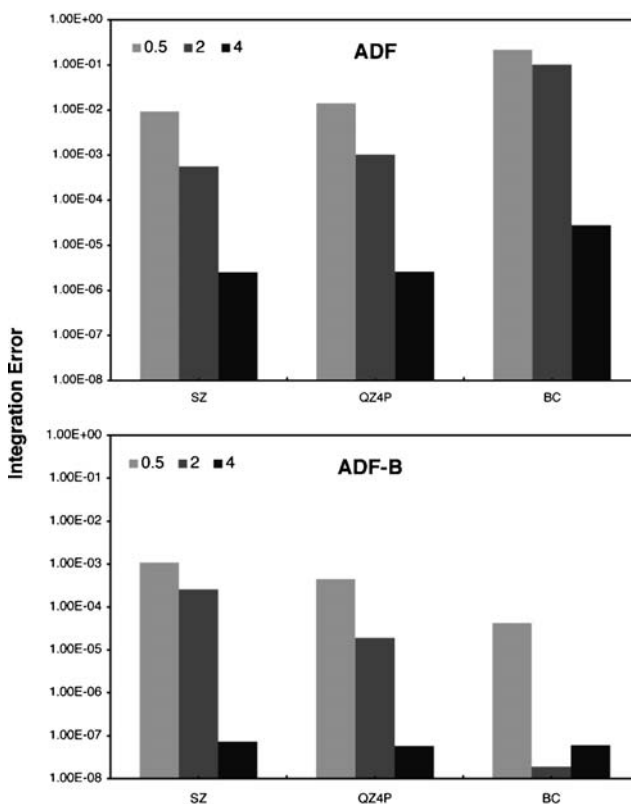


Fig. 5. Absolute error in orbital energies due to numerical integration, for calculations on H_2^+ with various basis sets and integration grids. The error in the energy is taken relative to a calculation with the integration accuracy parameter (accint) set to 12. The basis sets are described in the text. The *top panel* is for calculations with ADF, and the *bottom panel* for calculations performed with ADF-B. For each basis set, *bars* are presented that correspond to different values of accint. Values of 0.5, 2.0, and 4.0 were used

dard ADF basis sets. For a hydrogen atom, SZ includes a single $1s$ STO with exponent 1.24. The QZ4P basis set includes the following STOs: $1s$ (3.300), $1s$ (2.000), $1s$ (1.400), $1s$ (1.000), $1s$ (0.710), $2p$ (2.000), $2p$ (1.000), $3d$ (2.500), $3d$ (1.500).

The last basis set shown in Fig. 5 is labeled ‘BC’, for bond-centered. This is a very large basis set, with functions centered not only on the atoms, but also at the center of the bond. The STOs assigned to each atom were $1s$ (5.000), $1s$ (3.300), $1s$ (2.000), $1s$ (1.500), $1s$

(1.000), $2p$ (5.000), $2p$ (3.000), $2p$ (2.000), $3d$ (7.500), $3d$ (4.500), $3d$ (3.000), $4f$ (10.000), $4f$ (6.000), and $4f$ (4.000). At the bond center, the following STOs were included in the BC basis: $1s$ (5.000), $1s$ (3.300), $1s$ (2.000), $1s$ (1.400), $1s$ (1.000), $1s$ (0.710), $2s$ (0.750), $2s$ (0.500), $3s$ (0.500), $4s$ (0.500), $3d$ (4.000), $3d$ (2.000), $3d$ (1.400), $3d$ (1.000), $3d$ (0.500), $4d$ (0.500).

The results in Fig. 5 show that the error due to numerical integration is significantly smaller for ADF-B than for ADF. This holds for each of the three basis sets, and for the three grids used (accint = 4 is the default grid; accint = 2.0 and 0.5 are small and very small grids, respectively). As expected, the improvement is larger for larger bases: the ADF error actually increases slightly with basis set size, whereas the ADF-B error decreases. The distinction becomes striking for the largest basis, BC, where ADF-B achieves a numerical precision of around 10^{-4} on a very small grid (accint = 0.5). This is comparable to the precision achieved with the default grid (accint = 4) in ADF. ADF-B also significantly improves upon ADF for the larger accint = 2.0 and accint = 4.0 grids.

For any given basis set, the error for ADF-B decreases as the basis size is increased with accint equal to 0.5 and 2.0. For accint = 4.0, the ADF-B results show little dependence on basis set, probably because the calculations are very accurate for all of the basis sets, leaving little room for improvement over this particular range of basis sets.

These results demonstrate that benefits could be realized through Boys' theorem for molecular systems. ADF-B yields improved numerical precision in all cases, but is truly impressive in calculations with very large basis sets, larger than those typically in use today. Calculations that, for whatever reason, demand a very large basis, would require large integration grids to match the precision afforded by the basis set. Such calculations could profit from implementing Boys' theorem, allowing smaller integration grids to be used.

3.5 CO_2 molecule

The results for the CO_2 molecule are given in Fig. 6. They show the root-mean squared error in the orbital energies for various ranges of orbitals, as a function of integration grid size. The basis set QZ4P is the largest available set for routine calculations.

There is a striking difference between the lowest orbitals (see curves for orbitals 1–3, the $1s$ orbitals on C and O) and the valence orbitals (see curves for orbitals 7–9 as a representative set). Beyond a grid size of 125 points there are only minor differences between ADF and ADF-B (350 points corresponds approximately to accint = 4, the ADF default for a single point calculation). Both ADF and ADF-B are more accurate for the lowest orbitals (i.e., 1–3) than for the highest occupied orbitals (i.e., 7–9). Results for ADF-B are really only distinguishable from those for ADF on the smallest

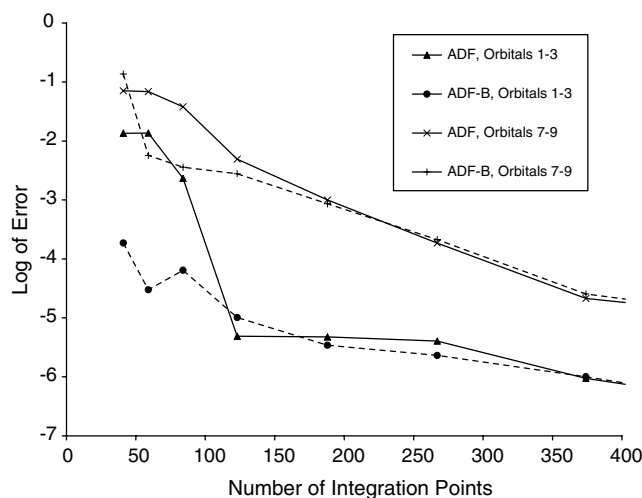


Fig. 6. Root-mean-squared (RMS) errors in the occupied orbital energies of CO_2 . All results are for the QZ4P atomic basis sets of C and O. Results are given for ADF and ADF-B, as a function of the number of integration points used. The RMS errors used are for different ranges of orbitals, ordered according to energy. The results are shown for RMS errors taken over the lowest three orbitals (i.e., 1–3) only, and taken over the highest three occupied orbitals (i.e., 7–9) only

integration grids shown. Where ADF errors tend to blow up as the integration grid shrinks below 125 integration points, the ADF-B errors continue to follow the trend set for larger grids, with error increasing more gradually.

For the lowest occupied orbitals, ADF-B maintains an acceptable level of error right down to the smallest grids tested. For calculations where only the lowest-lying orbitals are of importance, Boys' theorem offers clear advantages in terms of the grid size reduction.

For higher-lying occupied orbitals, the advantage is also more than an order of magnitude for certain grid sizes; however, for the high-energy orbitals it is not true that errors for small integration grids are comparable in magnitude to those for larger grids. Results for 50 integration points, for example, are at least two orders of magnitude worse than those for 350 points. Even though ADF-B outperforms ADF for these orbitals, it is still unable to compete with larger grid sizes. These results indicate, in line with what we found for the oxygen atom, that the QZ4P basis is still too far from completeness for the higher orbitals to fully benefit from Boys' theorem.

4 Conclusion

In the sense that Boys' theorem has been 'experimentally' confirmed, this study has succeeded. The theorem states that the error in orbital energy that arises out of utilizing a numerical integration scheme is proportional to the least-squares error of the best fit of the basis set to the orbital in question. In the limit of a solution-spanning basis set, orbital energies will be exact, regardless of

the numerical integration grid used. We have been able to show that for a hydrogen atom at least this is so.

However, our objectives were broader than simply affirming the truth of Boys' finding in the limiting case. We were equally interested in the behavior of the energy errors in the approach to that limit. Most of this study has concentrated on how the integration error varies as the basis set is expanded.

Tests on the hydrogen atom revealed that the error in orbital energy decreased for a Boys-conforming version of ADF as the least-squares basis set error decreased. This was true whether a single basis function was used, or multiple functions.

The Results for the oxygen atom demonstrated that the benefit from Boys' theorem is only substantial for larger basis sets than even the largest ones commonly used today. For the lowest orbital, O $1s$, where the basis error is relatively small, results for the Boys-conforming code were significantly better on small integration grids, but little improvement was seen for the high-energy occupied orbitals, where the basis set error was larger.

Calculations for polynuclear potentials (H_2^+ and CO_2) showed a similar trend: Boys' theorem was effective in reducing the error of the occupied orbital of H_2^+ and the lowest lying orbitals ($1s$ -based) of CO_2 , such that a significant reduction in grid size could be realized without large reductions in accuracy. For high-energy occupied orbitals in CO_2 , and a QZ4P basis, while Boys' theorem did contribute to improved accuracy on small

integration grids, the reduction was not nearly enough to bring errors in line with those for integration grids typically used in calculations.

In conclusion, it is quite difficult to enter the regime of near-complete basis sets where significant benefit can be derived from Boys' theorem. In this study, we have been able to do this for the occupied orbital of H_2^+ and the lowest-lying orbitals of CO_2 , but for high-energy occupied orbitals we would have to go considerably beyond even the best basis sets (e.g., QZ4P) commonly in use today. In practice, this means the applicability of Boys' theorem is to a select subset of problems, where very large basis sets are a requirement. Larger basis sets place increasingly higher demands on the numerical integration grid; application of Boys' theorem would act to negate this effect, significantly reducing the numerical grid, and making such calculations more tractable.

References

1. te Velde G, Bickelhaupt FM, Baerends EJ, Fonseca Guerra C, van Gisbergen SJA, Snijders JG, Ziegler T (2001) *J Comput Chem* 22:931
2. Boys SF (1969) *Proc R Soc Lond Ser A* 309:195
3. Osinga VP, Oostenbrink BC, van Lenthe E, Guerra CF, Baerends EJ (2002) *Chem Phys* 284:565
4. Perdew JP (1986) *Phys Rev B* 33:8822; Erratum (1986) *Phys Rev B* 34:7406; Becke A (1988) *Phys Rev A* 38:3098